DISCRETE GEOMETRIC AND FUSION BASED TECHNIQUES FOR OBJECT DETECTION AND DECISION SUPPORT

András Hajdu

Department of Computer Graphics and Image Processing
Faculty of Informatics, University of Debrecen

Dissertation for the Doctoral Degree of the Hungarian Academy of Sciences

2015

To my family

I gratefully acknowledge the contributions of my MSc and PhD students from University of Debrecen, Hungary: Bálint Antal, Henrietta Tomán, Balázs Harangi, László Kovács, János Tóth, István Lázár, Brigitta Nagy, Roland Harangozó, Ignác Csősz, László Szakács, József Szakács, Renátó Besenczi, Kristóf Szitha, Dávid Hornyik, László Farkas, László Dancsi; as well as Stylianos Asteriadis, Charalambos Giamas, Athanasios Roubies from Aristotle University of Thessaloniki, Greece. My colleagues at various institutions also provided valuable contributions: Ioannis Pitas and Nikos Nikolaidis from Aristotle University of Thessaloniki, Greece; Lajos Hajdu, János Kormos†, Attila Fazekas, György Kovács, Attila Szalkay and Adrienne Csutak from University of Debrecen, Hungary; Robert Tijdeman from Leiden University, The Netherlands; Rashid Jalal Qureshi from Islamia College Peshawar, Pakistan; Tünde Pető, Moorfields Eye Hospital London, UK; Fabricé Mériaudeau and Désiré Sidibé, University of Bourgogne, France; Adrián Cristal and Osman Ünsal, Barcelona Supercomputing Center; Spain.

Thermal videos were captured and provided by the Fire Department of Dortmund, Germany within the confines of the project FP6-004218, SHARE: Mobile Support for Rescue Forces, Integrating Multiple Modes of Interaction. Retinal images were collected and annotated by the Moorfields Eye Hospital Londok, UK and the Ophthalmology Clinics of the University of Debrecen, Hungary.

Contents

List of Figures

List of Tables

Introduction

This dissertation presents some novel results concerning object detection tasks in digital images. All the investigations were motivated by practical problems with the corresponding applied techniques belonging to the fields of discrete mathematics/geometry and information fusion. However, besides providing efficient practical tools, our aim was to discover also the theoretical background of the problems to be able to characterize and validate our solutions in an objective way. The work is built upon results [1–55] published after September 2003 – the time, when the author obtained his PhD degree. Among these publications [1–16] were published in journals covered by the Science Citation Index (SCI), while [17–55] are other indexed works. The structure of the dissertation basically follows a chronological order, which also reflects the logic how the results are built upon each other. This form of the presentation of the content also helps the reader understand why and how specific activities initiated each other. In this introduction, we shortly describe the practical problems that motivated the presented research and development activities. For each chapter, we give some details about the challenges of the specific task investigated there with a more comprehensive introduction is enclosed at the beginning of each chapter. Similarly, we highlight here the importance of our related results with including the complete presentation in the corresponding chapters. In general, our new results presented in the dissertation proved to be competitive in the corresponding fields, in several cases outperforming the previously used or other state-of-the-art approaches/algorithms/techniques. It is supported also by the fact that these results were published in leading journals, e.g. of the IEEE[1]. All fields we investigate are of high relevance with important applications attracting the attention of many experts, having a vast literature. The description of the related works are outlined in the corresponding chapters, as well.

The practical problems addressed in the dissertation primarily come from two fields. Our first application relates to the project SHARE[2], which aimed to develop an efficient rescue system for firefighters. Our task in the project was to extract interesting objects (e.g. humans) from videos acquired with thermal cameras by a rescue team. Thermal video analysis has great importance in such scenarios, since the infrared light has better penetration performance through smoke than the visible one. Thus, the firefighters can gather such information (e.g. door frames for exit, injured people on the floor, rescue team members) that they could not see by themselves or by using a camera working in the visible light domain. As a specific task, we focus on the recognition of human silhouettes in such videos.

Our second application relates to the project DRSCREEN[3] supervised by the author addressing the development of an automatic screening system for diabetic retinopathy (DR) based on the processing of retinal images. As for the clinical background of this field, more than 382 million people are suffering from diabetes in 2015 worldwide and the number of the diagnosed cases has been growing rapidly (e.g. in 2012 this figure was 360 million). Long-term diabetes also affects

[1]Institute of Electrical and Electronics Engineers.

[2]EU FP6 Information Society Technologies, FP6-004218, SHARE: Mobile Support for Rescue Forces, Integrating Multiple Modes of Interaction.

[3]TECH08-2 grant of the Hungarian National Office for Research and Technology (NKTH), DRSCREEN - Developing a computer-based image processing system for diabetic retinopathy screening.

the eye, resulting in a disease called diabetic retinopathy (DR). Automatic screening systems for DR are of great importance and our presented results also demonstrate our efforts to develop such a system. The corresponding tasks include the detection of the anatomic parts of the retina as well as the DR specific lesions. Using our object detection findings, we have developed a screening system to support everyday clinical routine.

As for the structure of the dissertation, some basic concepts and notations are introduced first in **chapter 1**. Besides the general formalism regarding digital images, we recall the necessary elements of discrete geometry and information fusion. From the tools of discrete geometry, we recall some graph theoretical concepts that are needed for the analysis of general digital curves representing e.g. object boundaries. In some of our fusion-based approaches, we also compose graphs having the candidates of object detectors as its vertices and apply weighted graph theoretical methods to extract the final suggested locations of desired objects. Since the dissertation focuses on fusion-based techniques in a large extent, we also recall the necessary concepts of this field. Namely, we recall decision rules, whose extension to the spatial domain are considered for data aggregation in our models, and error measures to evaluate the performance of our corresponding ensemble-based systems. We also describe briefly the clinical field of retinal image analysis with highlighting those anatomical parts and lesions, whose detection will be primarily addressed. As certain general methodological components that will be considered in several chapters later on, we give some details on databases used for training and evaluation, and also on preprocessing and object detection algorithms that will be considered to build up our ensemble-based systems. However, those algorithms and databases that are cited only in one chapter are going to be introduced there.

The novel contributions are started to be discussed from **chapter 2** with the presentation of a method for object simplification [1, 17]. This technique can be considered as an optimal sampling providing an output, which can represent the original object with smallest error in a Hausdorff distance-based object matching scenario. As for methodology, we extend the sampling methods using the centroidal Voronoi tessellation (CVT) framework in a theoretical basis to reach our aim. Our sampling methods are suitable for the simplification of both contour- and region-like objects. It has been tested on human silhouettes extracted from thermal videos captured during rescue actions, to recognize firefighters and injured people in different poses. We have also demonstrated how the sampling can be directed by including a weight function in the model [18]. In this way, we can focus more on the main shape characteristics with suppressing the focus on the more various peripheral behavior of similar objects. As for their exploitations, our methods have been integrated in the project SHARE. The importance of our contributions lies in the fact that we can save computation time during the matching process, since it linearly drops with the level of simplification. Fast, online methods are highly welcome in rescue scenarios, where time is a key factor. Besides speeding up computations, our approach helps in composing such template databases that need smaller storage space to be carried more easily by devices having limited resources during a fire/rescue scene. Besides the chamfer matching-based approach, we investigate how the sampling strategies can be considered to speed up computations in learning-based segmentation [19]. Namely, we check whether an appropriate sampling strategy may help in giving a better representation of vessel points to separate them better from the background to segment the vascular system of the retina within our project DRSCREEN.

Still based on discrete geometric approaches, **chapter 3** describes a method for digital curve compression [2, 20]. The motivation of this study originally was to compress the human body contours described in chapter 2 to support compact portability. We provide a graph theoretical approach to trace curves having arbitrary topology with assigning a graph to the digital curve and using letters from a Bezier-alphabet of linear line segments to approximate it. Because of the tracing step, the proposed method has better compression performance than the current state-of-

the-art approaches. This approach have also been incorporated in the SHARE system mentioned above to efficiently store human body contours. As a corresponding field, we also present some results regarding the possible improvement of the skeletonization of vessel intersections [21] with exploiting this result in our system DRSCREEN. Namely, when we extract a curve as the skeleton of a thicker object, the thinning process usually distorts the intersections, which can be improved by our proposed local method.

Though we also demonstrate the applicability of the above approaches there, from **chapter 4** on we present our results focusing on the specific clinical challenge of developing an automatic screening system (DRSCREEN) for diabetic retinopathy (DR). To be competitive in this highly investigated field, we have considered approaches based on information fusion by combining the output of different image processing algorithms dedicated to specific object detection tasks. In this way, we generate ensemble-based systems, which are generally known to be more accurate and balanced. Moreover, this approach leads to very flexible systems, since any future individual algorithms can be integrated to improve the performance of the ensembles further. From both theoretical/practical points of view a serious challenge is to measure up/assure the diversity (independency) of the member algorithms, since a more diverse system is naturally expected to perform better, even if the individual accuracies are lower. We start the discussion with presenting our results [3, 22] in chapter 4 for the detection of the optic disc (OD) and macula which are two normal anatomical parts of the retina. The detection of the OD is important to avoid its mis-recognition as a bright lesion (like exudate), while the macula has a significant role as being the center of the sharp vision. We propose approaches to automatically combine different OD and macula detectors, to benefit from their strengths while overcoming their weaknesses. In this study, we apply simple majority voting rule for the detection of the desired objects with selecting the region, where the largest number of the single candidates of the member algorithms fall. The dependency of the members is also addressed by assigning weights to their candidates considering appropriate pairwise statistics. To improve the detection accuracy, we include a strict geometric constraint on the mutual location of the OD and macula in our framework. Besides the simple majority voting-based rule, we apply weighted graph theoretical algorithms to perform the fusion of the individual detector outputs [23], when they are allowed to have multiple candidates. Moreover, we explain how to take advantage of all the available information from the output of the member algorithms supplied in terms of confidence maps [4]. We apply axiomatic and Bayesian approaches, as in the case of aggregation of judgments of experts in decision and risk analysis, to combine these confidence values. With the machine learning-based Bayesian models, we can also make a great effort to discover the dependencies among the member algorithms. Exhaustive experimental tests on publicly available datasets prove the competitiveness of all of these approaches against the state-of-the-art individual algorithms. Our experimental results verify the natural expectation that involving more information in the final decision raises accuracy and reliability, as well. These methods have also been considered in the automatic screening system DRSCREEN.

After reaching improvement in object detection using fusion-based approaches, we performed a thorough theoretical investigation to describe more precisely the behavior of voting models in the spatial domain. Accordingly, in **chapter 5**, we present our results on the generalization of the classic voting models to the spatial domain [5, 24], where the voters are detector algorithms with giving their votes in terms of single pixels for the location of the desired object. Beyond the generalization of the simple majority voting model, we explain how to assign weights to the member algorithms [25] based on their individual accuracies and the shape of the object to improve system performance. Concentrating only on independent components/variables is also a bottleneck in the theoretical investigations in this field. Since this approach can hardly be accepted in practice, we put effort to give an interval, where the accuracy of the system falls considering dependencies among the members in our models. Moreover, to address this issue better, we extend some

measures dedicated to discover the diversity of the members [26] to the spatial domain. This approach helps in composing efficient ensembles not purely based on the individual accuracies of the members, but also with taking the diversity issue also in account. The suitability of our models have been demonstrated in the OD detection task with empirical tests in publicly available databases showing improvement against the classic models.

Besides the detection of the normal anatomical components, the efficient detection of specific lesions is also a crucial task in an automatic screening system. In the case of DR, the most important lesions are the microaneurysms (MAs) and exudates. Regarding their detection with image processing techniques, the main difference is that MAs can be represented by single pixels, while exudates with regions. From the theory of information fusion, it is known that the combined system is more efficient in general, if the members are diverse. In **chapter 6**, we present our related results considering the fusion of lesion candidate extractors in a voting-based environment. The larger diversity among the members is reached by applying different preprocessing algorithms before the lesion candidate extractor algorithms. This idea can be easily reasoned by the observation that different preprocessors enhance image content differently. However, this approach is primarily meaningful in an ensemble-based scenario, where the possible deteriorations of the members can be compensated with increasing divergence between them. In this way, we create ⟨preprocessor, candidate extractor⟩ (shortly ⟨PP, CE⟩) pairs and compose combined systems from them [6, 27–29]. Then, we demonstrate how to create very competitive detectors in this way to recognize MAs [7, 30, 31]. This result of ours is internationally also admitted, as from 2012 till the submission of this dissertation our method is ranked as first in the Retinopathy Online Challenge[4], which is the primary benchmark forum of MA detector algorithms. We also show that the detection of MAs can be improved further by adding some contextual information regarding them [8, 32] in the extracting process. Namely, the detection of MAs highly depends on the characteristics of the imaging device and other image properties (e.g. type of compression). As a result, some MAs can be easily spotted on the background of the retina, while the recognition of others are more difficult. Besides image characteristics, the spatial location also has influence on the detection of MAs (e.g. proximity of vessel parts, etc.).

The main purpose behind our activities dedicated to the processing of retinal images was to develop a decision support system for the automatic screening of diabetic retinopathy. The motivation for creating such reliable systems is to reduce the manual effort of mass screening, which also raises a financial issue. While several studies focus on the recognition of patients having DR and considering the specificity of the screening as a matter of efficiency, in **chapter 7**, we show how both sensitivity and specificity can be kept at high level by combining novel screening features and a decision-making process. Regarding decision making, automatic DR screening systems either partially follow clinical protocols or use a machine learning classifier. A common way to improve reliability in machine learning-based applications is to use ensemble-based approaches. For medical decision support, ensemble methods have been successfully applied to several fields. The proposed system [9, 33, 34] is ensemble-based at more levels: we consider ensemble systems both in image processing tasks and decision making. In chapter 7, we present how the characteristic features extracted by lesion detection and anatomical part recognition algorithms described in the subsequent chapters and others [35, 36] are integrated in a screening system. As for lesions, besides investigating fusion-based techniques, we have developed an individual MA detector [10, 37] based on intensity profile analysis, as well. This method – as a member algorithm – has also been incorporated in our detector ensemble considered for MA detection. Using similar principles as in our MA detector presented in chapter 6, we compose a fusion-based system also for the detection of exudates [38], which is a competitive approach for this task based on our experimental

[4]Retinopathy Online Challenge (ROC), http://webeye.ophth.uiowa.edu/ROC/

studies. All these features are then classified by using an ensemble of classifiers to reach the final decision. Besides the flow of decision making, we also present how to prefilter severely diseased cases or low quality images. The system has already been evaluated on publicly available databases providing high performance compared with other state-of-the-art techniques. These experimental tests suggest that our system is capable of performing automatic screening beyond simple decision support. Especially, our results are very close to meet the recommendations of the British Diabetic Association with its sensitivity 80% and specificity 95%.

Partially due to space reasons and to keep the content of the dissertation more coherent, some results are presented in a short form in **chapter 8**. These works are either loosely related or have just finished recently showing how the results of the dissertation are going to be improved or integrated in future activities. Regarding discrete geometry, we mention that in some of our works [11–14, 39] we investigate efficient ways of digital distance measurement. We also apply digital geometric approaches to improve the performance of active contour (snake) models in human body extraction tasks [40–42]. As for our clinical application, we have shown that the intensity profile analysis considered for MA detection can be successfully applied in retinal vessel segmentation, as well [43]. As a more theoretical approach, we investigate kernel functions leading to translation invariance in intensity [15] with demonstrating our findings in MA detection. Besides ensemble-based approaches, we have also developed a competitive methods for exudate detection in [16, 44–47]. We determine clusters of retinal images coming from different sources to optimize MA [48] and exudate [49] detectors accordingly. As a natural step forward, we determine the optimal parameter settings in our ensemble-based MA [50] and exudate [51] detectors and give suggestions to make the stohastic search procedure computationally more efficient. XML-based metadata schemes have been created for the content description in both of our systems SHARE [52] and DRSCREEN [53]. To increase the performance of our system in DR screening, in [54, 55] we have shown that the inclusion of proteomic data gathered from tear as a secondary modality is a promising approach.

New scientific results are enclosed in terms of thesis points in the **Summary of new scientific results**. To preserve the readability of the dissertation, some proofs of technical nature have been moved to the **Appendix**. The author's own papers cited in the dissertation are enclosed in chapter **Author's publications**, while others' works in the **Bibliography**.

Chapter 1

Basic concepts and notations

In this chapter, we recall some basic concepts, notations and results that will be needed later on. After the introduction of some general formalism, we collect the necessary elements of graph theory and information fusion according to the content of the dissertation. We also give a brief description of clinical concepts, databases and detector algorithms regarding the specific application fields.

1.1 General formalism

Let the sets of natural, integer and real numbers be denoted by $\mathbb{N} = \mathbb{Z}_{\geq 0}$, $\mathbb{Z}$, and $\mathbb{R}$, respectively. To refer to their positive (resp. non-negative) subsets, we will use the notations $\mathbb{Z}_{>0}$, $\mathbb{R}_{>0}$ (resp. $\mathbb{Z}_{\geq 0}$, $\mathbb{R}_{\geq 0}$). With an $m \in \mathbb{N}$, the m-dimensional (mD for short) number sets will be denoted by $\mathbb{Z}^m$, $\mathbb{R}^m$, respectively. From 2D on ($m \geq 2$), the elements of $\mathbb{Z}^m$, $\mathbb{R}^m$ will be written using the vector notation $\boldsymbol{x} = (\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(m))$. For scalars, we omit the vector format (e.g. $k \in \mathbb{Z}, p \in \mathbb{R}$). The spatial subsets will be denoted by capital letters as $A, B, C \in \mathbb{Z}^m$ or $\mathbb{R}^m$; the cardinality of a set A is written as $|A|$. To measure the distance of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^m$ or $\mathbb{R}^m$, usually their Euclidean distance $d(\boldsymbol{x}, \boldsymbol{y})$ will be considered.

This dissertation focuses on 2D digital images, where an intensity image I of l-levels ($l \in \mathbb{N}$) over $A \subset \mathbb{Z}^2$ is defined as $I : A \to \{0, \ldots, l-1\}$. For a point $\boldsymbol{x} \in A$, the pair $(\boldsymbol{x}, I(\boldsymbol{x}))$ is called a pixel. As most general instances, we will consider 8-bit ($l = 256$) intensity images of resolution $r \times c$ ($r, c \in \mathbb{N}$) $I : \{1, \ldots, c\} \times \{1, \ldots, r\} \to \{0, \ldots, 255\}$. Notice that this definition will allow us to refer to image point coordinates using the horizontal–vertical order. 24-bit RGB color images will be represented by a triplet (I_R, I_G, I_B) of 8-bit intensity images of the same resolution dedicated to the red, green and blue color channels, respectively. As another family, we will also consider binary images with $l = 2$.

1.2 Graph theory

In this dissertation, graphs will be considered in digital images, so we introduce the corresponding notations accordingly. A multigraph $\mathcal{G}$ is defined as a pair (V, E), where $V \subset \mathbb{Z}^2$ is a set of vertices, and $E \subseteq V \times V = \{(\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{u}, \boldsymbol{v} \in V\}$ is a multiset of edges between the vertices. We focus on undirected graphs, so $\forall \boldsymbol{u}, \boldsymbol{v} \in V : (\boldsymbol{u}, \boldsymbol{v}) \in E$ implies $(\boldsymbol{v}, \boldsymbol{u}) \in E$, and thus, we will write $\{\boldsymbol{u}, \boldsymbol{v}\}$ for the edges from now on. In our graphs, we allow loops (edges of type $\{\boldsymbol{u}, \boldsymbol{u}\}$) and multiple edges (more edges between two vertices), which also reasons why we consider the more general formalism of multigraphs. If there are no multiple edges, we will use the classic graph concept, where E is supposed to be a set instead of a multiset. The degree of a vertex is the number of edges containing the vertex. A walk is a list of vertices $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ with $\{\boldsymbol{u_1}, \boldsymbol{u_2}\}, \{\boldsymbol{u_2}, \boldsymbol{u_3}\}, \ldots, \{\boldsymbol{u_{n-1}}, \boldsymbol{u_n}\} \in E$, with $\boldsymbol{u_1} = \boldsymbol{u_n}$ in the case of a cycle (closed walk). $\mathcal{G}$ is connected, if any two of its vertices have a walk connecting them. A walk which includes every edge of $\mathcal{G}$ exactly once is called an Eulerian walk (or an Eulerian cycle, if the start and end vertices coincide). Notice that any Eulerian cycle is also an Eulerian walk. $\mathcal{G}$ is an Eulerian graph, if it has an Eulerian walk containing all of its edges. An Eulerian decomposition of $\mathcal{G}$ has the form $\mathcal{G} = \bigcup_{i=1}^{n} \mathcal{G}_i$ such that all the $\mathcal{G}_i$'s are edge-disjoint Eulerian graphs i.e. they do not contain common edges.

1.3 Information fusion

The results in the dissertation relating to information fusion are based on the consideration that merging the output of member components may lead to applications of higher accuracy. In our case, the member components are primarily image processing algorithms, whose outputs are merged for object detection purposes. Thus, besides introducing the general corresponding formalism, we also specialize our notation to this scenario.

As classic formulation [56], let $\mathcal{D}$ be a set (ensemble) of classifiers (voters) $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$, $\mathcal{D}_i : \Lambda \subseteq \mathbb{R}^m \to \mathbb{R}^M_{\geq 0}$ ($i = 1, \ldots, n$), and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ is a set of finite class labels.

The classifier $\mathcal{D}_i$ assigns the support values $\mathcal{D}_i(\boldsymbol{\chi}) = (d_{i,1}(\boldsymbol{\chi}), \ldots, d_{i,M}(\boldsymbol{\chi}))$ to a feature vector $\boldsymbol{\chi} \in \Lambda$ describing the opinion of the classifier on what degree $\boldsymbol{\chi}$ should be labeled by $\omega_1, \ldots, \omega_M$, respectively. Then, in a fusion-based scenario, the final class label is determined for $\boldsymbol{\chi}$ by applying some rule to the individual labels supported by the classifiers $\mathcal{D}_1, \ldots, \mathcal{D}_n$. Namely, as a general formulation, for each j $(j = 1, \ldots, M)$ a discriminator function $g_j(\boldsymbol{\chi})$ is calculated as

$$g_j(\boldsymbol{\chi}) = \mathcal{F}\left(d_{1,j}(\boldsymbol{\chi}), \ldots, d_{n,j}(\boldsymbol{\chi})\right), \tag{1.1}$$

where $\mathcal{F}$ is a combination function. According to the selection of the support values $d_{i,j}(\boldsymbol{\chi})$ and the combination function $\mathcal{F}$, we can set up several decision rules and derive different ensemble classifiers like the following algebraic ones:

$$\mathcal{D}_{avg}(\boldsymbol{\chi}) = \omega_k \iff g_k(\boldsymbol{\chi}) = \max_{j=1}^{M}\left(g_j(\boldsymbol{\chi}) = \frac{1}{n}\sum_{i=1}^{n} d_{i,j}(\boldsymbol{\chi})\right), \tag{1.2}$$

$$\mathcal{D}_{pro}(\boldsymbol{\chi}) = \omega_k \iff g_k(\boldsymbol{\chi}) = \max_{j=1}^{M}\left(g_j(\boldsymbol{\chi}) = \prod_{i=1}^{n} d_{i,j}(\boldsymbol{\chi})\right), \tag{1.3}$$

$$\mathcal{D}_{min}(\boldsymbol{\chi}) = \omega_k \iff g_k(\boldsymbol{\chi}) = \max_{j=1}^{M}\left(g_j(\boldsymbol{\chi}) = \min_{i=1}^{n}\left(d_{i,j}(\boldsymbol{\chi})\right)\right), \tag{1.4}$$

$$\mathcal{D}_{max}(\boldsymbol{\chi}) = \omega_k \iff g_k(\boldsymbol{\chi}) = \max_{j=1}^{M}\left(g_j(\boldsymbol{\chi}) = \max_{i=1}^{n}\left(d_{i,j}(\boldsymbol{\chi})\right)\right). \tag{1.5}$$

Notice that these properties give constraints on which class label ω_k should be selected for $\boldsymbol{\chi}$. Here, we apply the intuitive notation $\mathcal{D}(\boldsymbol{\chi}) = \omega_k$ instead of $\mathcal{D}(\boldsymbol{\chi}) = (\underbrace{0, \ldots, 0}_{k-1}, 1, \underbrace{0, \ldots, 0}_{M-k})$ with the only non-zero support 1 is put on the k-th label, since the overall aim for any ensemble classifier is to select a single class label as a final decision.

The simple majority voting based classic ensemble classifier can be derived by restricting the support of the individual classifiers with $d_{1,j}(\boldsymbol{\chi}) = 1$, if the classifier $\mathcal{D}_i$ labels $\boldsymbol{\chi}$ in the class ω_j and $d_{1,j}(\boldsymbol{\chi}) = 0$, otherwise. The final labeling of the ensemble is based on determining the class received the largest support in terms of the number of votes:

$$D_{maj}(\boldsymbol{\chi}) = \omega_k \iff g_k(\boldsymbol{\chi}) = \max_{j=1}^{M}\left(g_j(\boldsymbol{\chi}) = \sum_{i=1}^{n} d_{i,j}(\boldsymbol{\chi})\right). \tag{1.6}$$

From the simple majority voting model we can easily derive a weighted one with assigning weights $w_i \in \mathbb{R}_{\geq 0}$ to the classifiers $\mathcal{D}_i$ implying the following final decision rule:

$$D_{wmaj}(\boldsymbol{\chi}) = \omega_k \iff g_k(\boldsymbol{\chi}) = \max_{j=1}^{M}\left(g_j(\boldsymbol{\chi}) = \sum_{i=1}^{n} w_i d_{i,j}(\boldsymbol{\chi})\right). \tag{1.7}$$

In contrast to classic majority voting, here we consider each classifier output equipped with different weights w_i $(0 \leq w_i \leq 1,\ i = 1, \ldots, n)$. It seems natural to give the classifiers with higher accuracies greater importance in making the final decision. Notice that the classic majority voting scheme can be considered as a special case of the weighted one, since in the majority rule the weight of each vote given by a classifier is constrained to 1 i.e. $w_i = 1$ for all $i = 1, \ldots, n$.

Several results of the dissertation correspond to ensembles $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n\}$ of object detector algorithms $\mathcal{A}_i$ $(i = 1, \ldots, n)$. Based on the nature of the problem, we separate the cases when a single or multiple objects are to be detected in an image. Moreover, we focus on such objects that can be represented by single pixels with e.g. their centers. Thus, the object detector algorithms give their votes in terms of pixels as candidates for the location of the object. According to these considerations, we introduce the related notations as follows.

Let I be a digital image of size $r \times c$. A candidate extractor algorithm for a single object detection scenario is defined as $\dot{\mathcal{A}}_i : I \to P(\{1,\dots,c\} \times \{1,\dots,r\})$ with $i \in \{1,\dots,n\}$ and $\dot{\mathcal{A}}_i(I) = \{\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,1}}, \dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,2}}, \dots, \dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,k}}\}$, where $k \in \mathbb{N}$ and $P(A)$ denotes the power set of a set A. Notice that this definition allows a candidate extractor algorithm to give more candidates within an image for the possible location of a single object. However, we will investigate also such cases, when each ensemble member can have only one candidate for the location of the object. Thus, when with $k = 1$ algorithm $\dot{\mathcal{A}}_i$ has a single candidate, we will write $\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i}}$ instead of $\{\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,1}}\}$ for its candidate set.

If multiple objects may appear in the image, the definition of a candidate extractor algorithm is modified accordingly as $\ddot{\mathcal{A}}_i : I \to P(\{1,\dots,c\} \times \{1,\dots,r\})$ with $\ddot{\mathcal{A}}_i(I) = \{\ddot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,1}}, \ddot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,2}}, \dots, \ddot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,k}}\}$, where $k \in \mathbb{N}$. Note the difference between the notation of the candidates corresponding to the single and multiple objects detection scenarios. Namely, $\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,j}}$ for some $j \le k$ is the j-th guess of $\dot{\mathcal{A}}_i$ for a single object, while $\ddot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,j}}$ $(j \le k)$ predicts the appearance of a desired object in the corresponding location. Since the chapters of the dissertation separate the single and multiple object detection scenarios, we will use the simple notations $\mathcal{A}_i$ and $\boldsymbol{c}^{\boldsymbol{I}}_{\boldsymbol{i,\cdot}}$ for both cases if it does not hurt clarity. As a further simplification of notation, we will omit the symbol I from the upper index of candidates, when only one image is concerned and write $\boldsymbol{c}_{\boldsymbol{i,\cdot}}$ for short.

As from the candidate extractor algorithms $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ ensemble candidates are composed via the fusion of their candidate sets $\mathcal{A}(I) = \cup_{i=1}^{n} \mathcal{A}_i(I)$, we define a confidence measure to describe the rate of agreement of the members on the specific candidates. To do so, first we introduce a proximity relation $\cong$ to decide whether two candidates indicate the same object or not. With $\boldsymbol{c_1}, \boldsymbol{c_2} \in \{1,\dots,c\} \times \{1,\dots,r\}$, we say that $\boldsymbol{c_1} \cong \boldsymbol{c_2}$ if $d(\boldsymbol{c_1}, \boldsymbol{c_2}) < \mathcal{T}_d$ with some distance threshold $\mathcal{T}_d \in \mathbb{R}_{\ge 0}$. As in our applications the objects to be detected are circular, $\mathcal{T}_d$ can be selected as the diameter of the desired object. Now, the confidence of the ensemble on any of its candidates $\boldsymbol{c} \in \mathcal{A}(I)$ is defined as

$$conf_{\mathcal{A}}(\boldsymbol{c}) = |\{\mathcal{A}_i \in \mathcal{A} \ : \ \exists \boldsymbol{c'} \in \mathcal{A}_i(I) \text{ such that } \boldsymbol{c} \cong \boldsymbol{c'}\}| \ / \ |\mathcal{A}|. \tag{1.8}$$

Notice that $conf_{\mathcal{A}}(\boldsymbol{c}) \in \{k/|\mathcal{A}| \ : \ k = 1,\dots,|\mathcal{A}|\}$. We also classify the ensemble candidates based on the degree of confidence. Namely, the α–level candidates of $\mathcal{A}$ are defined as

$$\mathcal{A}(I)_\alpha = \{\boldsymbol{c} \in \mathcal{A}(I) \ : \ conf_{\mathcal{A}}(\boldsymbol{c}) \ge \alpha\}, \text{ where } 1/|\mathcal{A}| \le \alpha \le 1. \tag{1.9}$$

As specific cases, 1–level, α–level with $\alpha > 1/2$, and $1/|\mathcal{A}|$–level candidates are selected by each of, the majority of, and at least one of the members of the ensemble, respectively. For the latter case it should be noted that $\mathcal{A}(I) = \mathcal{A}(I)_{1/|\mathcal{A}|}$. Single algorithms are formally represented by ensembles consisting of one member providing $|\mathcal{A}| = 1$ and $\alpha = 1$–level confidence for all the candidates.

1.3.1 Error measurement

We also must set up a framework to measure the accuracy of ensemble-based approaches discussed in the dissertation for object detection in practical problems. We start with making the decision whether the candidates found by a specific member algorithm are true or false ones regarding some ground truth usually provided in terms of manual annotations.

Let $\mathcal{GT}(I) \subseteq \{1, \dots, c\} \times \{1, \dots, r\}$ be a so called ground truth set of candidates for an image I. For the classification of each candidate $\boldsymbol{c} \in \mathcal{A}(I)$ of an ensemble $\mathcal{A}$ in the same image regarding $\mathcal{GT}(I)$ and confidence level $1/|\mathcal{A}| \le \alpha \le 1$ we apply the following:

- $\boldsymbol{c}$ is an α–true positive (TP_α), if $\boldsymbol{c} \in \mathcal{A}(I)_\alpha$ and $\exists \boldsymbol{c_{gt}} \in \mathcal{GT}(I)$ such that $\boldsymbol{c} \cong \boldsymbol{c_{gt}}$;
- $\boldsymbol{c}$ is an α–false positive (FP_α), if $\boldsymbol{c} \in \mathcal{A}(I)_\alpha$ and $\nexists \boldsymbol{c_{gt}} \in \mathcal{GT}(I)$ such that $\boldsymbol{c} \cong \boldsymbol{c_{gt}}$.

Regarding a candidate $\boldsymbol{c_{gt}} \in \mathcal{GT}(I)$, we apply:

- $\boldsymbol{c_{gt}}$ is an α–false negative (FN_α), if $\nexists \boldsymbol{c} \in \mathcal{A}(I)_\alpha$ such that $\boldsymbol{c} \cong \boldsymbol{c_{gt}}$.

Finally,

- each point in $\{1, \ldots, c\} \times \{1, \ldots, r\} \backslash (\mathcal{GT}(I) \cup \mathcal{A}(I)_\alpha)$ is an α–true negative (TN_α).

The set of all true positives, false positives, false negatives, and true negatives for a given image I will be denoted by $TP(I)_\alpha, FP(I)_\alpha, FN(I)_\alpha$, and $TN(I)_\alpha$, respectively. Notice that $\mathcal{GT}(I)$ is usually a manually annotated set created by experts of the application field. Moreover, since performance evaluation is usually expected to be given at database level, digital images are often organized into an image database DB. Now, to calculate the performance of an ensemble $\mathcal{A}$ regarding some ground truth, we introduce the following classic measures at both image and image database level [57, 58]:

- Sensitivity

$$SEN(I)_\alpha = \frac{|TP(I)_\alpha|}{|TP(I)_\alpha| + |FP(I)_\alpha|}, \quad SEN(DB)_\alpha = \sum_{I \in DB} \frac{SEN(I)_\alpha}{|DB|}, \tag{1.10}$$

- Specificity

$$SPE(I)_\alpha = \frac{|TN(I)_\alpha|}{|FN(I)_\alpha| + |TN(I)_\alpha|}, \quad SPE(DB)_\alpha = \sum_{I \in DB} \frac{SPE(I)_\alpha}{|DB|}, \tag{1.11}$$

- False Positive Rate

$$FPR(I)_\alpha = 1 - SPE(I)_\alpha, \quad FPR(DB) = \sum_{I \in DB} \frac{FPR(I)_\alpha}{|DB|}, \tag{1.12}$$

- Positive Predictive Value

$$PPV(I)_\alpha = \frac{|TP(I)_\alpha|}{|TP(I)_\alpha| + |FN(I)_\alpha|}, \quad PPV(DB)_\alpha = \sum_{I \in DB} \frac{PPV(I)_\alpha}{|DB|}, \tag{1.13}$$

- F-Score

$$\textit{F-Score}(I)_\alpha = \frac{2SEN(I)PPV(I)}{SEN(I) + PPV(I)}, \quad \textit{F-Score}(DB)_\alpha = \sum_{I \in DB} \frac{\text{F-Score}(I)_\alpha}{|DB|}, \tag{1.14}$$

- Accuracy

$$ACC(I)_\alpha = \frac{|TP(I)_\alpha| + |TN(I)_\alpha|}{|TP(I)_\alpha| + |FN(I)_\alpha| + |TN(I)_\alpha| + |FP(I)_\alpha|}, \tag{1.15}$$

$$ACC(DB)_\alpha = \sum_{I \in DB} \frac{ACC(I)_\alpha}{|DB|}, \tag{1.16}$$

- Average False Positives per Image

$$FPI(DB)_\alpha = \sum_{I \in DB} \frac{FP(I)_\alpha}{|DB|}, \tag{1.17}$$

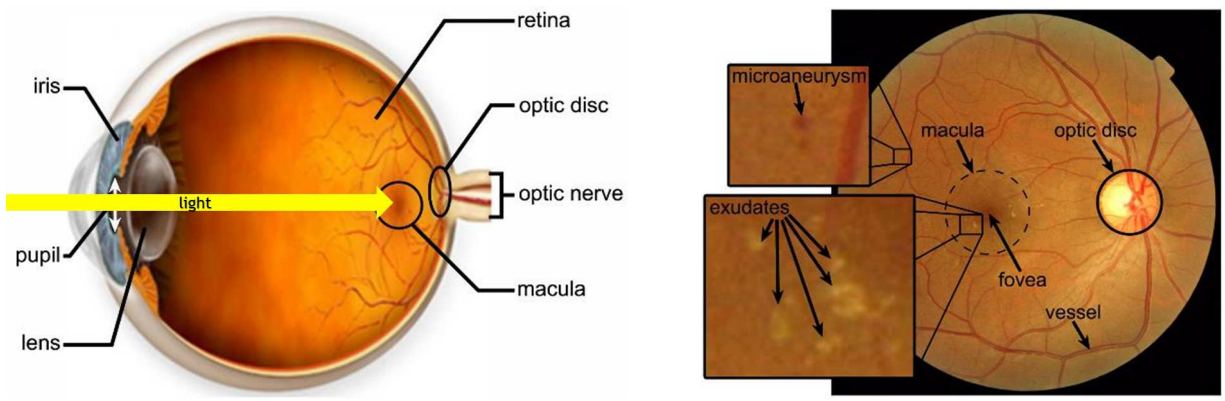
- Competition Performance Metrics

$$CPM(DB) = \sum_{g \in G} \frac{\{SEN(DB)_\alpha : FPI(DB)_\alpha = g\}}{|G|}, \text{ where } G = \left\{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8\right\}. \tag{1.18}$$

As a specific case, we will omit the index α in the footnote of the above performance measures, if $|\mathcal{A}| = 1$ and simply write e.g. $SEN(I)$ instead of $SEN(I)_{1/|\mathcal{A}|}$, when our aim is to evaluate a single algorithm only.

1.4 Clinical concepts

In this section, we briefly describe those concepts of a clinical field that will be regularly referred in some chapters of the dissertation. Namely, some of our single and multiple objects detection results relate to some anatomical parts and diabetes specific lesions appearing in the human retina (fundus). In Figure 1.1(a), we can see the anatomic location of the retina as a layer in the eye dedicated to percept the light in the human vision system. The retinal anatomical parts and lesions are shown in Figure 1.1(b), whose detections are addressed in the dissertation. Namely, as single object detection problems, we will give results according to the localization of the optic disc (OD) and the macula. The localization of such lesions as microaneurysms (MAs) and exudates belong to the family of multiple objects detection problems, since usually several such lesions appear in a diseased case. More specific introductions regarding the detections of these objects will be given in the corresponding chapters. The region of interest (ROI) concept in the case of a retinal image generally corresponds to the useful content, that is, the image without its black background.



(a) (b)

Figure 1.1: *Basic concepts of retinal image analysis; (a) the structure of the human eye and the location of the retina, (b) anatomical parts and diabetes related lesions of the retina.*

1.5 Databases

In this section, we list those databases and give the corresponding characteristics of images belonging to them which will be considered for training and evaluation in several chapters of the dissertation. Some other datasets that are used only in one chapter are going to be introduced there. The databases listed here relate to the field of retinal image analysis.

1.5.1 Retinopathy Online Challenge (ROC) database

Retinopathy Online Challenge (ROC) [58] is a worldwide online competition dedicated to measure the accuracy of microaneurysm (MA) detectors. The ROC database consists of 50 training and 50 test images having different resolutions (768×576, 1058×1061 and 1389×1383 pixels), 45° field-of-view (FOV) and JPEG compression. The average number of MAs in an image for the training and test sets are 6.72 and 6.86, respectively. There are 13 and 10 images of the training and test sets, respectively, where no MAs are marked by the clinical experts.

1.5.2 DiaretDB0 database

The DiaretDB0 [59] database contains 130 losslessly compressed color fundus (retinal) images with a resolution of 1500×1152 pixels and 50° FOV. Among the images, 20 are normal and 110 contain signs of diabetic retinopathy (DR), like hard exudates, soft exudates, micronaneuyrysms, hemorrhages and neovascularization. For every fundus image, a corresponding ground truth file is available containing the OD/macula centers and all lesions appearing in the specific retinal image.

1.5.3 DiaretDB1 database

The DiaretDB1 v2.1 [60] database contains 28 losslessly compressed training and 61 test images, respectively with a resolution of 1500×1152 pixels and 50° FOV. As for ground truth an expert in ophthalmology marked the OD/macula centers and the areas related to microaneurysms, hemorrhages, and hard/soft exudates. The average number of MAs for the training and test sets are 4.34 and 3.91, respectively. There are 15 and 39 images of the training and test sets, respectively, where no MAs are marked by the experts.

1.5.4 Database provided by the Moorfields Eye Hospital, London

This database consists of 327 losslessly compressed images with resolutions of 3072×2048 and 1360×1024 pixels and 50° FOV for which manually drawn regions of the OD are also available. For testing MA detection performance, 60 from the images were annotated accordingly by an expert in ophthalmology. The average number of MAs for the training and test sets are 8.67 and 8.87, respectively. There are 10 and 8 images in the training and test sets, respectively, where no MAs are marked by the experts.

1.5.5 Messidor database

This database consists of 1200 losslessly compressed 24-bits RGB images with 45° FOV at different resolutions (1440×960, 2240×1488 and 2304×1536 pixels). For each image, a grading score ranging from R0 to R3 is also provided. These grades correspond to the following clinical conditions: a patient with a grade R0 has no DR. R1 and R2 are mild and severe cases of non-proliferative retinopathy, respectively. Finally, R3 is the most serious condition (proliferative retinopathy). The grading is based on the appearance of MAs, haemorrhages and neovascularization. The corresponding proportion of the images in the Messidor dataset: 540 R0 (46%), 153 R1 (12.75%), 247 R2 (20.58%) and 260 R3 (21.67%). This database is kindly provided by the Messidor program partners[1].

[1] Available at: `http://messidor.crihan.fr`.

1.5.6 DRIVE database

The database DRIVE [61] contains 40 JPEG-compressed color fundus images of resolution 768×584 pixels and $45°$ FOV. The set of 40 images has been divided into a training and a test set, both containing 20 images. For the training images, a single manual segmentation of the vessel system is available. For the test cases, two manual segmentations are available; one is used as gold standard, the other one can be used to compare computer generated segmentations with those of an independent human observer.

1.6 Algorithms to create ensembles

In this section, we collect those state-of-the-art image preprocessing and object detector algorithms that will be considered in more chapters of the dissertation. Those algorithms that are cited only in one chapter are going to be introduced there. The algorithms listed here relate to retinal image analysis with the main motivation to form member algorithms by the help of theirs for our detector ensembles. Thus, they have been collected not only based on their reported performances, but also based on the consideration to be built upon different principles; this behavior can be expected to have positive effect to compose efficient ensembles from more diverse members.

1.6.1 Preprocessing methods

In image processing problems, preprocessing algorithms are applied to enhance some characteristics in the image to support further extraction steps. These algorithms transform the original image I into an image I' having the same size. In object detection problems, only such transforms are allowed that do not modify the spatial locations of the desired objects. However, beyond this intuitive definition, it is meaningless to introduce a rigorous definition for preprocessing algorithms, since no clear rule can be given to decide whether objects remain detectable afterwards.

No preprocessing (PP_1)

We consider the results of the candidate extractors obtained for the original images without any preprocessing. That is, we formally consider a "No preprocessing" operation, as well.

Walter-Klein contrast enhancement [62] (PP_2)

This preprocessing method aims to enhance the contrast of fundus images by applying a gray level transformation in the following way:

$$I'(\boldsymbol{x}) = \begin{cases} \dfrac{(I'_{max} - I'_{min})\,(I(\boldsymbol{x}) - I_{min})^s}{2\,(\mu - I_{min})^s} + I'_{min}, & \text{if } I(\boldsymbol{x}) \leq \mu, \\ \dfrac{(I'_{max} - I'_{min})\,(I_{max} - I(\boldsymbol{x}))^s}{2\,(I_{max} - \mu)^s} + I'_{max}, & \text{otherwise,} \end{cases} \tag{1.19}$$

where I_{min}, I_{max} and I'_{min}, I'_{max} are the minimal and maximal intensity levels of the original and the enhanced images, respectively, μ is the mean value of the original intensity image and $s \in \mathbb{R}$ is a transition parameter.

Contrast limited adaptive histogram equalization [63] (PP_3)

Contrast limited adaptive histogram equalization (CLAHE) is a popular technique in biomedical image processing, since it is very effective in making the salient parts more visible. The image is

split into disjoint regions and in each region a local histogram equalization is applied. Then, the boundaries between the regions are eliminated with a bilinear interpolation.

Vessel removal and extrapolation [64] (PP_4)

We investigate the effect of processing images with the complete vessel system being removed based on the idea proposed in [64]. We extrapolate the missing parts to fill in the holes caused by the removal using the inpainting algorithm presented in [65]. MAs appearing near vessels become more easily detectable in this way.

Illumination equalization [66] (PP_5)

This preprocessing method aims to reduce the vignetting effect caused by non-uniform illumination of retinal images. Each pixel intensity is set as $I'(\boldsymbol{x}) = I(\boldsymbol{x}) + \mu_d - \mu_l$, where μ_d is the desired average intensity and μ_l is the local average intensity. Small dark objects (like MAs) appearing on the border of the retina are enhanced by this step.

Gray-world normalization [66] (PP_6)

Each pixel on the green channel of the image is transformed via $I'(\boldsymbol{x}) = I(\boldsymbol{x})/\mu$, where μ is the average intensity of the green channel I_G.

Intensity adjustment [67] (PP_7)

This preprocessing method enhances the contrast of a grayscale image by saturating the lowest and highest 1% of the intensity values.

Background subtraction of retinal blood vessels [68] (PP_8)

This method is recommended for vascular system detection. Blood vessels on retinal images show similar local appearance to MAs. This approach considers the vessel system as the foreground of the image. The background is extracted by applying an averaging filter, which is followed by threshold averaging for smoothing. The background image is then subtracted from the original one.

Histogram equalization (HE) [66] (PP_9)

This preprocessing method aims to enhance the global contrast of the image by redistributing the intensity values of the image. First, the accumulated normalized histogram of the image is created. Then, the histogram is transformed to reflect uniform distribution.

1.6.2 Single object (optic disc) candidate extractors

In the single object detection based scenarios discussed in the dissertation, the corresponding detector algorithms selects one point in the image as the center of the desired object. However, we will also let these algorithms to extract more candidates for the same object as discussed previously. The corresponding algorithms listed here will aim to detect the optic disc (OD) in retinal images.

Lalonde et al. [69] (OD_1)

This algorithm generates a pyramid with simple Haar-based discrete wavelet transform. The pixel with the highest intensity value in the low-resolution image (4th or 5th level of decomposition) is considered as the center of the OD.

Lalonde et al. 2 [69] (OD_2)

This method uses an edge detection algorithm which is based on Rayleight-based CFAR threshold. Next, Hausdorff distance is calculated between the set of edge points and a circular template like the average OD. The pixel with the lowest distance value is selected for OD center.

Sopharak et al. [70] (OD_3)

This algorithm applies a median and a CLAHE filter on the retinal image. In a neighborhood of each pixel, the entropy of intensity is calculated; the pixel with the largest entropy value is selected as the OD center.

Niemeijer et al. [71] (OD_4)

In this approach, specific vessel features are extracted (number, width, orientation and density of vessels and their combination), and a kNN classifier is applied to decide whether a pixel belongs to the OD region. The centroid of the largest extracted component is considered as the OD center.

Ravishankar et al. [64] (OD_5)

This algorithm was proposed to fit lines to the thinned vessel system by Hough transformation. The intersection of these lines results in a so-called probability map. A weighting is also applied considering the intensity values corresponding to the intersection points. The pixel having the highest probability is considered as the OD center.

Hoover et al. [72] (OD_6)

This method thins the vessel system and models each line-shape segment with a fuzzy segment. A voting map of these fuzzy segments is created and the pixel receiving the most votes is considered as the OD center.

Zhu et al. [73] (OD_7)

This approach locates the border of the OD in terms of a circle with a given diameter using the circular Hough transformation. For this aim, edge detection is applied and the circle containing the most edge points is selected. The center of this circle is considered to be the OD center.

1.6.3 Single object (macula) candidate extractors

Similarly to OD detection, we list some algorithms dedicated to the detection of the macula, which is the center of the sharp vision in the retina. The macula can be represented by its center (the fovea) as a single point (see also Figure 1.1(b)).

Petsatodis et al. [74] (MAC_1)

In this approach, a region of interest (ROI) is defined to process macula detection. A Gaussian low-pass filter is applied to smooth the image. The statistical mean and standard deviation of the ROI area is used to compute a threshold for segmentation to get binary objects. The object that is located nearest to the center of the ROI is labeled as the macula. Its center of mass is considered to be the center of the macula.

Sekhar et al. [75] (MAC_2)

Here, a region of interest (ROI) for the macula is defined by means of its spatial relationship with the OD. That is, the portion of a sector is subtended at the center of the OD by an angle of 30° above and below the line between this center and the center of the image. The macula is identified within this ROI by iteratively applying a threshold, and then performing a morphological opening on the resulting blob. The fovea is simply determined as the centroid of this blob.

Fleming et al. [76] (MAC_3)

This algorithm is proposed to identify the macular region based on the information of the temporal arcade (thickest vessel of the retina) and the OD center. First, the arcade is found by using semi-elliptical templates having a range of sizes, orientations and eccentricities. Next, the OD is detected by using a Hough transformation with circular templates having diameters from 70% to 125% of the average OD diameter. Finally, the fovea is detected by finding the maximum correlation coefficient between the image and a foveal model.

Zana et al. [77] (MAC_4)

This is a region merging algorithm based on watershed cell decomposition and morphological operations for macula detection. After noise removal, morphological closing followed by opening is performed to remove the small dark holes and white spots. A watershed based decomposition of the gradient image into cells is done, and the cell with darkest gray level inside the macula is selected as the first step of a merging algorithm. A complex criterion based on the gray values and edges of the filtered image is calculated to merge the cells of the macula while rejecting perifoveal inter-capillary zones in order to produce the contour of the macula.

Antal et al. [36] (MAC_5)

Besides the algorithms discussed so far, we have also tested an own macula detector [36] and used its results in our combined system. First, we extract the green plane from the color fundus image. Then, we generate the background image by applying a median filter and subtract it from the green plane, resulting in a shade corrected image. Next, we binarize the image by considering all-non zero pixels as foreground pixels and others as background ones. Finally, we apply an image labeling procedure and select the largest component as the macula.

1.6.4 Multiple objects candidate extractors

As for the multiple objects detection scenario based on ensemble-based methods, we will consider the detection of MAs in retinal images. Individual MA detectors consider different principles to extract MA candidates with the aim to detect any objects in the image showing MA-like characteristics.

Walter et al. [78] (MA_1)

Candidate extraction is accomplished by grayscale diameter closing. That is, this method aims to find all sufficiently small dark patterns on the green channel. Finally, a double threshold is applied.

Spencer et al. [79] (MA_2)

From the input fundus image, the vascular map is extracted by applying twelve morphological top-hat transformations with twelve rotated linear structuring elements (with a radial resolution of 15 °). Then, the vascular map is subtracted from the input image, which is followed by the application of a Gaussian matched filter. The resulting image is then binarized with a fixed threshold. Since the extracted candidates are not precise representations of the actual lesions, a region growing step is also applied to them. While the original paper [79] aims to detect MAs on fluorescein angiographic images, our implementation was based on the modified version published by Fleming et al. [80].

Circular Hough transformation [81] (MA_3)

Following the idea presented in [81], we established an approach based on the detection of small circular spots in the image. Candidates are obtained by detecting circles in the images using circular Hough transformation. With this technique, a set of circular objects can be extracted from the image.

Zhang et al. [82] (MA_4)

In order to extract candidates, this method constructs a maximal correlation response for the input retinal image. This is accomplished by considering the maximal correlation coefficient with five Gaussian masks with different standard deviations for each pixel. The maximal correlation response image is thresholded to obtain the candidates. Vessel detection and region growing is applied to reduce the number of candidates and to determine their precise size, respectively.

Lázár et al. [37] (MA_5)

In this own detector [37], pixel-wise cross-section profiles with multiple orientations are used to construct a multi-directional height map. This map assigns a set of height values that describe the distinction of the pixel from its surroundings in a particular direction. In a modified multilevel attribute opening step, a score map is constructed from which the MAs are extracted by thresholding.

Chapter 2

Optimal approach for fast object-template matching

In this chapter, we propose a novel way to simplify object description for faster matching. Our approach is based on centroidal Voronoi tessellation methods and can be applied to any dimensions. Furthermore, we show that it is optimal for chamfer matching purposes. We prove some fundamental statements regarding the proposed approach and recommend corresponding algorithms to obtain the optimal set of points. The results of this chapter have been published in [1, 17–19] and are also incorporated in the project SHARE: Mobile Support for Rescue Forces, Integrating Multiple Modes of Interaction, EU FP6 Information Society Technologies, Contract Number FP6-004218.

Object detection and classification have been challenging problems in digital image analysis and computer vision. Usually, some image features (e.g. edges) are extracted and then a matching procedure is applied to find the occurrences of predefined templates. Chamfer matching is one of these methods [83]. Here, usually an edge detection step is applied first to obtain a binary image, for which a distance map is calculated [84, 85]. If there are no additional possibilities (e.g. background subtraction [86, 87], intensity-based presegmentation [88]) to restrict the search range, a scanning step is applied to the entire image to match predefined templates, like in [84, 89].

Chamfer matching is a popular method in real-time object matching applications. Several supplementary approaches have been recommended to speed up its computation. One of them is to simplify the description of the objects to be matched by representing them with a smaller number of points. This can be achieved by a coarse-to-fine approach [90] regarding object resolution, or by selecting some significant points on the object boundary. For the latter case, a natural consideration was proposed in [91], by selecting points with high curvature (e.g. corners). However, actual optimization of such a selection was not given.

In this chapter, we propose a novel way of simplifying the object description for faster matching. Our approach, which is based on centroidal Voronoi tessellation (CVT) methods, can be applied to any dimensions [92, 93]. Furthermore, we show that it is optimal for chamfer matching purposes. We prove some fundamental statements regarding the proposed approach and recommend corresponding algorithms to obtain the optimal set of points.

In one of our human person detection/tracking/recognition systems we use snakes to describe the target objects to be recognized. Recognition is performed by matching contour templates from a database representing whole objects or object parts. Consequently, we shall present our experimental studies for the classic application setup of chamfer matching, when contour object descriptions are matched. Moreover, we shortly deal with the natural extension of our approach to any other kind of object representation. Namely, besides the extension of the naive simplification to region-like objects, we present how to focus more on some object regions during their simplification with defining a specific weight function accordingly. For example, larger uncertainty closer to the boundary of the object can be compensated by a weight function concentrating on the skeleton.

Our simplification approach can be applied in two ways in the matching process. On the one hand, the database templates can be simplified for faster performance. On the other hand, we can consider a simpler description of the target objects. For example, in the case of snakes, we can iterate with a smaller number of snake points. Furthermore, we investigate how the matching accuracy drops, when the database templates are simplified and also how the number of snake points affects the reliability of the matching. We also compare our technique with naive representation reduction approaches.

The structure of the chapter is as follows. In section 2.1, we describe chamfer matching. Then, section 2.2 explains how the objects to be matched can be simplified optimally for chamfer matching based on a CVT approach. We also present efficient algorithms for this simplification. Section 2.3 contains our experimental results on the proposed simplification. Here, we also explain how snake representation speeds up chamfer matching techniques and illustrate the efficiency of the simplification in a region-based object matching approach. We also present how sampling strategies can be exploited in a learning-based object detection scenario.

2.1 Chamfer matching

Chamfer matching was proposed in [83] for measuring the distance between binary images. While Hamming or other distances are used on exactly matched pixel pairs only, chamfer matching is

a more robust approach for a less accurate matching. Let S be a binary image ($S \subseteq \mathbb{Z}^m$), and $T = \{\boldsymbol{t}_i \in \mathbb{Z}^m \mid i = 1, \ldots, K;\ m \in \mathbb{N};\ \boldsymbol{t}_1 = \mathbf{O}\}$ be a binary template consisting of K ordered points starting at the origin $\mathbf{O}$ of $\mathbb{Z}^m$. The distance map of S is defined as $d_S : \mathbb{Z}^m \to \mathbb{R}_{\geq 0}$, such that, for any $\boldsymbol{x} \in \mathbb{Z}^m$, $d_S(\boldsymbol{x})$ is the distance of $\boldsymbol{x}$ and the closest element of S. By definition, $d_S(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in S$. To create distance maps, either the Euclidean distance [94, 95], or city-block/chessboard/chamfer distances [96, 97] can be used for faster implementation. Moreover, we can find robust variants of distance maps that weigh distance values, like the truncated linear/quadratic [89, 98], or the exponentially decreasing one [99] in order to concentrate more on the regions closer to the edges. As introduced in section 1.1, d will refer to the Euclidean distance function throughout the chapter.

When the distance map d_S is obtained, a natural way to determine how good the fit is between S and T at $\boldsymbol{x} \in \mathbb{Z}^m$ is to consider the chamfer distance

$$d_{\boldsymbol{x}}(S, T) = \frac{1}{K} \sum_{i=1}^{K} d_S(\boldsymbol{x} + \boldsymbol{t}_i). \tag{2.1}$$

Obviously, smaller $d_{\boldsymbol{x}}$ values correspond to better fit between T and S at $\boldsymbol{x}$. To get an impression of the basic steps of chamfer matching, see Figure 2.1.

(a) (b) (c) (d)

Figure 2.1: *Basic steps of chamfer matching; (a) extracted binary contour S, (b) the distance map d_S of S, (c) a template T (magnified) to be matched against S, (d) finding the best matching position of T on d_S.*

Besides the arithmetic mean in (2.1), one can use the root mean square average, the median or the maximum of the corresponding distance values, as well [100]. Moreover, to suppress the influence of outliers (due to noise) or missing data (due to occlusion or segmentation errors), the α-trimmed mean (where the large distance values are excluded from the summation) or the truncated mean (where a global threshold is applied for the distance map) can be employed [89, 100, 101]. Note that the chamfer distance is in close relation to the Hausdorff distance [89]. In our experiments, we considered the 2D case, and used the $\langle 3, 4 \rangle$ distance map proposed in [97], which is fast to generate. However, as a comparative analysis shows, other distance maps lead basically to the same results with our approach.

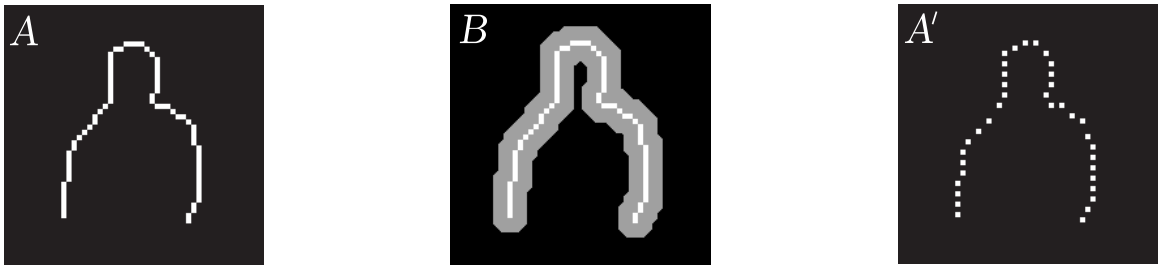
Usually, the target object goes under some geometric transformation (e.g. translation, rotation, scaling) in the image. Thus, in the worst case, we have to search for the optimal parameters for translation, rotation, magnification, that is, for an optimal affine transformation. A direct search for all these transformation parameters is very time-consuming. Thus, algorithmic speed ups have been proposed in the literature. Borgefors [90] suggested a hierarchical edge matching based on a "coarse to fine" spatial resolution approach. However, Huttenlocher et al. [89] warned that this method has some risk of losing information when reducing resolution. They suggested searching the transformation space using a cell decomposition (divide and conquer) strategy [101]. This method is based on the fact that if a wrong match is found for a given transformation, then similar transformations (being close in the parameter space) can be excluded, as well. This idea has become popular in applications. However, as it still needs quite a lot of computations

or parallelization, the transformation space is often reduced, like e.g. in [84], using an a priori estimation of the scale parameter.

2.2 Reducing the number of object points involved in chamfer matching

As mentioned above, there are two ways to reduce the number of points involved in the chamfer matching procedure. We can adjust both the density of the points describing the target objects and the database templates. The database templates can be reduced in a preprocessing step, as it is independent of the feature extraction procedure. We should remove points from the template in an optimal way, with respect to the distance measure used. To set up our optimality criterion we also note that in chamfer matching, the target object and the template to be matched can be exchanged. In other words, we follow the same procedure when matching either the template against the target object or the target object against the template. Thus, any optimization approach for the target object must also be valid for the template, and vice versa. It is already clear from our previous discussions that, if the target object description is simplified, the generated distance map should remain similar to the original one. Now, by the invertibility of the matching, to simplify the template, we should keep those points that generate a distance map, which is "close" to the one generated by the original template. Similar goals are set in [91], but no actual optimization is given.

As it was discussed in section 2.1, to solve this simplification problem, there must be a search region defined within which the distance maps are compared with each other. An obvious search region can be a bounding box or some dilated version of the original template. More precisely, we solve the following problem: reduce a discrete set $A \subseteq B$ with cardinality $|B| = M$ to A' with $A' \subseteq A$, and $|A'| = K \leq M$, so that the distance map generated by A' is closest to the distance map generated by A within B; Figure 2.2 shows an example.



(a) (b) (c)

Figure 2.2: *Simplification of an object for fast matching by keeping its distance map close to the original one; (a) object A to be simplified, (b) the region B (in gray) within which the distance map should be preserved, (c) the simplified object A'.*

Distance maps can be compared simply by summing up the differences between their corresponding values within B. The following lemma shows that this comparison can also be performed by comparing the sum of the values of the distance maps generated by the reduced variants of A.

Lemma 2.2.1. *The discrete set A' is an optimal reduction of A, if we have*

$$\sum_{\boldsymbol{x} \in B} d_{A'}(\boldsymbol{x}) = \min_{\substack{A'' \subseteq A \\ |A''| = K}} \left(\sum_{\boldsymbol{x} \in B} d_{A''}(\boldsymbol{x}) \right). \tag{2.2}$$

Proof. If we remove points from A, the corresponding distance values in the distance map generated by the remaining subset will be greater than in d_A. Thus for any $A'' \subseteq A$ with $|A''| = K$,

$$\sum_{\boldsymbol{x}\in B} |d_{A''}(\boldsymbol{x}) - d_A(\boldsymbol{x})| = \sum_{\boldsymbol{x}\in B} d_{A''}(\boldsymbol{x}) - \sum_{\boldsymbol{x}\in B} d_A(\boldsymbol{x}) \geq \sum_{\boldsymbol{x}\in B} d_{A'}(\boldsymbol{x}) - \sum_{\boldsymbol{x}\in B} d_A(\boldsymbol{x}) \tag{2.3}$$

holds, if A' is defined according to (2.2). □

The solution of the previous problem could obviously be found by a very expensive "brute force" algorithm, by checking $\binom{N}{K}$ cases, assuming that $|A| = N$. To find a more robust solution for this problem, let us re-formulate it in the Euclidean space. For a more general setup, we also define a weight function ϱ on B, to keep the relative importance of the different parts of A adjustable for the simplification process. Such weight functions can be probability kernels, like the uniform or Gaussian ones, or can be adjusted to the shape e.g. with concentrating to its skeleton. As the distance values within a distance map are based on the closest points of the generator object, the problem can be interpreted in the continuous case using the elements of the Voronoi tessellation framework.

Let $A \subseteq B \subseteq \mathbb{R}^m$, such that A is compact and convex and B is bounded. Moreover, let $K \in \mathbb{N}$ and $\varrho : B \to \mathbb{R}_{\geq 0}$. Find the set of points $A' = \{\boldsymbol{y}_i \in A \mid i = 1, \ldots, K\}$, which minimizes

$$\sum_{i=1}^{K} \int_{V_i(A')} \varrho(\boldsymbol{y}) |\boldsymbol{y} - \boldsymbol{y}_i|^2 d\boldsymbol{y}, \tag{2.4}$$

where $\{V_i(A')\}_{i=1}^{K}$ denotes the Voronoi tessellation of B generated by A'.

Though it is already suggested in Lemma 2.2.1, the subsequent construction of the solution to problem (2.4) will clarify the equivalence of the two problems above. For simplicity, A^K will denote the Cartesian product $\underbrace{A \times \cdots \times A}_{K\ times}$. The following lemma explains why A needs to be convex and compact, while B must obviously be bounded, so that the integral has a finite value in (2.4).

Lemma 2.2.2. *Let $A \subseteq \mathbb{R}^m$ be compact and convex, and let $\boldsymbol{x} \in \mathbb{R}^m$ be arbitrarily chosen. Then, a unique $\boldsymbol{y} \in A$ exists, such that $d(\boldsymbol{x}, \boldsymbol{y}) = \min_{\boldsymbol{z}\in A} d(\boldsymbol{x}, \boldsymbol{z})$.*

Proof. The compactness (closedness) of A obviously implies the existence of such a $\boldsymbol{y} \in A$. Now, to show that $\boldsymbol{y}$ is unique, let us assume that $\boldsymbol{y}_1, \boldsymbol{y}_2 \in A$, $\boldsymbol{y}_1 \neq \boldsymbol{y}_2$ exist such that $d(\boldsymbol{x}, \boldsymbol{y}_1) = d(\boldsymbol{x}, \boldsymbol{y}_2) = \min_{\boldsymbol{z}\in A} d(\boldsymbol{x}, \boldsymbol{z})$. Now, as $d(\boldsymbol{x}, \frac{\boldsymbol{y}_1+\boldsymbol{y}_2}{2}) < d(\boldsymbol{x}, \boldsymbol{y}_1)$, and the convexity of A gives $\frac{\boldsymbol{y}_1+\boldsymbol{y}_2}{2} \in A$, we have a contradiction. □

We note here that problem (2.4) has been thoroughly investigated when $A = B$. It has been shown that the optimal solution should define a centroidal Voronoi tessellation (CVT) on A [92], which means that the generators of the Voronoi cells are also their mass centers (centroids), respectively. Such a distribution of points can be achieved using CVT algorithms [92, 102–104]. The CVT approach together with the generating algorithms were extended to the tessellation of surfaces by considering constrained CVT (CCVT) [93]. In this chapter, we consider the case $A \subset B$. Though it has some aspects in common with the existing CCVT techniques, it is a different problem. We will call our approach region-influenced centroidal Voronoi tessellation, or shortly RCVT. Several theoretical results for CVT and CCVT can be shown to hold for RCVT, as well.

We begin by defining the same distortion functional as in [93, 105] for the pair (Y, Z) with $Y, Z \in A^K$ as

$$\mathcal{E}(Y, Z) = \sum_{i=1}^{K} \int_{V_i(Y)} \varrho(\boldsymbol{y})|\boldsymbol{y} - \boldsymbol{z}_i|^2 d\boldsymbol{y}, \tag{2.5}$$

with a density function ϱ defined over B. Let us now consider a modification of the fixed point iteration of the so called Lloyd map [92]. Namely, for a set of Voronoi generators $Z \in A^K$, let $T : A^K \to A^K$, with $T(Z) = Z'$, such that $d(\boldsymbol{z}_i', \boldsymbol{z}_i^*) = \min_{\boldsymbol{a} \in A} d(\boldsymbol{a}, \boldsymbol{z}_i^*)$, for $i = 1, \ldots, K$, where

$$\boldsymbol{z}_i^* = \frac{\int_{V_i(Z)} \boldsymbol{z}\varrho(\boldsymbol{z})d\boldsymbol{z}}{\int_{V_i(Z)} \varrho(\boldsymbol{z})d\boldsymbol{z}}. \tag{2.6}$$

That is, $Z \in B^K$ is the set of mass centers of the corresponding Voronoi cells $\{V_i(Z)\}_{i=1}^K$. In other words, the mapping T moves the current generators Z to those points Z' of A which are closest to the corresponding mass centers Z^*, respectively. The existence of such points Z' is guaranteed by Lemma 2.2.2. We shall show that the optimally selected points of A regarding problem (2.4) must be the ones closest to the centroids of the generated Voronoi tessellation. The generators of such a tessellation obviously form a fixed point of the mapping T, which (as we shall show) can be reached via a fixed point iteration of T

$$Z_n = T(Z_{n-1}), \quad n \geq 1. \tag{2.7}$$

The above iteration process can be interpreted as a modification of the Lloyd algorithm [103], like in the CCVT case [105]. Thus we can summarize our proposed algorithm in the following way.

Algorithm 2.2.3. Lloyd algorithm for computing RCVT.

Input:

$B \subseteq \mathbb{R}^m$, selected search region;
$A \subseteq \mathbb{R}^m$, the set to be simplified (with $A \subseteq B$);
$K \in \mathbb{N}$, the number of generators;
ϱ, a density function over B;
$Z \in A^K$, an initial set of generators;

Output:

$\{V_i\}_{i=1}^K$, a RCVT with K generators $Z \in A^K$;

Iteration:

1. *Construct the Voronoi tessellation $\{V_i\}_{i=1}^K$ of B with generators $Z \in A^K$;*
2. *Define the new set of generators as the points of A closest to the centroids of $\{V_i\}_{i=1}^K$;*
3. *Repeat steps 1 and 2 until some terminating criterion is met.*

Now we show that some basic properties of CVT [105] remain valid also for RCVT.

Lemma 2.2.4. *Let ϱ be a positive and smooth density function defined on a smooth bounded domain B. Then:*

1. *$\mathcal{E}$ is continuous and differentiable in $\overline{B}^K \times \overline{B}^K$ (where $\overline{B}$ stands for the closure of B);*

2. $\mathcal{E}(Z, T(Z)) = \min\limits_{Y \in \overline{B}^K} \mathcal{E}(Z, Y)$;

3. $\mathcal{E}(Z, Z) = \min\limits_{Y \in \overline{B}^K} \mathcal{E}(Y, Z)$.

Proof. The first statement of Lemma 2.2.4 can be proven in a similar way as in the CVT case [105].

The proof of the second statement is based on the fact that the distortion value $\mathcal{E}$ increases with shoving off the mass centers of the tessellation $\{V_i(Z)\}_{i=1}^K$. Let us fix an $i \in \{1, \dots, K\}$. For simpler calculations, without loss of generality, we may assume that $\boldsymbol{z}_i^* = \mathbf{O}$. Let us choose now $\boldsymbol{z}_i', \boldsymbol{z}_i'' \in V_i(Z)$ such that $|\boldsymbol{z}_i'| \leq |\boldsymbol{z}_i''|$. Then,

$$\int\limits_{V_i(Z)} \varrho(\boldsymbol{y})|\boldsymbol{y} - \boldsymbol{z}_i'|^2 d\boldsymbol{y} - \int\limits_{V_i(Z)} \varrho(\boldsymbol{y})|\boldsymbol{y} - \boldsymbol{z}_i''|^2 d\boldsymbol{y} = (|\boldsymbol{z}_i'|^2 - |\boldsymbol{z}_i''|^2) \int\limits_{V_i(Z)} \varrho(\boldsymbol{y}) d\boldsymbol{y} \leq 0. \tag{2.8}$$

Thus, for any $Z' = \{\boldsymbol{z}_i'\}_{i=1}^K$, $Z'' = \{\boldsymbol{z}_i''\}_{i=1}^K \in A^K$ with $|\boldsymbol{z}_i'| \leq |\boldsymbol{z}_i''|$ for all $i \in \{1, \dots, K\}$, we have $\mathcal{E}(Z, Z') \leq \mathcal{E}(Z, Z'')$, which completes the proof of the second statement of the lemma.

Finally, notice that the third statement of the lemma has already been proven for the CVT case in [105] without the extra restriction $T_i(Z) \in A$. □

Given these basic properties, we can formulate the important monotone decreasing behavior of the distortion functional $\mathcal{E}$ for the proposed iterative algorithm.

Lemma 2.2.5. *Let $\{Z_n\}_{n=1}^\infty$ be the sequence of generating sets produced by Algorithm 2.2.3. Then, we have:*

1. $Z_n = T(Z_{n-1})$;

2. $\mathcal{E}(Z_n, Z_n) \leq \mathcal{E}(Z_{n-1}, Z_{n-1})$.

Proof. The first statement of the lemma is obvious, as it is the formal description of how Algorithm 2.2.3 operates through T.

As to the second statement, by Lemma 2.2.4, we have

$$\mathcal{E}(Z_n, Z_n) = \min_{Y \in \overline{B}^K} \mathcal{E}(Y, Z_n) \leq \mathcal{E}(Z_{n-1}, Z_n) = \min_{Y \in \overline{B}^K} (Z_{n-1}, Y) \leq \mathcal{E}(Z_{n-1}, Z_{n-1}). \tag{2.9}$$

□

It has been shown in [105] for the CVT case that, if the density function is positive except on a set of measure zero, stationary points of the distortion $\mathcal{E}$ are given by fixed points of the Lloyd map T. The result below justifies that fixed points can be attained as a limit of Lloyd iterations in our case, as well.

Theorem 2.2.6. *Any limit set Z of Algorithm 2.2.3 is a fixed point of the Lloyd map, and thus, (Z, Z) is a critical point of $\mathcal{E}$. Moreover, for an iteration starting from a given point, all elements in the set of its limit points share the same distortion value $\mathcal{E}$.*

Proof. Algorithm 2.2.3 produces a bounded sequence $\{Z_n\}_{n=1}^{\infty}$ in $\overline{B}^K$ and thus it has a convergent subsequence. Let Z be a limit point of $\{Z_n\}_{n=1}^{\infty}$ with a subsequence $\{Z_{n_j}\}_{j=1}^{\infty}$ such that $\lim_{n_j\to\infty} Z_{n_j} = Z$. Since the distortion values are monotone decreasing, all the limit points have the same distortion value

$$\mathcal{E}(Z, Z) = \lim_{n_j\to\infty} \mathcal{E}(Z_{n_j}, Z_{n_j}) = \inf_n \mathcal{E}(Z_n, Z_n). \tag{2.10}$$

Let $\mathcal{E}_1$ and $\mathcal{E}_2$ denote the partial derivatives of $\mathcal{E}$ with respect to all the variables of the first and second arguments, respectively. Then Lemma 2.2.4 implies

$$\mathcal{E}_1(U, Z_n)_{|U=Z_n} = 0, \tag{2.11}$$

and by continuity

$$\mathcal{E}_1(Z, Z) = 0. \tag{2.12}$$

Now, if $\mathcal{E}_2(Z, U)_{|U=Z} = 0$, then (Z, Z) is a critical point of $\mathcal{E}$ and the proof is complete. Otherwise, a Y exists such that

$$\mathcal{E}(Z, Y) < \mathcal{E}(Z, Z). \tag{2.13}$$

Then, for every sufficiently small δ, there is an index n_j such that

$$\begin{gathered}\mathcal{E}(Z_{n_j}, Y) < \mathcal{E}(Z, Y) + \delta < \mathcal{E}(Z, Z) \leq \\ \mathcal{E}(Z_{n_j+1}, Z_{n_j+1}) \leq \mathcal{E}(Z_{n_j}, Z_{n_j+1}),\end{gathered} \tag{2.14}$$

contradicting the fact that

$$\mathcal{E}(Z_{n_j}, Z_{n_j+1}) = \min_{Y\in\overline{B}^K} \mathcal{E}(Z_{n_j}, Y). \tag{2.15}$$

□

As an immediate consequence of Lemma 2.2.5, the points solving problem (2.4) can be found by Algorithm 2.2.3.

Corollary 2.2.7. *Let $Z \in A^K$ be a point set that optimizes problem (2.4). Then $T(Z) = Z$.*

The Lloyd algorithm has the disadvantage that the Voronoi regions must be computed. So, equivalent iterative statistical methods based on random sampling were proposed [102, 105]. They are initialized with a random selection of K points. Then, in every iteration step, a Monte-Carlo sampling is executed to update the centroids of the cells. These methods can easily be adopted for our task, as well.

Algorithm 2.2.8. Random sampling algorithm for computing RCVT

---

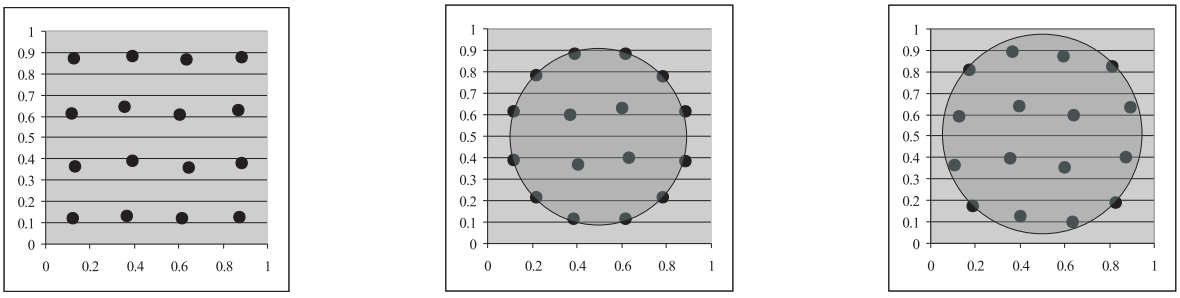
1. Choose a $q \in \mathbb{N}$ and constants $\alpha_1, \alpha_2, \beta_1, \beta_2$, such that $\alpha_2, \beta_2 > 0$, $\alpha_1 + \alpha_2 = \beta_1 + \beta_2 = 1$; choose an initial set of K points $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$ in A, e.g. by using a Monte Carlo method; set $j_i = 1$ for $i = 1, \ldots, K$;
2. Choose q points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q$ in B at random, e.g. by a Monte Carlo method, according to some probability density function (uniform one in this description);
3. For $i = 1, \ldots, K$, collect in the set W_i all sampling points $\boldsymbol{y}_r$ closest to $\boldsymbol{z}_i$ among $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$,

i.e. the ones lying in the Voronoi region of $\boldsymbol{z}_i$ *w.r.t.* B*; if the set* W_i *is empty, do nothing; otherwise, compute the* $\boldsymbol{u}_i$ *average of* W_i *and set*

$$\boldsymbol{z}_i \leftarrow \frac{(\alpha_1 j_i + \beta_1)\boldsymbol{z}_i + (\alpha_2 j_i + \beta_2)\boldsymbol{u}_i}{j_i + 1}, \quad j_i \leftarrow j_i + 1. \tag{2.16}$$

The new set of $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$ *along with the unchanged* $\boldsymbol{z}_j$*'s (i.e. when* W_i *is empty), form the new set of points* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$*;*
4. If for some i*,* $\boldsymbol{z}_i \notin A$*, then* $\boldsymbol{z}_i \leftarrow \boldsymbol{z}$*, where* $d(\boldsymbol{z}_i, \boldsymbol{z}) = \min_{\boldsymbol{y} \in A}\{d(\boldsymbol{z}_i, \boldsymbol{y})\}$*, i.e.,* $\boldsymbol{z}$ *is the nearest point to* $\boldsymbol{z}_i$ *in* A*.*
5. If the new points meet some convergence criterion, terminate; otherwise, return to step 1.

The actual modification of the basic CVT algorithm [92] is done in step 4, where the centroids are mapped into A, in a similar way as in the corresponding CCVT algorithm [93]. Note that it would not be sufficient to apply this step only once in the end, as the above iteration also influences the positions of the centroids better inside A. Figure 2.3 shows how the final positions of the centroids change according to the radius of the disc A centered within the unit square B.



(a) (b) (c)

Figure 2.3: *The result of the modified CVT algorithm when* B *is the unit square; (a)* $A = B$*, (b)* $A \subset B$ *is the disc of radius 0.4 centered at (0.5,0.5), (c)* $A \subset B$ *is the disc of radius 0.45 centered at (0.5,0.5).*

An example of the reduction of a contour like object is shown in Figure 2.4. Here, the object A to be simplified is a circle of radius 0.45. Figure 2.4(a) shows the result of the CCVT algorithm, while Figure 2.4(b) depicts the outcome of the proposed RCVT method, when B is chosen to be the unit square with $B \supset A$. In the CCVT case, equidistant points are preserved, since no outside region is considered there. For RCVT, it can be observed that more points are selected closer to the directions $\pm\frac{\pi}{2}$, $\pm\frac{3\pi}{4}$, since the square B has a larger spread along them. Consequently, more points of A are needed to represent these larger diagonal zones.

The global convergence of the corresponding versions of Algorithms 2.2.3 and 2.2.8 were thoroughly investigated only in 1D for the CVT/CCVT case. Regarding CVT, the global convergence is shown for a closed interval [92], while for CCVT, it is shown for smooth bounded curves [93], in the case of any positive and smooth density function. Higher dimensional convergence is still an open issue, and so it is for RCVT, as well.

(a) (b)

Figure 2.4: *Simplification of the circle A having radius 0.45 centered at (0.5,0.5) to 100 of its points; (a) the result of the CCVT algorithm, (b) the result of the RCVT algorithm, when B is the unit square.*

2.3 Experimental results

In our experiments, we considered the discrete object simplification problem on digital images. In the discrete case, we consider the following variant of the continuous distortion functional (2.5):

$$\mathcal{E}(Y,Z) = \sum_{i=1}^{K} \sum_{\boldsymbol{y} \in V_i(Y)} \varrho(\boldsymbol{y}) |\boldsymbol{y} - \boldsymbol{z}_i|^2, \tag{2.17}$$

with $B = \cup_{i=1}^{K} V_i(Y)$ being a set of M points in $\mathbb{R}^m$, and $Y, Z \in A^K$. Following a similar argument to that given in [92] for the CVT case, it can be shown that the above distortion can be minimized by a Voronoi tessellation generated by its centroids Z. Similarly to the CVT case, as the number of sample points M tends to infinity, the convergence of the energy [104] and of the centroids [106] can be proven under certain conditions.

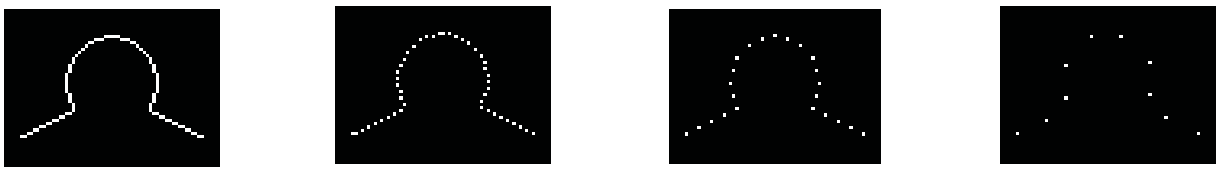
During the RCVT iteration, the generators were supposed to have real coordinates. After the last iteration step they were rounded to the closest point of object A, when they were lying outside A. Additionally, for a more flexible application, we ignored the convexity criteria for the objects without encountering problems in our tests. To overcome possible concavity problems in our implementation, we initialize the RCVT algorithm to start from a point set obtained by uniform sampling of A. In this way, if A has a relatively smooth boundary, the closest point of object A was found within the same Voronoi cell. If possible infinite loops caused by concavity do not let the iteration converge, repetitions can be expected for the alternating closest point assignment caused by concavity. Though we did not need to do so, in case of more severe concavity problems, a promising approach might be the subdivision of A into convex (or smoother) subsets, with separate optimization. However, in this case, the borders of the subsets must be handled with additional care.

We present two experimental setups with the corresponding results of our optimization approach in chamfer matching. First, we show our results in template matching on snake representations for contour simplification. Then we also consider the possible simplification of region object representations.

2.3.1 Simplifying object contour representations

In the case of a simple one-pixel wide digital template A, when the search set B is chosen to be at its vicinity, possible naive alternatives can keep randomly selected or equidistant points instead of applying the RCVT algorithm. For a practical example, see Figure 2.5 for a head template (set

A) used in our system, where we applied various percentages of contour point reduction. However, notice that equidistant simplification is hard to interpret for non-contour objects, while RCVT can be applied for arbitrary sets without problems. As for random selections, we considered the average performance of some random sampling in our experiments.



(a) (b) (c) (d)

Figure 2.5: *Reduction of the number of template points for chamfer matching; (a) original template, (b) 50%, (c) 75%, (d) 90% reduction of the template points.*

The search regions (sets B), for checking the change in the distance map, can be obtained e.g. by successive dilations of the original template A. Here we used the 3×3 square structuring element C [107]. Figure 2.6 shows the result of some dilations.

(a) (b) (c)

Figure 2.6: *Dilations of the head template to create the search region $B = A \oplus nC$ for (a) $n = 1$, (b) $n = 3$, (c) $n = 12$.*

Figure 2.7 depicts some quantitative results to compare the accuracy of the random, equidistant and RCVT based reduction. In this experiment 25% of the template points were kept to form a new template A'. The calculation was made for dilations $B = A \oplus nC$ of the original head template for several $n \in \mathbb{N}$. The horizontal axis is the number of dilation steps n, whereas the vertical axis shows the distance map error $E = \sum_{\boldsymbol{x} \in B} (d_{A'}(\boldsymbol{x}) - d_A(\boldsymbol{x}))$. We can see that the proposed RCVT approach gives remarkably better distance map approximations in the case of larger search regions. Besides the $\langle 3, 4 \rangle$ distance map considered primarily in the chapter, we include the results corresponding to the distance map $\langle 5, 7, 11 \rangle$, which is known to be computationally more demanding, but a more accurate approximation of the Euclidean distance [97]. According to this test, we can expect to derive basically the same results in case of any other approximation of the Euclidean distance.

Another important issue is to investigate how the distance map error E depends on the simplification of the template, in order to determine the acceptable reduction level. According to the results of our experiments shown in Figure 2.8, we can conclude that the accuracy falls exponentially with the percentage of the retained template points. Severe inaccuracy can be experienced in the case of excessive (80%, 90%) simplification.

Chamfer matching with Gradient Vector Flow (GVF) snakes

In one of our developed person detection/tracking/recognition systems[1], we analyze infrared images captured in a fire scene, like the ones shown in Figure 2.9.

[1]EU FP6 Information Society Technologies, FP6-004218, SHARE: Mobile Support for Rescue Forces, Integrating Multiple Modes of Interaction.

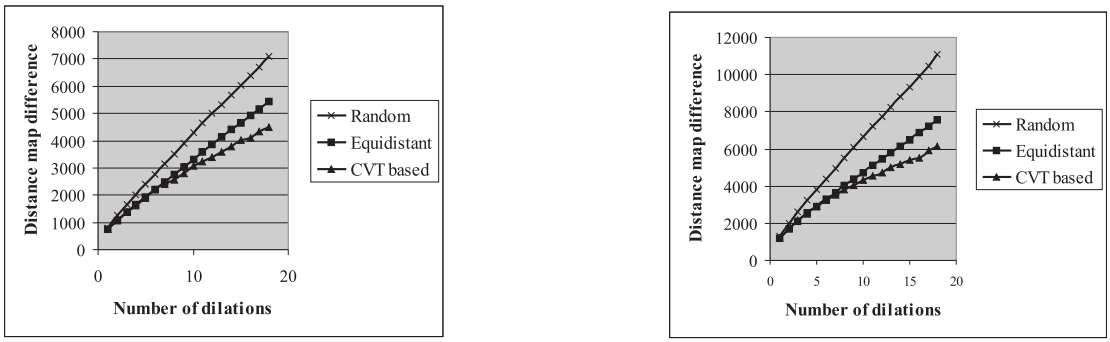


Figure 2.7: *The comparison of the equidistant and RCVT based reduction of one-pixel wide objects for chamfer matching for different distance maps: (a)* $\langle 3, 4\rangle$ *distance map, (b)* $\langle 5, 7, 11\rangle$ *distance map.*

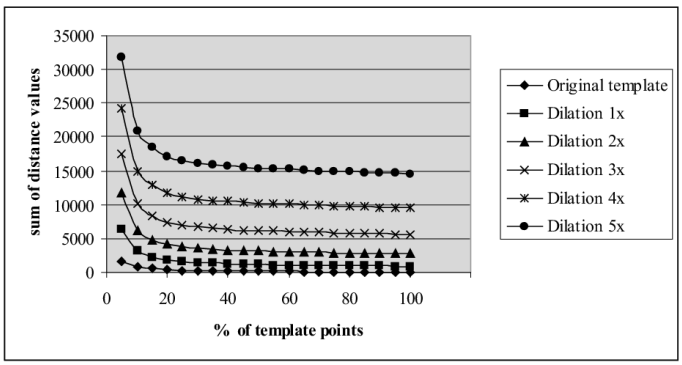


Figure 2.8: *The change of the distance map error at different levels of reduction of the points of the head template.*

Figure 2.9: *Thermal images captured under varying temperature conditions.*

We cannot use background subtraction here (no prerecorded background data exist), and we can hardly use the infrared intensity data (due to varying temperature values) to locate objects, e.g. humans. Accordingly, as a robust active contour [108] technique, the GVF snake has been chosen to extract object boundaries [109]. A very useful outcome of the snake algorithm is that, in the case of a closed snake, we have the snake points in an indexed sequence $[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p]$, with $\boldsymbol{s}_p = \boldsymbol{s}_1$. The process for evolving the GVF snake considers two additional parameters for the density of the points composing the snake. Namely, $d_{\max}$ and $d_{\min}$ denote the maximum and minimum distance allowed between two snake points, respectively. It can be easily seen that by requiring $d_{\max} < 1$, we can guarantee an 8-connected snake. The main disadvantage of considering

small $d_{\max}$ values (a dense snake) is that the iterative process can become very time-consuming. Therefore, an important point in our approach is to investigate how dense the snake points can be for a reliable chamfer matching.

To recognize the object represented by the snake, we match whole object contours or parts of them. For example, for human body detection, the object can be classified as a human, based either on successful whole human contour or on head and limb matching.

Matching along the snake

In section 2.1, we discussed the difficulties rising from the necessary geometric transformations between the target object and database object representations. This usually results in an exhaustive search for the appropriate transformation parameters. Using snakes, we are in quite a comfortable position to make obvious restrictions to this parameter space. First of all, we can avoid translating the template "blindly" over the entire image, as we adjust its origin to snake points. After translation, we can utilize the sequentiality of the snake points in finding the suitable rotation angle. Namely, we can consider consequent snake points to estimate the direction of the snake by comparing a given snake point with some subsequent ones. Naturally, this method can be adopted easily only for templates having a straight starting segment, in which case the straight segment can be aligned to the estimated (or close) angle. Otherwise, the divide and conquer strategy can be used here, for rotation. The magnification parameter can be bounded easily by adjusting it, regarding the spatial "size" (e.g. perimeter, area, bounding box) of the snake.

Moreover, as now we have a closed boundary with no outliers, it is less important to involve edge direction information [110] in measuring how good the fit is. The "blind" methods also suffer from the problem of selecting suitable threshold(s) for (e.g. the Canny) edge detection and giving many false negatives in the case of a cluttered/noisy scene. Taking all these factors into consideration, we can perform the matching steps in an obviously shorter time than in the general case. Thus, considering the geometric transformations summarized in section 2.1 regarding scaling, translating and rotating, now the best match at $\boldsymbol{s_i} \in S$ can be defined as

$$\widehat{d_{\boldsymbol{s}_i}(S,T)} = \min_{\substack{\lambda \in \Lambda \\ \Theta \in [0,2\pi[}} \left(d_{\boldsymbol{s}_i}(S, \lambda T_\Theta) \right), \tag{2.18}$$

where Λ is a set of possible scaling values, and T_Θ denotes the set T rotated around its origin by Θ. Consequently, the best matching value can be given as

$$d(S,T) = \min_{\boldsymbol{s}_i \in S} \left(\widehat{d_{\boldsymbol{s}_i}(S,T)} \right), \tag{2.19}$$

and the best matching position is the snake point where (2.19) is taken.

Matching human body parts

In this section, we present some experimental results regarding matching human body parts, like the head and limbs. Our matching process is based on the above described steps, and we tested how the density of the snake S and the reduction of the templates affected matching reliability. The templates T_j (head and leg) and their simplified versions were matched along the snake, as it was described in the previous section. Figure 2.10 shows examples for matching the head and leg template.

We experienced that simplifying the template and considering a less dense snake representation speed up matching computations. Figure 2.11 shows $\widehat{d_{\boldsymbol{s}_i}(S,T)}$ for various densities of the snake and head template points. The correct position for the template (shown in Figure 2.10(a)) was

(a) (b)

Figure 2.10: *Best fitting positions of templates for a human silhouette represented by a snake; (a) head, (b) leg.*

found in all these cases according to the minimal distance value (normalized to 1) indicated by an arrow in Figure 2.11. However, the reliability of matching naturally deteriorates when less snake/template data are used, as it can be checked in Figure 2.12, as well. Here the corresponding Receiver Operating Characteristic (*ROC*) curves [111] for the cases shown in Figure 2.11(a)-(d) are presented, respectively. The curves show how true (acceptable matching positions) and false positives are found by raising the threshold value for the distance map error. For this experiment the snake points were manually preclassified as acceptable/unacceptable matching positions, that is, true/false positives.

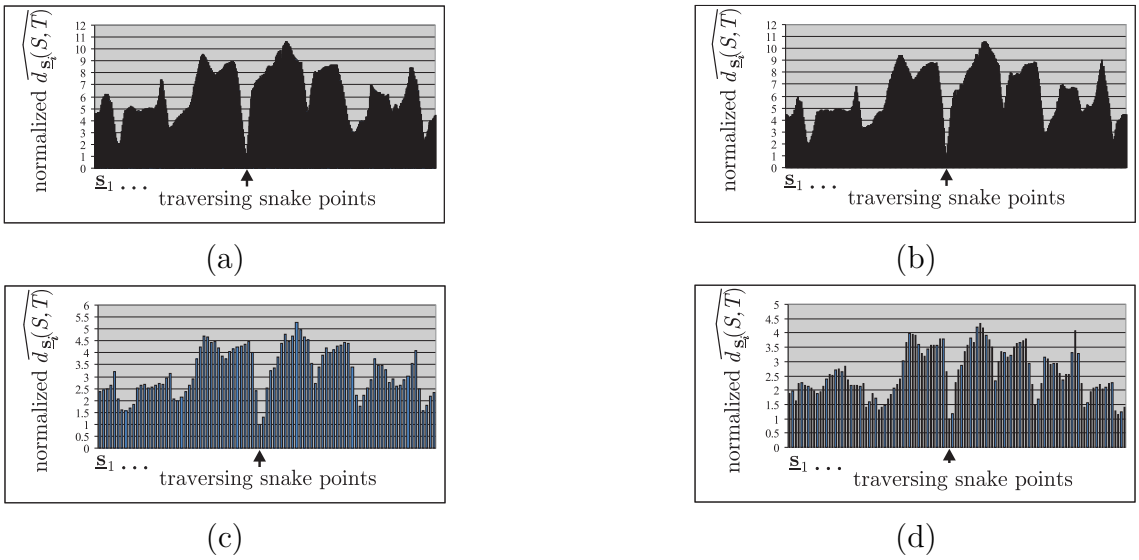


Figure 2.11: *Point-wise goodness of fit distance profile for matching the head template against the body contour at different levels of template simplification and snake density. Best match is found at the normalized sum of distance values 1; (a)* 100% *retained template points, dense snake ($d_{\max} < 1$), (b)* 100% *retained template points, less dense snake ($d_{\max} < 4$), (c)* 25% *retained template points, dense snake ($d_{\max} < 1$), (d)* 25% *retained template points, less dense snake ($d_{\max} < 4$).*

Moreover, in order to have an experimental comparison between the RCVT and the naive contour reducing approaches (random and equidistant sampling), we set up a test environment. In this experiment, we performed the chamfer matching of head templates against a database of 60 elements containing the original head template distorted by several geometric transformations (stretching and skewing) together with some head silhouettes extracted from real videos. Besides

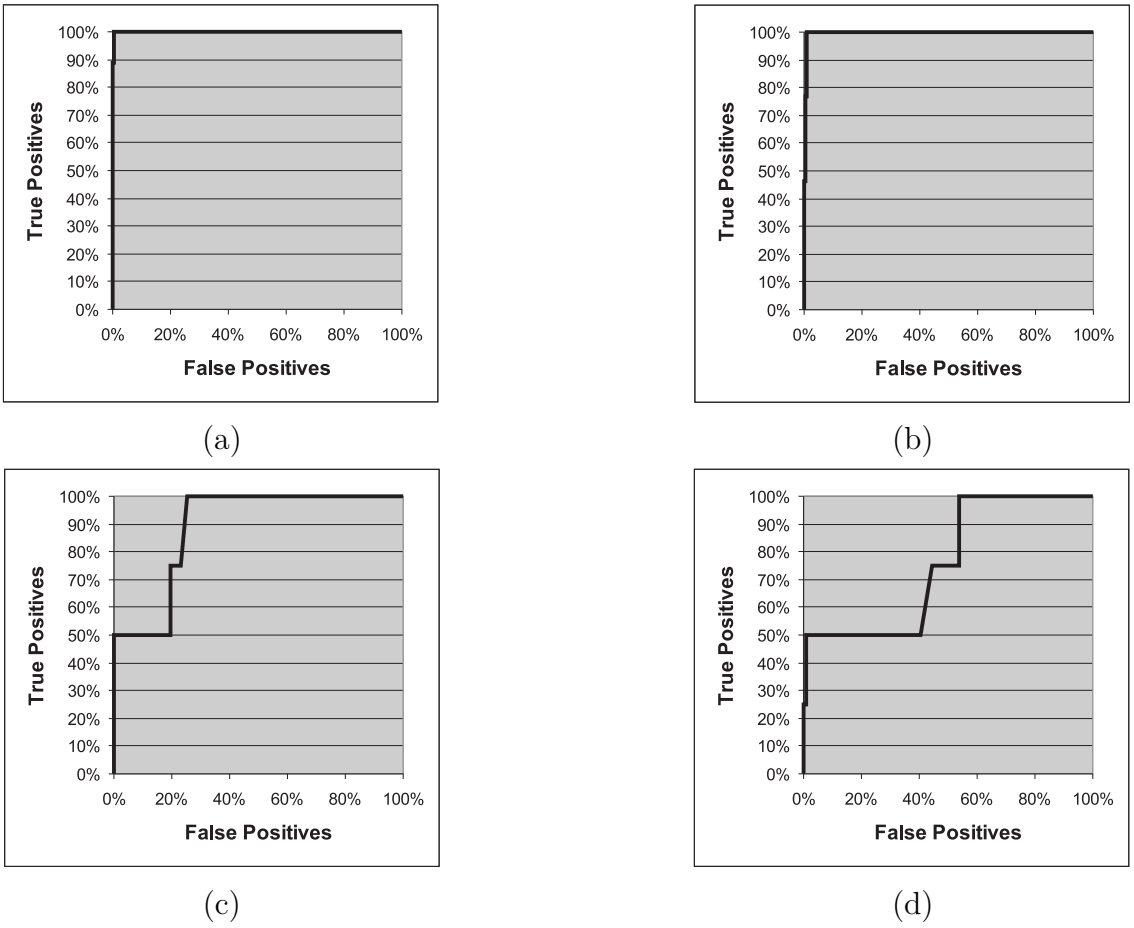


Figure 2.12: *ROC curves for matching the head template against the body contour at different levels of template simplification and snake density; (a)* 100% *retained template points, dense snake* ($d_{\max} < 1$)*, (b)* 100% *retained template points, less dense snake* ($d_{\max} < 4$)*, (c)* 25% *retained template points, dense snake* ($d_{\max} < 1$)*, (d)* 25% *retained template points, less dense snake* ($d_{\max} < 4$)*.*

making a test without any reductions, we considered random, equidistant, and RCVT-based simplification of the head template. Naturally, in case of a simplification, the same number of points were retained.

Our main aim here was to experimentally validate the assumption that the original head template can be replaced more reliably with applying RCVT instead of some other naive simplification approach. As an easily obtainable result for the simplified objects, first we checked the deviation of the best matching positions of the simplified templates from the best matching position of the original head template. The deviation was calculated considering all the database elements as the sum of the squared distance between the best matching positions found for the original (non-simplified), and simplified templates, respectively. This analysis gives a preliminary impression on which simplification approach can lead to the most valid replacement of the original template. In the way discussed before, we considered RCVT simplification of the head template regarding several search regions, which naturally had no effect on the performance of the equidistant and random sampling. The deviations are shown in Figure 2.13 for this experiment.

We can see that RCVT provides improvement regarding both the random and equidistant sampling. We note two things here. On the one hand, this analysis gives a quick impression about the possible improvement obtainable by RCVT. On the other hand, we have to keep in mind that the selection of a larger search region does not lead necessarily to better matching performance.

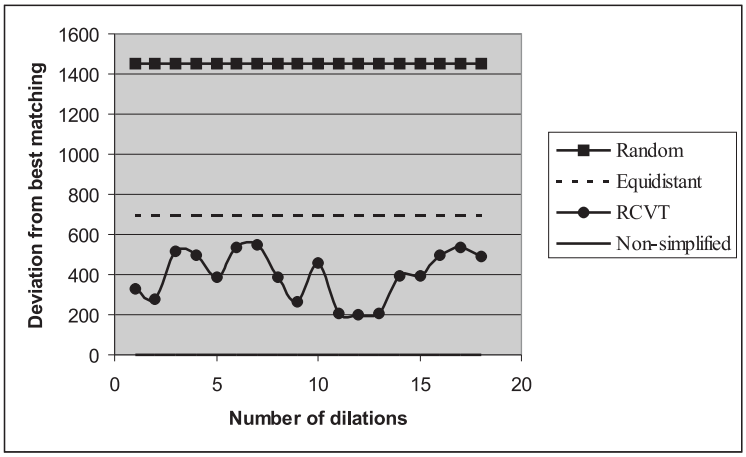


Figure 2.13: *Comparing the performance of simplified head templates with considering the sum of the distances between their best matching positions from those of the original head template for a dataset of head silhouettes.*

With the selection of the search region size for RCVT, we can respond to the expected spatial deviation of the template from the object. In other words, if less precise matching is expected, a larger search region can be used to try to cover a larger area.

To confirm this hypothesis and obtain more detailed comparative results, we performed a second test using the test database. Namely, a *good match* region was defined as a neighborhood of the best matching position of the original template. This way, we can create *ROC* curves to see how similarly the simplified templates behave compared to the original one. To perform this analysis and also to validate Figure 2.13, we considered RCVT simplifications belonging search regions of 5, 12, 18 dilations, respectively. The results are shown in Figure 2.14.

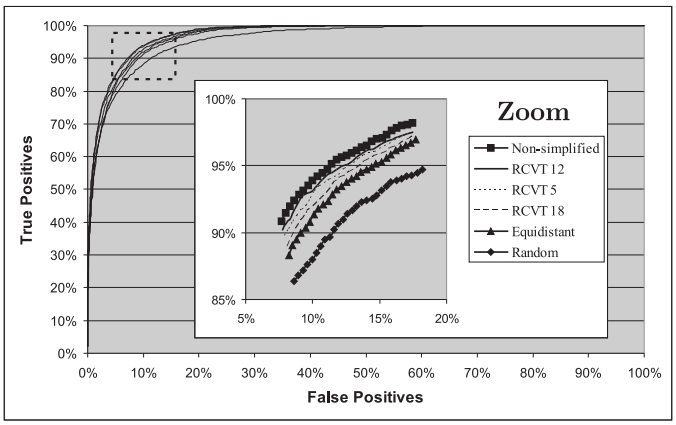


Figure 2.14: *ROC curves to measure the performance of the simplification methods applied to the head template on an experimental dataset of target head objects. The numbers 5, 12, 18 assigned to the RCVT simplifications refer to the size of the search region in terms of dilation steps.*

As it can be seen in Figure 2.14, the results suggested by Figure 2.13 are confirmed corresponding to the deviation from the best matching positions. For simplicity, in this approach we excluded all geometric distortion and snake-based matching issues, and performed pure chamfer matching with the translation of the head template on the input distance maps.

The computation time increases linearly with the percentage of the template points retained, since the same family of operations should be performed for a larger point set. Our experiments also reflected this behavior, as it can be seen in Figure 2.15 for the head template. Similar results were found for the leg templates.

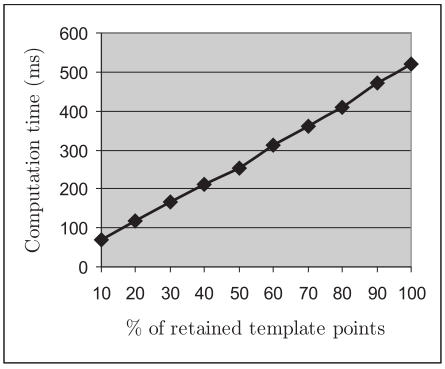
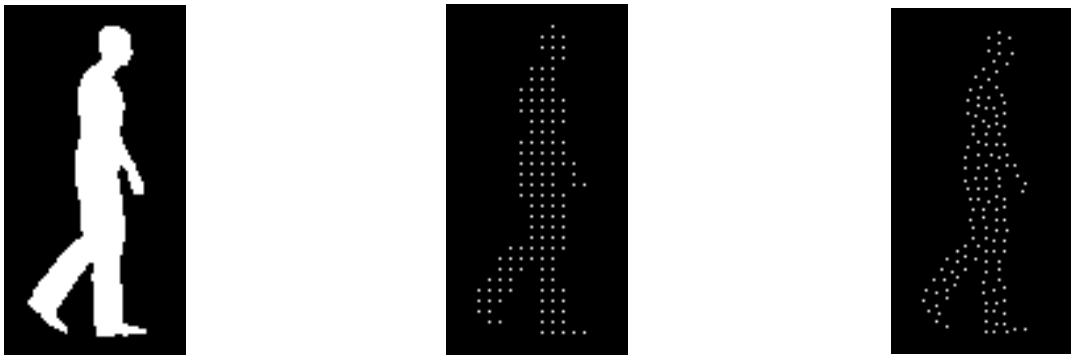


Figure 2.15: *Computation time of object matching vs. percentage of retained points of the original head template.*

2.3.2 Simplifying object region representations

Though chamfer matching is usually employed to match object contour representations, it is also possible to apply it to region matching, e.g. for human body part matching [112]. After applying RCVT optimization to simplify regions, we obtained similar experimental results to the contour case described in section 2.3.1. Such an example for a human body silhouette is shown in Figure 2.16.



(a) (b) (c)

Figure 2.16: *The result of the RCVT algorithm in 2D; (a) the object A to be simplified (walking human), (b) the initial generators chosen by naive equidistant sampling, (c) the generators achieved by the RCVT algorithm for $A = B$.*

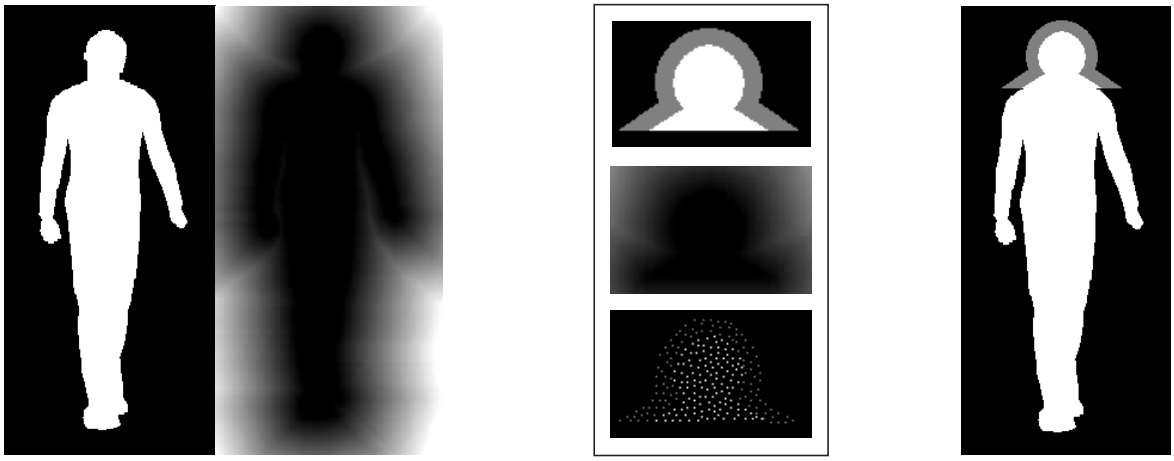
If the regions to be matched are approximately of the same size (whole shape matching), then it can be sufficient to execute the matching as we discussed for the contour objects. However, if we want to match object parts, as we discussed it in section 2.3.1, then we need to modify our approach. Because, with the same setup, the object part would give very good matches anywhere inside the target object, as the distance values are 0 there (or close to 0 in case of simplification). To overcome this difficulty, we can consider a subregion $\widehat{A}$ of the compliment A^c of the object part (its background). A natural selection for $\widehat{A}$ is $\widehat{A} = (A \oplus nC) \cap W$ for some $n \in \mathbb{N}$, and $W \subseteq A^c$,

as shown in Figure 2.17(b). The role of W is to define an area around the object part A within its background. If desired, we can simplify A and $\widehat{A}$ using the RCVT approach, obtaining A' and $\widehat{A}'$, respectively.

Furthermore, we determine the distance maps both for the target object S (see Figure 2.17(a)) and for A (see Figure 2.17(b)). To determine how good the fit is at a point $\boldsymbol{x} \in \mathbb{Z}^2$, we calculate

$$\sum_{\boldsymbol{y} \in A' \cup \widehat{A}'} |d_S(\boldsymbol{x}+\boldsymbol{y}) - d_A(\boldsymbol{y})|. \tag{2.20}$$

The best match is found when this sum is minimal. Notice that in this example we ignored all the geometric transformation issues discussed in section 2.1. To match body parts based on regions, another approach could consider signed distance maps [113].



(a) (b) (c)

Figure 2.17: *Chamfer matching of object parts; (a) target object, and its distance map, (b) from top to bottom: head template region A (white) with a subset $\widehat{A}$ of its complement (gray), the distance map of A, and the simplification of A and $\widehat{A}$, (c) best matching position for the simplified template.*

Skeleton-based simplification

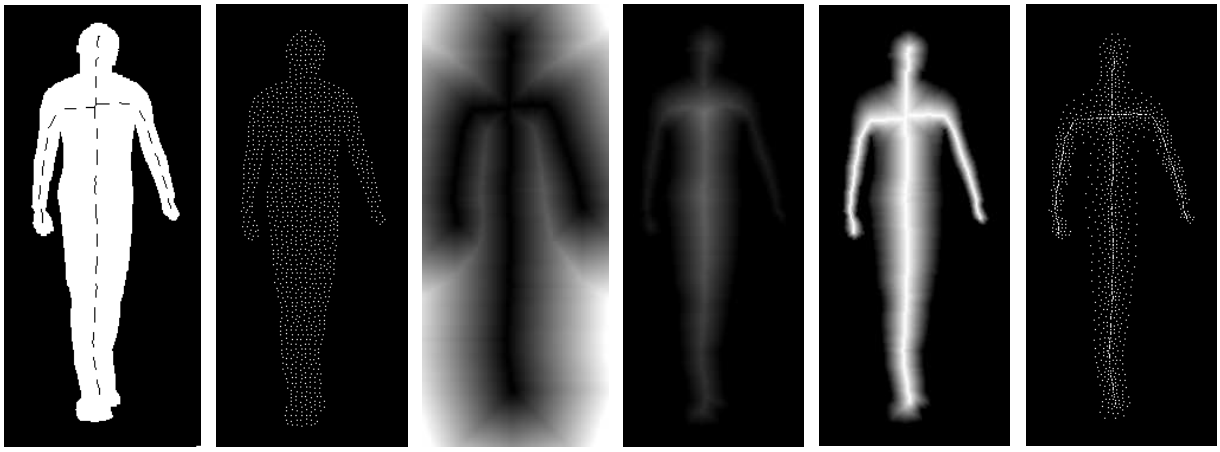
To realize a more adaptive simplification approach than the naive ones illustrated in Figure 2.16, it is also possible to focus more on some object regions during their simplification, e.g. if we expect larger uncertainty closer to the boundary of the object in the matching procedure. In this case, we can define a weight function $\widetilde{\varrho}$ according to problem (2.4), which concentrates on the object morphological skeleton [114]. Thus, for every $\boldsymbol{x} \in B$, we define

$$\widetilde{\varrho}(\boldsymbol{x}) = \frac{d(\boldsymbol{x}, \overline{B^c})}{d(\boldsymbol{x}, \mathrm{sk}(A)) + d(\boldsymbol{x}, \overline{B^c})}, \tag{2.21}$$

where $\mathrm{sk}(A)$ is the skeleton of A [114], B^c is the complement of B, and d is some common distance function. Note that for all $\boldsymbol{x} \in B$, $0 \leq \widetilde{\varrho}(\boldsymbol{x}) \leq 1$, such that $\widetilde{\varrho}$ vanishes at the boundary of B, then monotonously increases till reaching $\mathrm{sk}(A)$, where it takes value 1. In this way, the weight for the points within B is adjusted according to their relative distance from the "center" of A and from $\overline{B^c}$. Now, to derive a weight (density) function over B, put

$$\varrho(\boldsymbol{x}) = \frac{\widetilde{\varrho}(\boldsymbol{x})}{\int\limits_B \widetilde{\varrho}(\boldsymbol{y}) d\boldsymbol{y}} \quad \text{for } \boldsymbol{x} \in B. \tag{2.22}$$

In practice, this weight function can be easily derived using the distance maps [90] of sk(A), and B^c to approximate d. Figure 2.18 depicts such an example, where $A = B$, the basic CVT case is used. We also found the pruning (removal of small branches) of the skeleton to be useful.



(a) (b) (c) (d) (e) (f)

Figure 2.18: *Simplification result of an object according to a weight function concentrating on its skeleton; (a) target object $A = B$ (sk(A) is shown by dashed line), (b) result of simplification in case of uniform weighting, (c) distance map of sk(A), (d) distance map of $\overline{B^c}$, (e) weight function $\widetilde{\varrho}$ (higher intensities show larger weight values), (f) result of CVT simplification using ϱ derived from $\widetilde{\varrho}$.*

To be able to begin the template matching, first we need to extract the target object from the input image. In our system SHARE, one of the desired tasks is the detection of human (victim or firefighter) appearance within thermal videos from a rescue scenario. To reach this aim, we considered a fuzzy segmentation technique [115], which is known to be robust also for medical (e.g. CT) images. The extracted binary region usually needs some simple postprocessing to smoothen the boundary, eliminate gaps and holes, etc. These minor refinements can be achieved by some classic elements of mathematical morphology [114]. Some input test images together with the result of the fuzzy segmentation is shown in Figure 2.19.

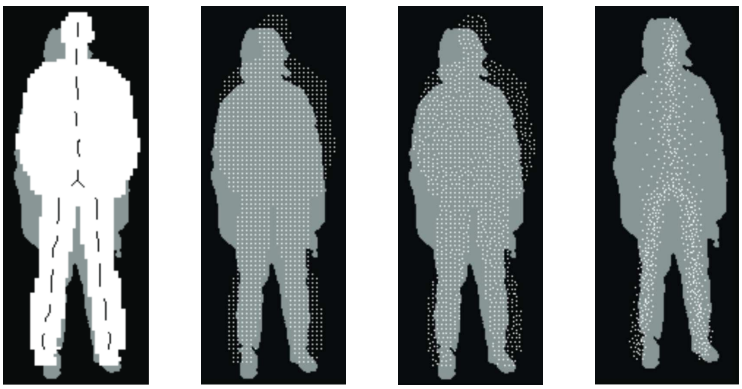
(a) (b)

Figure 2.19: *The result of fuzzy segmentation; (a) standing human pose, (b) walking human pose.*

To create templates to be matched, we created simulated (artificial) data using the realistic

3D human motion software Poser®[2]. As it is not a crucial point in our present experiments, we ignored all the geometric transformation issues on how to align the target and template objects. As a guideline, we recommend to consult with [101] to gather information on a robust search of the affine parameter space.

To demonstrate the main idea of our approach, we show the performance of such templates that do not completely fit the target. As our main goal is to introduce the novel idea on skeleton-based simplification and matching, the corresponding test set of two objects is just to reflect the possible improvements and relations. With the skeleton-based simplification, in the case of a regular human pose/motion (like standing/walking), we can expect better matching for more template elements, since the skeleton does not change drastically with the boundary e.g. for close phases of the same motion. The goodness of fit value of a template against the target object is calculated as the percentage of the matching pixels of the template at the best matching position. Figures 2.20 and 2.21 show the best matching positions of a standing and walking template together with their discussed simplifications, respectively. The skeleton of the original template is also marked.



(a) (b) (c) (d)

Figure 2.20: *Best matching position (shown in white) for standing template using (a) original template (skeleton is marked), (b) trivial uniform simplification, (c) CVT-based uniform simplification, (d) skeleton-based simplification. Target object is shown in gray.*

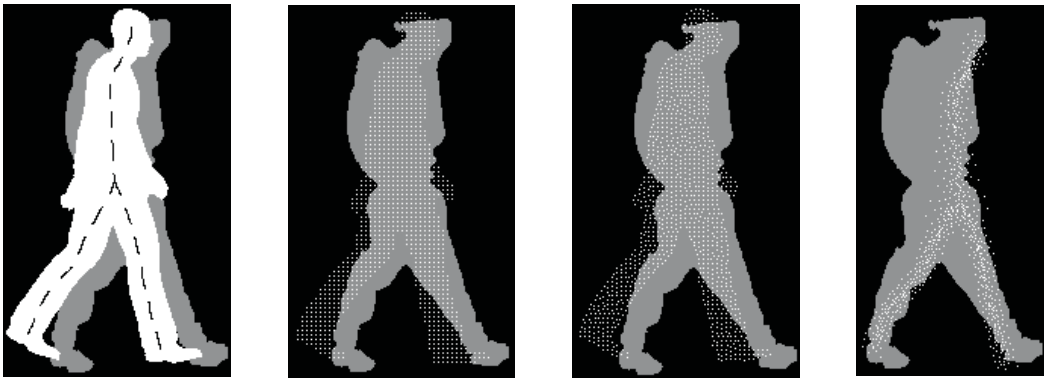
We can see that the skeleton-based simplification has better performance than the uniform ones, since the main skeleton of the target objects and the templates did not differ that much. The proper quantitative results are given in Table 2.1.

	Simplification			
	No (original)	Unifom (trivial)	Unifom (CVT)	Weighted (skeleton)
Standing pose	80,9%	81,2%	81,1%	94,0%
Walking pose	85,8%	86,2%	86,2%	94,6%

Table 2.1: *Goodness of fit of simplified templates given in terms of the percentage of matching template points.*

This analysis also validates the natural assumption that the uniformly sampled template behaves similarly to the original one.

[2]Poser® is a registered trademark of Smith Micro Software, Inc.



(a) (b) (c) (d)

Figure 2.21: *Best matching position for walking template using (a) original template, (b) trivial uniform simplification, (c) CVT-based uniform simplification, (d) skeleton-based simplification.*

2.3.3 Subsampling strategies in learning-based vessel segmentation

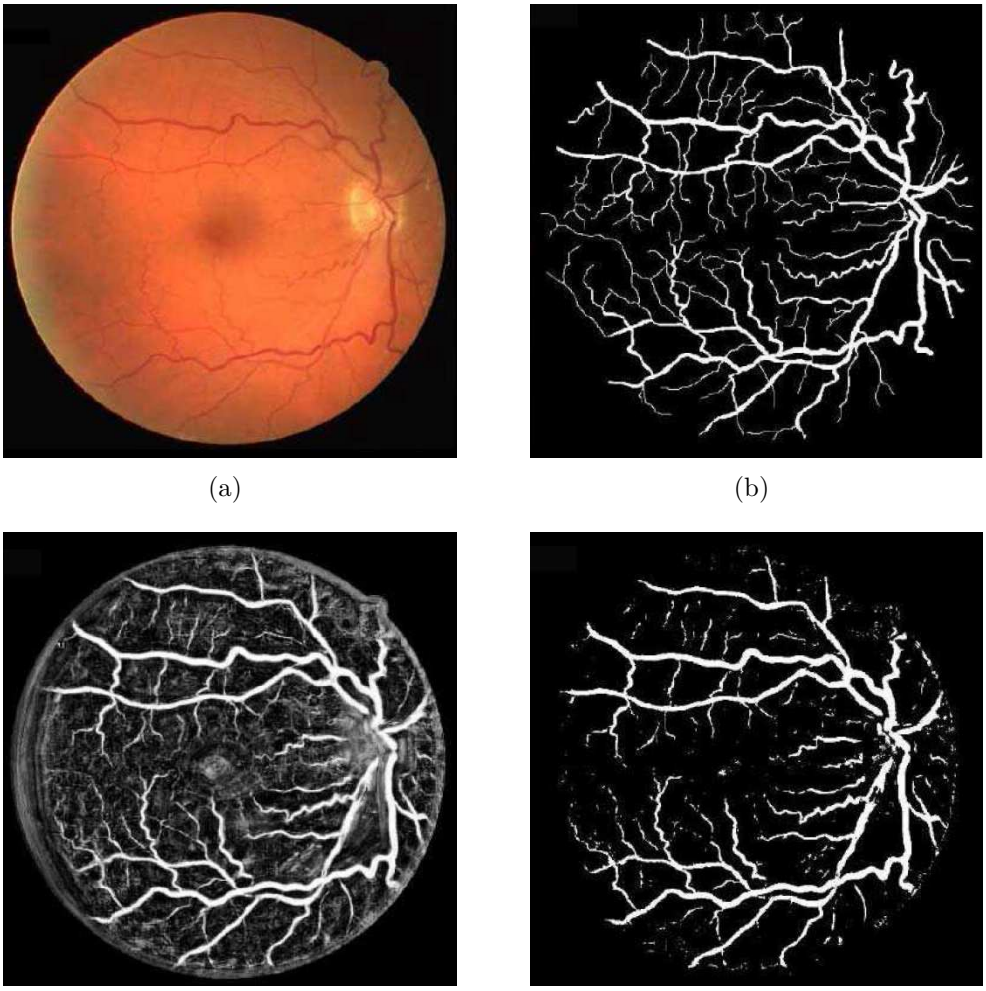
The original practical motivation of the efforts discussed so far in this chapter was the recognition of human appearance in thermal videos. However, we have investigated whether these sampling strategies can be exploited also in other application fields to speed up computations. Accordingly, in this section, we present some results on applying these approaches in a learning-based environment dedicated to the segmentation of retinal vessels (see also section 1.4).

Algorithms for segmenting the vascular system of the retina have high importance in automatic systems for detecting diseases (such as diabetic retinopathy) based on digital retinal (fundus) images. Using an accurately segmented vascular system, we have better chances to locate other anatomical parts (e.g. optic disc and macula/fovea) of the retina and explore the disorders of the vascular system itself. State-of-the-art segmentation algorithms are usually based on machine learning, with a subsequent classification step to decide whether the pixels belong to the vascular system or to its retinal complement. We can highlight the method presented in [116] as a typical approach in this field with good performance having reported. Figure 2.22 depicts an original fundus image, a manually segmented vascular system for the training phase, the confidence results of the classification step and the detected vascular system, respectively. The classification of a pixel is done by thresholding the confidence map.

In our screening system DRSCREEN (see also the Introduction) for diabetic retinopathy, we investigated how this type of segmentation algorithms can be further improved. For efficient classification, a representative sample of vascular pixels is needed. However, the corresponding literature lacks the suitable selection of the training sample. Both to avoid overtraining and to save computational time, a sampling of the manually segmented vascular systems is desired during the training step. The most obvious selection of the points is random sampling. However, it seems also natural that a more adaptive sampling should lead to better overall segmentation performance. Thus, we have tested such approaches that can guarantee more homogeneous representation of the original object. For an objective comparison, we calculated the symmetric difference of the test images and the vascular system found by involving the investigated sampling strategies.

Sampling strategies

We have tested whether the random sampling can be outperformed by using more homogeneous sampling approaches based on the centroidal Voronoi (CVT) or the constrained centroidal Voronoi



(a) (b) (c) (d)

Figure 2.22: *Components of a learning-based vessel segmentation algorithm; (a) input retinal image, (b) manual segmentation for training, (c) confidence values to classify as vessel, (d) segmented vascular system.*

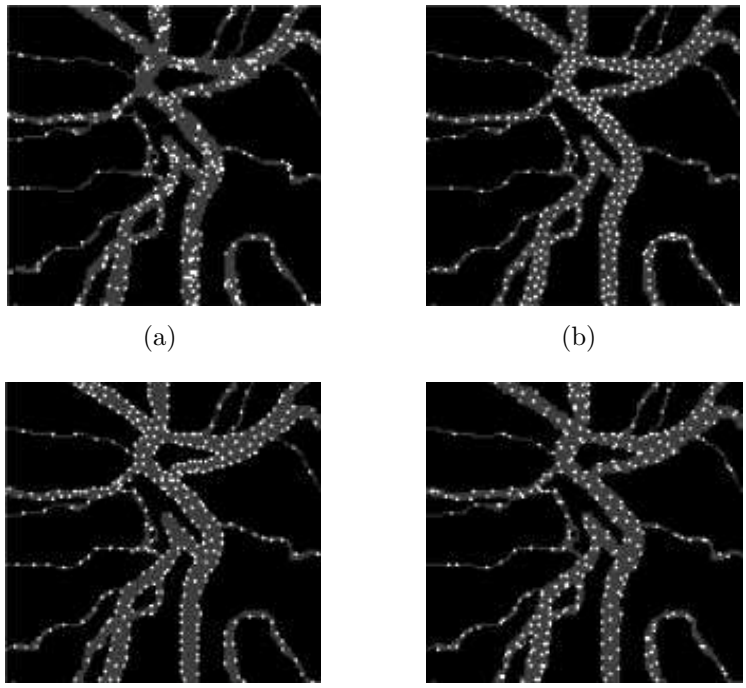
(CCVT) tessellation. Since CVT does not sample the boundary of the sets, and CCVT may emphasize it too much, we also combined them with having a union of half number of points using CVT and CCVT, separately. For an illustrative result of these strategies on retinal vessels see Figure 2.23.

Experimental setup

As for preprocessing, the segmentation algorithm we have selected works on the green channel of the fundus images. Moreover, we applied adaptive histogram equalization (CLAHE) described in section 1.6.1.

Our training database contained 10 fundus images from the publicly available DRIVE database (see section 1.5.6). This database also contains the manually segmented vascular system for each of the images to support quantitative comparison of different segmentation algorithms. For the elements of the test database we have selected 4 other images from the DRIVE database.

For vessel extraction, we considered the robust vessel segmentation method presented in [116]. This approach considers a kNN classifier based on learnt feature vectors for both vascular and



(a) (b) (c) (d)

Figure 2.23: *Result of sampling strategies shown for a part of the vascular system; (a) random sampling, (b) CVT, (c) CCVT, (d) CVT and CCVT combined.*

retinal background pixels. The method composes the feature vectors from the green level intensity of the pixels, and from the responses of 0th, 1st and 2nd order derivatives of Gaussian masks having different standard deviations.

In our experimental tests, we compared the performance of the following sampling strategies:

- random,
- centroidal Voronoi tessellation based (CVT),
- constrained centroidal Voronoi tesselation based (CCVT),
- a combination of the CVT and CCVT sampling.

We performed our tests using the following levels of sampling: 0.05%, 0.5%, 1%, 5%, 10%. For example, level 5% means that 5% of the pixels are retained from the object for training. These sampling levels were applied to both the vessel and background pixels.

We defined a quantitative measure for an objective comparison of the segmentation results of different sampling strategies, and levels. Namely, we calculated the normalized symmetric difference S of the vascular system A found by the segmentation algorithm and the manually segmented one B that we had for all the test images

$$S(A,B) = \frac{|A \setminus B| + |B \setminus A|}{|A \cup B|}, \tag{2.23}$$

where the symbol $\setminus$ denotes the set difference operator. Notice that $S(A,B) = 0$, if A and B coincide, while $S(A,B) = 1$, if they have no common points at all. To show the performance of segmentation, we consider the goodness value $G(A,B) = 1 - S(A,B)$. Thus, to summarize our experimental setup, it was prepared in a way to be able to compare sampling strategies and levels objectively.

Results

In our experiments, we were looking for the answer to the following questions:

- which sampling approach from the investigated ones leads to the best segmentation results?
- how does the segmentation accuracy drop with the level of sampling?
- how does the computational time drop with the level of sampling?

Figure 2.24 depicts the comparative results of the investigated sampling approaches (random, CVT, CCVT, combined) for several levels (0.05%, 0.5%, 1%, 5%, 10%) of sampling.

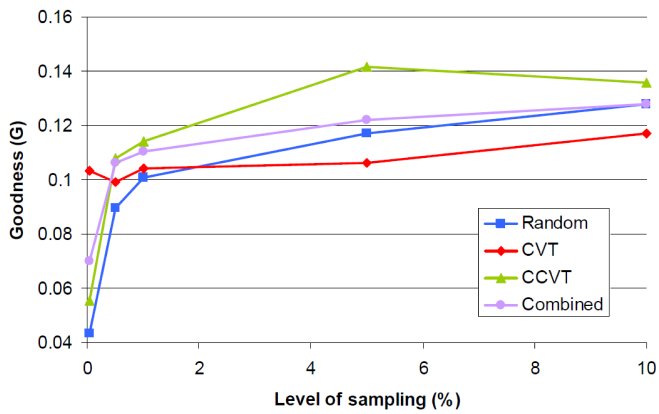


Figure 2.24: *Segmentation performance of different sampling strategies regarding the level of simplification.*

From Figure 2.24 we can see that CVT outperformed the other strategies in the case of the smallest level of sampling. However, as the sampling level increased, in all of our test cases CCVT provided the best performance. Moreover, as expected, random sampling is the less reliable one and the performance of the combination of CVT and CCVT is somewhere between these two. It is also to be noticed that random sampling is able to outperform CVT in some cases that suggests the importance of training pixels close to the object boundary. From Figure 2.24 we can also see that below the sampling level 1% the accuracy falls considerably.

In Figure 2.25, we present the global performance of the sampling approaches (random, CVT, CCVT, combined). To have an overall goodness value, we calculated the average performance for the sampling levels. We can conclude that CCVT provided the most representative sampling.

We collected the computational times of the vessel segmentation algorithm for all the test images. Naturally, we ran all of our tests on the same computer with a 2.4GHz Intel Pentium Dual-Core CPU and 2GB RAM. Our experiments indicated that only the level of sampling affected the segmentation time. Thus, we do not present computational results for the sampling methods separately. Instead, Figure 2.26 depicts the average segmentation times of all the test images and sampling methods for fixed sampling levels. We can see that a reasonable amount of computation can be saved by using a smaller number of training pixels.

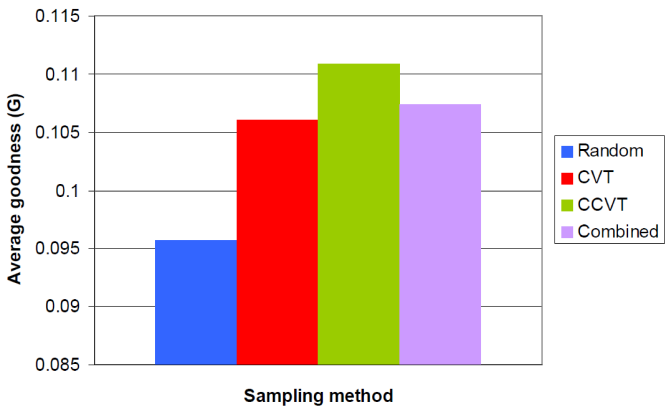


Figure 2.25: *Average segmentation performance of different sampling strategies.*

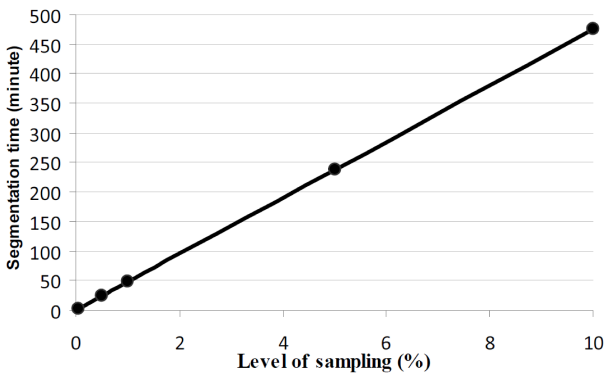


Figure 2.26: *Segmentation times regarding the level of simplification.*

Chapter 3

Piecewise linear digital curve representation and compression

In this chapter, we propose a graph theoretical approach to trace curves having arbitrary topology to obtain better compression performance, when splitting the curve into straight line segments. Because of the tracing step, the proposed method has better compression performance than the current state-of-the-art approaches. The main improvement lies in the fact that we perform a complete tracing of the curve instead of decomposing its storing canvas recursively, while only line segments remain in the quadtree cells. The results of this chapter have been published in [2, 20, 21] and are also incorporated in the project SHARE: Mobile Support for Rescue Forces, Integrating Multiple Modes of Interaction, EU FP6 Information Society Technologies, Contract Number FP6-004218.

Besides detection, a common task related to objects is their efficient representation and storage. In this chapter, we present a method for the compression of digital curves using graph-based representations and algorithms.

Digital planar curves are used in several fields of computer graphics, discrete geometry and digital image analysis. Many results have been produced regarding their geometric behavior since [117]. A special topic is digital curve compression. Besides simple techniques like chain coding, a usual way is to partition the curve into straight line segments [118] for compression. These methods usually focus on simple curves with no self-intersections, and assume the preliminary knowledge on the sequential order of the curve points. The state-of-the-art approach JBEAM [119] considers an alphabet of short line segments (called beamlets) to compose the curve. This method divides the binary image containing the curve using quadtree decomposition till having a single linear curve segment in every quadtree cell that can be substituted by a beamlet. The advantage of this approach is that any curve can be handled by sufficiently fine quadtree decomposition. However, a drawback is the obligation of decomposing subsequently, when a cell contains such segments that already could have been coded separately.

In this chapter, we propose a graph theoretical approach to trace curves having arbitrary topology to obtain better compression performance, when splitting the curve into straight line segments. Because of the tracing step, the proposed method has better compression performance than JBEAM [119]. The main improvement lies in the fact that we perform a complete tracing of the curve instead of decomposing its storing canvas recursively, while only line segments remain in the quadtree cells.

The structure of this chapter is as follows. In section 3.1, we recall the graph theoretical background that serves as a basis for our approach in tracing curves. We also explain how the suitable graph representation of the digital curve is obtained. Section 3.2 describes how the tracing is optimized regarding coding the curve with straight line segments. The method selected for compression is presented in section 3.3. Section 3.4 contains our comparative analysis regarding other state-of-the-art approaches. We explain some variants of the basic approach in section 3.5 with highlighting their advantages and drawbacks. Some open issues and other possible applications are discussed in section 3.6 with presenting some of our later results regarding the improvement of the skeletonization of intersections of retinal vessels in section 3.7.

3.1 Tracing curves using graph theory

In this section, we recall some concepts of curve theory that we apply to trace a curve and also some techniques that were considered to obtain the corresponding graph representation of the curve. The necessary graph theoretical notations have been introduced in section 1.2. Additionally, we recall some well-known facts on Eulerian graphs and their decomposition (see e.g. [120, 121]):
i) Every Eulerian graph is connected.
ii) A connected graph contains an Eulerian cycle if and only if all of its vertices have even degree.
iii) A connected graph contains an Eulerian walk if and only if at most two of its vertices have odd degree. If there are two vertices with odd degree, the walk starts from either of them and ends in the other.
iv) Every connected graph has an Eulerian decomposition.

3.1.1 Assignment of a graph to a digital curve

The definition of simple curves in the Euclidean space was given by P. Urysohn in 1923 and K. Menger in 1932 independently (see [122] for a review). The curves were classified based on the branching indices of the curve points, where a branching index of a curve point is equal to the

number of curve segments meeting at the given point. The adequate mathematical formulation for the Euclidean space can be found in [122, 123]. For the discrete domain $\mathbb{Z}^2$, this definition can also be adapted using the well-known 8-neighboring relation. More precisely (see [122, 123]):

- a digital curve C has branching index $B(\boldsymbol{p}) > 0$ at its point $\boldsymbol{p} \in C$, if and only if exactly $B(\boldsymbol{p})$ 8-neighbors of $\boldsymbol{p}$ belong to C,
- $\boldsymbol{p} \in C$ is a regular point, if and only if $B(\boldsymbol{p}) = 2$,
- $\boldsymbol{p} \in C$ is a branch point, if and only if $B(\boldsymbol{p}) \geq 3$,
- $\boldsymbol{p} \in C$ is an end point, if and only if $B(\boldsymbol{p}) = 1$,
- the 2D digital curve C is simple, if and only if all of its points are regular.

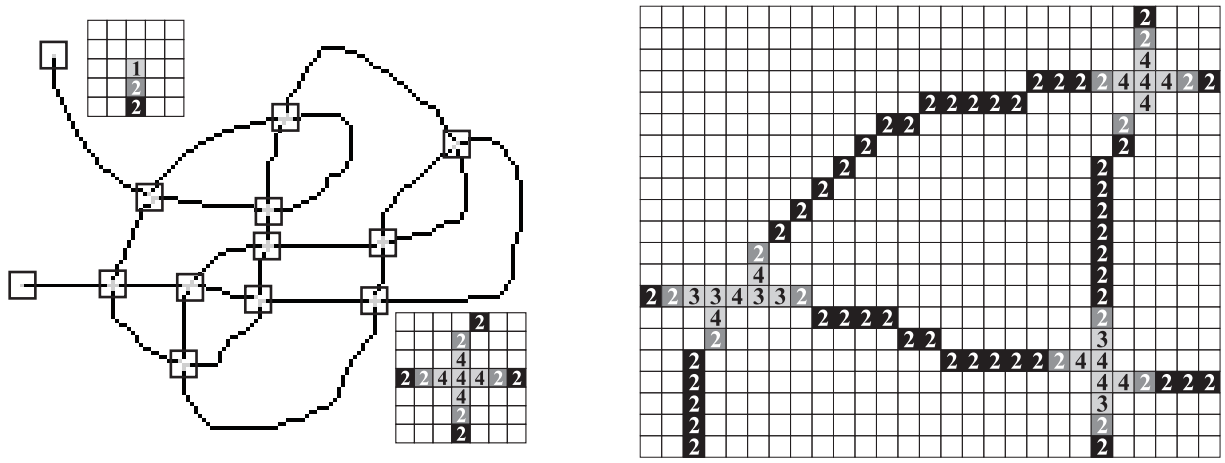
To make graph theoretical algorithms applicable to digital curves, we also need a precise concept for a junction (see [122, 123]). An 8-connected region of branch points is called a junction. The branching index of a junction $J \subseteq C$ is the number of regular or end points of C being 8-neighbors of any of the branch points of J. We mention here that this classic approach for defining the junctions of digital curves might be restrictive in some applications, which issue will be discussed in section 3.6.

Using all the above definitions, we are ready to assign an abstract curve graph to a curve (see [123]). The abstract curve graph $\mathcal{G}_C = (V_C, E_C)$ of the curve C is an undirected multigraph, where the vertices in V_C are either junctions or end points of C. Two vertices are connected by an edge, if and only if the corresponding junctions or end points are 8-connected. Thus, the degree of a vertex is just the branching index of the corresponding junction.

As for the technical details of the extraction of an abstract curve graph, we consider 1-pixel wide curves. If the input curve is not 1-pixel wide, we can apply a preliminary thinning step on it. To determine the edge set E_C, we locate the end points of the edges as regular points being 8-neighbors to junctions (if both of their 8-neighbors are branch points, the edge is degenerated having length 1). Then, the edge end points are organized into pairs (edges) based on the condition that an 8-connected path can be found between them, whose elements are regular points. Figure 3.1(a) depicts the result of locating end points and junctions (shown framed, in light gray), while Figure 3.1(b) takes a closer look for the selection of edge end points (dark gray), and for the edges defined by them. These figures also indicate the branching indices of the curve points.

Notice that loops and multiple edges are also handled by this approach without any difficulties. To find the 8-paths between edge end points we can use the recursive Floodfill8 algorithm [124] starting from the edge end points. The simplified abstract curve graph representation of the curve in Figure 3.1(a) is shown in Figure 3.2. The vertex indices are assigned in the order of the vertex scanning procedure in the figure.

Now we are ready to summarize our main approach in the following curve tracing (CT) algorithm.



(a) (b)

Figure 3.1: *Locating vertices and extracting the edges for the abstract curve graph $\mathcal{G}_C = (V_C, E_C)$; (a) input test planar curve C with its end points and junctions to compose V_C are framed, (b) extracting edges for E_C via locating edge end points (dark grey) and connecting them with 8-paths.*

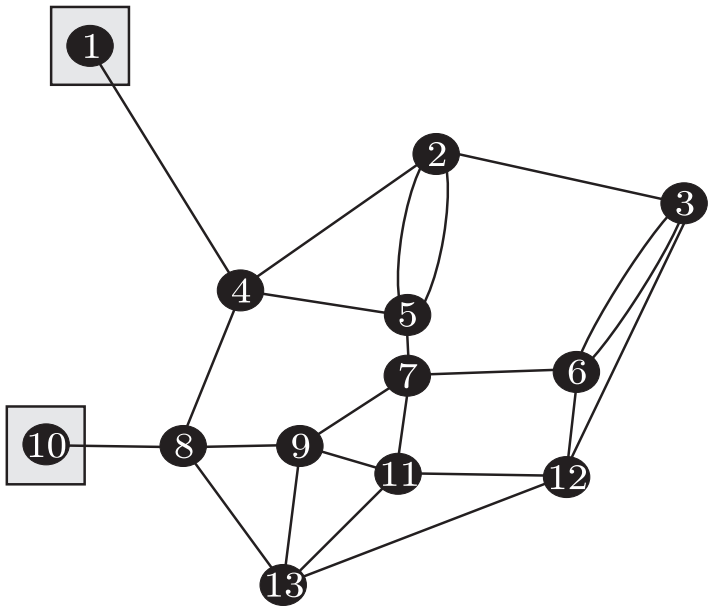


Figure 3.2: *The simplified abstract curve graph of the curve shown in Figure 3.1(a) with vertices of odd degree framed.*

Algorithm 3.1.1. Curve tracing (CT) algorithm.

1. *Extract the abstract curve graph* $\mathcal{G}_C = (V_C, E_C)$ *of the curve* C.
2. *Create an Eulerian decomposition* $\bigcup_{i=1}^{n} C_i$ *of* C *based on* $\mathcal{G}_C$.
3. *Trace all the* C_i*'s separately through their Eulerian paths.*

3.2 Optimized tracing for compression

The first step to trace an Eulerian curve is to locate a starting vertex according to statement *iii)* at the beginning of section 3.1. We check the vertices and select one having odd degree. If all the degrees are even, we can choose an arbitrary vertex to start from. Then we take an edge from the starting vertex to initialize the tracer. For example, in the graph shown in Figure 3.2 two vertices (1, and 10) have odd degrees. Thus, the Eulerian path should start from vertex 1 to finish at vertex 10, or vice versa.

As more Eulerian paths may exist, we have to decide which edge to take next, when reaching junctions. As our intention is to substitute the curve part with straight line segments for curve compression, the natural decision is to go on straight ahead at junctions. Thus, let us assume that we arrive at an edge end point $\boldsymbol{e_0}$ at a junction that has edge end points $\boldsymbol{e_0}, \boldsymbol{e_1}, \ldots, \boldsymbol{e_k}$ not visited yet. We calculate the centroid of the junction by

$$\overline{\boldsymbol{e}} = \frac{1}{k+1} \sum_{j=0}^{k} \boldsymbol{e_j}. \tag{3.1}$$

The centroid $\overline{\boldsymbol{e}}$ can be rounded to have integer coordinates, or considered as a real valued vector, as well. Let α_i denote the angle $\angle \boldsymbol{e_0}\overline{\boldsymbol{e}}\boldsymbol{e_i}$ for $i = 1, \ldots, k$. The traversal of the curve is chosen to correspond to the edge end point $\boldsymbol{e_l}$ for which

$$|180^\circ - \alpha_l| = \min_{j=1}^{k}\{|180^\circ - \alpha_j|\}. \tag{3.2}$$

To extract a path between $\boldsymbol{e_0}$ and $\boldsymbol{e_l}$ we can use the Floodfill8 algorithm again. Now, we have to start Floodfill8 from $\boldsymbol{e_0}$ to flood the vertex points of the junction with selecting the path with minimal length. See Figure 3.3 for an example on how the decision is made to trace through a junction based on the above discussion. The selected path through the junction contains $\boldsymbol{e_0}$, $\overline{\boldsymbol{e}}$, and $\boldsymbol{e_2}$ and indicated by $\times$ marks in the figure.

Using this junction traversal decision, we are able to trace the whole curve along one of its Eulerian paths. The traced curve is composed by concatenating the edges with the short segment going through the vertices. For the complete tracing see Figure 3.4, where beside the start and end point of the Eulerian path, arrow heads show the optimal directions at the curve junctions. Borrowing the vertex numbering from Figure 3.2, the Eulerian path is:

$$\{1, 4, 5, 2, 4, 8, 13, 11, 7, 5, 2, 3, 6, 7, 9, 13, 12, 6, 3, 12, 11, 9, 8, 10\}.$$

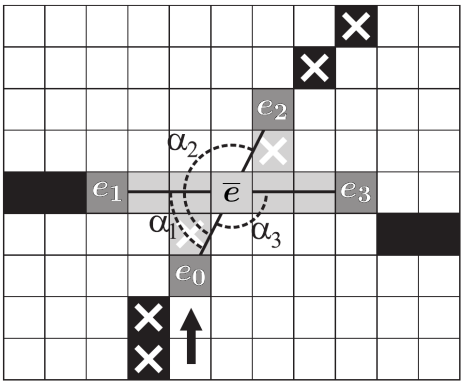


Figure 3.3: *Optimal curve tracing through a junction with finding the most straight direction and connecting corresponding edge end points.*

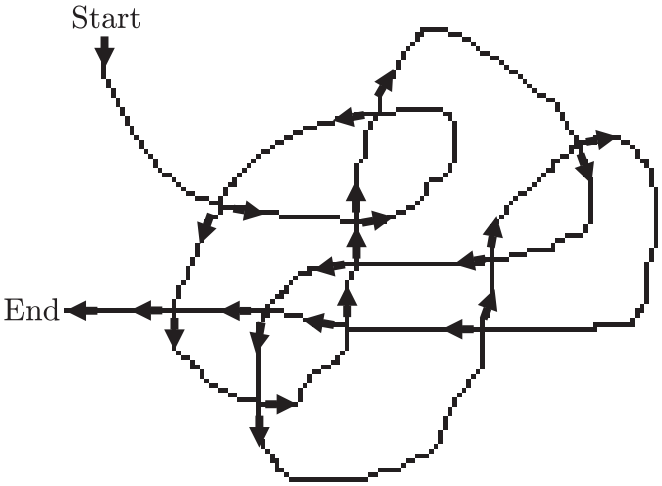


Figure 3.4: *Tracing the whole curve by choosing optimal directions at junctions.*

During the extraction of an Eulerian path, it should be noted that we are not always free in choosing the most straight direction at junctions. Fleury's algorithm [125] guarantees to find an Eulerian path, and we have to combine our junction traversal method with the classic graph theoretical recommendations (see e.g. [125, 126]), which briefly are:

- always leave one edge available to get back to the starting vertex or to the other odd vertex,
- do not use an edge to go to a vertex unless there is another edge available to leave it.

Notice that the graph theoretical recommendations have higher priority than optimal traversal selection to guarantee the proper Eulerian path.

3.3 Curve compression by an alphabet of line segments

We can choose from a vast number of techniques to partition a curve into digital straight line segments [118]. These techniques can be classified as offline (the curve is examined globally to find an optimal partitioning), or online (the curve is decomposed into line segments during its traversal). Though our proposed approach is suitable for both tasks, we discuss an online coding possibility here. To partition the curve into digital straight segments we use a linear online method presented in [127].

To obtain a coding scheme from the straight line decomposition, we replace all the produced linear curve segments by elements of an alphabet of line segments. We create a finite alphabet Λ whose letters are digital line segments of all possible orientations having length at most T pixels. As an obvious consequence, we have to stop processing the curve when the maximal segment length T is reached and we have to look for the next segment, even if the coded one would continue straight. Moreover, to keep the cardinality of Λ small, we consider unique straight line segments to connect two points. For this purpose, we consider the Bresenham line drawing algorithm [128] to create the letters of Λ. Notice that this way we allow some information loss, since the Bresenham segments may slightly differ from the ones extracted during the online curve segmentation process. On the other hand, these differences are really minor perceptually, since digital straightness is our essential requirement. It is easy to prove that the cardinality of Λ, and the number of bits needed for coding a letter can be calculated as

$$|\Lambda| = 4T(T-1), \text{ and } \log_2 |\Lambda| \leq 2(\log_2 T + 1), \tag{3.3}$$

respectively. As an example for such an alphabet see Figure 3.5 for $T = 6$ with the letters shown only for the Cartesian coordinate domain $0 \leq y \leq x$. Thus, the alphabet also contains the letters obtained from the shown set by rotations of $0°$, $\pm 90°$, $180°$, and by mirroring to the lines $x = 0$, $y = 0$, $y = x$, and $y = -x$, respectively.

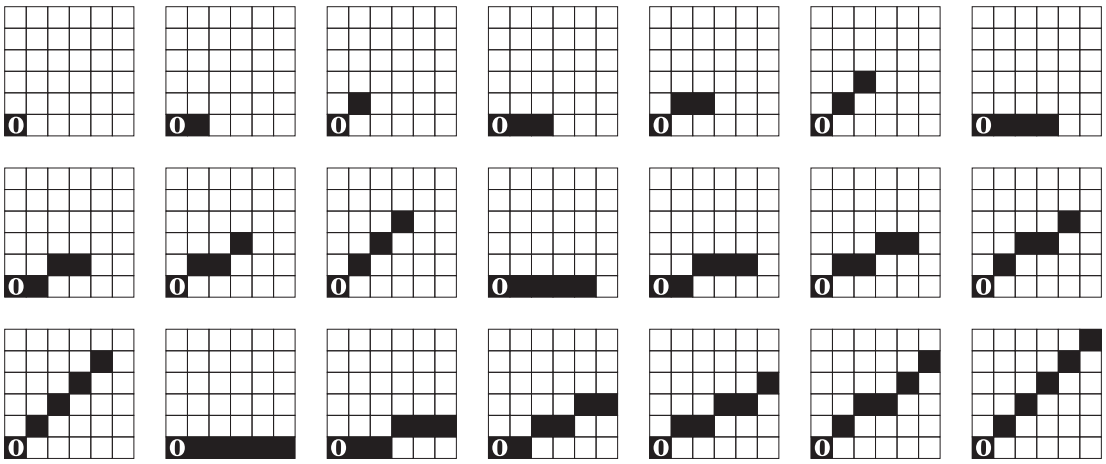


Figure 3.5: *An example alphabet Λ for $T = 6$ (line segments of length at most 6). Only the letters belonging to the domain $0 \leq y \leq x$ are depicted.*

To check the compression efficiency of our method in comparison with other state-of-the-art methods, we considered a dataset of typical test curves shown in Figure 3.6(a)-(d). To demonstrate the extendibility of our method to such curves that cannot be traversed with a simple Eulerian path, we consider another example shown in Figure 3.6(e).

Notice that the graph representation of Figure 3.6(e) contains four vertices having odd parities. Consequently, it should be decomposed at least into two Eulerian paths. In our example, starting from any of the odd vertices and judging by the linearity criteria at the junctions, we decompose the original curve into the two segments shown in Figure 3.6(f) and Figure 3.6(g), respectively. To compress the original curve, we compress these Eulerian paths separately.

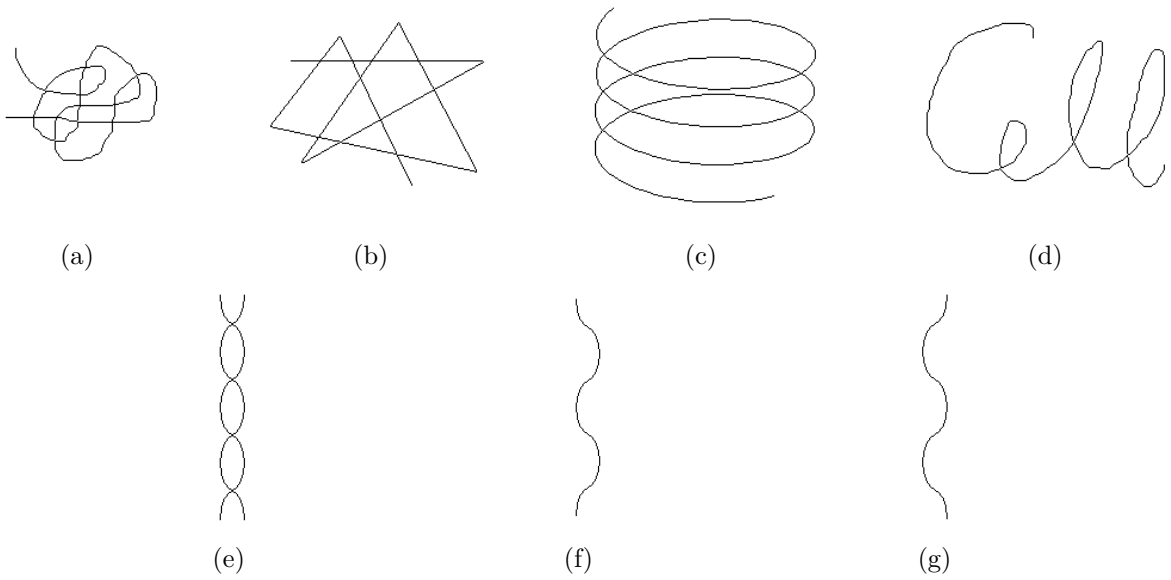


Figure 3.6: *Test curves of different types; (a) General, (b) Lines, (c) Spring, (d) Script, (e) Non-Eulerian, (f)-(g) Eulerian paths to compose the Non-Eulerian curve.*

3.4 Comparative analysis

3.4.1 Comparing with JBEAM

To test and compare the compression efficiency of our method, we fixed the following setup. We used $T = 32$ as a threshold for the maximum line segment length for all the test curves, and considered the default lossy JBEAM parametrization [119]. Our experimental results are shown in Table 3.1.

Test curve	# of pixels	JBEAM (# of bits)	Proposed method (CT)	
			# of bits	# of segments
General	2127	1586	744	62
Lines	2745	1398	468	39
Spring	4113	2308	1224	102
Script	2511	1419	828	69
Non-Eulerian	1242	834	$2 \times 240 = 480$	$2 \times 20 = 40$

Table 3.1: *Comparative quantitative results with JBEAM.*

We can conclude that the proposed method has a 50% improvement on average in compression against JBEAM. Figure 3.7 depicts the coding results for our sample curves. We marked the end points of the line segments found by our coding method. For the sake of completeness, we mention that the coordinates of the starting point of every Eulerian path should be stored, as well. However, we ignored this issue in our calculations, since it produces only insignificant increase in the number of bits of the compressed curve.

3.4.2 MPEG-4 contour-based shape coding

Within MPEG-4, a vertex-based shape approximation was developed to code the outline of shapes [129]. For image/video transmission, the usual task is to transmit region-like shapes, thus, their

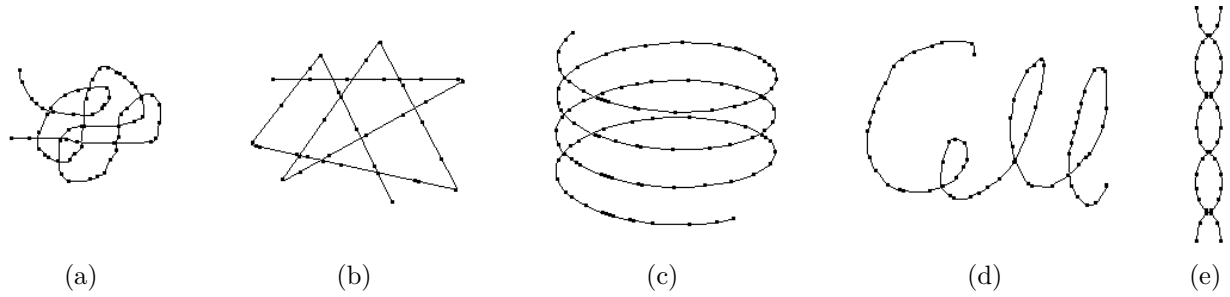


Figure 3.7: *The partitioning of the test curves into line segments; (a) General, (b) Lines, (c) Spring, (d) Script, (e) Non-Eulerian.*

boundaries can be represented by simple closed curves. In MPEG-4, the boundary of the shape is approximated by a polygon for lossy shape coding. For lossless shape coding, the polygon approximation degenerates to chain coding [130]. The polygon is found through a recursive splitting process that starts with the longest axis (diameter) of the shape as an initial polygon $\overline{\boldsymbol{v_0 v_1}}$. A polygon segment $\overline{\boldsymbol{v_k v_{k+1}}}$ is associated with the curve part C_k composed by the points $\boldsymbol{t_{k,i}}$ with $i = 0, \ldots, |C_k| - 1$. The approximation error at C_k is defined as

$$d_{\max}(k) = \max_{i=0}^{|C_k|-1} d(\boldsymbol{t_{k,i}}, \overline{\boldsymbol{v_k v_{k+1}}}) \tag{3.4}$$

using the Euclidean distance d. Now, if $d_{\max}(k) > d^*_{\max}$ for a fixed threshold $d^*_{\max}$, then $\boldsymbol{t_{k,j}}$ is selected as a new polygon vertex, where $d(\boldsymbol{t_{k,j}}, \overline{\boldsymbol{v_k v_{k+1}}}) = \max_{i=0}^{|C_k|-1} d(\boldsymbol{t_{k,i}}, \overline{\boldsymbol{v_k v_{k+1}}})$. In other words, we recursively split those polygon segments which are not sufficiently close to the curve. The boundary point having largest distance from the polygon segment is selected as a new vertex. For better coding performance, we shift the polygon vertices to re-index them so that the first and last vertices have largest difference between their horizontal or vertical coordinates. After storing the position of $\boldsymbol{v_0}$, each remaining vertex position is encoded by a difference vector from its predecessor in the form $\boldsymbol{v_d} = \boldsymbol{v_k} - \boldsymbol{v_{k+1}}$. Finally, the components of the difference vectors are coded further by variable-length (e.g. Huffmann [131]) coding tables.

Though our main intention in the chapter is to code non-simple curves, our scheme can be naturally applied to simple closed contours, as well. Technically, the obvious way is to break the connection between any two points of the curve, which leads to an abstract curve graph representation having a single edge. Since straightness is the main condition for the proposed method, we have to find a corresponding MPEG-4 threshold $d^*_{\max}$. From [118] we know that a finite arc is a digital straight segment, if and only if, its points are between or on a pair of parallel lines having a main diagonal distance of at most $\sqrt{2}$. Thus, $d^*_{\max} \leq \frac{\sqrt{2}}{2}$ should hold to make the approximated curve partitions be straight line segments, as well. However, in our experiments we found that the above MPEG-4 approach leads to worse performance than the proposed technique with respect to the number of polygon segments with $d^*_{\max} = \frac{\sqrt{2}}{2}$. The main reason is that the selection of new vertices is ambiguous, since more curve points can reside at the largest distance from the approximating polygon. We found that approximately the same performance can be achieved by having $d^*_{\max} = 1$. Though with $d^*_{\max} = 1$ we slightly hurt the criterion of straightness, the subjective perceptual performance is reported to be acceptable for a human observer up to $d^*_{\max} = 1.4$ [129]. For a comparative analysis, we consider some simple closed curves (shape boundaries) shown in Figure 3.8. The polygon vertices found by the MPEG-4 method with $d^*_{\max} = 1$ are also marked with large dots in the figure.

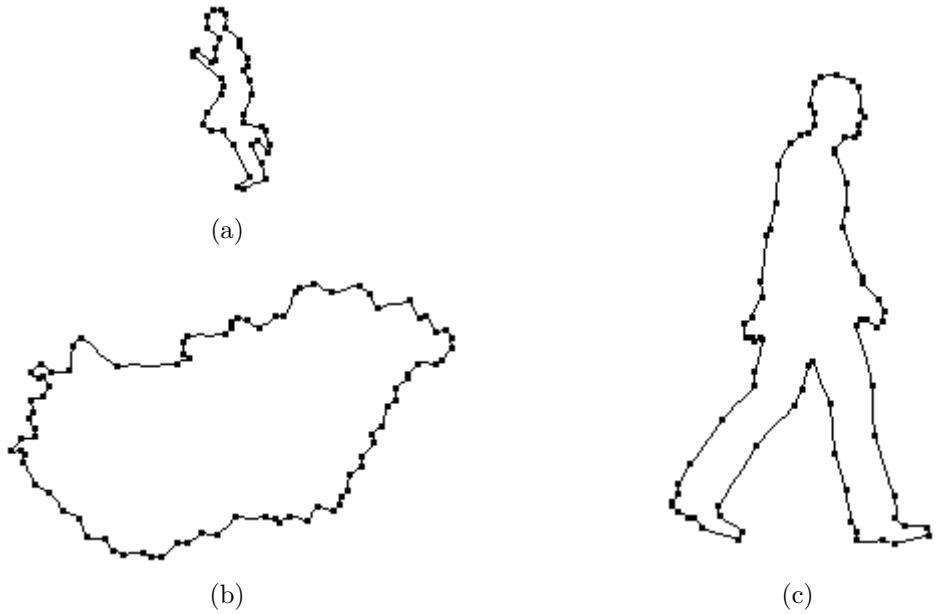


Figure 3.8: *Closed curves for a comparative study with MPEG-4 technique. The vertices of the approximating MPEG-4 polygons are also indicated; (a) Running, (b) Hungary, (c) Walking.*

Test curve	# of pixels	# of segments		# of bits		# of bits (Huffmann coding)	
		MPEG-4	Proposed	MPEG-4	Proposed	MPEG-4	Proposed
Running	238	41	36	492	408	239	207
Hungary	593	88	86	1056	1032	439	435
Walking	753	81	72	972	864	410	378

Table 3.2: *Comparative quantitative results with MPEG-4 coding.*

Table 3.2 contains the corresponding quantitative results. We also present data to see the additional coding improvement we can gain by applying a consequent Huffmann coding.

There are some more recommendations to achieve minor improvements using the MPEG-4 coder which would be adaptable to the proposed scheme, as well. On the one hand, for lossy shape coding, the selection of the vertices on the object boundary might not be optimal. Therefore, the vertices can be shifted by 1 pixel within a neighborhood of size 3×3. On the other hand, the maximal length of the alphabet elements can be fixed dynamically as the length of the longest polygon segment found. Notice that in this case, this length information should be transmitted, as well.

3.5 Alternative compression approaches

We can apply slightly different approaches for the compression of the curves regarding the ones presented in the previous section. In all the cases, we consider the same alphabet Λ of line segments introduced previously, and the same online method to partition the curves into line segments.

3.5.1 Compressing edges separately

The process presented in section 3.1.1 is suitable to extract the edges of an abstract curve graph, where the edges are actually curve segments. We might as well execute our compression approach

at edge level without the intention to perform any tracing of the curve. Though this simple approach avoids the oversegmentation of the quadtree cells considered by JBEAM [119], it has several drawbacks. In this case, curve segments (edges) are compressed separately, so we have to store the coordinates of the start pixel of each curve segment for appropriate geometric positioning. Moreover, since no junction point information is stored, we have to connect the consecutive curve segments during the decoding process (e.g. using the Bresenham algorithm [128]), which may lead to curve distortion at junctions.

The compression rate can be calculated easily, as now we have to summarize simply the number of line segments that are needed to compress the separate curve segments. Also for further use, we introduce the corresponding weight function $w : E_C \to \mathbb{N}$ for the edges. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{G}_C$. Then the weight (cost) of the edge $\{\boldsymbol{u}, \boldsymbol{v}\}$ is defined as

$$w(\{\boldsymbol{u}, \boldsymbol{v}\}) = \# \text{ of line segments needed to compress } \{\boldsymbol{u}, \boldsymbol{v}\}. \tag{3.5}$$

Thus, the curve can be stored using

$$\sum_{\{\boldsymbol{u},\boldsymbol{v}\} \in E_C} w(\{\boldsymbol{u}, \boldsymbol{v}\}) \tag{3.6}$$

letters from Λ. For an example, see Figure 3.9, where the weights w are shown for the General test curve.

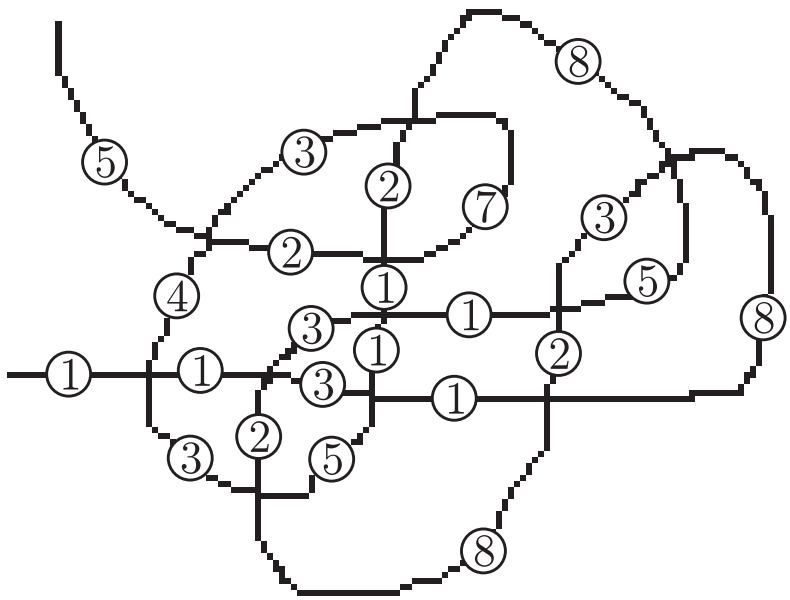


Figure 3.9: *Weighting the abstract curve graph $\mathcal{G}_C$. The weights of the edges (curve segments) correspond to the number of line segments needed to code the edges, respectively.*

Moreover, check Table 3.3 to have a comparison for all our test curves. From the table we can see that besides covering junctions, the CT method presented in section 3.3 has better compression performance.

3.5.2 Curve compression without partitioning

As we discussed in section 3.3, our basic approach considers an Eulerian decomposition of the curve to avoid redundancy in coding edges. However, we can skip this step by finding a path that

Test curve	Compressing edges separately		
	# of edges	# of bits	# of segments
General	23	948	79
Lines	21	600	50
Spring	15	1236	103
Script	8	864	72
Non-Eulerian	10	672	56

Table 3.3: *Experimental results of separate (curve part) edge compression.*

may contain some edges more than once. This sort of graph theoretical problem is known as the Chinese Postman Problem (CPP for short) [132, 133].

In this case, we consider the abstract curve graph $\mathcal{G}_C$ as a weighted one, where as a natural choice we can use the weight function defined in (3.5). In general, we can solve a CPP problem through the following steps.

1. Calculate the shortest length of a path between every two vertices with odd parities using the Floyd algorithm [134].

2. Perform a perfect maximum matching [131] for the set of vertices with odd parities. That is, organize these vertices into pairs in such a way that the sum of the corresponding path lengths is minimal.

3. Find the shortest paths using Dijkstra's algorithm [131, 135] between the vertex pairs found in step 2), and add these paths as artificial edges to the graph.

4. Now the graph becomes an Eulerian one, so an Eulerian path can be found, as discussed in section 3.2.

Let us consider the example in Figure 3.10(a), which actually depicts a slightly different variant of the General test curve.

Now, as can be seen from its simplified graph representation in Figure 3.10(b), this curve is not an Eulerian one, as it has six odd degree vertices (1,2,3,12,16,17). After calculating edge weights according to (3.5), we can organize these vertices into the optimal pairs (1,12), (2,3), (16,17). The respective distance of these pairs are 10, 3, 6 (with summing the edge weights in the shortest path between them). Thus step 3) of the CPP algorithm will provide the (redundant) artificial edges $\{2, 3\}$ and $\{16, 17\}$ to be added to the graph. These edges are shown with dashed lines in Figure 3.10(b). Notice that this way, the graph becomes an Eulerian one, having vertices 1 and 12 of odd degree. Finally, using the approach discussed in section 3.2, the following Eulerian path is found in our example:

$$\{1, 6, 7, 4, 6, 10, 15, 13, 9, 7, 4, 3, 2, 3, 5, 8, 9, 11, 15, 16, 17, 16, 14, 8, 5, 14, 13, 11, 10, 12\}.$$

This path is also shown in Figure 3.11 using bidirectional arrows, where the path travels edges $\{2, 3\}$ and $\{16, 17\}$ twice.

The compression performance obviously drops when the CPP approach is used in comparison with the CT one, as some curve segments are stored more than once. For example, while the CT algorithm needs $62 + 6 + 3 = 71$ line segments to code the curve in Figure 3.10(a), the CPP approach needs $62 + 2 \times 6 + 2 \times 3 = 80$ ones. Notice that according to Table 3.1, the CT algorithm needed 62 line segments without edges $\{2, 3\}$ and $\{16, 17\}$ for the compression. Besides its simple idea, the advantage of the CPP method lies in the fact that the curve does not need to be broken into Eulerian parts, and thus only the coordinates of the start pixel should be stored.

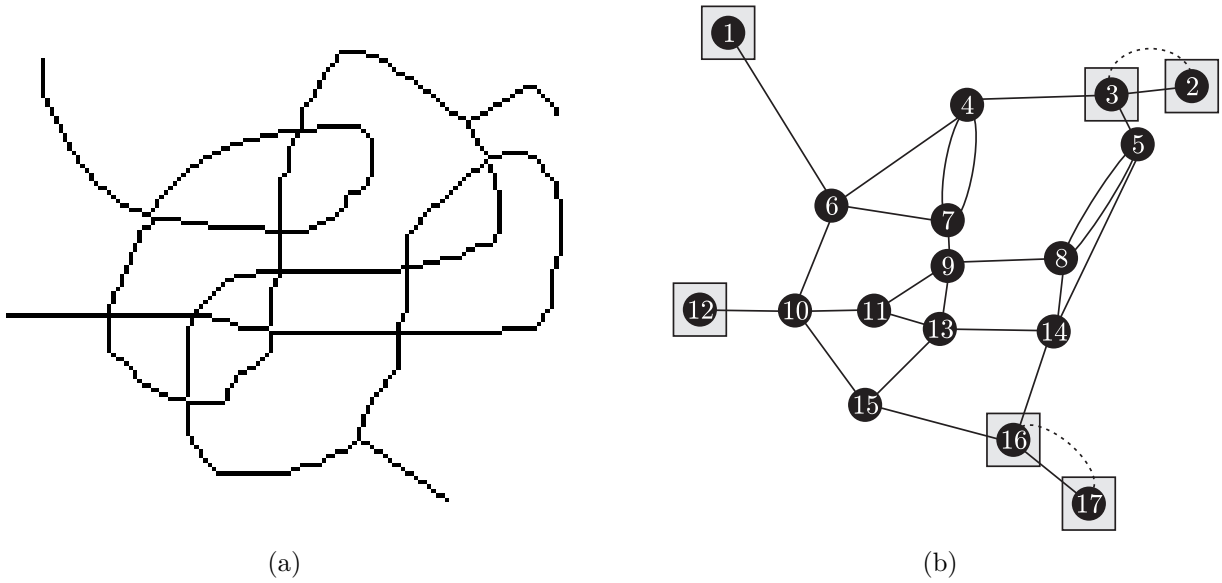


Figure 3.10: *An example for the CPP algorithm; (a) Non-Eulerian variant of the General curve, (b) adding new edges (dashed) to have an Eulerian abstract curve graph.*

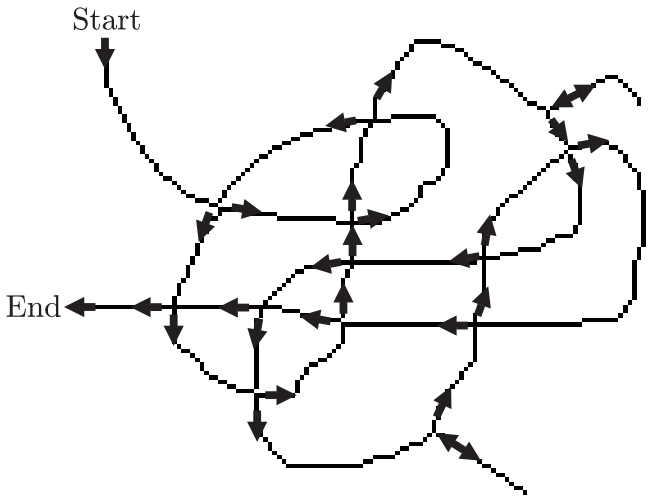


Figure 3.11: *Tracing a non-Eulerian curve using the CPP algorithm (bidirectional arrows correspond to edges to be taken twice).*

3.6 Discussion on curve compression

We must mention that JBEAM [119] also realizes a progressive approach in curve compression. That is, when only a part of bits has been received, the digital curve can be recovered approximately. Our method does not have a direct support for progressive encoding, however, some kind of progressivity could be reached by re-ordering the transmitted Bresenham segments. For example, to have an impression about the shape of the curve, we can send the segments having indices equidistantly sampled along the graph route. Another approach may relate to the case, when junctions are more important. Then we can transmit those segments first which contain

or touch branch points. In all these cases, additionally data are needed to be transmitted, since the segments are no longer consequent ones. The necessary extra amount of such data is an open issue.

In our approach, it is a key consideration to keep linearity when replacing curve segments with Bresenham ones. Minor deficiencies may occur from the fact that we use different methodology for judging on linearity and replacing with fixed linear line segments. In other words, if we fix two endpoints, more curve segments between them can be judged as linear but replaced with the same Bresenham one. This approach naturally leads to some kind of distortion, however, the linear behavior is preserved perceptually. For a given Bresenham segment, the expected rate of the distortion is related to the possible number of segments judged as linear between its endpoints. Consequently, longer segments could cause larger distortion on average, however, we need smaller number of longer segments to code the curve, since the number of the curve points is fixed. Accordingly, though it might be a future topic to clarify the precise relation between the maximal length of the Bresenham alphabet and the distortion, we do not expect remarkable differences.

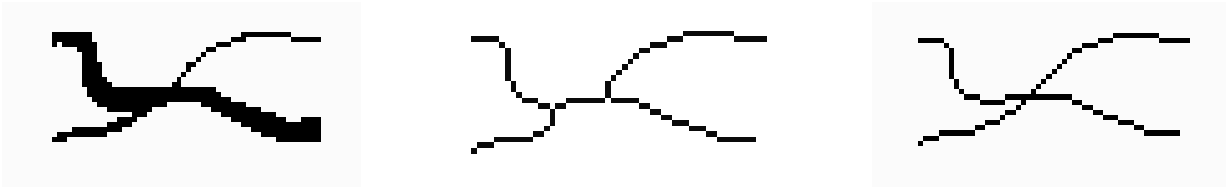
In several applications, we can encounter with such intersections where the local tangents of the intersecting curve segments almost coincide. For example, we can think of an intersection of nearly horizontal lines. In the Euclidean case, we would not have any difficulties, but in the digital domain the intersection consists of more than one pixel. One possibility to overcome this difficulty is to revise the definition of a junction, and to apply a distance threshold $\mathcal{T}$. Then we can collect branch points closer than $\mathcal{T}$, and merge into a junction the regular points between them. In other words, we decide upon the allowed closeness of the tangents. The threshold value $\mathcal{T}$ can be adjusted according to the expected behavior of the curves in the application. A possible improvement of this simple procedure is described in section 3.7.

The steps we considered in our approach can have individual importance in other application fields, as well. Our intention to "untie" curves can have impact in curve watermarking [136, 137], where the capability to provide the input data in terms of few large blocks is highly welcome. The ability of tracing complex curves looks feasible in reconstructing hand-written text or figures, similarly e.g. to [138]. The compression method can also be used to efficiently store the boundaries of shapes, e.g. the templates of a human silhouette database.

3.7 Thickness-based binary morphological improvement of line intersections

The motivation of the efforts discussed so far in this chapter has been the efficient representation of arbitrary digital curves with a specific interest in closed object boundaries (e.g. human silhouettes). However – similarly to section 2.3.3 –, we have investigated whether some of our findings can be exploited in other fields, as well. Accordingly, in this section, we present some results on the improvement of distorted digital line intersections appearing at the crossings of retinal vessels. Namely, we have worked out a more precise discrete geometric approach to improve the quality of the intersections of digital curves extracted by skeletonization instead of using the simple definition of a junction given in section 3.6.

Skeletonization [114] (see also section 2.3.2) is a binary morphological operation to extract the centerline of an object. The skeleton is often found by thinning the object with removing pixels without affecting the general shape. The skeleton is a popular object representation, since pixel-wise observations can be used to characterize the spatial behavior of the objects. For example, junctions can be located as pixels with at least three 8-neighbors. However, the skeleton is usually distorted as it is illustrated in Figure 3.12, with having two junction pixels instead of a single one. In such cases, the intersecting elements cannot be tracked properly later on.



(a) (b) (c)

Figure 3.12: *Distortion of an intersection during skeletonization; (a) original image, (b) its skeleton, (c) desired skeleton.*

Several approaches have been released to overcome this problem. Basically, these methods can be grouped as skeleton-based or global approaches. In the skeleton-based case, the distorted locations are found at the skeletonized images, while in the global case the intersections are tried to be localized in the original line drawing image. The usual drawback of a skeleton-based decision is that it needs a proximity parameter to differentiate between true intersections and enclosed segments connecting intersection points and the definition of this parameter is challenging for the whole image. Global approaches are usually based on corner detection [139, 140] inheriting the typical related problems, like too many/few corner candidates. However, their strength is that they try to locate and improve distorted intersections without altering the original image. We have encountered with this intersection distortion problem during the improvement of the skeleton of the retinal vascular system. As a specific field of automatic screening of diabetic retinopathy, the proper mapping of the vascular system has very important role in gathering information to diagnose diseases. For example, a proper traversal of the vessels gives information about how the thickness of the vessels' changes, or about the artery/vein ratio. After applying a segmentation method [116], a binary image (see Figure 3.13) can be extracted with similarities to line drawing ones.

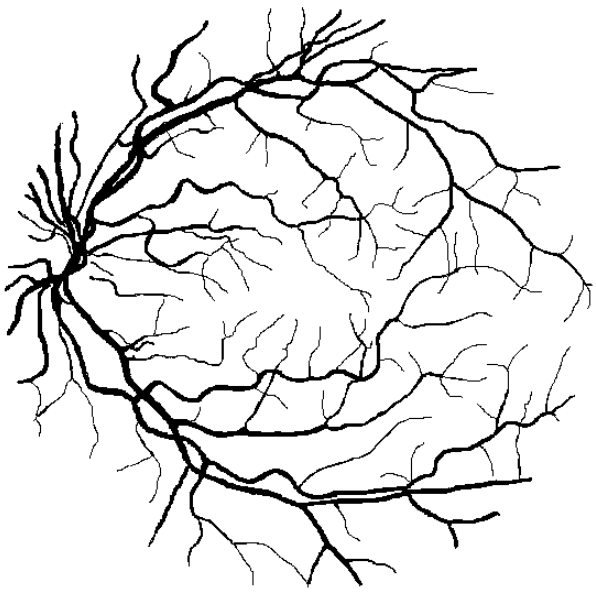


Figure 3.13: *Binary vessel map of the retina.*

In this section, we propose a global technique for suppressing distorted intersections in the

skeleton that can be considered as both a stand-alone or a complementary approach to others mentioned before. Our method is based on the separation of the thick and thin vessels of the complete vascular system (input image) before applying skeletonization. The main motivation behind this approach is to be able to extract the precise skeletonization of the thick vessels at the intersection, while the skeletonization of the thin components is executed separately. However, since the vessel system is split into two parts, a consecutive reconnecting step is needed to connect the thin skeletal elements to the thick ones.

In [21], we have given the theoretical model that is considered to measure the distortion as a function of the thickness of the intersecting elements. Here, for the intersection of two vessels, we consider two stripes bounded by pairs of parallel lines. That is, our model considers three parameters: the width of the intersecting stripes and the angle enclosed by them. To measure the degree of the distortion (DD) of the intersection, a natural error function can be defined as the distance between the desired and actual junction points. The theoretical calculations to determine this error for given width and angle parameters are omitted here; see [21] for more details. We have found that the degree of distortion is large for stripes having large width differences. This observation motivated the separation of the vessels based on their widths.

Our approach to improve the skeleton of binary vascular images is based on two major steps. First, we split the vascular system into two parts containing the thick and thin vessel segments and perform the skeleton of the two sets. Then, we reconnect the components disconnected in the splitting step. The separation of the thick and thin vessels of the whole vascular system is performed by applying erosion steps that remove pixels having less than four 8-neighboring pixels. We repeat the procedure recursively until no more changes are found in the image. By this process, thin blood vessels are erased with the thick vessel subsystem preserved. Then, the thin vascular subsystem can be trivially generated by subtracting the thick one from the original binary image. The result of this splitting procedure is shown in Figure 3.14 for the input binary vascular system depicted in Figure 3.13.

(a) (b)

Figure 3.14: *The result of the splitting step; (a) the thick vessels subsystem, (b) the thin vessels subsystem.*

We note that our approach for erosion is slightly different from the classic morphological operation. Namely, our approach preserves all the vessels above a given width to guarantee that the thick vessels remain connected. After successfully splitting the vascular system into two parts, we perform the skeletonization of the thick and thin subsystems. For this purpose, any skeletonization algorithm can be used. In our implementations, we considered the one recommended by Deutsch [141].

When taking the union of the skeletons of the thick and thin vessel subsystems, several discontinuities (gaps) remain to be filled in. This phenomenon is also depicted in Figure 3.15. To fill in such gaps, we perform direction estimations at the corresponding endpoints of the thin skeleton to connect them to the thick skeleton subsystem. The direction estimation is performed by following the classic recommendation of discrete geometry [122] for the calculation of a tangent at the given endpoint. That is, the reconnecting line is calculated from the endpoint and the pixel on the skeleton of thin vessel having distance of P pixels from the endpoint. The value of P is reduced iteratively from a threshold until a successful reconnection is found. There is a wide range of vessel widths from (less than) a pixel to a maximum width. Since direction estimation may miss the thick vessel, a parameter M for the maximum allowed reconnection distance between the endpoint and a thick vessel is considered. M should be set to the half of the maximum vessel width. We accept the thin vessel to be connected to the potential thick vessel, if the thick skeletal pixel and the thin endpoint pixel can be connected in this way. If the algorithm is not able to find a thick skeletal pixel for a thin endpoint, then it resumes the skeleton extracted from the whole vascular system for that location.

(a) (b)

Figure 3.15: *The reconnection step to fill in the gaps between the skeletons of the thick and thin vessels; (a) a gap to be filled in, (b) result after reconnecting the separate skeletons.*

For our experimental tests we considered a database of 130 vessel intersections extracted from the 19 binary vascular images of the database DRIVE described in section 1.5.6. At the resolution of the images in DRIVE the corresponding vessel widths are between 1 and 10 pixels. Considering this dataset, the parameters of our algorithm have been adjusted for optimal performance as: maximum allowed reconnection distance $M = 5$, distance from the endpoint for direction estimation $P = 6$ pixels.

To see the improvement, we have compared the skeleton resulted by our approach with the classic skeleton extracted from the original intersection image without any splitting/reconnection step. For a quantitative comparison, we considered the degree of distortion term DD defined before. From the 130 intersections, our algorithm generated 47 different skeletons than classic skeletonization. That is, basically in 36% of the cases the segments had sufficiently different widths for a reliable split. Comparing the results with respect to the DD term, we got that our algorithm gave better results for 29 intersections. In 14 cases, the detected junction points were different, but DD remained the same. There were 2 intersections, where our algorithm gave worse

results than classic skeletonization. In these cases, the thin vessels can be considered rather as circular segments than linear ones causing a failure for direction estimation. Considering all the 130 intersections, our algorithm reduced the total DD error from 140 to 72 pixels with providing a 48% suppression of distortion for these cases. Our results are also summarized in Table 3.4.

Splitable intersections (47/130)		
Skeletonizaton:	Classic	Proposed
Total DD error (pixels):	140.0	72.0
Average DD error (pixels):	3.11	1.6
Distortion decreased to: 51.4286%		
All intersections (130/130)		
Unsuccessful splitting		85
Better results with proposed method		29
Same results with proposed method		14
Worse results with proposed method		2

Table 3.4: *Improvement of the proposed method against classic skeletonization for 130 vessel intersections.*

Chapter 4

Combining algorithms for automatic detection of optic disc (OD) and macula in fundus images

In this chapter, we propose approaches to localize the optic disc and macula by the combination of the outputs of individual detectors in retinal images. We introduce techniques for fusing both single and multiple candidates as well as whole confidence maps of the member algorithms. The results of this chapter have been published in [3, 4, 22, 23] and are also incorporated in the project DRSCREEN: Developing a computer-based image processing system for diabetic retinopathy screening, TECH08-2 grant of the Hungarian National Office for Research and Technology (NKTH).

From this chapter on, we will focus on ensemble-based methods in object detection problems. We start with presenting our results in single object detection scenarios dedicated to the detection of the optic disc (OD) and macula in retinal images.

An important prerequisite to reach automation in diabetic retinopathy screening is the accurate location of the main anatomical features in the fundus image, notably the OD and the macula (see also section 1.4). The OD is a circular shaped anatomical structure with a bright appearance, so its localization is important not to misrecognize it as a bright lesion like exudate or cotton-wool spot. It is the location where the optic nerve enters the eye. If the position and the radius of the OD are detected correctly, then they can be used as references for locating other anatomical parts like the macula/fovea. The macula resides roughly in the center of the retina, temporal to the optic nerve. This small and highly sensitive anatomical part is responsible for detailed central vision. The macula allows us to appreciate details and perform tasks that require central vision such as reading. The fovea is the very center of the macula, the site of our sharpest vision. Information about the locations of these features is necessary because the severity and characterization of abnormalities in the eye partially depend on their distances from the fovea. OD detection is important in the computer-aided analysis of retinal images. It is crucial for the precise identification of the macula to make possible the successful grading of macular pathology such as diabetic maculopathy (clinically significant macular edema, macular ischemia). Much work has been carried out in this field and series of interesting algorithms [69, 70, 72, 142, 143] have been proposed in the recent past using various methods ranging from filtering [144] and threshold methods [75] to kNN regression [145]. Performance is generally good, but each method has situations, where it fails. However, there is in fact no reason to assume that a single algorithm would be optimal for the detection of various anatomical parts of the retina. It is difficult to determine which one is the best approach, because good results were reported for healthy retinas, but accuracy may drop on more challenging datasets containing diseased retinas with variable appearance of ODs in terms of intensity, color, contour definition etc.

We propose an approach to automatically combine different OD and macula detectors to benefit from their strengths while overcoming their weaknesses. In particular, we propose a flexible fusion scheme which considers the region with maximum number of algorithms' outputs. For OD and macula detection the respective algorithms OD_1, OD_2, OD_3, OD_4, OD_5, OD_6, OD_7, and MAC_1, MAC_2, MAC_3, MAC_4, MAC_5 described in details in sections 1.6.2 and 1.6.3 are considered in this chapter. In order to balance the contribution of individual algorithms participating in determining the final decision, weights are assigned to the outputs. The criterion for the selection of the candidate algorithms to be combined is based on different principles, good detection accuracy or low computation time. We show that our combination approach outperforms the individual algorithms for both the detection of OD and the localization of the fovea. These comparisons are made separately for the two anatomical parts. Moreover, we show how the detection performance can be increased further if we incorporate some geometric considerations regarding the relative spatial locations of the desired objects; the final candidates are determined via an iterative graph reduction method. As an extension of this approach, we also investigate the case, when a member algorithm can have more candidates for the OD. As we will see, we can reach further improvement in detection accuracy with incorporating this larger amount of information. The final candidate is selected again by the help of graph theory via solving a maximum weighted clique problem. Next, we also examine how to apply in our ensemble-based framework all the accessible information supplied by the member algorithms by making them return confidence values for each image pixel. These values inform us about the confidence of each algorithm that a given pixel is the center point of the object. We apply axiomatic and Bayesian approaches, as in the case of aggregation of judgments of experts in decision and risk analysis, to combine these confidence maps. Besides increasing the accuracy of the detection of the OD center, we can also extract the OD region with

this approach, so that it can be eliminated before the detection of bright lesions like exudates. Experimental results and analysis are provided on the three publicly available databases DiaretDB0 [59], DiaretDB1 [60] and DRIVE [61] described in section 1.5. We note here that any possible deviations in the accuracy figures of the algorithms are due to reimplementation or ground truth issues.

The rest of this chapter is organized as follows. In section 4.1, we describe how we combine the candidates of the OD and macula detectors, when they can vote with single candidates. Our corresponding experimental results are presented in section 4.2. Then, in section 4.3, we explain how our approach can be extended with letting the member algorithms give more candidates for a desired object. As a further improvement, we introduce our methodology for the detection of the center and the region of the OD based on the combination of confidence maps provided by the individual algorithms in section 4.4.

4.1 Combination of single candidates of member algorithms

The main steps involved in the proposed system are depicted in the flowchart shown in Figure 4.1 that presents a compact summary of our approach. In the subsequent sections, we describe these steps in details. As chronologically the first approach in this field, in this study we have included only OD_1, OD_2, OD_3, OD_4 and OD_5 from the OD detectors, since OD_6 and OD_7 have been implemented later.

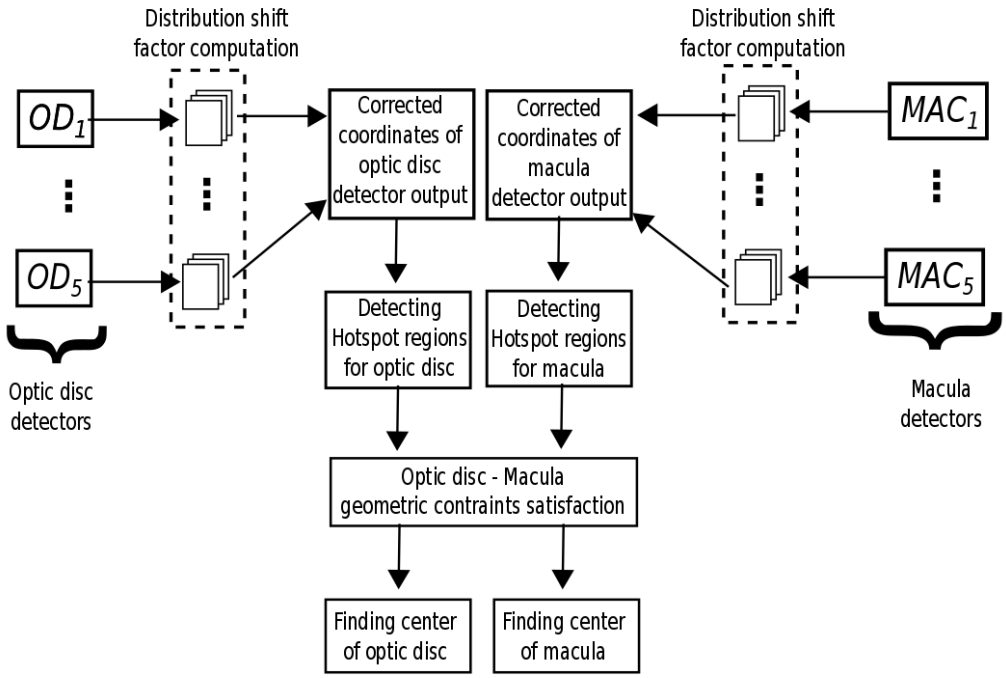


Figure 4.1: *Flowchart showing the steps of the proposed technique.*

4.1.1 Computing shift vectors for the output of detectors

The OD/macula centers detected by a particular algorithm in all test images are mapped onto a single image to check the distortion of its distribution. We observed that the outputs generated by the algorithms are quite dispersed and deviated from the manually selected (ground truth) macula

center. In Figure 4.2(a), the Gaussian kernel density estimation of the MAC_1 algorithm outputs is shown for the macula and of the OD_5 algorithm for the OD, respectively. Therefore, we propose to compute the distortion in the data and to apply a shift operation prior to actual combination of outputs for finding macula center to make the individual algorithms less biased. To compute the shift factor, we calculate the average difference between the candidate of the algorithm and the manually selected macula center for that image set $\mathcal{I}$, where the algorithm successfully found the OD/macula region. That is, the shifting vector is calculated as

$$\sum_{I \in \mathcal{I}} \frac{\boldsymbol{c^I} - \boldsymbol{c^I_{gt}}}{|\mathcal{I}|}, \tag{4.1}$$

where $\boldsymbol{c^I}$ stands for the OD/macula center candidate of the corresponding algorithm for image I, while $\boldsymbol{c^I_{gt}}$ is for the manually selected OD/macula center on the same image. The new output is generated by applying the distortion shift factor on each output pixel coordinate. The improvement in the result can be seen in Figure 4.2.

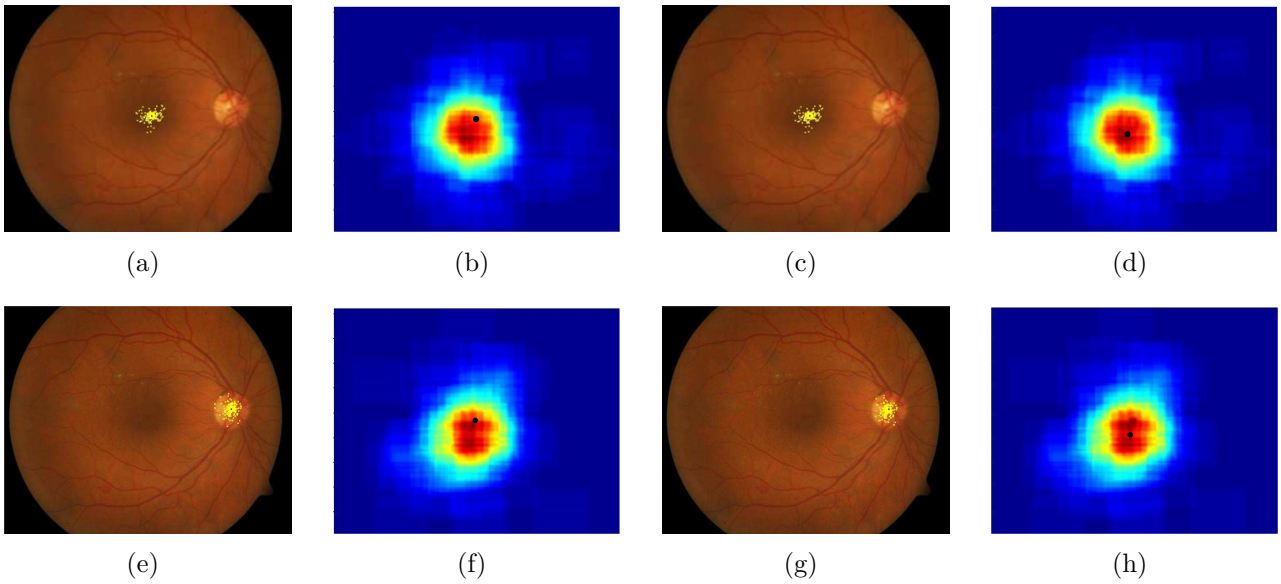


Figure 4.2: *Refining the output of the detectors by shifting; (a)-(d) the distortion in macula detector MAC_1 outputs before and after applying shifting, (e)-(h) the distortion in OD detector OD_5 outputs before and after applying shifting. The black dots indicate the manually selected centers.*

4.1.2 Detecting the OD and macula separately

We first devised a circular template voting scheme to determine the hotspot region i.e., an area in the image, where the maximal number of outputs fall. A circular template of diameter $d_{OD} \in \mathbb{R}_{>0}$ is fit on each pixel in the image and outputs of candidate algorithms that fall within the diameter of the predefined circular template are counted. The circular template covering the maximum number of OD detector outputs is considered to be a hotspot. There can be more hotspots covering the maximum number of detector outputs; hence they together define a hotspot region. The diameter d_{OD} of the template was set to 102 pixels, keeping in view the fact that clinically this is the average OD diameter at the investigated resolution (FOV corresponds to 1432 pixels). A circular template voting scheme has an intuitive appeal, and our results show it works well in practice.

Overall, in our combined system we have used the outputs of the five algorithms OD_1, OD_2, OD_3, OD_4, OD_5 for OD detection and the five algorithms MAC_1, MAC_2, MAC_3, MAC_4, $MAC5$ for macula detection listed in the introduction of this chapter. Following the principle of majority voting, the center of the circular templates covering the maximum number of OD and macula detector outputs is considered to be the OD and macula hotspots, respectively. If there is a tie, then such a conflict is handled using an additional postprocessing. For example, for OD detection a Gaussian filter is applied on the green channel of the image with a large sigma ($\sigma = 300$). The smoothened image is then subtracted from the original one to get a rather dark result in which OD appears as a brighter patch in comparison with the background. The average intensity near the detector outputs is computed using the same circular template, and the area with the highest average intensity is selected as the final OD hotspot region.

In the case of macula detection, we have found that the algorithm MAC_3 influenced the most the combination result, therefore, whenever there is a tie among hotspot regions, the hotspot region containing the output of MAC_3 is selected for further processing. If the output of MAC_3 does not belong to any hotspot, then the hotspot region being the closest to the image center is selected. Sometimes, majority voting may fail to detect the correct OD or macula center, especially in the case of a diseased retina (see e.g. Figure 4.3, where the template with three outputs is selected as the final hotspot for the OD).

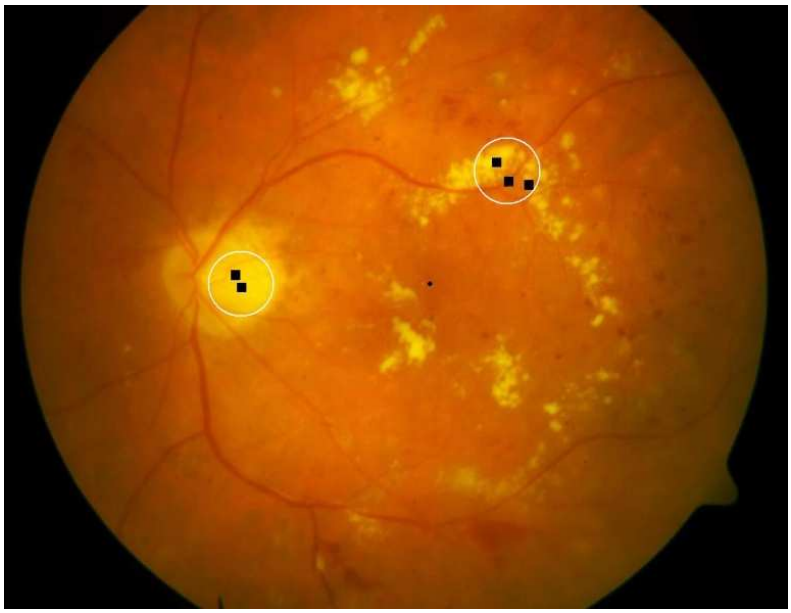


Figure 4.3: *False positive reported in the majority voting during OD detection.*

4.1.3 Detection based on mutual information

The accuracy of detection can be further raised by exploiting the anatomical constraints between the OD and macula. Particularly, the information about distance and angle criteria can be used to choose the best candidate hotspot region pair. The minimum radius circle around a set of points is a simple measure of the area they occupy as well as a useful tool in graphical applications. Welzl [146] proposed a simple randomized algorithm for the minimum covering circle problem. Therefore, to reduce the image processing needs instead of scanning the whole image, a minimum radius bounding circle is found to locate the hotspot region as being candidate of OD/macula. For this purpose, an unordered collection of different OD detectors and macula detectors are produced

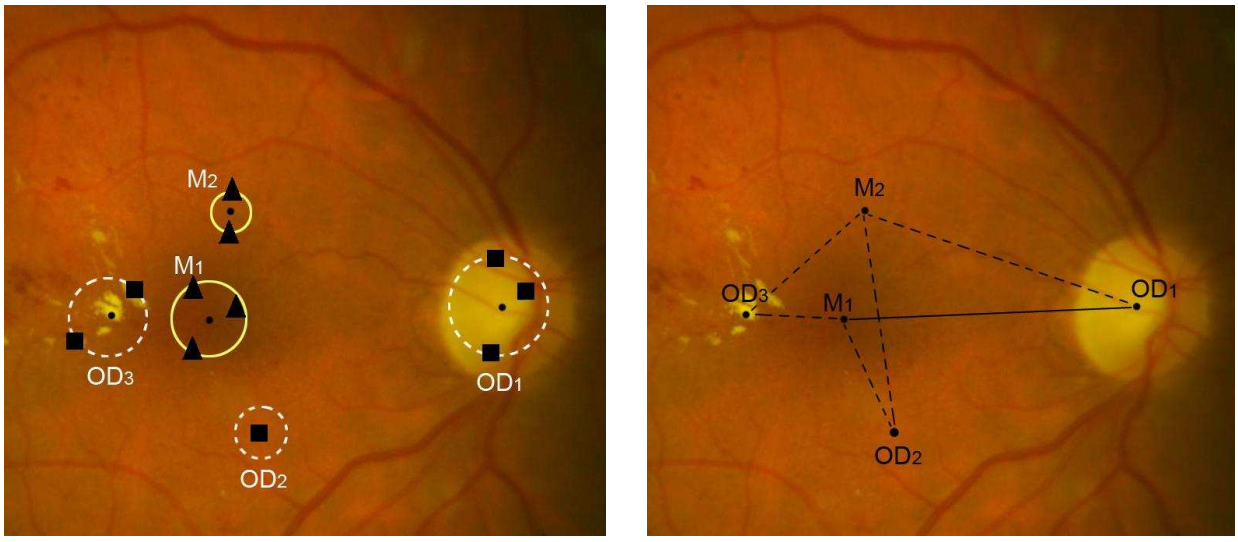
separately using the combination rule

$$C(n, n') = \frac{n!}{n'!(n-n')!}, \tag{4.2}$$

where $n, n' \in \mathbb{Z}_{\geq 0}$ and $n' \leq n$. Here, n is the number of detectors and n' is the number of items taken from them. The minimal enclosing circle algorithm considers every circle defined by these set of detectors generated using combination without replacement. The algorithm returns center and radius of the circle defined by the detectors. There can be many such circular regions, however, to validate a region as being a hotspot one (candidate for OD/macula), the OD radius $d_{OD}/2$ is used as a threshold to accept or discard a circle. Namely, if the radius of the minimal enclosing circle is less than or equal to the OD radius, then such a circle is considered as a candidate for OD/macula. For all the hotspot regions for the macula and OD detected in this way, we calculate a score using their mutual spatial information and the candidate hotspot pair with maximum score is selected as the final hotspot region for macula and OD. The score of a pair is based on the number of outputs present in a hotspot and penalties in terms of distance and angle errors:

$$Score(\mathrm{M}_i, \mathrm{OD}_j) = |\mathrm{M}_i||\mathrm{OD}_j| - \frac{D_{err}}{D_{avg}} - \frac{A_{err}}{A_{avg}}, \tag{4.3}$$

where M_i and OD_j stand for the i-th macula and j-th OD candidate hotspot regions, respectively. D_{err} and A_{err} are the respective distance and angle errors of M_i and OD_j regarding the average ones. Using manually marked macula and OD centers, we found that the average distance D_{avg} is 114 pixels and the average angle A_{avg} is 0°. As an example, Figure 4.4(a) shows some macula and OD hotspot regions. In Figure 4.4(b), all possible OD/macula candidates represented by the centroids of the corresponding hotspots are connected with dashed lines; the solid line between M_1 and OD_1 has the maximum score.



(a) (b)

Figure 4.4: *Selecting final OD and macula candidates based on mutual spatial information; (a) a retinal image with spotted hotspot regions for macula and OD, (b) hotspot pair connected with solid line providing the largest score regarding the geometric and cardinality constraints.*

4.1.4 Defining the object center as the weighted combination of candidates

The center of the final hotspot regions could be found by averaging the included candidates, however, for a more accurate estimation, weights can be associated with detector outputs to determine the final location. The principal difficulty is to choose the proper weights. By following classic statistical recommendations, one possibility is to assume the elements of the observed dataset to be linear combinations of some basis.

If we denote the output centers of the included individual $n' \leq n$ algorithms by random variables $\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_{n'}}$ with variances $Var(\boldsymbol{c_i}) = \sigma_i > 0$, then the problem of output combination is to reduce these n' outputs to one final center $\boldsymbol{c_{wres}}$. An appropriate combination of the outputs can be the weighted linear combination

$$\boldsymbol{c_{wres}} = \sum_{i=1}^{n'} w_i \boldsymbol{c_i}, \tag{4.4}$$

where $w_1, \ldots, w'_n$ are non-negative weights constrained to $\sum_{i=1}^{n'} w_i = 1$. This constraint guarantees the combined estimate to remain unbiased. The variance of the estimator $\boldsymbol{c_{wres}}$ is determined by the choice of weights and the variances of the individual outputs. When the random variables are independent, then the variance of the estimator $\boldsymbol{c_{wres}}$ is determined only by the choice of weights and the variance of individual outputs. In this case, we have to minimize the expression

$$Var(\boldsymbol{c_{wres}}) = \sum_{i=1}^{n'} w_i^2 Var(\boldsymbol{c_i}). \tag{4.5}$$

In this way, the choice of weights that minimize the variance of the combined estimate is

$$w_i = \sigma_i^{-2} / \sum_{i=j}^{n'} \sigma_j^{-2}, \; i = 1, \ldots, n', \tag{4.6}$$

as a result of [147]. Choosing these weights leads to small variance, hence, the expected Euclidean distance from the true center of the combined estimate can be minimized. The higher is the weight for an output, the more that detector is trusted to provide the correct answer. In the other case, when the random variables are dependent, we can consider the pairwise covariances, as well. We minimize

$$Var(\boldsymbol{c_{wres}}) = \sum_{i=1}^{n'} w_i^2 Var(\boldsymbol{c_i}) + 2 \sum_{1 \leq i < j \leq n'} w_i w_j Cov(\boldsymbol{c_i}, \boldsymbol{c_j}) \tag{4.7}$$

to get more appropriate weights. We can also write (4.7) as $Var(\boldsymbol{c_{wres}}) = \boldsymbol{w}^T D \boldsymbol{w}$, where $\boldsymbol{w} = (w_1, \ldots, w'_n)$ and D is a matrix containing the pairwise covariances. This minimum problem can be addressed via solving $\partial Var(\boldsymbol{c_{wres}}) / \partial w_i = 0$ for each $i = 1, \ldots, n'$ to determine the proper weights.

In our setup, we have compared separately all the macula and OD detectors with χ^2–test to check their independency. Then, according to the above description, we removed the term $Cov(\boldsymbol{c_i}, \boldsymbol{c_j})$ from (4.7), when the i-th and j-th algorithms were found to be independent according to this test.

In several results of the chapter presented later in section 4.3, we follow a simpler way to calculate the final candidate $\boldsymbol{c_{res}}$ as the centroid of the included individual detector outputs

$$\boldsymbol{c_{ares}} = \frac{\sum_{i=1}^{n'} \boldsymbol{c_i}}{n'}. \tag{4.8}$$

The reason of this simpler derivation of the final candidate is that for the weighted final result $\boldsymbol{c_{wres}}$, we have to calculate the corresponding weights for all the possible combinations of the detectors. However, this approach is hard to realize in practice, especially, if we add a new member algorithm to our system.

4.2 Results for single candidates of members

We have evaluated our proposed combination of algorithms to localize the OD and macula (fovea) on three publicly available databases DiaretDB0, DiaretDB1 and DRIVE described in section 1.5 in details.

The proposed algorithm should be trained so these databases have been split into training/test parts. The training database contains 40% of the images from each database, thus, the test database contains the rest 60% of the images. To reduce variability, multiple rounds of cross-validation are performed using different partitions, performing the analysis on one subset (the training set), and validating the analysis on the other subset (the validation set or testing set).

4.2.1 OD detection results

The individual OD detectors and their ensemble have been evaluated on the basis of two criterions i.e., to fall inside the manually selected OD patches ODR (see Figure 4.5), and to be close to the manually selected OD center. More precisely, the final OD center candidate is considered as true positive if it falls in the manually selected OD region ODR, otherwise it is false positive. Then, using the notations from section 1.3.1, we can calculate the Positive Predicted Value $PPV(DB)$ to measure the primary error for a database DB as

$$PPV(DB) = \frac{\sum\limits_{I \in DB} TP(I)}{|DB|}. \tag{4.9}$$

If the involved database is obvious, we will shortly write PPV.

As for the secondary error function Err_2, we calculate the average Euclidean distance of the candidates from the centroids of the corresponding manually selected ODs. Here, for a fair comparison we exclude those images, where the detected OD center falls outside the OD region:

$$Err_2(DB) = \frac{\sum\limits_{\substack{I \in DB \\ \boldsymbol{c^I_{wres}} \in ODR^I}} d(\boldsymbol{c^I_{wres}}, \boldsymbol{c^I_{gt}})}{\sum\limits_{I \in DB} TP(I)}, \tag{4.10}$$

where $\boldsymbol{c^I_{wres}}$ is the final OD candidate and ODR^I is the manually selected OD region for the image I. The short notation Err_2 is considered for a fixed database, and also notice that the same secondary error function is applied, when we will consider $\boldsymbol{c_{ares}}$ (see (4.8)) instead of $\boldsymbol{c_{wres}}$.

Table 4.1 shows the primary error (PPV) figures as the correct detection of the OD location based on the OD patch; the percentage detection rate of the proposed approach is higher than any of the individual algorithms. In Table 4.2, we can see the secondary error term Err_2 that measures the average Euclidean distance of the candidates from the manually selected OD centers. Overall, we can observe that the combined system led to more accurate localization than any of the individual members.

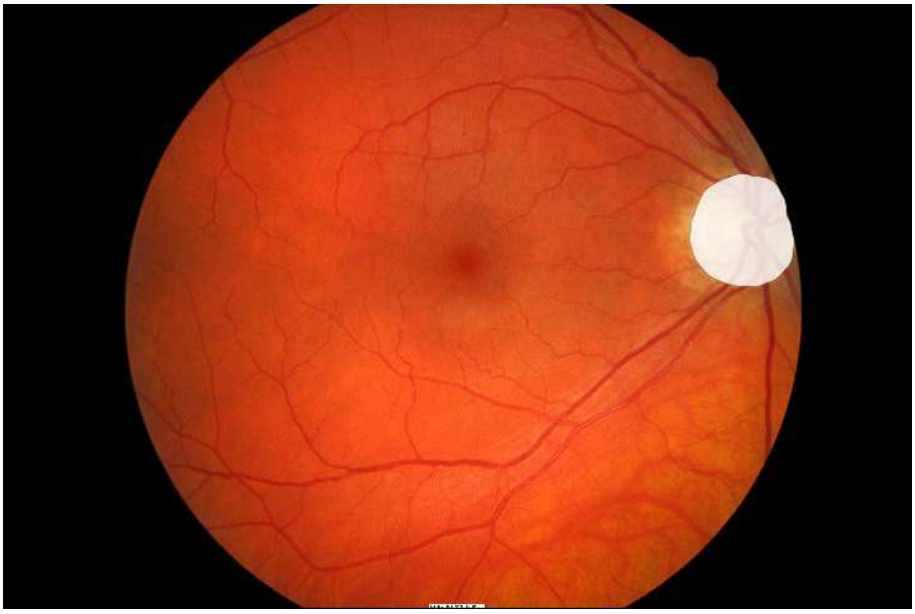


Figure 4.5: *A retinal image with manually selected OD region.*

Test databases	OD detectors					
	OD_1	OD_2	OD_3	OD_4	OD_5	Ensemble
DiaretDB0	89.52%	77.56%	78.20%	95.29%	80.12%	96.79%
DiaretDB1	88.99%	75.46%	77.04%	93.70%	76.41%	94.02%
DRIVE	80.55%	97.22%	67.35%	98.61%	86.10%	100%
Total average	87.95%	79.88%	76.12%	95.26%	79.78%	96.34%

Table 4.1: *Candidates falling inside the manually selected OD patch; primary error PPV.*

Test databases	OD detectors					
	OD_1	OD_2	OD_3	OD_4	OD_5	Ensemble
DiaretDB0	15.89	20.63	18.55	11.89	11.41	12.23
DiaretDB1	17.38	20.84	20.41	13.05	11.95	11.95
Drive	23.31	13.21	20.69	16.98	14.67	15.95
Total average	17.54	19.55	19.52	13.07	12.10	12.71

Table 4.2: *Average Euclidean error of the OD candidates; secondary error Err_2.*

4.2.2 Macula detection results

The methods dedicated to the localization of the fovea as the center of the macula have been evaluated regarding the same primary (PPV) and secondary (Err_2) error functions as the OD detectors. Namely, the macula error (see Table 4.3) is the number of times the algorithm's output falls within the $d_{OD}/2$ radius of the manually selected macula center; the fovea error (see Table 4.4) is the average Euclidean distance of these candidates and the manually selected centers. The Err_2 figures are given in pixels at the resolution of the average ROI diameter of DiaretDB0 and DiaretDB1 (1432 pixels). Weights are calculated with considering dependencies of the algorithms. Similarly to the OD results, the combined system provided more accurate overall results than any of the individual algorithms for the localization of the fovea.

Test databases	Macula (fovea) detectors					
	MAC_1	MAC_2	MAC_3	MAC_4	MAC_5	Ensemble
DiaretDB0	67.94%	70.29%	82.47%	86.10%	60.67%	93.58%
DiaretDB1	60.373%	77.66%	78.29%	91.50%	71.69%	99.68%
DRIVE	72.06%	81.07%	58.55%	81.07%	81.07%	91.88%
Total average	65.77%	74.24%	78.03%	87.39%	67.10%	95.53%

Table 4.3: *Candidates falling in the macula region; primary error PPV.*

Test databases	Macula (fovea) detectors					
	MAC_1	MAC_2	MAC_3	MAC_4	MAC_5	Ensemble
DiaretDB0	29.98	31.63	52.13	35.57	29.28	29.64
DiaretDB1	27.05	29.93	47.88	28.86	27.75	24.26
DRIVE	16.52	24.19	61.68	36.08	22.19	26.80
Total average	27.27	30.11	51.81	33.25	27.86	27.38

Table 4.4: *Average Euclidean error of the fovea candidates; secondary error Err_2.*

4.2.3 Joint OD-macula detection results

Research in ensemble methods has largely revolved around designing ensembles consisting of competent yet complementary models. The goal of the proposed approach is to leverage the power of multiple OD and macula detectors that are competent, but also complementary. The results demonstrate that the use of mutual geometric constraints between OD and macula has improved upon traditional approaches like simple averaging or voting. Weighted averaging has proven to be very effective and has achieved better prediction accuracy than any of the individual algorithms could have on their own. The use of geometric constraints between OD and macula is a unique contribution which has been rewarded as it outperforms all individual decisions (Table 4.1) and separate combination (Table 4.3) as shown in Table 4.5. In this comparison, we have considered the primary error function PPV measuring the relative numbers of candidates falling inside the desired regions.

Test databases	Joint OD-macula detection	
	OD	Fovea
DiaretDB0	97.64%	96.79%
DiaretDB1	97.79%	98.74%
DRIVE	100%	91.73%
Total average	98.06%	96.87%

Table 4.5: *OD/macula (fovea) candidates falling in OD/macula regions using mutual spatial information; primary error PPV.*

4.3 Improving detection accuracy by using multiple candidates of members

This chapter focuses on single image detection, and so far we let the member algorithms to give only one candidate for the desired object. Now, we apply further improvements to this model and demonstrate it on OD detection. Namely, on the one hand, we extract more than one candidate

from the output of each algorithm to increase the chance of getting the true one among the candidates for the OD. Furthermore, we assign weight to each candidate to replace simple majority voting by a weighted one and treat them as weighted nodes of a complete graph. For the location of the OD, we select that subgraph of the candidates which meets the OD geometry constraint and has a maximal total weight. For this selection, we borrow a graph theoretical approach providing the optimal solution.

As further improvement, for this analysis we also extend our OD detector ensemble. Namely, besides the detectors OD_1, OD_2, OD_3, OD_4, OD_5 considered so far, we include the detectors OD_6 and OD_7 described in details in section 1.5.

4.3.1 Extraction of multiple candidates

In section 4.1, we have examined how the results of individual algorithms can be combined, with selecting that OD-sized disc, where the most algorithms voted to. Here, all the individual detectors gave their single vote for the OD center. Now, we let them suggest multiple candidates and also assign weights to them.

Multiple candidates and OD-geometry constraint

All the individual algorithms OD_1, OD_2, OD_3, OD_4, OD_5, OD_6, and OD_7 are capable of assigning a confidence value $Conf(\boldsymbol{x}) \in [0, 1]$ to each image pixel $\boldsymbol{x}$ showing the probability of being the OD center according to the algorithm. These values together compose a so-called probability map for each algorithm for the image; see Figure 4.6 for the probability maps of some of the OD detectors. To have multiple candidates from an algorithm, we can simply select not only one, but more pixels as OD center candidates having large confidence values (see e.g. Figure 4.7 for OD_5). As a geometric constraint, the candidates of the same algorithm are not allowed to be closer than the average OD diameter d_{OD}. In this way, we can assure that a possible OD region is suggested by at most only one candidate of each algorithm.

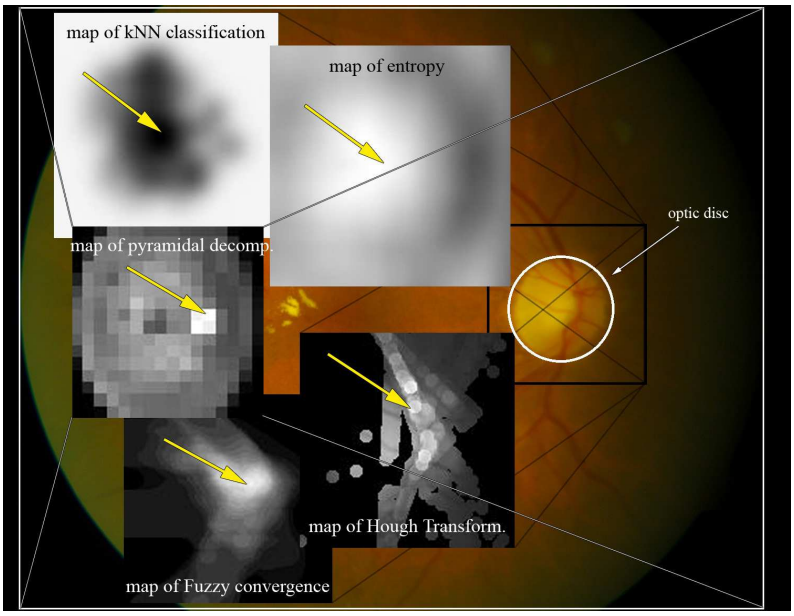


Figure 4.6: *Confidence maps of individual algorithms; arrows show the maximal values.*

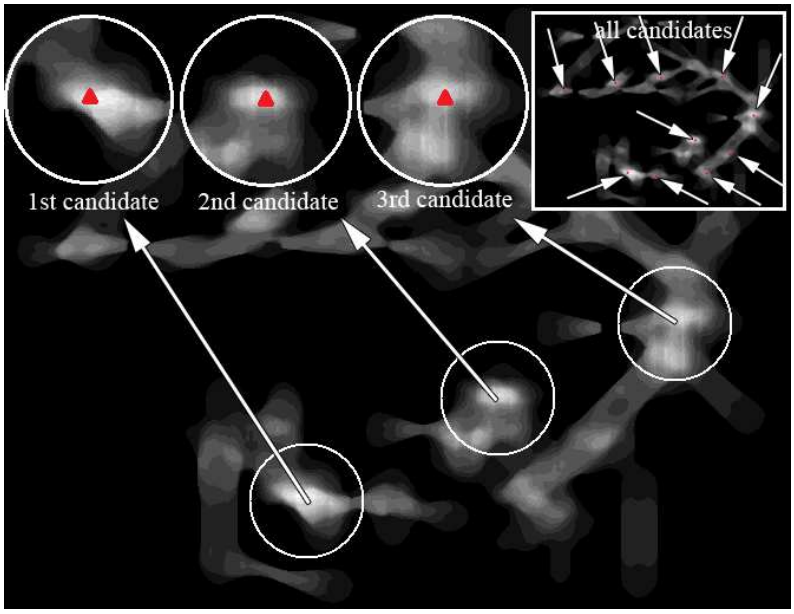


Figure 4.7: *The confidence map of algorithm OD_5 with candidates meeting the OD-geometry constraint.*

Selecting the number of candidates of members

It is a key issue, how many candidates of a specific member algorithm we should consider. By following the notations introduced in section 1.3, let $OD_i(I)$ denote the set of candidates considered for the i-th member algorithm ($i = 1, \ldots, 7$). Since we keep the same number of candidates for each image in a database, for the i-th member algorithm let $|OD_i|$ denote this figure uniformly for each image of the database. We have to keep the $|OD_i|$ ($i = 1, \ldots, 7$) figures relatively small to avoid too many false positive candidates confusing the final combination. On the other hand, we try to have the true OD location among these candidates of each individual OD detector. Thus, for each individual algorithm we have investigated how probable it is that the true location of the OD is among the candidates as we increase $|OD_i|$ ($i = 1, \ldots, 7$). These probability values were calculated using a training dataset and checking the ratio of the images, where the true OD was among the candidates. We have found that for each individual detector OD_i we can fix a small $|OD_i|$ figure above which this probability value will increase only slowly (see Figure 4.8 for this elbow behavior).

As a quantitative validation for this selection, we measured the performance of the maximal weighted clique-based ensemble when each member algorithm provides one more and one less candidate than the values corresponding to the elbows. We have found that additional 3 and 2 false detections appeared in the merged DiaretDB0 and DiaretDB1 datasets, respectively, which result supports our proposed selection strategy.

Assigning weights to the candidates of members

Let us assume that the set of candidates $OD_i(I)$ of the i-th member algorithm is sorted according to the probability of the candidates for each image I. That is, using the notation for candidates from section 1.3, $Conf(\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,j}}) \geq Conf(\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,k}})$ hold for all $\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,j}}, \dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,k}} \in OD_i(I)$ with $1 \leq j < k \leq |OD_i(I)|$.

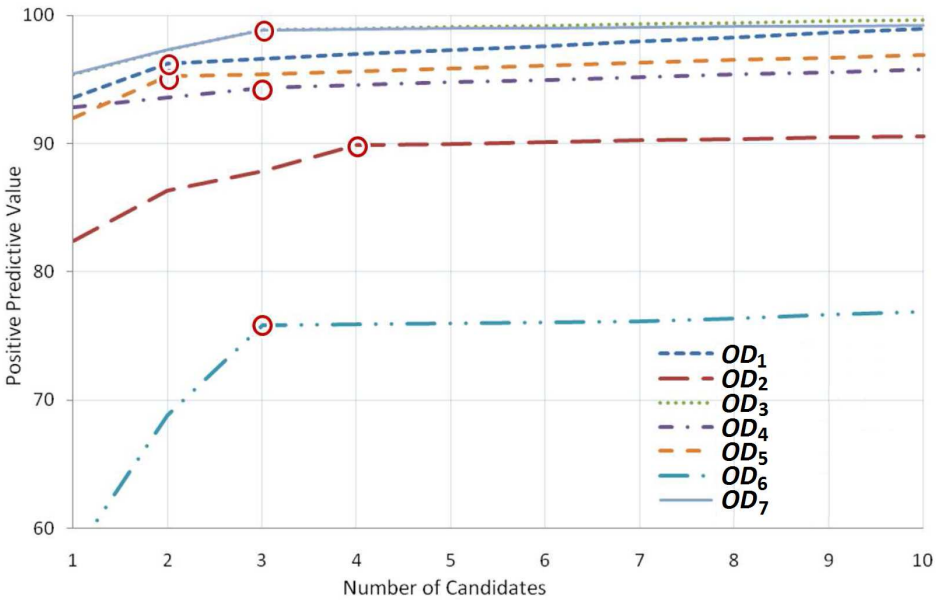


Figure 4.8: *The increase of the Positive Predictive Value (see section 1.3.1) of the OD detector algorithms regarding the number of their candidates. Red circles show the chosen numbers of candidates.*

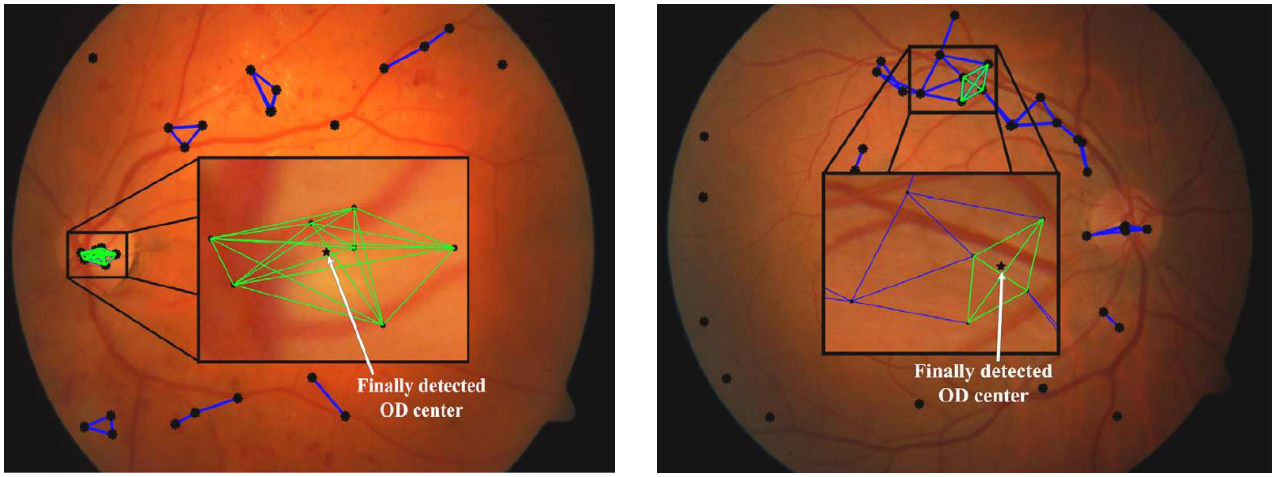
Then, we assign weights to the candidates as

$$w_{i,j} = \frac{Conf(\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,j}})}{\max\limits_{k=1}^{|OD_i|} Conf(\dot{\boldsymbol{c}}^{\boldsymbol{I}}_{\boldsymbol{i,k}})} \tag{4.11}$$

for $i = 1, \ldots, 7$, $j = 1, \ldots, |OD_i|$. Notice that this normalization will give $w_{i,1} = 1$ $(i = 1, \ldots, 7)$ for all the first candidates of the algorithms. Moreover, the weights also reflect the priority of the algorithms regarding their own candidates.

4.3.2 Finding a clique with maximal total weights

As a generalization of simple majority voting, we locate the OD as the region containing candidates having largest weight in total. Notice that the OD-geometry constraint provides that within an OD-sized disc only candidates belonging to different algorithms may fall. To solve this problem, we build up a graph, whose vertices are the $\sum_{i=1}^{7} |OD_i|$ candidates weighted as given in section 4.3.1. Moreover, we connect those vertices with edges that fall inside a possible OD region, that is, their distance is smaller than the average OD diameter d_{OD}; for such graphs see also Figure 4.9 for both a correct and incorrect detection. In this graph, we have to look for the maximal clique with maximal total sum of weights of nodes. The corresponding graph theoretical problem in this representation is the maximum-weight clique one, which is known to be NP-hard. For the solution of this problem, we borrow the algorithm [148], which is based on a heuristic vertex-coloring and a backtrack search. This method provides the maximal-weighted cliques and works quickly on a dense graph, too. The maximal-weighted clique obtained is considered as the solution of the multiple candidates ensemble-based OD detection system. The final OD center $\boldsymbol{c_{ares}}$ is calculated as the centroid of the selected clique using (4.8).



(a) (b)

Figure 4.9: *The clique meeting the geometric constraint and having maximal total weight; (a) correct detection, (b) incorrect detection.*

4.3.3 Results for multiple candidates of members

We have tested the proposed multiple candidates method on the publicly available databases DiaretDB0, DiaretDB1 and DRIVE described in section 1.5 in details. We consider two ways for accuracy measurement similarly to the single candidates case discussed in section 4.2. That is, as a primary error, we consider PPV as given in (4.9). On the other hand, we calculate the secondary error Err_2 defined in (4.10).

Table 4.8 compares the respective performances of the individual algorithms, and the ensembles based on simple majority voting and the proposed weighted majority voting method on the databases. Additionally, in Table 4.9 we give some more details by showing the accuracies of detection on the healthy/diseased images, separately. These tables will be included in section 4.4.3, since they will contain performance results of other models introduced is later sections of the chapter. In Tables 4.8 and 4.9 we can see that the solution based on multiple candidates outperforms both the individual detectors and the single candidates based ensemble and thus, lead to a true improvement.

4.4 Detection of the OD by combining probability models

Besides increasing the accuracy of the detection of the OD center, our aim is also the segmentation of the OD region so that it can be eliminated before the detection of bright lesions. In this chapter, first we have observed ensembles, where each member algorithms had a single candidate for the desired object. Then, in section 4.3, we have extended this approach to let each member indicate more possible candidates fo the OD. Now, we examine how to apply in our ensemble-based framework all the available information supplied by the member algorithms by aggregating the corresponding confidence values they provide for each image pixel. These confidence values inform us about the probability that a given pixel is the center point of the object. We apply axiomatic and Bayesian approaches, as in the case of aggregation of judgments of experts in decision and risk analysis, to combine these confidence values. Taking advantage of more information is supposed to lead to an improvement, which natural expectation is validated also by our experimental studies. Namely, we have found that the proposed method outperforms both the simpler ensemble-based

systems and the state-of-the-art individual member algorithms on publicly available datasets. Besides single localization, this approach can be adapted for the precise detection of the OD region, as well.

4.4.1 Fusion of probability maps for the OD center and region

The basic idea of the proposed method is to utilize as much information as possible about the location of a single object. Namely, we expect the member algorithms to assign a value $Conf(\boldsymbol{x})$ to each $\boldsymbol{x} \in I$ indicating their confidence that $\boldsymbol{x}$ is the center of the object. The most of the algorithms basically assign such a value to each pixel, but they apply a threshold to select only one location corresponding to the highest value. Thus, we can easily modify them with omitting their final thresholding step and can consider each pixel as a candidate equipped with confidence values by each member algorithm. These confidence values define probability maps (PMs) for the input image. Now, we introduce some possible approaches to fuse these maps in order to increase the accuracy of single object detection.

The fields of decision making and risk analysis, where information derived from several experts and aggregated by a decision maker, have a well-established literature [149–151]. In general, the aggregation of information increases the precision of the forecast. In our scenario, we can consider the confidence values assigned to each pixel $\boldsymbol{x} \in I$ as the opinion of the member algorithms on how probably the given pixel is the center point of the object. Based on the fact of the positive effect of the ensemble, if we consider the algorithms as experts with voting their confidence value and apply aggregation accordingly, the accuracy of the single object detection should improve.

As a short summary concerning the combination of information derived from experts, basically two approaches exist in the corresponding literature. One of them is based on clearly established mathematical rules, whereas the other one is entrusted to the interaction of experts, also known as a behavior-based method. In a behavior-based model, the experts contact the decision maker directly or indirectly to make him/her take their arguments and statements into consideration to reach consensus. In this approach, the quality of the individual experts and the dependencies among them are considered implicitly rather than explicitly. So, we examine only the applicability of strict theoretical approaches, which are widely available in the literature from the simple axiomatic methods to the processes requiring different information aggregation models. In the case of single object detection, axiomatic approaches can be applied easily to each $\boldsymbol{x} \in I$ to aggregate the probability values assigned by the member algorithms $\mathcal{A}_i$ $(i = 1, \ldots, n)$ to $\boldsymbol{x}$ using the general formulation given in section 1.3. Considering the more complex approaches, we should apply a training set to determine all the necessary parameters to set up the model for the ensemble.

To start the proper formalization of the proposed ensemble-based framework, let the true (ground-truth) center of the single object be denoted by $\boldsymbol{c_{gt}}$. Let L_1 denote the event when $\boldsymbol{x} = \boldsymbol{c_{gt}}$, while L_0 the one, when $\boldsymbol{x} \neq \boldsymbol{c_{gt}}$. Since most of the object detection algorithms consider various features of $\boldsymbol{x}$ and its neighborhood for localization, let $H_i(\boldsymbol{x})$ $(i = 1, \ldots, n)$ be the set of these features used by $\mathcal{A}_i$ to assign a probability value to $\boldsymbol{x}$. To show the confidence of $\mathcal{A}_i$ that $\boldsymbol{x}$ is the center of the object, a probability map PM_i $(i = 1, \ldots, n)$ can be defined as

$$PM_i(\boldsymbol{x}) = P(L_1|H_i(\boldsymbol{x})). \tag{4.12}$$

Here, we omit the details on the feature sets H_i, since they are completely algorithm dependent, and focus on the aggregation of the PMs instead.

From the maps PM_i $(i = 1, \ldots, n)$ we derive probability density functions to make the following two conditions hold:

$$PM_i(\boldsymbol{x}) > 0 \text{ for all } \boldsymbol{x} \in I, \tag{4.13}$$

$$\sum_{\boldsymbol{x} \in I} PM_i(\boldsymbol{x}) = 1. \tag{4.14}$$

Notice that we would not need to require strict inequality in condition (4.13) considering probability density functions in general. However, we have to avoid the case, when the confidence value equals 0, because the axiomatic combination rules considering the product of the confidence values would become meaningless. Thus, for a fair comparison of all the rules, technically we exclude the confidence value 0. Condition (4.13) can be reached with assigning a very small probability value $\epsilon > 0$ to each position, which originally has zero confidence:

$$PM_{\epsilon_i}(\boldsymbol{x}) = \max(PM_i(\boldsymbol{x}), \epsilon). \tag{4.15}$$

Next, to meet condition (4.14), we perform the following normalization step:

$$PDF_i(\boldsymbol{x}) = \frac{PM_{\epsilon_i}(\boldsymbol{x})}{\sum\limits_{\boldsymbol{x} \in I} PM_{\epsilon_i}(\boldsymbol{x})}. \tag{4.16}$$

In this way, the probability maps PM_i $(i = 1, \ldots, n)$ are transformed to the probability density functions PDF_i. After we have these PDFs, we can fuse them by applying standard axiomatic approaches or more complex aggregation models.

Aggregation based on axiomatic approaches

The product, sum, minimum, maximum of the probability density functions are the simplest approaches of aggregation in the corresponding literature [149, 150]. These techniques are realized by simple arithmetic operations performed between two or more PDFs given by the experts. One of the most commonly used axiomatic approaches is the linear opinion pool published by Stone in [152]. This method calculates the weighted sum of the probability density functions rendered by the experts

$$PDF_{LINOP} = \sum_{i=1}^{n} w_i PDF_i, \tag{4.17}$$

where PDF_{LINOP} represents the combined probability density and $w_i \in \mathbb{R}_{\geq 0}$ $(i = 1, \ldots, n)$ the weights assigned to the experts provided that we have information on their reliability. As a natural condition, $\sum_{i=1}^{n} w_i = 1$ must hold. If $w_i = 1/n$ for all $i = 1, \ldots, n$, we have a simple linear combination (average), otherwise a weighted linear one.

Multiplicative averaging (also known as logarithmic opinion pool) is another commonly used fusion approach [149]. In this case, probability density functions are combined as

$$PDF_{LOGOP} = k \prod_{i=1}^{n} PDF_i^{w_i}, \tag{4.18}$$

where k is a normalizing constant and w_i $(i = 1, \ldots, n)$ represent the same weights as above. If $w_i = 1/n$ for all $i = 1, \ldots, n$, (4.18) returns essentially the geometric mean of the individual distributions. These axiomatic approaches combine the PDFs in a simple way with ignoring the quality of the members and the dependencies among them. Now, we start discussing the Bayesian models of the information aggregation process, which require input regarding bias and dependencies of the experts.

Aggregation based on Bayesian models

In [153] and [154], Morris formally laid the foundation of the Bayesian paradigm to aggregate the information collected from different experts. The Bayesian models operate on the individual probability density functions to aggregate them. In the case of single object detection, according to these models the image pixels can be considered as the center of the desired object (event L_1) or not (event L_0). Thus, using Bayes' theorem the ensemble-based labeling of each pixel $\boldsymbol{x} \in I$ to be the object center is determined based on the probability maps PDF_i $(i = 1, \ldots, n)$ as

$$\boldsymbol{x} = \boldsymbol{c_{gt}}, \text{ if } P\left(L_1 | PDF_1(\boldsymbol{x}), \ldots, PDF_n(\boldsymbol{x})\right) > P\left(L_0 | PDF_1(\boldsymbol{x}), \ldots, PDF_n(\boldsymbol{x})\right). \tag{4.19}$$

As only two cases are possible, we have $P(L_0) = 1 - P(L_1)$ for each pixel. Thus, in our case, it is sufficient to determine the probability of L_1 for the pixels with the help of Bayes' theorem. For each $\boldsymbol{x} \in I$ we calculate the posterior probability in (4.19) by the help of Bayes' rule as

$$P(L | PDF_1, \ldots, PDF_n) = \frac{P(L) P(PDF_1, \ldots, PDF_n | L)}{P(PDF_1, \ldots, PDF_n)}, \text{ where } L \in \{L_0, L_1\}. \tag{4.20}$$

Notice that L does not appear in the denominator $P(PDF_1, \ldots, PDF_n)$, so this term is applied only for normalization. Thus, it can be omitted by following the general recommendations [151].

The *a priori* probability in the numerator of (4.20) can be easily estimated from the training database. The calculation of the joint probability density function $P(PDF_1, \ldots, PDF_n | L)$ depends on whether the model takes the dependencies of the member algorithms into account or not. In this respect, there are two basic approaches in the relevant literature as discussed next.

Naïve Bayes combination. In our first Bayesian approach, we suppose that the experts do not influence each other, there is no connection between them, so they give their opinion or forecast completely independently. That is, according to this naive hypothesis, the decision maker manages the information collected from the experts independently. This type of aggregation is known as the Naïve Bayes model and the joint density function in (4.20) can be separated according to the conditionally independent assumption based on the formula

$$P(PDF_1, \ldots, PDF_n | L) = \prod_{i=1}^{n} P(PDF_i | L). \tag{4.21}$$

Consequently, the aggregation of the probability density functions PDF_i $(i = 1, \ldots, n)$ can be derived based on (4.20) and (4.21) as

$$PDF_{NB}(\boldsymbol{x}) = P(L_1) \prod_{i=1}^{n} P(PDF_i(\boldsymbol{x}) | L_1), \tag{4.22}$$

where $P(PDF_i(\boldsymbol{x}) | L_1)$ $(i = 1, \ldots, n)$ are estimated on the basis of the probability values assigned by the algorithms to the pixels within the manually segmented object in the images of the training set. Since all the terms of (4.22) can be estimated from the training examples, the Naïve Bayes model can be easily constructed and adopted, as well. However, this model ignores the dependencies among the members, although the assumption on the conditional independence of the experts is fulfilled very rarely in practice.

To discover the dependencies of the member algorithms $\mathcal{A}_i$ and $\mathcal{A}_j$ $(i, j = 1, \ldots, n)$, we can calculate Pearson's correlation coefficient $\varrho_{i,j}$ [155]. The coefficients $\varrho_{i,j}$ are calculated pairwise for the member algorithms through comparing all the pairs $PDF_i(\boldsymbol{x}), PDF_j(\boldsymbol{x})$ $(i, j = 1, \ldots, n)$ as

$$\varrho_{i,j} = \frac{\mathbb{E}\left[\left(PDF_i - \mathbb{E}(PDF_i)\right)\left(PDF_j - \mathbb{E}(PDF_j)\right)\right]}{\sigma(PDF_i)\sigma(PDF_j)}, \tag{4.23}$$

where $\mathbb{E}$ and σ stand for the mean and standard deviation of their arguments, respectively. The coefficients $\varrho_{i,j}$ describe the dependencies between $\mathcal{A}_i$ and $\mathcal{A}_j$. Non-zero coefficients show dependencies suggesting that the model can be improved further as presented in the next section.

Augmented Naïve Bayes combination. The problem that experts do not provide their opinions or forecasts entirely independently from each other is well-known in the corresponding literature [156]. So, combining their input in a way that the decision maker considers the experts independent have a negative effect on the result. Thus, a Bayesian model is required, which is able to take all the dependencies between the experts into account. To address this issue, the optimal Augmented Naïve Bayes (ANB) model has been suggested [157], where during the learning phase the dependencies of the members are also incorporated. However, creating such an ANB model is an NP-hard problem [158], so it is recommended to choose an alternative approach which takes the dependencies into consideration, however, does not try to disclose them entirely. One of these models is the Tree Augmented Naïve Bayes (TAN) [157] one, which has the disadvantage that only the most dependent pairs are kept and the effects of the less dependent experts are omitted. As a trade-off, the complexity of the creation of the TAN model is significantly reduced. Contrarily, the Hidden Naïve Bayes (HNB) model developed by Zhang et al. [159] is capable of taking all the dependent experts into account collectively. Thus, the HNB model approximates the precision of the optimal ANB model better, while its time complexity for training is only polynomial.

The basic idea of the HNB model is that a hidden expert HE is created for each expert which can affect it. Thus, the i-th expert depends only on the i-th HE (HE_i), where HE_i contains all the dependency relations between the i-th and the other experts. That is, the joint probability in the numerator of (4.20) can be calculated by the HNB model considering the dependencies among the experts as

$$P(PDF_1, \ldots, PDF_n|L) = \prod_{i=1}^{n} P(PDF_i|HE_i, L), \tag{4.24}$$

where

$$P(PDF_i|HE_i, L) = \sum_{j=1,\ j\neq i}^{n} W_{i,j} P(PDF_i|PDF_j, L) \tag{4.25}$$

with $W_{i,j} \in \mathbb{R}_{\geq 0}$ $(i, j = 1, \ldots, n, i \neq j)$ and $\sum_{j=1,\ j\neq i}^{n} W_{i,j} = 1$.

As it can be seen, HE_i is the hidden expert of PDF_i $(i = 1, \ldots, n)$ and is basically a mixture of the weighted dependencies with other experts. The weights $W_{i,j}$ $(i, j = 1, \ldots, n, i \neq j)$ are determined using the training set based on the Conditional Mutual Information CMI of PDF_i and PDF_j as

$$W_{i,j} = \frac{CMI(PDF_i; PDF_j|L)}{\sum\limits_{j=1,\ j\neq i}^{n} CMI(PDF_i; PDF_j|L)}, \tag{4.26}$$

where for the discrete random variables X, Y, Z the term CMI is calculated as

$$CMI(X; Y|Z) = \sum_{x\in X} \sum_{y\in Y} \sum_{z\in Z} P_{X,Y,Z}(x, y, z) \log \frac{P_Z(z) P_{X,Y,Z}(x, y, z)}{P_{X,Z}(x, z) P_{Y,Z}(y, z)}, \tag{4.27}$$

where the marginal, joint, and/or conditional probability mass functions are denoted by P with the appropriate subscript.

Using the weights $W_{i,j}$ $(i, j = 1 \ldots, n, i \neq j)$, the hidden experts HE_i $(i = 1, \ldots, n)$ can be determined. Thus, the HNB model incorporates all the dependencies among experts similarly to the optimal ANB one. However, the time complexity of the training phase of HNB is $\mathcal{O}(tn^2 +$

$kn^2\nu^2$), where t is the number of training pixels of the training images, n is the number of member algorithms, k is the number of classes and ν is the average number of values an attribute can take.

After defining the weights $W_{i,j}$ $(i, j = 1 \ldots, n, i \neq j)$, the probability maps PDF_i $(i = 1, \ldots, n)$ via the HNB model are aggregated on the basis of the formula

$$PDF_{HNB}(\boldsymbol{x}) = P(L_1) \prod_{i=1}^{n} P(PDF_i(\boldsymbol{x}) | HE_i(\boldsymbol{x}), L_1). \tag{4.28}$$

Localization of the OD by the fusion of probability maps

Till this point of section 4.4, we have proposed a general ensemble-based framework for single object detection, when we have more than one member algorithms which can generate probability maps to locate the object. Now we show how to apply these approaches for OD detection and observe their performance considering the accuracy of detection. For visual and precise comparison of the proposed ensemble-based approaches, we use the same input image as in Figure 4.9(b) with considering the same detectors OD_1, OD_2, OD_3, OD_4, OD_5, OD_6, OD_7, as well.

The OD detector algorithms are slightly modified since they are not allowed to threshold the confidence values $Conf(\boldsymbol{x})$ to extract a single pixel having the highest value as the final center candidate. Instead, all the image pixels $\boldsymbol{x} \in I$ are equipped by a confidence value for each member. These confidence values together compose probability maps for I corresponding to each OD detector (see Figure 4.10 for all these PMs).

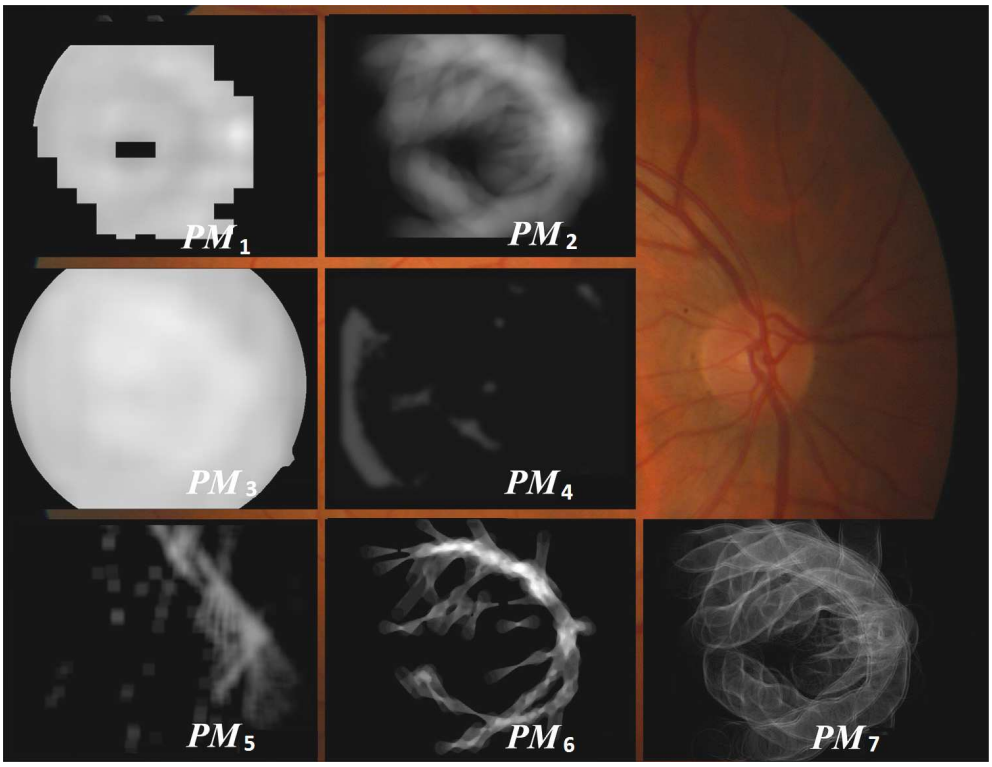


Figure 4.10: *Probability maps (PMs) of member algorithms showing their confidence on whether an image pixel corresponds to the OD center or not.*

As we have discussed in section 4.4.1, the proposed ensemble approaches can be applied if the PMs fulfill conditions (4.13) and (4.14). For this aim, the PMs are transformed to probability density functions PDFs via (4.15) and (4.16). In Figure 4.11, we can see a visual representation of the PDFs derived from the PMs of Figure 4.10.

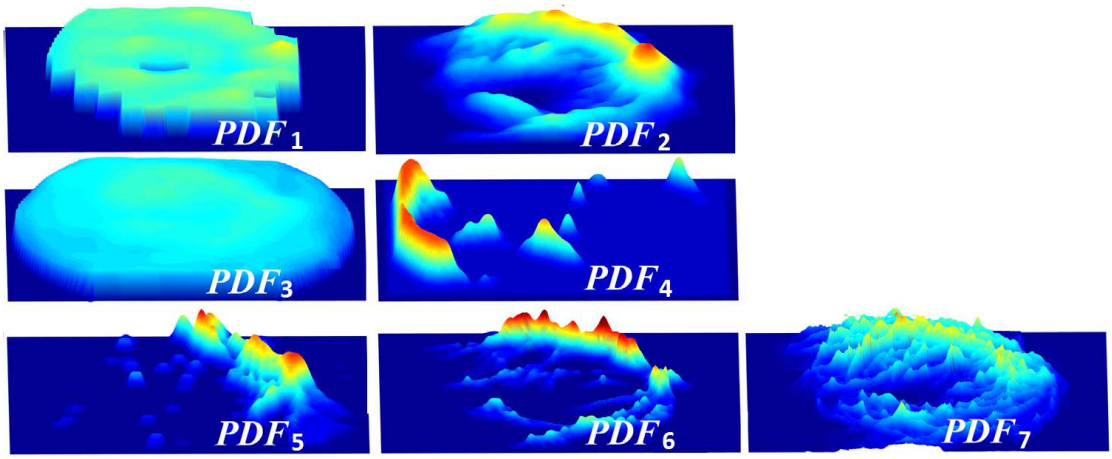


Figure 4.11: *The probability density functions (PDFs) of the member OD detectors.*

After constructing the PDFs, they can be fused by applying the standard axiomatic or Bayesian model-based approaches. For axiomatic ensembles of PDFs (see (4.17) and (4.18)), the weights w_i $(i = 1, \ldots, 7)$ are calculated from the individual accuracies of the members as $w_i = p_i / \sum_{i=1}^{7} p_i$, where p_i denotes the accuracy of the detector OD_i on the training set. The proper details of the databases used for training and testing will be given in section 4.4.2. After making tests on a training set, we adjust the following weights: $w_1 = 0.16$, $w_2 = 0.18$, $w_3 = 0.17$, $w_4 = 0.16$, $w_5 = 0.16$, $w_6 = 0.04$, $w_7 = 0.13$. The result of the combination of PDF_i $(i = 1, \ldots, 7)$ by the weighted linear opinion pool and weighted logarithmic opinion pool can be seen in Figure 4.12(a) and (b), respectively.

Now we turn to the Naïve Bayes model for OD detection. During the training stage, we determine the probability of OD center pixels among all the pixels of the training images. However, there is only one OD center point in the image, and considering the number of all the image pixels, $P(L_1)$ is a very small value. Since L_1 would be very under-represented in this way in a training dataset, we interpret L_1 in a bit wider sense. Namely, we let L_1 represent not only the case when $\boldsymbol{x} = \boldsymbol{c_{gt}}$, but also when $\boldsymbol{x}$ falls inside the OD region ($\boldsymbol{x} \in ODR$). In other words, we do not restrict our attention to the center, but we accept any OD pixels. Notice that in this way $P(L_1)$ becomes sufficiently large, and from now on we work in this extended context. Besides the *a priori* probability $P(L_1)$, the conditional probabilities $P(PDF_i(\boldsymbol{x})|L_1)$ $(i = 1, \ldots, 7)$ are also calculated inside and outside the region of the OD. A sample result for the combination based on the Naïve Bayes model can be seen in Figure 4.12(c).

As we have mentioned earlier, the assumption on the conditional independence of the experts is fulfilled very rarely in practice. This assumption does not hold for the involved OD detectors either. To confirm this hypothesis, we calculate Pearson's correlation coefficients $\varrho_{i,j}$ defined in (4.23) for all the possible pairs of member algorithms with enclosing them in Table 4.6.

A smaller (close to 0) correlation value corresponds to smaller dependency of the given algorithms. For instance, OD_1 and OD_4 seem to be the most diverse algorithms compared to each other regarding this measure. There are no zeros in Table 4.6 showing the trivial fact that the members cannot be completely independent. Thus, we can apply the HNB model which takes the dependencies among the detectors also into consideration. The sample result of the HNB combination can be seen in Figure 4.12(d).

In all the sub-images of Figure 4.12, very high probability values can be observed for the region of the OD, so its location can be found with a simple extra step. Namely, the center of the average OD size disc template D_{OD} of diameter d_{OD} (see also section 4.1.2) is matched on each pixel of

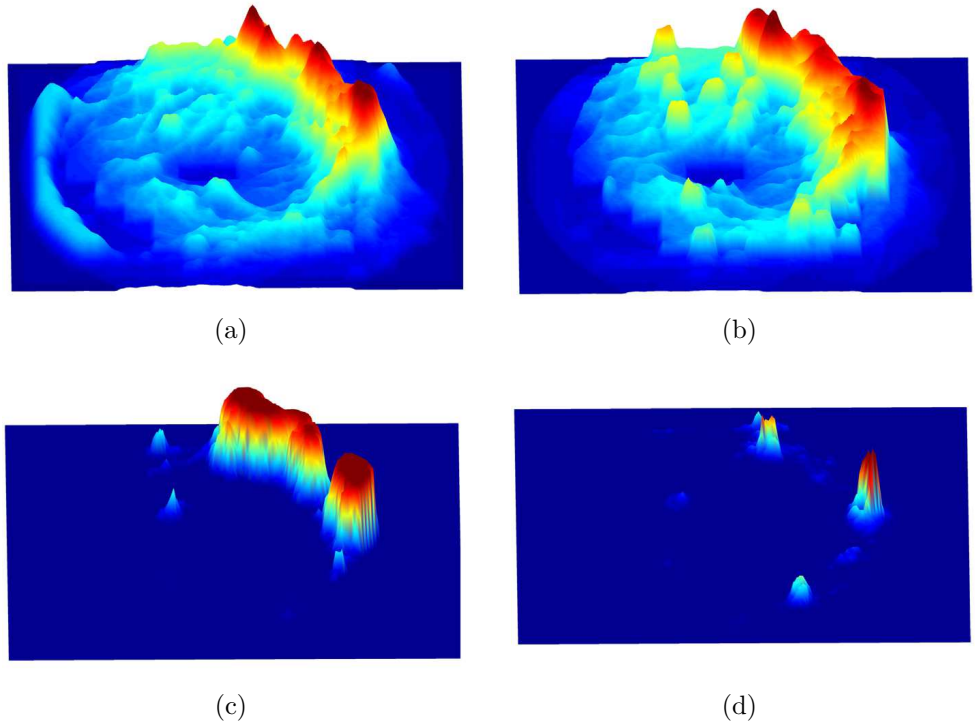


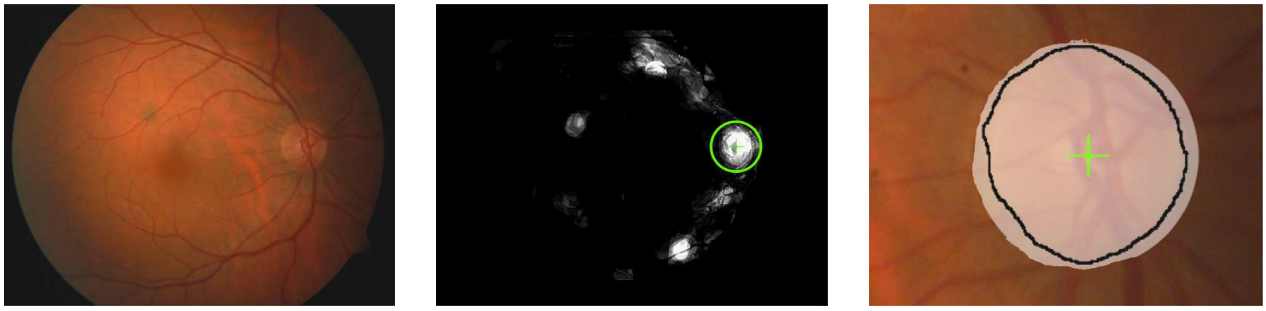
Figure 4.12: *Results of combination of the probability density functions PDF_i $(i = 1, \ldots, 7)$ by (a) linear opinion pool, (b) logarithmic opinion pool, (c) Naïve Bayes model, (d) Hidden Naïve Bayes model.*

ϱ	OD_1	OD_2	OD_3	OD_4	OD_5	OD_6	OD_7
OD_1	1.0000	0.6034	0.6339	0.1122	0.2645	0.3235	0.5634
OD_2	0.6034	1.0000	0.6612	0.4070	0.6150	0.6070	0.8172
OD_3	0.6339	0.6612	1.0000	0.2094	0.3607	0.3934	0.6437
OD_4	0.1122	0.4070	0.2094	1.0000	0.3733	0.1928	0.2427
OD_5	0.2645	0.6150	0.3607	0.3733	1.0000	0.5553	0.4793
OD_6	0.3235	0.6070	0.3934	0.1928	0.5553	1.0000	0.4733
OD_7	0.5634	0.8172	0.6437	0.2427	0.4793	0.4733	1.0000

Table 4.6: *Pearson's correlation coefficients of the member algorithms.*

the resulted ensemble PDF image and that pixel is selected as the final OD center $\boldsymbol{c_{res}}$, where we find the maximum sum of PDF values for the pixels within the matched D_{OD}.

If we compare the results of the axiomatic models (Figure 4.12(a), (b)) with the Bayesian ones (Figure 4.12(c), (d)), we can see significant difference between the areas, where the OD is detected with lower probability. There are high peaks at the possible OD locations, but the rest of Figure 4.12(c), (d) are flat because of the refined aggregation methods which take the dependencies among the members also into account. Furthermore, as we consider L_1 in a wider sense with letting it represent the event $\boldsymbol{x} \in ODR$, the resulted PDFs of the Bayesian models show high probability also at the pixels falling inside the OD region not just at the center of it. Thus, we can determine the final OD region if we consider the area corresponding to the peak found by matching the template D_{OD}. If D_{OD} expands this region, their union is considered as the segmentation of the OD region, as it can be seen also in Figure 4.13.



(a) (b) (c)

Figure 4.13: *Locating the OD as the strongest peak covered by the OD template; (a) original image, (b) result of the combination by HNB model, (c) segmented OD (white region) with marking the center point (green cross) and the manually drawn OD contour (black).*

4.4.2 Databases used for training and evaluation

We evaluated our proposed combination approaches to localize the OD on publicly available image sets. Namely, the databases DiaretDB0, DiaretDB1 and DRIVE described in section 1.5 were considered for performance analysis. The proposed Bayesian approaches for combining the member algorithms require training images to construct the Bayesian models. So the NB and HNB models should be trained on the training set without the influence of the test set. Therefore, our private database described in section 1.5.4 is used during the training phase. The images from our database were captured at resolutions of 3072×2048 and 1360×1024 pixels with the average diameters of the ROI are 2287 and 1340 pixels, respectively. Table 4.7 contains the summary of databases used for training and the evaluation of the proposed combinations of individual algorithms for OD localization.

Database	# of images	Healthy	Unhealthy	FOV	ROI	Resolution
DiaretDB0	130	20	110	50°	1410	1500×1152
DiaretDB1	89	5	84	50°	1410	1500×1152
DRIVE	40	33	7	45°	540	768×584
TRAININGSet	327	241	86	50°	2288 1340	3072×2048 1360×1024

Table 4.7: *Summary of the test datasets for OD detection using probability models.*

To collect the training dataset for Bayesian models, we calculated PDF_i $(i = 1, \ldots, 7)$ for all the 327 images. Then, we selected 400,000 pixels randomly from these images together with their corresponding PDF_i $(i = 1, \ldots, 7)$ values. In this way, 4% of the pixels belonged to OD regions in the training set.

4.4.3 Experimental results

For the measurability of the accuracies of the algorithms, manually selected ODs are also supplied for the images of the DiaretDB0, DiaretDB1 and DRIVE databases. We consider three ways for measuring the accuracy. The primary and secondary error terms are just the same ones described in sections 4.2.1 and 4.3.3. Namely, the primary error is calculated in terms of the Positive Predictive Value PPV defined in (4.9). As secondary error measurement we calculate Err_2 defined in (4.10). Additionally, as a third error $Err_3(DB)$ for a database DB, we measure

the overlapping area between the manually labeled OD region and the one determined by the aggregation of the outputs of the individual algorithms by applying the Bayesian approach

$$Err_3(DB) = \frac{1}{|DB|} \sum_{I \in DB} \left| \frac{R^I_{res} \cap ODR^I}{R^I_{res} \cup ODR^I} \right|, \tag{4.29}$$

where R^I_{res} denotes the automatically segmented OD region for image I. If the applied database is obvious, we will shortly write Err_3.

In this chapter, we demonstrate how to compose ensembles with using more and more information from the members. Table 4.8 compares the respective performances of the member algorithms, simple majority voting, OD detection by finding maximal weighted clique, detection on the aggregated probability map by the axiomatic approaches and by the Bayesian models on the DiaretDB0, DiaretDB1 and DRIVE databases. For simplicity, we normalized for the highest image resolution when calculating Err_2. As we can see, the highest accuracy for PPV can be achieved, when the aggregation model uses the largest amount of information about the OD location provided by the algorithms and the dependencies between them. The HNB model has a solid performance also for Err_2.

OD detector	**DiaretDB0**		**DiaretDB1**		**DRIVE**	
	PPV	Err_2	PPV	Err_2	PPV	Err_2
OD_1	86.92%	26.45	91.01%	27.64	80.00%	17.05
OD_2	90.77%	20.06	91.01%	17.72	92.50%	11.48
OD_3	87.69%	36.91	85.39%	36.73	85.00%	18.70
OD_4	85.38%	20.35	80.91%	21.79	70.00%	9.65
OD_5	84.62%	42.49	84.27%	47.66	72.50%	17.62
OD_6	30.11%	45.80	40.45%	44.23	82.50%	15.41
OD_7	63.08%	39.04	68.54%	37.49	92.50%	14.79
Majority voting	94.62%	15.68	94.38%	17.40	100.00%	11.26
Maximal-weighted clique	95.38%	9.38	95.51%	9.19	100.00%	6.08
Weighted LINOP	93.08%	16.76	95.51%	16.14	100.00%	12.12
Weighted LOGOP	91.54%	24.53	93.26%	24.31	100.00%	12.59
NB model	96.15%	22.36	96.63%	21.24	100.00%	11.62
HNB model	98.46%	16.86	98.88%	14.49	100.00%	9.12

Table 4.8: *OD detection results of member algorithms and ensembles on the databases DiaretDB0, DiaretDB1 and DRIVE.*

For the sake of completeness, we have included the PPV results of OD detection for our private database (TRAININGSet) based on probability map combination by the NB and HNB models which are 93.49% and 96.24% using 10 folds cross-validation tests, respectively. The performance results found for the TRAININGSet are slightly worse than for the others. However, notice that we have relatively more diseased cases in this set making the OD not to be the brightest region, which naturally makes its detection more challenging.

Moreover, to confirm the usefulness and positive contribution of each involved individual algorithm, we made an exhaustive evaluation and tested all the possible $2^7 - 1 = 127$ combinations of them as an ensemble system. The TRAININGSet was used for this evaluation with dividing it into a test (216 images) and training (111 images) part and the HNB model-based ensembles were tested on the test part. The average and the range of PPV values of these ensembles regarding their cardinality (i.e., the number of algorithms composing them) are shown in Figure 4.14.

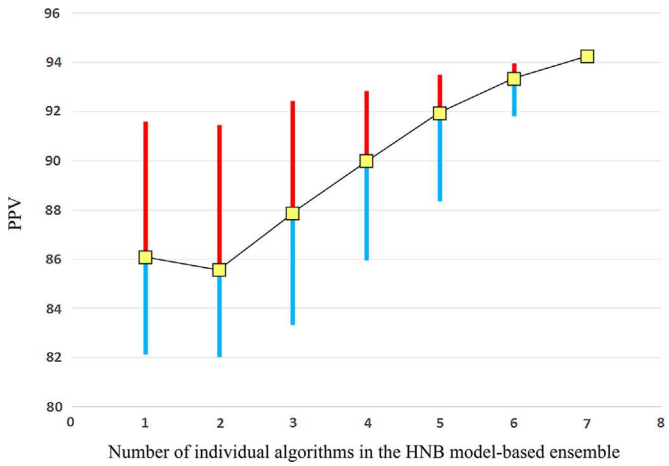


Figure 4.14: *Performance of possible OD detector ensembles regarding their cardinality. Yellow squares show the average PPV of ensembles having the same number of member algorithms. Red/blue line indicates the maximum/minimum PPV values.*

As it can be seen, the cardinality of the ensemble grows together with the average PPV. A small fallback can be observed between the performance of the individual algorithms and the two-components ensembles which can be explained by the small and even number of the experts. Regarding two-components ensembles, a high confidence value at a false OD position could not be corrected by another lower confidence value at the right position. Also note that the highest PPV is achieved by the ensemble consisting of all the individual members.

None of the member algorithms involved in the study is suitable for segmenting the region of the OD. Therefore, we did not evaluate the individual OD detectors for the error function Err_3. We merely notice that the segmented OD region based on the combined probability maps by the NB model against the manually drawn OD region achieved the following corresponding precisions: $Err_3 = 0.66$ on DiaretDB0, $Err_3 = 0.71$ on DiaretDB1 and $Err_3 = 0.66$ on DRIVE. When the combination was made by the HNB model, then we got $Err_3 = 0.84/0.87/0.73$ for DiaretDB0, DiaretDB1, and DRIVE, respectively. Although, many state-of-the-art algorithms (e.g. [160–162]) are available for the segmentation of the OD boundary which can outperform the proposed ensemble of the algorithms regarding Err_3, we have to take into consideration the fact that none of the member algorithms can detect the region of the OD. However, their combination is suitable for this task as it can be observed in Figure 4.15.

The ensemble of the member algorithms can localize the center of the OD also in images with signs of parapapillary atrophy and myopia which signs cause problem for many detectors. Although the detected region is extended as it can be seen in Figure 4.16, it causes smaller problem in our case, where the main motivation is to remove the OD before the subsequent exudate detection step.

For the sake of completeness, we insert Table 4.9 to show that the proposed ensembles of the member algorithms can improve OD detection performance in both healthy and unhealthy images. Especially, the unhealthy cases are more challenging for individual algorithms because of the large variability of the appearance of different diseases. The table shows the numbers of correctly detected cases against the total number of images in the included databases.

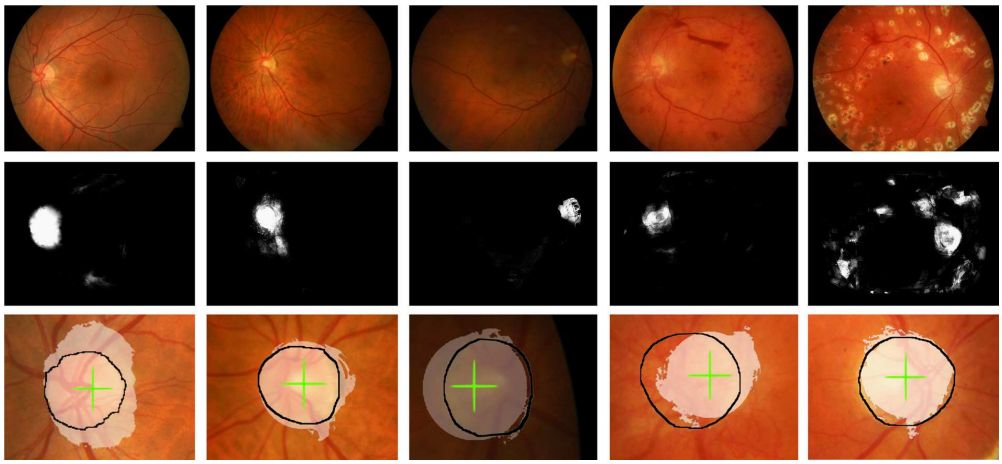


Figure 4.15: *Results of the segmentation of the OD region based on aggregated probability maps using the HNB model. The three rows shows sample images, the results of aggregation by HNB and a closer look of the finally segmented OD regions, respectively. The black boundary in the third row shows the manually segmented OD regions, the white patch the automatically segmented ones and the green cross the automatically detected OD center.*

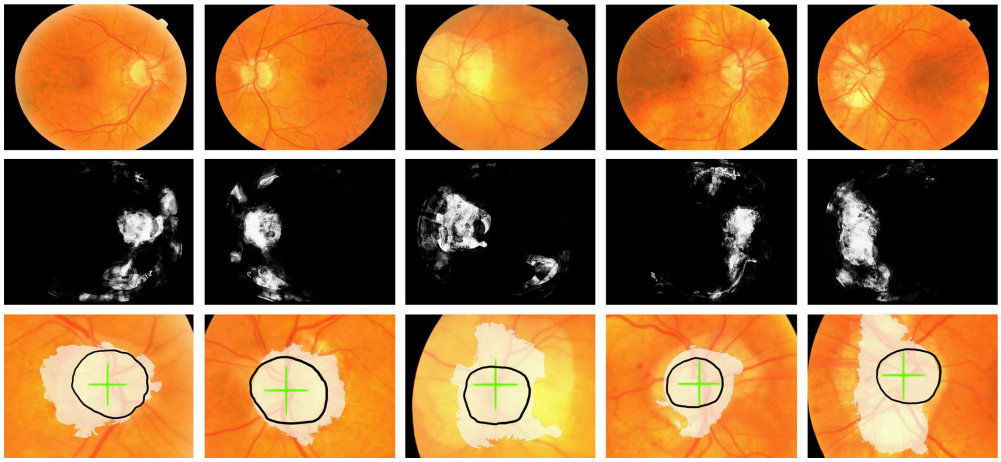


Figure 4.16: *Results of the segmentation of the OD region by HNB model in images, where signs of parapapillary atrophy and myopia can be found.*

4.4.4 Discussion on the combination of probability models

We note that the proposed ensemble framework can be further improved by the selection of the optimal subset of the involved member algorithms. This selection is a further issue and goes beyond this study, because it can be considered as an optimization problem, which takes the dependencies among the members, the individual accuracies, the possibility of redundant information, the computational time, the parameters of the individual algorithms and the values of the error functions into consideration. As a demonstrative example, if we exclude the two algorithms OD_6 and OD_7 having the lowest PPV performances, the ensemble of the remaining algorithms results additional 4 and 1 false images in DiaretDB0 and DiaretDB1, respectively, comparing with the ensemble of all the algorithms.

Each of the OD detector algorithms involved in the demonstration of the single object detection

OD detector	DiaretDB0		DiaretDB1		DRIVE	
	Healthy	Unhealthy	Healthy	Unhealthy	Healthy	Unhealthy
OD_1	18/20	95/110	5/5	76/84	26/33	6/7
OD_2	19/20	99/110	5/5	76/84	30/33	7/7
OD_3	18/20	96/110	5/5	71/84	28/33	6/7
OD_4	18/20	93/110	5/5	67/84	24/33	4/7
OD_5	17/20	93/110	5/5	70/84	27/33	6/7
OD_6	7/20	33/110	3/5	33/84	29/33	4/7
OD_7	13/20	69/110	4/5	57/84	30/33	7/7
Majority voting	19/20	104/110	5/5	79/84	33/33	7/7
Maximal-weighted clique	20/20	104/110	5/5	80/84	33/33	7/7
Weighted LINOP	19/20	102/110	5/5	80/84	33/33	7/7
Weighted LOGOP	19/20	104/110	5/5	78/84	33/33	7/7
NB model	20/20	105/110	5/5	81/84	33/33	7/7
HNB model	20/20	108/110	5/5	83/84	33/33	7/7

Table 4.9: *OD detection performance regarding healthy and unhealthy cases.*

task was reimplemented in MATLAB®[1] by following the respective instructions given in their corresponding original papers. For testing purposes, we used a single core 2.4GHz CPU with 2GB memory. The computational times of the original algorithms are the same with a negligible difference as their variants which are capable of assigning probability values to each pixel with producing the maps PDF_i $(i = 1, \ldots, 7)$ for an input image. The reason for the same demands of computations is that the original algorithms also calculate confidence values for each pixel and select the maximum to return the center of the OD. The computational times of the member algorithms were OD_1: 190 seconds (s), OD_2: 237 s, OD_3: 93 s, OD_4: 214 s, OD_5: 1.36 s, OD_6: 5.57 s, OD_7: 4.76 s, respectively. The aggregation of the maps PDF_i $(i = 1, \ldots, 7)$ is done in 40 milliseconds (ms) for axiomatic approaches. The computational times for the Naïve Bayes model-based aggregation approaches also depend on the size of the input image. Excluding the training phase, we have measured the following times: simple NB: 220 ms, TAN: 200 ms, HNB: 250 ms. For the applied training dataset, the number of operations to train the HNB model was 25,972,450.

[1]MATLAB® is a registered trademark of The MathWorks, Inc.

Chapter 5

Generalizing the majority voting scheme to spatially constrained voting

In this chapter, we generalize classic majority voting by incorporating terms $p_{n,k}$ for the probability that a good decision is made, if we have k correct votes out of n ones. With a geometric constraint, our model is specialized to be applicable for the detection of objects having special geometry. The model is analyzed for both dependent and independent voting algorithms, and improved further for weighted models. Corresponding results have been published in [5, 24–26] and incorporated in the project DRSCREEN: Developing a computer-based image processing system for diabetic retinopathy screening, TECH08-2 grant of the Hungarian National Office for Research and Technology (NKTH).

In this chapter, we investigate how voting systems behave if we apply some further constraints on the votes. Namely, we generalize the classic majority voting scheme by introducing real values $0 \leq p_{n,k} \leq 1$ for the probability that a good decision is made if we have k correct votes out of the n ones. In other words, in our case it will be possible that a good decision is made even if the good votes are in minority (less than half).

The creation of this new model is motivated by a retinal image processing problem – the detection of the optic disc (OD), which appears as a bright circular patch within the region of interest (ROI) in a retinal image (see Figure 5.1). Namely, in sections 4.1 and 4.2 we have observed that organizing more individual OD detector algorithms into an ensemble may raise detection accuracy. In this voting system, each individual OD algorithm votes in terms of a single pixel as its candidate for the OD center. The application of existing majority voting models are not completely adequate here, since they consider only the correctness of the votes, which concerns falling into the true OD region in this scenario. However, in our case, the spatial behavior of the votes is also important, since they vote together for a specific location of the OD, only if they fall within a region matching the OD geometry. Consequently, we should consider discs of diameter of the OD $d_{OD} \in \mathbb{R}_{>0}$ covering the candidates of the individual detector algorithms as shown in Figure 5.1.

The diameter d_{OD} can be derived by averaging the manual annotations made by clinical experts on a dataset and can be adjusted to the resolution of the image as described in section 4.1.2. As a final decision, the disc having diameter d_{OD} with maximal number of candidates included is chosen for the OD location. In this combined system, we can make a good decision even if the false candidates have majority such as in the case illustrated in Figure 5.1. A wrong decision is made only when a subset of false candidates with larger cardinality than the number of correct ones can be covered by a disc having diameter d_{OD}.

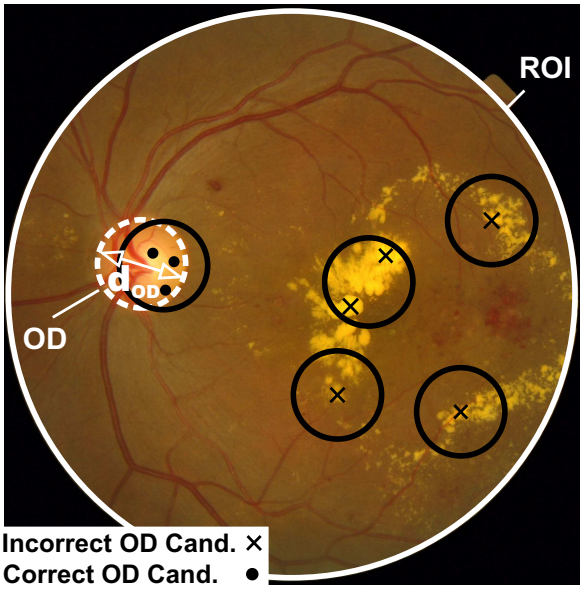


Figure 5.1: *The optic disc (OD) of diameter d_{OD} in a retinal image and the OD center candidates (3 correct, 5 false) of individual detector algorithms. Candidates inside the black circles can vote together for possible OD locations.*

In this chapter, we propose the generalization of the classic majority voting model by incorporating the probability terms $p_{n,k}$ mentioned before. With an appropriate geometric constraint, our generalized model can be specialized to be applicable for the OD detection scenario, as well. Namely, the corresponding values $p_{n,k}$ will be adjusted by requiring that the candidates should

fall inside a disc of a fixed diameter d_{OD} to vote together. With the help of this model, we can also characterize our detector ensemble and gain information on further improvability issues. As a different approach, it would be possible to require more than half of the votes to fall inside such a disc. However, this strict majority voting rule is rather unnatural in the spatial domain, which impression has been also confirmed during our empirical studies.

The rest of the chapter is organized as follows. With recalling the basic concepts of classic majority voting from section 1.3, in section 5.1 we show how to incorporate the probability terms $p_{n,k}$ to constrain the basic formulation. We present theoretical results for the case of independent voters. Since in applications independent detector algorithms can hardly be expected, we also generalize the method to the dependent case in section 5.2. As a main focus, we investigate the possible lowest and highest accuracies of constrained ensembles. Moreover, we both consider equal and different individual accuracies for the members of the ensemble. From the practical point of view, the further improvability of an ensemble is of great importance, so in section 5.3 we give the theoretical background on how an ensemble behaves if a new member is added to it. Section 5.4 contains our empirical results regarding real world applications (OD and macula detection), where we apply this new model to characterize our current detector ensembles and to analyze its further improvability by adding a new algorithm. As complementary research results, in section 5.5 we present how classic weighted majority voting can be generalized to our spatial voting case regarding independent classifiers. To address the issue of dependency among the members more, in section 5.6 we generalize classic diversity measures and show that the corresponding modification of the weights assigned to the member algorithms can raise ensemble accuracy.

5.1 Generalization to constrained voting

For classic majority voting, in [163] Kuncheva et al. discuss exhaustively the following special case. Let n be odd, $|\Omega| = 2$ (each classifier has a binary (correct/false) output value) and all classifiers are independent and have the same classification accuracy $p \in \mathbb{R}_{\geq 0}$. A correct class label is given by majority voting if at least $\lceil n/2 \rceil$ classifiers give correct answers. The majority voting rule with independent classifier decisions gives an overall correct classification accuracy calculated by the formula

$$Q = \sum_{k=\lceil n/2 \rceil}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \tag{5.1}$$

Several interesting results can be found in [164] applying majority voting to pattern recognition tasks. This method is guaranteed to give a higher accuracy than the individual classifiers, if the classifiers are independent and $p > 0.5$ holds for their individual accuracies.

As it has been discussed in the introduction of the chapter, we generalize the classic majority voting approach by considering some constraints that must be also met by the votes. To give a more general methodology beyond geometric considerations, we model this type of constrained voting by introducing values $0 \leq p_{n,k} \leq 1$ describing the probability of making a good decision, when we have exactly k good votes from the n voters. Then, in section 5.4 we will adopt this general model to our practical problem with spatial (geometric) constraints.

As we have summarized in the introduction, several theoretical results are achieved for independent voters in the current literature, so we start with generalizing them to this case. However, in the vast majority of applications, we cannot expect independency among algorithms trying to detect the same object. Thus, later we extend the model to the case of dependent voters with generalizing such formerly investigated concepts that have high practical impact, as well.

5.1.1 The independent case

In our model, we consider classifiers $\mathcal{D}_i$ with accuracies p_i as random variables η_i of Bernoulli distribution, i.e.,

$$P(\eta_i = 1) = p_i, \quad P(\eta_i = 0) = 1 - p_i \quad (i = 1, \dots, n). \tag{5.2}$$

Here $\eta_i = 1$ means correct classification by $\mathcal{D}_i$. In particular, the accuracy of $\mathcal{D}_i$ is just the expected value of η_i, that is, $\mathbb{E}\eta_i = p_i$ $(i = 1, \dots, n)$.

Let $p_{n,k}$ $(k = 0, 1, \dots, n)$ be given real numbers with $0 \le p_{n,0} \le p_{n,1} \le \dots \le p_{n,n} \le 1$, and let the random variable ξ be such that

$$P(\xi = 1) = p_{n,k} \quad \text{and} \quad P(\xi = 0) = 1 - p_{n,k}, \tag{5.3}$$

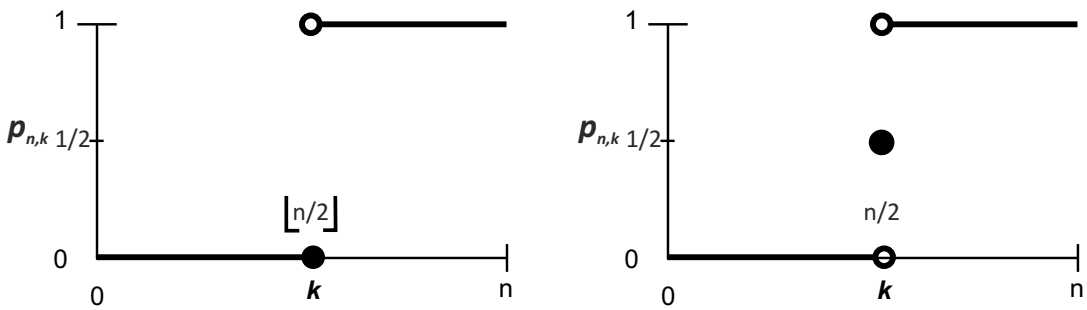
where $k = |\{i \ : \ \eta_i = 1\}|$. That is, ξ represents the modified majority voting of the classifiers $\mathcal{D}_1, \dots, \mathcal{D}_n$: if k out of the n classifiers give a correct vote, then we make a good decision (i.e., we have $\xi = 1$) with probability $p_{n,k}$.

Notice that in the special case, where

$$p_{n,k} = \begin{cases} 1, & \text{if } k > n/2, \\ 1/2, & \text{if } k = n/2, \\ 0, & \text{otherwise}, \end{cases} \tag{5.4}$$

we get back the classic majority voting scheme.

The values $p_{n,k}$ as a function of k corresponding to the classic majority voting can be observed in Figure 5.2 for both an odd and even n, respectively.



(a) n is odd (b) n is even

Figure 5.2: *The graph of $p_{n,k}$ for classic majority voting for (a) an odd, and (b) an even number of voters n.*

Table 5.1: *Ensemble accuracy for classic majority voting.*

	n=3	n=5	n=7	n=9
$p = 0.6$	0.6480	0.6826	0.7102	0.7334
$p = 0.7$	0.7840	0.8369	0.8740	0.9012
$p = 0.8$	0.8960	0.9421	0.9667	0.9804
$p = 0.9$	0.9720	0.9914	0.9973	0.9991

As the very first step of our generalization, we show that similarly to the individual voters, ξ is of Bernoulli distribution, as well. We also provide its corresponding parameter q that represents the accuracy of the ensemble in our model.

Lemma 5.1.1. *The random variable ξ is of Bernoulli distribution with parameter q, where*

$$q = \sum_{k=0}^{n} p_{n,k} \left(\sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\ldots,n\} \setminus I} (1 - p_j) \right). \tag{5.5}$$

Proof. Since for any $k \in \{0, 1, \ldots, n\}$ we have

$$P(|\{i \ : \ \eta_i = 1\}| = k) = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\ldots,n\} \setminus I} (1 - p_j), \tag{5.6}$$

the statement immediately follows from the definition of ξ. □

The special case assuming equal accuracy for the classifiers received strong attention in the literature, so we investigate this case first. That is, in the rest of section 5.1, we suppose that $p = p_1 = \ldots = p_n$. Then, (5.5) reads as

$$q = \sum_{k=0}^{n} p_{n,k} \binom{n}{k} p^k (1 - p)^{n-k}. \tag{5.7}$$

Thus, if n is odd then by the particular choice (5.4) for the values $p_{n,k}$, we get $q = Q$, where Q is given in (5.1). In order to make our generalized majority voting model more accurate than the individual decisions, we have to guarantee that $q \geq p$. The next statement yields a guideline along this way.

Proposition 5.1.2. *Let $p_{n,k} = k/n$ $(k = 0, 1, \ldots, n)$. Then, we have $q = p$, and consequently $\mathbb{E}\xi = p$, where $\mathbb{E}\xi$ is the expected value of ξ.*

Proof. Since by Lemma 5.1.1 ξ is of Bernoulli distribution with parameter q, we have $\mathbb{E}\xi = q$. Thus, we just need to show that $q = p$ whenever $p_{n,k} = k/n$ $(k = 0, 1, \ldots, n)$. By our settings, from (5.7) we have

$$q = \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} p^k (1 - p)^{n-k} = \frac{1}{n} \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k}. \tag{5.8}$$

Observe that the last sum just expresses the expected value np of a random variable of binomial distribution with parameters (n, p). Thus, we have $q = p$, and the statement follows. □

Figure 5.3 also illustrates the special linear case for $p_{n,k} = k/n$.

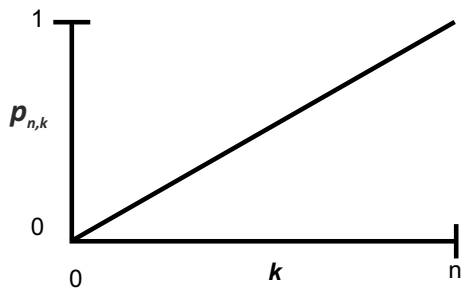


Figure 5.3: *The graph of $p_{n,k} = k/n$ providing $p = q$.*

The above statement shows that if the probabilities $p_{n,k}$ increase uniformly (linearly), then the ensemble has the same accuracy as the individual classifiers. As a trivial consequence we obtain the following corollary.

Corollary 5.1.3. *Suppose that for all $k = 0, 1, \ldots, n$ we have $p_{n,k} \geq k/n$. Then $q \geq p$, and consequently $\mathbb{E}\xi \geq p$.*

The next result helps us compare our model constrained by $p_{n,k}$ with the classic majority voting scheme.

Theorem 5.1.4. *Suppose that $p \geq 1/2$ and for any k with $0 \leq k \leq n/2$ we have:*

(i) $p_{n,k} + p_{n,n-k} \geq 1$,

(ii) $p_{n,n-k} \geq (n-k)/n$.

Let q be given by (5.7). Then, $q \geq p$, and consequently $\mathbb{E}\xi \geq p$.

Proof. See the Appendix. □

As a specific case, we obtain the following corollary concerning the classic majority voting scheme [163].

Corollary 5.1.5. *Suppose that n is odd, $p \geq 1/2$ and for all $k = 0, 1, \ldots, n$ we have*

$$p_{n,k} = \begin{cases} 1, & \text{if } k > n/2, \\ 0, & \text{otherwise.} \end{cases} \tag{5.9}$$

Then, $q \geq p$, and consequently $\mathbb{E}\xi \geq p$.

Proof. Observing that by the above choice for the values $p_{n,k}$ both properties (i) and (ii) of Theorem 5.1.4 are satisfied, the statement immediately follows from Theorem 5.1.4. □

Of particular interest is the case, when the ensemble makes exclusively good decisions after t executions. That is, we are curious to know the conditions to have a system with accuracy 100%. So write $\xi^{\otimes t}$ for the random variable obtained by repeating ξ independently t times, and counting the number of one values (correct decisions) received, where t is a positive integer. Then, as it is well-known, $\xi^{\otimes t}$ is a random variable of binomial distribution with parameters (t, q) with q given by (5.7). Now we are interested in the probability $P(\xi^{\otimes t} = t)$. In case of using an individual classifier $\mathcal{D}_i$ (that is, a random variable η_i) with any $i = 1, \ldots, n$, we certainly have $P(\eta_i^{\otimes t} = t) = p^t$. To make the ensemble better than the individual classifiers, we need to choose the probabilities $p_{n,k}$ so that $P(\xi^{\otimes t} = t) \geq p^t$. In fact, we can characterize a much more general case. For this purpose we need the following lemma, due to Gilat [165].

Lemma 5.1.6. *For any integers t and l with $1 \leq l \leq t$ the function*

$$f(x) = \sum_{k=l}^{t} \binom{t}{k} x^k (1-x)^{t-k} \tag{5.10}$$

is strictly monotone increasing on $[0, 1]$.

Notice that for any $x \in [0, 1]$ we have

$$\sum_{k=0}^{t} \binom{t}{k} x^k (1-x)^{t-k} = 1. \tag{5.11}$$

As a simple consequence of Lemma 5.1.6, we obtain the following result.

Theorem 5.1.7. *Let t and l be integers with $1 \le l \le t$. Then, $P(\xi^{\otimes t} \ge l) \ge P(\eta_1^{\otimes t} \ge l)$, if and only if $q \ge p$, i.e., $\mathbb{E}\xi^{\otimes t} \ge tp$.*

Proof. Let t and l be as given in the statement. Then, we have

$$\begin{aligned} P(\xi^{\otimes t} \ge l) &= \sum_{k=l}^{t} \binom{t}{k} q^k (1-q)^{t-k}, \\ P(\eta_1^{\otimes t} \ge l) &= \sum_{k=l}^{t} \binom{t}{k} p^k (1-p)^{t-k}. \end{aligned} \tag{5.12}$$

Thus, by Lemma 5.1.6, we obtain

$$P(\xi^{\otimes t} \ge l) \ge P(\eta_1^{\otimes t} \ge l), \tag{5.13}$$

if and only if $q \ge p$, and the theorem follows. □

5.2 The dependent case

In this section, we investigate how dependencies among the voters influence the accuracy of the ensemble; for related results, see e.g. [166, 167]. For this purpose, we generalize some concepts that were introduced for classic majority voting to measure the extremal behavior (minimal and maximal accuracies) of an ensemble. First we consider *pattern of success* and *pattern of failure* [166] which are such realizations of the votes in a series of experiments that lead to the possible highest and lowest accuracy of the ensemble, respectively. It is worth noting that to define these measures, a rather serious restriction considering discretization of the model is needed to be applied. Namely, not only the accuracies of the individual classifiers are given, but also the precise numbers of successful decisions during the experiment are fixed. E.g., for a classifier having accuracy $p = 0.6$ we consider 6 correct votes in 10 experimental runs.

Though there are some results in the literature for the case of different accuracies p_i of the classifiers $\mathcal{D}_i$ (or, in other words, for the case $\mathbb{E}\eta_i = p_i$ $(i = 1, \ldots, n)$), see e.g. [168–170] and the references there, the vast majority of the results (such as e.g. in [163]) concern the case $p = p_1 = \ldots = p_n$. So in section 5.2.1, we shall make the latter assumption, too. However, in section 5.2.2, we give a much more general framework which handles both dependencies without the restriction considering discretization, and also different accuracies of classifiers that makes the model realistic for applications.

5.2.1 Pattern of success and pattern of failure

In this section, we suppose that the individual classifier accuracies coincide ($p = p_1 = \ldots = p_n$). Repeat the experiments $\eta_1, \ldots, \eta_n$ t times, with some positive integer t, and write $\eta_i^{(j)}$ for the j-th realization of η_i $(j = 1, \ldots, t)$. Suppose (as a rather strong, but standard assumption) that we have

$$|\{j \ : \ \eta_i^{(j)} = 1\}| = r \quad \text{for all } i = 1, \ldots, n. \tag{5.14}$$

Here r is a positive integer with $r = np$. We are interested in the behavior (accuracy) of ξ repeated t times, or in other words in the value $\mathbb{E}\xi^{\otimes t}$, under the condition (5.14). Write $\xi^{(j)}$ for the j-th realization of ξ $(j = 1, \ldots, t)$. Then, we clearly have $\mathbb{E}\xi^{\otimes t} = \mathbb{E}\xi^{(1)} + \ldots + \mathbb{E}\xi^{(t)}$.

The number of one values is fixed for η_i, however, their positions can freely change. For simplicity, we shall describe the situation by a table T of size $n \times t$: in the (i, j)-th entry $T(i, j)$ of T we write 0 or 1, according to the actual value of $\eta_i^{(j)}$ $(1 \le i \le n, 1 \le j \le t)$. Our first result in this interpretation concerns the case of linear $p_{n,k}$.

Proposition 5.2.1. *If $p_{n,k} = k/n$ for all $k = 0, 1, \ldots, n$, then $\mathbb{E}\xi^{\otimes t} = r$.*

Proof. Denote by u_j the number of ones in the j-th column of the table T for $j = 1, \ldots, t$. Then, we have $\mathbb{E}\xi^{(j)} = u_j/n$. Thus,

$$\mathbb{E}\xi^{\otimes t} = \mathbb{E}\xi^{(1)} + \ldots + \mathbb{E}\xi^{(t)} = u_1/n + \ldots + u_t/n. \tag{5.15}$$

Since $u_1 + \ldots + u_t$ is just the total number of ones in T, we have

$$u_1 + \ldots + u_t = nr. \tag{5.16}$$

Combining (5.15) and (5.16) we obtain $\mathbb{E}\xi^{\otimes t} = r$, and the statement follows. □

In view of the proof of Proposition 5.2.1, we see that in case of a general system $p_{n,k}$ we have

$$\mathbb{E}\xi^{\otimes t} = \sum_{j=1}^{t} p_{n,u_j}. \tag{5.17}$$

So to describe the pattern of success (highest accuracy) and the pattern of failure (lowest accuracy), we need to maximize and minimize the above quantity, respectively.

Our next result concerns the pattern of success. Here we consider the problem only under some further assumptions, which in fact are not necessary to study and describe the situation as it will be shown in section 5.2.2. However, on the one hand, the statement together with its proof already show the basic idea for construction. On the other hand, former results usually consider these assumptions, so in this way our model can be fitted to the existing literature, as well. In section 5.2.2, we describe the general method, which works without any technical restrictions.

Theorem 5.2.2. *Let the probabilities $p_{n,k}$ be arbitrary, up to $p_{n,0} = 0$. Let $k_1 \neq 0$ be an index such that $p_{n,k_1}/k_1 \geq p_{n,k}/k$ for all $k = 1, \ldots, n$. Then, $\mathbb{E}\xi^{\otimes t} \leq nrp_{n,k_1}/k_1$. Further, if $tk_1 = nr$ then the maximum can be attained.*

Proof. Using (5.17) and our assumption $p_{n,k_1}/k_1 \geq p_{n,k}/k$ for all $k = 1, \ldots, n$, we get

$$\mathbb{E}\xi^{\otimes t} = \sum_{j=1}^{t} p_{n,u_j} = \sum_{\substack{j=1 \\ u_j \neq 0}}^{t} u_j p_{n,u_j}/u_j \leq \sum_{j=1}^{t} u_j p_{n,k_1}/k_1 = (p_{n,k_1}/k_1) \sum_{j=1}^{t} u_j = nrp_{n,k_1}/k_1, \tag{5.18}$$

which implies the first part of the statement.

Assume now that we also have $tk_1 = nr$. Fill in the $n \times t$ table T with zeros and ones arbitrarily, such that we have r ones in each row. If there is a column containing less than k_1 ones, then by $tk_1 = nr$ there is another column with more than k_1 ones. Write j_1 and j_2 for the indices of these columns, respectively. Then, there is a row say with index i, such that $T(i, j_1) = 0$ and $T(i, j_2) = 1$. Change these zero and one values, and continue this process as long as possible. Since $tk_1 = nr$, finally we end up with a table T containing r ones in each row and k_1 ones in each column. Then, we have

$$\mathbb{E}\xi^{\otimes t} = \sum_{j=1}^{t} p_{n,k_1} = tp_{n,k_1} = tk_1 p_{n,k_1}/k_1 = nrp_{n,k_1}/k_1, \tag{5.19}$$

and the theorem follows. □

Our next theorem describes the pattern of failure, in a similar fashion as the previous statement.

Theorem 5.2.3. *Let the probabilities $p_{n,k}$ be arbitrary, up to $p_{n,0} = 0$. Let $k_2 \neq 0$ be an index such that $p_{n,k_2}/k_2 \leq p_{n,k}/k$ for all $k = 1, \ldots, n$. Then, $\mathbb{E}\xi^{\otimes t} \geq nrp_{n,k_2}/k_2$. Further, if $tk_2 = nr$ then the minimum can be attained.*

Proof. Since the proof follows the same lines as that of Theorem 5.2.2, we omit the details. □

Similarly to the independent case in section 5.1.1, we also investigate the case, when only good decision is made by the ensemble. In other words, we would like to describe the situation, where

$$P(\xi^{\otimes t} = t) = \prod_{j=1}^{t} p_{n,u_j} \tag{5.20}$$

is maximal. Notice that in this case one can easily obtain a table T with $P(\xi^{\otimes t} = t) = 0$. So now finding the minimum (i.e., investigating the pattern of failure) does not make sense. For the special case of $p_{n,k} = k/n$, we have the following result.

Theorem 5.2.4. *Let $p_{n,k} = k/n$ for all $k = 0, 1, \ldots, n$, and assume that $nr \geq t$. Then $P(\xi^{\otimes t} = t)$ is maximal for the tables T in which*

$$\lfloor nr/t \rfloor \leq u_j \leq \lceil nr/t \rceil \quad (1 \leq j \leq t), \tag{5.21}$$

where u_j denotes the number of ones in the j-th column of T. Further, all these tables T can be explicitly constructed.

Proof. Let T be an arbitrary table having r ones in each row such that T has no column consisting only of zeros. Since $nr \geq t$, such a T exists (and can be easily constructed). In view of the proof of Proposition 5.2.1, for the corresponding $\xi^{\otimes t}$ we have

$$P(\xi^{\otimes t} = t) = (1/n^t) \prod_{j=1}^{t} u_j. \tag{5.22}$$

If for some indices $1 \leq j_1, j_2 \leq t$ we have $u_{j_1} - u_{j_2} \geq 2$, then $(u_{j_1} - 1)(u_{j_2} + 1) > u_{j_1}u_{j_2}$ clearly holds. Hence moving a one from the j_1-th column to the j_2-th column (keeping its row; just as at the end of the proof of Theorem 5.2.2), the new value for $P(\xi^{\otimes t} = t)$ will be larger than the previous one. Continuing this process as long as possible, finally we obtain a table T, where for any indices $1 \leq j_1, j_2 \leq t$ we have $|u_{j_1} - u_{j_2}| \leq 1$. Obviously, this is equivalent to

$$\lfloor nr/t \rfloor \leq u_j \leq \lceil nr/t \rceil \quad (1 \leq j \leq t). \tag{5.23}$$

Observing that for all such tables T the values $P(\xi^{\otimes t} = t)$ coincide, and these tables differ from each other only by a permutation of their columns, the theorem follows. □

Notice that if $t > nr$ then T necessarily has a column with all zero entries, whence $P(\xi^{\otimes t} = t) = 0$ in this case. For general values $p_{n,k}$, we have the following result.

Theorem 5.2.5. *Let the probabilities $p_{n,k}$ be arbitrary, up to $p_{n,0} = 0$ and $p_{n,k} > 0$ for $0 < k \leq n$. Let $k_0 \neq 0$ be an index such that $(\ln p_{n,k_0})/k_0 \geq (\ln p_{n,k})/k$ for all $k = 1, \ldots, n$. Then, $P(\xi^{\otimes t} = t) \leq p_{n,k_0}^{(nr/k_0)}$. Further, if $tk_0 = nr$ then the maximum can be attained.*

Proof. First, we have

$$P(\xi^{\otimes t} = t) = \prod_{j=1}^{t} p_{n,u_j} = \exp\left(\sum_{j=1}^{t} \ln p_{n,u_j}\right). \tag{5.24}$$

On the other hand, by our assumption $(\ln p_{n,k_0})/k_0 \geq (\ln p_{n,k})/k$ for all $k = 1, \ldots, n$

$$\sum_{j=1}^{t} \ln p_{n,u_j} = \sum_{\substack{j=1 \\ u_j \neq 0}}^{t} \frac{u_j \ln p_{n,u_j}}{u_j} \leq \sum_{j=1}^{t} \frac{u_j \ln p_{n,k_0}}{k_0} = \frac{\ln p_{n,k_0}}{k_0} \sum_{j=1}^{t} u_j = \frac{nr \ln p_{n,k_0}}{k_0} \tag{5.25}$$

holds. Thus,

$$P(\xi^{\otimes t} = t) \leq p_{n,k_0}^{(nr/k_0)}, \tag{5.26}$$

which implies the first part of the statement. The second part can be proved by following the argument at the end of the proof of Theorem 5.2.2. □

5.2.2 Extremal accuracies by linear programming

In this section, we drop the condition (5.14), and give a compact tool based on linear programming to calculate the minimal and maximal ensemble accuracies. We assumed earlier that the random variables η_i $(i = 1, \ldots, n)$ are independent. In our application, we consider different algorithms detecting the OD as voters. These algorithms cannot be assumed to be independent in all cases, because it can happen that the operations of the algorithms are based on very similar principles. In case of dependent algorithms, we have to decide how to measure the dependencies of the algorithms. For this aim, we can investigate the joint distribution of the outputs of the algorithms. So let

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1, \ldots, \eta_n = a_n), \tag{5.27}$$

where $a_i \in \{0, 1, *\}$. The star denotes any of the possible correctness values, that is, $* = 0$ or 1, but it is not specified. The probabilities $c_{a_1,\ldots,a_n}$ can be considered as the entries of the contingency table of $\eta_1, \ldots, \eta_n$. The problem to determine the combination of voters achieving the best/worst ensemble performance is equivalent to maximize/minimize the function

$$q(c_{a_1,\ldots,a_n}) = \sum_{k=0}^{n} \left(p_{n,k} \sum_{a_1+\ldots+a_n=k} c_{a_1,\ldots,a_n} \right) \tag{5.28}$$

under the conditions

$$\begin{aligned} &\sum_{a_i=1} c_{*,\ldots,*,a_i,*,\ldots,*} = p_i \quad (i = 1, \ldots, n), \\ &\sum_{a_1,\ldots,a_n} c_{a_1,\ldots,a_n} = 1, \\ &c_{a_1,\ldots,a_n} \geq 0, a_i \in \{0, 1\} \quad (i = 1, \ldots, n), \end{aligned} \tag{5.29}$$

where $\mathbb{E}\eta_i = p_i$ $(i = 1, \ldots, n)$ is the accuracy of the i-th detecting algorithm. Observe that this is just a linear programming problem for the variables $c_{a_1,\ldots,a_n}$, which can be solved by standard tools.

In the special case, when $(\eta_1, \ldots, \eta_n)$ are totally independent, we have

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1) \ldots P(\eta_n = a_n). \tag{5.30}$$

That is, the entries of the contingency table can be determined by the probabilities $p_1, \ldots, p_n$. In this case, the ensemble performance q is simply given by (5.5).

5.3 Extending the ensemble by adding a new classifier

From the practical point of view, it is very important to study the improvability of an existing ensemble regarding its accuracy. To address this issue, we investigate to what extent the addition of a new classifier $\mathcal{D}_{n+1}$ with accuracy p_{n+1} may improve the system. For this study, we observe both the change of the system accuracy q and the interval $[q_{min}, q_{max}]$ for the minimal and maximal system accuracy. More precisely, we will consider the following cases.

A. We fix the individual accuracies and output of the algorithms of the current ensemble for an experiment in terms of a contingency table, and:

   1. add a new independent algorithm and check how the ensemble accuracy (q) changes,
   2. add a new dependent algorithm and check how the minimal (q_{min}) and maximal (q_{max}) ensemble accuracy change, respectively.

B. We fix the individual accuracies, but ignore the output of the algorithms of the current ensemble for an experiment, add a new algorithm and check the minimal (q_{min}) and maximal (q_{max}) ensemble accuracy.

After adding a new algorithm to the existing system, the new system accuracy depends not only on the accuracies $p_1, \ldots, p_{n+1}$, but also on the values $p_{n+1,k}$. As an estimation for $p_{n+1,k}$, from the definition of $p_{n,k}$ we have

$$p_{n,k} \geq p_{n+1,k}, \tag{5.31}$$

$$p_{n,k} \leq p_{n+1,k+1}. \tag{5.32}$$

In (5.31), the added vote is supposed to be false, so the probability of good decision after the extension cannot be greater than in the existing system. The estimation (5.32) describes the case of adding a correct vote to the system. To sum up (5.31) and (5.32), we get the following properties for $p_{n+1,k}$:

$$p_{n,k-1} \leq p_{n+1,k} \leq p_{n,k}. \tag{5.33}$$

Applying inequalities (5.33), the values $p_{n+1,k}$ can be estimated from the values $p_{n,k}$.

If a new member is added to an existing ensemble, the accuracy of the extended ensemble is affected by two main properties of the new voter: its accuracy and its correlation with the members of the existing system. Let η_{n+1} be a random variable with $\mathbb{E}\eta_{n+1} = p_{n+1}$. To determine the best/worst choice for the new member to achieve the best (q_{max})/worst (q_{min}) performance for the extended ensemble the following linear optimization problem has to be solved in the general case B. Maximize/Minimize the function

$$q(c_{a_1,\ldots,a_{n+1}}) = \sum_{k=0}^{n+1} \left(p_{n+1,k} \sum_{\substack{a_1+\ldots+a_n+\\ +a_{n+1}=k}} c_{a_1,\ldots,a_{n+1}} \right) \tag{5.34}$$

under the following conditions:

$$\begin{aligned} &\sum_{a_i=1} c_{*,\ldots,*,a_i,*,\ldots,*} = p_i \quad (i = 1, \ldots, n+1), \\ &\sum_{a_1,\ldots,a_{n+1}} c_{a_1,\ldots,a_{n+1}} = 1, \\ &c_{a_1,\ldots,a_{n+1}} \geq 0, \ a_i \in \{0,1\} \quad (i = 1, \ldots, n+1), \end{aligned} \tag{5.35}$$

where $\mathbb{E}\eta_i = p_i$ $(i = 1, \ldots, n+1)$, so the accuracy of the i-th classifier is p_i.

In case A.2., besides the objective function in (5.34) and the conditions in (5.35) are the same, we have an extra condition

$$c_{a_1,\dots,a_n} = c_{a_1,\dots,a_n,0} + c_{a_1,\dots,a_n,1}. \tag{5.36}$$

From the definition of $c_{a_1,\dots,a_n}$ given in (5.27) it follows that the term containing $c_{a_1,\dots,a_{n+1}}$ in (5.34) can be split as

$$\sum_{a_1+\dots+a_{n+1}=k} c_{a_1,\dots,a_{n+1}} = \sum_{a_1+\dots+a_n=k} c_{a_1,\dots,a_n,0} + \sum_{a_1+\dots+a_n=k-1} c_{a_1,\dots,a_n,1}. \tag{5.37}$$

Without having any further information about $p_{n+1,k}$, we can give an interval for q_{min} and q_{max}. Let $q^{\ominus}_{min}/q^{\ominus}_{max}$ and $q^{\oplus}_{min}/q^{\oplus}_{max}$ be the minimal/maximal value of the objective function (5.34) if we consider the estimations $p_{n,k-1} = p_{n+1,k}$ and $p_{n+1,k} = p_{n,k}$, respectively. From (5.33), we get

$$q^{\ominus}_{min} \le q_{min} \le q^{\oplus}_{min}, \quad \text{and} \quad q^{\ominus}_{max} \le q_{max} \le q^{\oplus}_{max}. \tag{5.38}$$

In the special case, when η_{n+1} is totally independent from $(\eta_1, \dots, \eta_n)$, the entries of the extended contingency table can be determined by $c_{a_1,\dots,a_n}$ and p_{n+1} as

$$\begin{aligned} c_{a_1,\dots,a_n,1} &= p_{n+1} c_{a_1,\dots,a_n}, \\ c_{a_1,\dots,a_n,0} &= (1-p_{n+1}) c_{a_1,\dots,a_n}. \end{aligned} \tag{5.39}$$

Considering the equations (5.34), (5.37) and (5.39) we get that the linear optimization problem can be solved by maximizing/minimizing the function

$$q(c_{a_1,\dots,a_{n+1}}) = \sum_{k=0}^{n+1} p_{n+1,k} \left(\sum_{a_1+\dots+a_n=k} (1-p_{n+1}) c_{a_1,\dots,a_n} + + \sum_{a_1+\dots+a_n=k-1} p_{n+1} c_{a_1,\dots,a_n} \right) \tag{5.40}$$

under the conditions given in (5.29).

If we consider that the entries of the contingency table of $\eta_1, \dots, \eta_n$ remain the same after adding an independent variable η_{n+1} to the ensemble (case A.1.), the solution of the problem in (5.40) under the conditions (5.29) depends only on p_{n+1} and $p_{n+1,k}$.

In the same way as in (5.38), from (5.33) we get

$$q^{\ominus} \le q \le q^{\oplus}, \tag{5.41}$$

where $q^{\ominus}$ and $q^{\oplus}$ denote the minimal/maximal value of the objective function (5.40) for a fixed p_{n+1} if we consider the estimations $p_{n,k-1} = p_{n+1,k}$ and $p_{n+1,k} = p_{n,k}$, respectively.

For the improvability of the system, we have the following proposition.

Proposition 5.3.1. *For the accuracy of the extended ensemble we have*

$$q(c_{a_1,\dots,a_{n+1}}) \ge q(c_{a_1,\dots,a_n}), \tag{5.42}$$

provided that

$$p_{n+1} \ge \frac{\sum_{k=0}^{n} \left(\sum_{a_1+\dots+a_n=k} c_{a_1,\dots,a_n} (p_{n,k} - p_{n+1,k}) \right)}{\sum_{k=0}^{n} \left(\sum_{a_1+\dots+a_n=k} c_{a_1,\dots,a_n} (p_{n+1,k+1} - p_{n+1,k}) \right)} \tag{5.43}$$

holds for the accuracy of the added member.

Proof. First, notice that the value of this fraction is non-negative, since $p_{n,k} \ge p_{n+1,k}$ and $p_{n+1,k} \le p_{n+1,k+1}$. Thus, from (5.33) and (5.39) the statement follows. □

In section 5.4.4, we will show some experimental results for the improvability of the accuracy of our OD detector ensemble with adding a new algorithm.

5.4 Application – OD detection

Now we turn to show how our generalized model supports real-world problems in a clinical field. Progressive eye diseases can be caused by diabetic retinopathy (DR) which can lead even to blindness. One of the first essential steps in automatic grading of the retinal images is to determine the exact location of the main anatomical features, such as the OD. The locations of these features play important role in making diagnosis in the clinical protocol. In this section, for the OD detection task, we start with showing how the general formulation considering the probabilities $p_{n,k}$ is restricted for this specific challenge using geometric constraints defined by anatomic rules. Then, we present the accuracy of our current ensemble, characterize it by the achieved results and discuss the possibilities of its further improvement.

5.4.1 Constraining by shape characteristics

In our application, the votes are required to fall inside a disc of diameter d_{OD} to vote together. For the calculation of the values $p_{n,k}$ for our proposed method, the k correct votes must fall inside the true OD region, however, the $n-k$ false ones can fall within discs with diameter d_{OD} anywhere else within the ROI (region of interest in the image). That is, more false regions are possible to be formed which gives the possibility to make a correct decision even if the true votes are in minority. Notice that a candidate of an algorithm is considered to be correct if its distance from the manually selected OD center is not larger than $d_{OD}/2$. For this configuration, see Figure 5.5.

If we assume independency among the algorithms, for our application the behavior of the values $p_{n,k}$ as a function of k for a given n is shown in Figure 5.4 for $n = 9$ and $p = 0.9$.

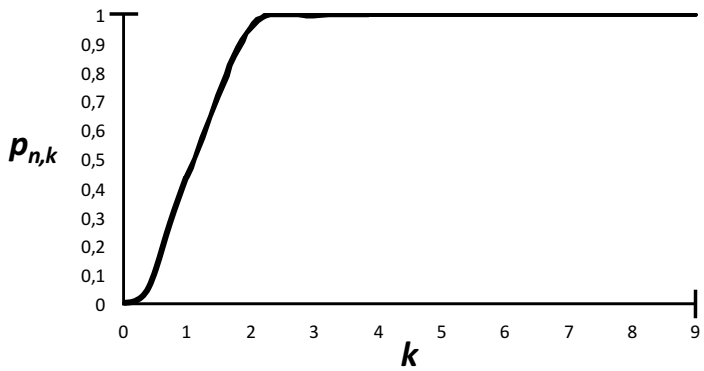


Figure 5.4: *The graph of $p_{n,k}$ for $n = 9$ and $p = 0.9$ with our geometric constraint to fall within a disc of diameter d_{OD}.*

This function has been determined empirically by dropping random pixels on the disc in a large number of experiments. Figure 5.4 shows that $p_{n,k}$ increases exponentially in k for a given n. This fact is also suggested by the results in [171, 172] saying that the probability that the diameter of a point set is not less than a given constant decreases exponentially if the number of points tends to infinity. Note that, this diameter corresponds again to the diameter d_{OD} of the OD.

The ensemble accuracy for our spatially constrained system is measured empirically by the help of a set of test images. The obtained data are enclosed in Table 5.2 for different number of independent classifiers (n) for some equal individual accuracies (p).

From Table 5.2 we can see a rapid increase in the ensemble accuracy. From trivial geometric considerations, it can be also seen why an ensemble with few members (e.g. $n = 3$) performs bad.

	n=3	n=5	n=7	n=9
$p = 0.6$	0.6435	0.9076	0.9654	0.9893
$p = 0.7$	0.7889	0.9631	0.9938	0.9985
$p = 0.8$	0.9029	0.9906	0.9986	0.9997
$p = 0.9$	0.9697	0.9994	1.0000	1.0000

Table 5.2: *Ensemble accuracy under the geometric constraint.*

Now, to describe the spatially constrained case in detail, let us assign the probability $(1-p_i)s_i$ with $s_i \in [0,1]$ to the i-th independent classifier. This probability means that the i-th voter makes false individual decision (term $1-p_i$) and participates in making a false ensemble decision (term s_i). For the algorithm $\mathcal{D}_i$ with accuracy p_i giving a false candidate having coordinates (x_i, y_i) for the OD center, we consider that the distribution of (x_i, y_i) is uniform outside the true OD region for all $i = 1, \ldots, n$. With this setup, we have

$$s_1 = \ldots = s_n = \frac{T_0}{T - T_0}, \tag{5.44}$$

where T_0 and T are the area of the OD and the ROI, respectively, so in this case s_i is the same predetermined constant for all $i = 1, \ldots, n$. For better understanding, see also Figure 5.5.

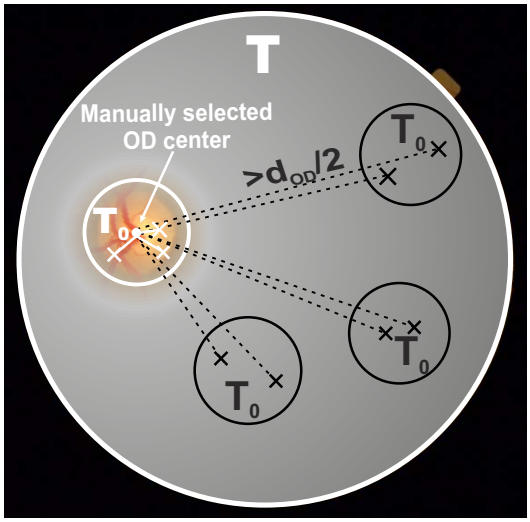


Figure 5.5: *The geometric constraint applied to the candidates of the algorithms: they should fall inside a disc of a fixed diameter d_{OD} to vote together.*

For the interpretation of the values $p_{n,k}$ for this case, let us consider the decomposition of the number of false candidates $n - k = k_1 + \ldots + k_l$, where all the false votes are covered by the l disjoint discs of diameter d_{OD}, and k_i is the cardinality of the false votes covered by the i-th disc. Without the loss of generality, we may assume that $k_1 \geq \ldots \geq k_l$. To determine the values $p_{n,k}$, we introduce the probability $P(n, k, k_1, \ldots, k_l)$ for the good decision in case of a concrete realization of the n votes as

$$P(n, k, k_1, \ldots, k_l) = \frac{n!}{k!k_1! \ldots k_l!} p_1 \ldots p_k (1 - p_{k+1}) \ldots (1 - p_n) \left(1 - \frac{T_0}{T}\right)^{k_1} \ldots \left(1 - \frac{lT_0}{T}\right)^{k_l}. \tag{5.45}$$

Applying the geometric constraint, false decision is made only when $k_1 > k$ so $p_{n,k} = 0$ for $k_1 > k$, while $p_{n,k} = 1$ for $k > k_1$ should hold. The case $k_1 = k$ is broken randomly. Based on these

considerations and summing for the possible distribution of the $n-k$ false votes among the discs, we can calculate the corresponding values $p_{n,k}$ as follows:

$$p_{n,k} = \sum_{k_1+\ldots+k_l=n-k, k>k_1} P(n,k,k_1,\ldots,k_l) + \frac{1}{2} \sum_{k_1+\ldots+k_l=n-k, k=k_1} P(n,k,k_1,\ldots,k_l). \tag{5.46}$$

The values $p_{n,k}$ calculated by (5.46) and the ones shown in Figure 5.4 slightly differ. The reason for this is that in our geometric derivation to have the closed form (5.46), we have considered only disjoint discs that completely fall inside the ROI, as well. However, these differences are minor, and both approaches have exponential trends.

From the basic results and concepts introduced in sections 1.3 and 5.1, strict majority voting scheme could be also applied as a decision rule, which means that at least $\lfloor n/2 \rfloor + 1$ votes should fall within a disc of diameter d_{OD} to make a good decision. However, this strict approach is much more unnatural than the proposed one confirmed by the experimental results presented in the next sections, as well.

5.4.2 An ensemble-based OD detector

To take advantage of the theoretical foundations of the previous sections for efficient OD detection, we have collected eight corresponding individual algorithms to create an ensemble from. Then, with a brute force approach (i.e., checking all the possible combinations) we select such an ensemble which maximizes the accuracy of the combined system. For measuring the accuracy of both the individual algorithms and the ensembles, we used the dataset Messidor described in section 1.5.5. As a result of brute force selection, we composed an ensemble from six OD-detectors, namely OD_1, OD_2, OD_3, OD_4, OD_5, and OD_6 described in section 1.6.2. Each individual accuracy (p_i) for these algorithms has been measured on the dataset Messidor and found to be $p_1 = 0.767$, $p_2 = 0.647$, $p_3 = 0.958$, $p_4 = 0.977$, $p_5 = 0.759$, and $p_6 = 0.315$, respectively.

As for the decision of the ensemble, we select the disc of the fixed diameter d_{OD} containing the largest number of algorithm candidates. Then, as the final OD center, we consider the centroid of these candidates. The final OD center is correctly found, if it falls inside the disc aligned to the manually selected OD center and having diameter d_{OD}.

5.4.3 Characterizing and comparing OD-ensemble accuracies

A natural question regarding the ensemble of the detectors is what accuracies we can expect as the best or worst based on the given individual detector accuracies. Then, we can see where the accuracy of our current ensemble falls within this interval, and can also check how it relates to a system which would contain independent ensemble members.

In our application, the values $p_{n,k}$ for calculating the above characterizing ensemble accuracies as a function of k for $n = 6$ is calculated empirically and shown in Figure 5.6. Notice that though our system naturally contains dependencies among its members, the exponential behavior of the independent ensemble (see Figure 5.4) can be observed here, as well.

Using the linear programming technique described in section 5.2.2, we have the following minimal and maximal ensemble accuracies, respectively:

$$q_{min} = 0.899, \qquad q_{max} = 1 \tag{5.47}$$

for the given individual accuracies.

Based on our experimental tests, the ensemble accuracy for our system has been found to be

$$q = 0.981, \tag{5.48}$$

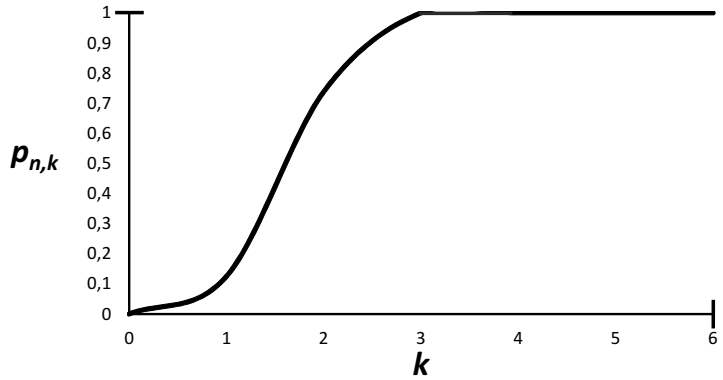


Figure 5.6: *The graph of $p_{n,k}$ for $n = 6$ in our OD detector ensemble.*

which is quite close to the possible maximal accuracy $q_{max} = 1$. However, if we calculate the system accuracy using (5.28) under the conditions (5.29) and with the assumption (5.30) on the independency of the detectors, we have

$$q_{ind} = 0.998. \tag{5.49}$$

That is, an ensemble of independent algorithms with the given individual accuracies $p_1, \ldots, p_6$ would lead to nearly perfect results regarding accuracy. On the other hand, it is not surprising that our current system performs worse, since in this specific detection task it is quite challenging to find algorithms based on different (independent/diverse) principles.

Similarly to our proposed method, we have also determined the highest ensemble accuracy regarding the strict majority voting scheme. In this case, the brute force search provided the highest accuracy for the ensemble of the five members having individual accuracies p_1, p_3, p_4, p_5, and p_6, respectively. The ensemble accuracy measured by following the strict decision rule (at least three votes should fall within a disc of diameter d_{OD}) has been found to be

$$q_{strict} = 0.944. \tag{5.50}$$

Comparing (5.50) with (5.48) confirms that the proposed spatially constrained voting model leads to remarkably higher accuracy than by simply extending the classic majority vote rule.

5.4.4 On adding algorithms to the detector

In section 5.3, we have laid the theoretical background to extend the ensemble with adding a new algorithm. Namely, we have formulated the ways of the calculation of ensemble accuracy for the cases, when the new member is dependent or independent from the ensemble, respectively. Besides the simple ensemble accuracy, we have also explained how the minimal and maximal accuracies of the ensemble would change. Now, we adopt these results to our specific application and investigate how our current OD detector ensemble is going to behave if a new detector algorithm is added. Besides the simple ensemble accuracy, we have also explained how the minimal and maximal accuracies of the ensemble would change.

To start our experimental discussion on this topic, we check the behavior of our current OD detector ensemble during its compilation. Namely, we measure the change of the ensemble accuracy, when the sixth member is added to the ensemble of five members. For this aim, we calculate the accuracy of each ensemble of five individual algorithms with the corresponding figures enclosed in

Index of excluded member	1	2	3	4	5	6
Ensemble accuracy (5 members)	0.980	0.976	0.957	0.961	0.976	0.979
Ensemble accuracy (6 members)	0.981					

Table 5.3: *Change of the ensemble accuracy, when the sixth member is added to the ensemble of five algorithms.*

Table 5.3. Thus, Table 5.3 contains the accuracies of the six possible ensembles of five members, where in the ith column the ith member is excluded having individual accuracy p_i, for $i = 1, \ldots, 6$.

From Table 5.3 we can see that the largest increase in accuracy (from 0.957 to 0.981) is reached not by adding the most accurate ($p_4 = 0.977$) member, but a slightly less accurate ($p_3 = 0.958$) one. Similarly, the smallest improvement (from 0.980 to 0.981) is found not by adding the least accurate ($p_6 = 0.315$) member, but by adding an individually more accurate ($p_1 = 0.767$) one. To understand these results we should realize that there are specific dependencies among the members. Thus, in general, it is not sufficient to simply compose an ensemble based on the individual accuracies.

Next, we adopt the results from section 5.3 to investigate how our current OD detector ensemble consisting of six algorithms is going to behave if a new detector algorithm is added. We start with the case A.1 from section 5.3, when the dependencies of the current ensemble members are considered as known in terms of a contingency table belonging to our experimental test on the dataset Messidor and the new algorithm is considered to be independent from the ensemble. For this case, through the solution of (5.40), we gain the numeric results enclosed in Table 5.4. Notice that in this case we can check the interval $[q^{\ominus}, q^{\oplus}]$ introduced in (5.41) where the ensemble accuracy will fall based on the lower and upper estimation that can be derived for $p_{n+1,k}$ as given in (5.33).

Accuracy of the new algorithm	$q^{\ominus}$	$q^{\oplus}$
$p_7 = 0.6$	0.957	0.989
$p_7 = 0.9$	0.975	0.995

Table 5.4: *The interval for the OD detector ensemble accuracy, if a new independent algorithm is added to a dependent system.*

From Table 5.4, we can see that in our application a new (independent) algorithm with accuracy approximately 0.9 is highly expected to improve the current system accuracy given in (5.48). The case A.1 in section 5.3 also includes the special scenario, when the existing ensemble contains independent members and we add an independent algorithm, as well. For this scenario, we can investigate the minimal and maximal accuracies of the new system by solving the problem in (5.40) under the extra condition (5.30). In Table 5.5, we enclosed the respective accuracy figures regarding the lower and upper estimations of the values $p_{n+1,k}$.

Accuracy of the new algorithm	$q^{\ominus}$	$q^{\oplus}$
$p_7 = 0.6$	0.975	0.997
$p_7 = 0.9$	0.984	0.999

Table 5.5: *The interval for the OD detector ensemble accuracy, if a new independent algorithm is added to an independent system.*

By comparing Table 5.4 with Table 5.5, we can see that if we assume total independency among the algorithms, we can expect higher ensemble accuracy. Since the original ensemble would lead to

very high accuracy with independent algorithms as given in (5.49), only in case of a very accurate new algorithm we can expect improvement.

Next, we analyze the case A.2 from section 5.3, when the dependencies of the algorithms are still considered, but the new algorithm does not need to be independent. In this setup, we can determine the accuracy interval introduced in (5.38) for the minimal (q_{min}) and maximal (q_{max}) ensemble accuracies, respectively, based on the estimation for the values $p_{n+1,k}$ as given in (5.33). The corresponding figures presented in Table 5.6 can be determined by the solution of (5.34) under the conditions (5.35), (5.36).

Accuracy of the new algorithm	$q_{min}^{\ominus}$	$q_{min}^{\oplus}$	$q_{max}^{\ominus}$	$q_{max}^{\oplus}$
$p_7 = 0.1$	0.920	0.981	0.981	0.995
$p_7 = 0.7$	0.920	0.981	0.981	0.995
$p_7 = 0.9$	0.942	0.981	0.981	0.995

Table 5.6: *The interval for the minimal and maximal OD detector ensemble accuracy, if a new dependent algorithm is added to a dependent system.*

Table 5.6 shows that an individually very weak, but diverse algorithm could lead to a remarkable improvement of the ensemble, however, this possibility is rather unrealistic. Moreover, since the current ensemble is not optimal regarding dependencies, even with a very diverse and accurate algorithm we cannot reach accuracy 100%. It is also visible from Table 5.6 that the original system accuracy (5.48) cannot be outperformed with the lower estimation for $p_{n+1,k}$, and cannot be degraded with its upper estimation, either.

Another point which is worth considering is that since the retinal databases are quite heterogeneous, we cannot go for sure regarding the dependencies of the algorithms of the ensemble found for a specific (in our case for the Messidor) database. Thus, if we kept the individual accuracies of the ensemble members, but drop the dependency relations, it would be useful to know to what extent a new algorithm may ruin or improve the ensemble accuracy. Consequently, we investigate the case B in section 5.3, when a new algorithm with accuracy p_7 is added to our current ensemble with no constraints given for the dependencies. In other words, we check the intervals for the minimal and maximal accuracies of the extended system regarding the lower and upper estimation of the values $p_{n+1,k}$, respectively. The corresponding figures enclosed in Table 5.7 can be determined by solving (5.34) under the conditions (5.35).

Accuracy of the new algorithm	$q_{min}^{\ominus}$	$q_{min}^{\oplus}$	$q_{max}^{\ominus}$	$q_{max}^{\oplus}$
$p_7 = 0.7$	0.764	0.899	1	1
$p_7 = 0.9$	0.908	0.934	1	1

Table 5.7: *The interval for the minimal and maximal OD detector ensemble accuracy, if a new dependent algorithm is added to a system with no dependency constraints.*

Table 5.7 indicates the natural fact that if the dependencies are unknown, the minimal and maximal accuracy can highly differ, and e.g. the ensemble performance can be worse than that of some of its members. However, it is also worth considering for our specific OD detector ensemble that a new algorithm of accuracy $p_7 = 0.9$ by all means will raise the minimal system accuracy given in (5.47). A comparison with Table 5.6 shows that if we do not assume any dependencies for the original ensemble, we can reach higher maximal and lower minimal system accuracies.

For the strict majority voting approach, an ensemble with even number of members is meaningless, since as it is also known from classic theory [164] ensemble accuracy always drops for even number of members n regarding the $n-1$ case. So we analyzed the change in accuracy, when

the ensemble containing five members is extended to seven members. First of all, we determined the most accurate ensemble with seven members from all the implemented eight algorithms. This ensemble includes the same six algorithms as listed before plus OD_7 described in section 1.6.2 having individual accuracy $p_7 = 0.320$ on the Messidor database. Then, we selected the most/least accurate ensembles with five members, respectively, and checked which members were added to compile the ensemble with seven members. The corresponding quantitative results are given in Table 5.8.

Indices of excluded members	3, 4 (lowest accuracy)	6, 7 (highest accuracy)
Ensemble accuracy (5 members)	0.626	0.944
Ensemble accuracy (7 members)	0.853	

Table 5.8: *Change of the ensemble accuracy for strict majority, when the sixth and seventh members are added to the ensemble of five algorithms.*

The results of Table 5.8 are quite obvious, since two individually highly accurate (p_3, p_4) and also two rather inaccurate (p_6, p_7) algorithms are present. Thus, their joint removal leads to a strong drop/increment regarding the ensemble accuracy, respectively.

5.4.5 Further testing and macula detection

Our detailed experimental studies were performed on the image dataset Messidor which is described in section 1.5.5. However, it is well-known that we can expect high variance among retinal image databases (see e.g. [7]), so tests on different datasets are recommended. Thus, to validate more its efficiency, we tested the proposed ensemble-based approach on a database containing 327 images provided by the Moorfields Eye Hospital, London described in section 1.5.4. The highest accuracy $q = 0.921$ has been found for the ensemble containing the four members having individual accuracies $p_1 = 0.798$, $p_4 = 0.835$, $p_5 = 0.801$, $p_6 = 0.150$, respectively (for the remaining three algorithms we measured $p_2 = 0.342$, $p_3 = 0.780$, and $p_7 = 0.297$, respectively). Similarly to Messidor, the ensemble performed better than any of its members for the Moorfields dataset, as well. Moreover, we can observe that the individual accuracies varied more among the different datasets than that of the ensemble. This observation suggests that we can expect a more stable and calculable behavior if we work with ensembles.

Our approach can be extended to other detection problems with keeping in mind that the presented results are suitable to handle such shapes that can be described by set diameter. To demonstrate the efficiency of our method, we considered another detection problem: the localization of the macula, which is the center of the sharp vision in the retina and appears as a dark, disc-like object of diameter approximately 6mm. That is, we have a very similar scenario to that of the OD detection problem. We set up an ensemble of the five macula detectors MAC_1, MAC_2, MAC_3, MAC_4, and MAC_5 having individual accuracies 0.583, 0.870, 0.714, 0.624, 0.962, respectively. By applying the proposed spatially constrained decision scheme, we found 0.968 for the accuracy of the ensemble for the dataset Messidor. From this result we can see that our ensemble-based approach led to improvement in this field, as well.

5.5 Generalizations to weighted voting systems

In this section, we modify the final decision rule of the ensemble which will result in further improvement of system accuracy. Our generalization is based on the assignment of weights to the ensemble members (classifiers). First, we recall the necessary procedure for finding the weights in

classic majority voting (see e.g. [166]). Then, we derive how the appropriate weights can be found in our generalized voting case.

5.5.1 Classic weighted voting system

For weighted voting system, first let us consider the classifiers $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$ with accuracies $p_1, p_2, \ldots, p_n$, respectively. For this case, from section 1.3 we recall the ensemble classifier $\mathcal{D}_{wmaj}$ corresponding this type of decision rule with the discriminant functions

$$g_j(\boldsymbol{\chi}) = \sum_{i=1}^{n} w_i d_{i,j}. \tag{5.51}$$

Notice that the following discriminant functions can be equivalently used as decision rules:

$$g_j(\boldsymbol{\chi}) = P(s|\omega_j)P(\omega_j), \;\; g_j(\boldsymbol{\chi}) = \log(P(s|\omega_j)P(\omega_j)), \tag{5.52}$$

where $s = [s_1, \ldots, s_n]$ is the vector with the label output of the ensemble, where $s_i \in \Omega$ is the label suggested for $\boldsymbol{\chi}$ by the classifier $\mathcal{D}_i$ and $P(\omega_j)$ is the prior probability for class ω_j. A natural problem arises in the weighted majority system how to choose the optimal weights for the classifiers. If we consider independent classifiers $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$ with accuracies $p_1, p_2, \ldots, p_n$, then the system accuracy is maximized by assigning weights

$$w_i \propto \log \frac{p_i}{1-p_i}, \;\; i = 1, \ldots, n, \tag{5.53}$$

where $\propto$ stands for the approximately proportional relation. Notice that conditional independence is considered here, that is

$$P(s|\omega_j) = \prod_{i=1}^{n} P(s_i|\omega_j), \tag{5.54}$$

where $s = [s_1, \ldots, s_n]$ is the vector with the label output of the ensemble, where $s_i \in \Omega$ is the label suggested for $\boldsymbol{\chi}$ by the classifier $\mathcal{D}_i$.

The weights $w_i \propto log \frac{p_i}{1-p_i}$ do not guarantee the minimum classification errors, because the prior probabilities $P(\omega_j)$ for the classes have to be taken into account, too. More precisely, if the individual classifiers are mutually independent, and the *a priori* likelihood is that each choice is equally likely to be correct, the decision rule that maximizes the system accuracy is a weighted majority voting rule, obtained by assigning weights $w_i \propto \log \frac{p_i}{1-p_i}$.

5.5.2 Assigning weights by adopting shape constraint

This method (using weights in voting rule) can be applied in our generalized voting scheme presented in section 5.5.1. If we consider the classifiers $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$ with respective accuracies $p_1, p_2, \ldots, p_n$ and weights $w_1, \ldots, w_n$, then the final decision is made by choosing the maximal sum of weights, where some additional (e.g. geometric) constraints have to be fulfilled by the classifier outputs. Let us consider the probability $(1-p_i)r_i$ for the i-th classifier that means that the i-th classifier makes wrong classification and participates in making a wrong decision (such as fulfills the additional constraints, as well.)

In our application, we choose the maximal sum of those weights of the algorithms, whose outputs can be bounded by a circle with an appropriate radius. An algorithm takes part in making a wrong decision if its output falling outside the OD is close to other wrong candidates.

For the algorithm $\mathcal{D}_i$ with accuracy p_i giving a wrong candidate $\boldsymbol{c_i}$ for the OD we assume that the distribution of $\boldsymbol{c_i}$ is uniform outside the OD for all i ($i = 1, \ldots, n$). In this case, we have

$$r_1 = \ldots = r_n = \frac{T_0}{T - T_0}, \tag{5.55}$$

where T_0 and T are the area of the OD and the ROI, respectively, so r_i is the same predetermined constant for all i ($i = 1, \ldots, n$).

Theorem 5.5.1. *If independent classifiers $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$ are given (conditional independence is considered), then the optimal weight w_i for the classifier $\mathcal{D}_i$ with accuracy p_i can be calculated as*

$$w_i \propto \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)}. \tag{5.56}$$

Proof. See the Appendix. □

Notice that the weights $w_i \propto \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)}$ do not always guarantee the minimum classification errors. Only if the individual classifiers are independent and the prior probabilities for the classes $P(\omega_j)$ are equal, the decision rule that maximizes the system accuracy is a weighted majority voting rule, obtained by assigning weights by $w_i \propto \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)}$.

5.5.3 The weighted majority voting in OD detection

In our application, the output of each OD detecting algorithm OD_i is the OD center as a single pixel $\boldsymbol{c_i}$. In our ensemble-based system we have the set of class labels $\{\omega_{\boldsymbol{x}} | \boldsymbol{x} \in \text{ROI}\}$. For the OD detector OD_i with its output $\boldsymbol{c_i}$, the class label $\omega_{\boldsymbol{c_i}}$ is assigned to the OD. In this case, the classification is correct if the output $\boldsymbol{c_i}$ falls inside the OD in the retinal image. We can define the decision rule as the sum of the weights of the OD detecting algorithms, whose outputs can be covered by a disc of radius $d_{OD}/2$. The disc with the maximal sum of weights is accepted as the final candidate for the OD.

In this application, the condition for the equal prior probabilities for the classes is fulfilled if we suppose uniform distribution of the candidates both inside and outside the OD.

In contrast to the non-weighted systems, less conflicting situations can be obtained when the decision is not exact because of the equal number of outputs falling inside the discs of the predetermined radius. Further improvement of this weighted system on majority voting is that there is no need for accuracy constraints $p > 0.5$ on individual algorithms to achieve larger system accuracy. It can be shown that this weighted voting rule always outperforms the classic majority rule, since in case of a conflict (when the same number of votes are reached in other discs of radius $d_{OD}/2$) majority rule decides randomly between the disc candidates, while the weighted voting system can handle the conflict determining to the sum of the weights corresponding the output votes falling inside the discs.

5.5.4 Experimental results

We compare the system accuracies of the classic and the weighted majority voting for different accuracies and weights. We consider three schemes of accuracies for the algorithms:

- $A_1 : p_1 = p_2 = \ldots = p_9 = 0.6$,
- $A_2 : p_i = 1 - 0.1i$, $i = 1, \ldots, 9$,

- $A_3 : p_1 = 0.767, p_2 = 0.647, p_3 = 0.958, p_4 = 0.977, p_5 = 0.759, p_6 = 0.315, p_7 = 0.320, p_8 = 0.228, p_9 = 0.643$.

The first case is often examined in the literature with equal weights, the second one describes an artificial scenario with linearly dropping accuracies, while the third case contains the true accuracies of our OD detecting algorithms. The accuracy values $p_1, \ldots, p_7$ correspond to the seven OD detector algorithms discussed in section 5.4, while p_8 and p_9 to two detectors (from now on denoted by OD_8 and OD_9, respectively) implemented based on [173]. These accuracies were found on the Messidor database described in section 1.5.5.

For the weighted voting system, we apply the following weights w_i for the i-th algorithm having accuracies p_i $(i = 1, \ldots, 9)$:

- $B_1 : w_i = p_i$,
- $B_2 : w_i = log\frac{p_i}{1-p_i}$,
- $B_3 : w_i = \frac{p_i}{(1-p_i)^2 r_i(1-r_i)}$.

That is, first we study the case when each weight is equal to the accuracy of the individual algorithm (such as taken the i-th algorithm with accuracy p_i, then it participates in the final decision with weight $w_i = p_i$.) The second weighting is suggested as optimal for the weighted majority voting, the third one is important for our generalized weighted majority voting. In this way, we give a practical example to confirm the theoretical derivation of the optimal weights given in section 5.5.2.

We apply OD detecting algorithms as classifiers, so we can test and compare the overall performance of the different voting systems on classifier output generated artificially. In lack of independent OD detecting algorithms providing these accuracies, we are not able to test and compare the voting systems on retinal images. We generate the datasets with elements in the following way: we consider a disc of radius r_{ROI} (ROI) and a disc of radius $r_{OD} = d_{OD}/2$ inside the ROI (OD), where $r_{ROI} = 716$ and $r_{OD} = 51$ pixels, respectively (for details on the adjustment of these figures, see section 4.1.2). We generate 9 output points $\boldsymbol{c_i}$ (for the artificial outputs of $\mathcal{D}_i$), where the probability that the point $\boldsymbol{c_i}$ falls inside the OD is p_i and the distribution of $\boldsymbol{c_i}$ is uniform outside it. Now, the probability r_i $(i = 1, \ldots, 9)$ can be determined as

$$r_1 = \ldots = r_n = \frac{T_0}{T - T_0} = \frac{r_{OD}^2}{r_{ROI}^2 - r_{OD}^2}. \tag{5.57}$$

The overall performance of the four voting systems (MV – majority voting, WMV – weighted majority voting, GMV – generalized majority voting, WGMV- weighted generalized majority voting) with 9 different combinations of the 3 accuracy (A_1, A_2, A_3) and 3 weighting (B_1, B_2, B_3) schemes are presented in Tables 5.9, 5.10 and 5.11, respectively.

Weighting	MV	WMV	GMV	WGMV
B_1	0.7323	0.7323	0.9948	0.9996
B_2	0.7380	0.7380	0.9941	0.9991
B_3	0.7326	0.7326	0.9948	0.9989

Table 5.9: *Overall system accuracies for the set of classifier accuracies A_1.*

If all weights are equal in weighted voting, then it naturally results in the same system accuracy as in the non-weighted voting scheme, otherwise weighted voting outperforms the non-weighted one. Our generalized (non-weighted/weighted) voting system when geometric constrains

Weighting	MV	WMV	GMV	WGMV
B_1	0.5012	0.8066	0.9889	0.9943
B_2	0.4965	0.9688	0.9901	0.8712
B_3	0.5009	0.7289	0.9877	0.9951

Table 5.10: *Overall system accuracies for the set of classifier accuracies A_2.*

Weighting	MV	WMV	GMV	WGMV
B_1	0.8241	0.9526	0.9996	1.0000
B_2	0.8260	0.9926	0.9989	0.9941
B_3	0.8258	0.9481	0.9989	0.9998

Table 5.11: *Overall system accuracies for the set of classifier accuracies A_3.*

are essential in making the final decision has better overall performance than the classic (non-weighted/weighted) majority voting scheme.

For the OD detection application, we can test and compare our generalized non-weighted and generalized weighted voting system on a real database of retinal images, as well. The Messidor dataset described in section 1.5.5 in details is considered for this aim. In this test, we assigned the optimal weights derived in section 5.5.2 to the participating algorithms (classifiers) having individual accuracies $p_1 = 0.767, p_2 = 0.647, p_3 = 0.958, p_4 = 0.977, p_5 = 0.759, p_6 = 0.315, p_7 = 0.320, p_8 = 0.228, p_9 = 0.643$ (as given in case $A3$). However, notice that we have no information about the dependencies among these algorithms. Despite the unknown dependencies of the algorithms, we have found that weighted majority voting (0.98) outperformed simple majority voting (0.974), while simple majority voting outperformed the individual accuracies of the member algorithms.

5.6 Diversity measures for majority voting in the spatial domain

In this chapter, we have already presented how the classic majority voting model can be extended to the spatial domain e.g. to solve object detection problems. As we have also pointed out, the detector algorithms cannot be considered as independent classifiers in applications, so a good ensemble cannot be composed by simply selecting the individually most accurate members. In classic theory, diversity measures are recommended to help explore the dependencies among the classifiers. In this section, we generalize the classic diversity measures for the spatial domain within a majority voting framework. We show that these measures fit better to spatial applications with a specific example of OD detection. Moreover, we show how a more efficient descriptor can be found in terms of a weighted combination of diversity measures which correlates better with the accuracy of the ensemble.

5.6.1 Diversity measures in classic voting theory

Depending on whether it assesses the pairwise or group-wise dissimilarity, two types of diversity measures are considered in classic voting theory. Referring to section 5.2.1, if a system of n classifiers $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$ is given, $\eta_i^{(j)} = 1$ indicates the case, when the i-th detector ($i = 1, \ldots, n$) has a correct vote on the j-th sample ($i = 1, \ldots, t$), while $\eta_i^{(j)} = 0$ shows a wrong

decision. To describe the joint behavior and derive measures for the comparison of two classifiers $\mathcal{D}_i, \mathcal{D}_{i'}$ $(i, i' \in \{1, \ldots, n\})$, let

$$N^{ab} = |\{\ j\ : a, b \in \{0,1\},\ j \in \{1, \ldots, n\}, \eta_i^{(j)} = a, \eta_{i'}^{(j)} = b\}|. \tag{5.58}$$

Moreover, let $N^{a*} = N^{a0} + N^{a1}$, and $N^{*b} = N^{0b} + N^{1b}$ for $a, b \in \{0,1\}$. Notice that $t = N^{00} + N^{01} + N^{10} + N^{10}$ and $N^{ab} \neq N^{ba}$ $(a, b \in \{0,1\})$. The number of classifiers producing error on the j-th input sample $(j = 1, \ldots, t)$ is denoted by $m(j)$ which can be expressed as $m(j) = \sum_{j=1}^{n} (1 - \eta_i^{(j)})$. Finally, the error rate of the i-th classifier can be calculated as $e_i = \frac{1}{t} \sum_{j=1}^{t} (1 - \eta_i^{(j)})$.

In the current literature (see e.g. [166, 174]) the following diversity measures are suggested: minimum individual error, mean error, majority voting error, majority voting improvement, correlation coefficient, product-moment correlation measure, Q-statistics, disagreement measure, double-fault measure, entropy measure, measure of difficulty, Kohavi-Wolpert variance, interrater agreement measure, fault majority measure. Now, as examples, we recall some diversity measures from those that can be considered for generalization to our spatial model. These measures can be calculated between two ensemble members $\mathcal{D}_i, \mathcal{D}_{i'}$ $(i, i' = 1, \ldots, n)$ or at ensemble level.

- *The correlation coefficient $C2$:* it is a well known and frequently used statistical measure. For binary classifier output its definition takes the form

$$C2(\mathcal{D}_i, \mathcal{D}_{i'}) = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{N^{1*}N^{0*}}N^{*1}N^{*0}}. \tag{5.59}$$

- *The disagreement measure $D2$:* it depends on the number of samples for which the classifiers disagreed and the total number of observations. It is calculated as

$$D2(\mathcal{D}_i, \mathcal{D}_{i'}) = \frac{N^{01} + N^{10}}{t}. \tag{5.60}$$

- *Mean error $\bar{e}$:* this measure takes the average of individual classifier error rates e_i $(i = 1, \ldots, n)$ within the ensemble and is defined by the formula

$$\bar{e}(\mathcal{D}) = \frac{1}{n} \sum_{i=1}^{n} e_i. \tag{5.61}$$

- *Interrater agreement measure IA:* this measure characterizes the level of agreement. With the notation presented above it can be expressed as

$$IA(\mathcal{D}) = 1 - \frac{\sum_{j=1}^{t} m(j)(n - m(j))}{n(n-1)\bar{e}(1-\bar{e})}. \tag{5.62}$$

5.6.2 Generalized diversity measures for the spatial domain

The diversity measures described in section 5.6.1 give useful information on how to select the members to achieve the highest ensemble accuracy. More specifically, efficient diversity measures should correlate with the system accuracy. In the literature, the case when the classifier decision making method is not the majority rule is rarely examined. Our aim is to modify these measures to be applicable in our spatial voting scenarios. The main difference is that in our case we must observe the incorrect votes further when checking the dependency of two (or more) member

algorithms. Namely, when two algorithms miss, the chance of the final wrong decision increases only if these incorrect votes meet the geometric constraint with e.g. falling in a disc of diameter d_{OD} in the OD detection task. By this consideration, we label the member outputs with 1 if the vote is correct, $0'$ if the vote is wrong and there is another wrong vote close to it meeting the geometric constraint, and 0 if the vote is wrong, but no other wrong vote can be found close to it. Thus, we generalize the term N^{ab} to let $a, b \in \{0, 0', 1\}$, where e.g. $N^{0'0'}$ is the number of cases, when both algorithms miss and their votes meet also the geometric constraint. With this generalization, we have $t = N^{00} + N^{0'0'} + N^{0'1} + N^{10'} + N^{01} + N^{10} + N^{11}$ for the total number of cases for any two classifiers $\mathcal{D}_i, \mathcal{D}_{i'}$. Now, the diversity measures listed in section 5.6.1 can be generalized as follows.

- *The generalized correlation coefficient $C2'$:*

$$C2'(\mathcal{D}_i, \mathcal{D}_{i'}) = \frac{N^{11}N^{0'0'} - N^{01}N^{10}}{\sqrt{N^{1*}N^{0*}N^{*0}}}. \tag{5.63}$$

The modification of the other diversity measures defined between two classifiers can be interpreted in the same way. For the generalization of the disagreement measure and the ones describing the whole ensemble (e.g. the interrater agreement measure), we need some more considerations as given next.

- *The generalized disagreement measure $D2'$:* it depends on the number of samples for which the classifiers disagreed and the total number of observations. In this case, all possible disagreement situations have to be described in the modified formula

$$D2'(\mathcal{D}_i, \mathcal{D}_{i'}) = \frac{N^{01} + N^{10} + N^{0'1} + N^{10'} + N^{0'0} + N^{00'}}{t}. \tag{5.64}$$

- *The interrater agreement measure IA':* this measure characterizes the level of agreement. With the notation presented above it can be expressed as

$$IA'(\mathcal{D}) = 1 - \frac{\sum_{j=1}^{t} m'(j)(n - m'(j))}{tn(n-1)\bar{e}(1-\bar{e})}. \tag{5.65}$$

  In the classic formula, $m(x_i)$ stands for the number of classifiers producing error for the input sample, and $m'(j)$ expresses the number of wrong votes which are relevant in making the final decision, so the wrong candidates fulfill the geometric constraints, as well.

All the diversity measures are determined at ensemble levels. In the case of pairwise measures, we derive the ensemble level measure with averaging the pairwise values for all the possible pairs. That is, for a pairwise diversity measure D put

$$D(\mathcal{D}) = \sum_{\substack{\mathcal{D}_i, \mathcal{D}_{i'} \in \mathcal{D} \\ i \neq i'}} D(\mathcal{D}_i, \mathcal{D}_{i'}) / |\mathcal{D}|. \tag{5.66}$$

We selected the measures $C2'$, $D2'$, IA' as examples for showing our generalization approach as we found that they correlate well with the system accuracy. In [26], we have shown that the generalized variants correlates more with the system accuracy than the classic ones. However, we can improve the performance further, if we consider a weighted linear combination of these measures and compose the ensemble according to it. To do so, suppose that n classifiers and h diversity measures are given and the aim is to compose a system from the classifiers having the

highest accuracy using the diversity measures. For all the possible $\binom{n}{k}$ $(k = 1, \ldots, n)$ ensembles, we calculate a weighted combination of the diversity measures as

$$GD_j = \sum_{i=1}^{h} \alpha_{i,j} D_{i,j}, \qquad j = 1, \ldots, \binom{n}{k}, \qquad k = 1, \ldots, n, \tag{5.67}$$

where $\alpha_{i,j} \in \mathbb{R}_{\geqslant 0}$ are some weights, $D_{i,j}$ is the value of the i-th diversity measure on the j-th ensemble. With assuming stronger correlation between the weighted combination and system accuracy, the ensemble with the maximal GD_j is selected

$$GD = \max_j (GD_j) = \sum_{i=1}^{h} \alpha_i D_i. \tag{5.68}$$

The appropriate selection of the weights α_i are well-known from the literature for independent feature selectors. Namely, the optimal weights can be determined from the individual accuracies of the feature selectors [166]. In this special case, the correlation values show the performance of the diversity measures as feature selectors. If we consider independent members $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_h$ with accuracies $p_1, p_2, \ldots, p_h$, then GD can be maximized by assigning the weights

$$\alpha_i = \ln \frac{acorr_i}{1 - acorr_i}, \quad i = 1, \ldots, h. \tag{5.69}$$

In our application, the accuracy $acorr_i$ is defined as the average correlation of the i-th diversity measure with the system accuracy regarding all possible assembled ensembles having the same number of members. For our OD detection task, in [26] we have found that an ensemble composed based on the GD figure outperforms the one collecting simply the individually most accurate members, or by considering only a single generalized diversity measure for this aim. That is, to compose the ensemble we can rely on the combined diversity measure GD, as it correlates quite well with the system accuracy demonstrated also in Table 5.12.

System accuracy	**Ensemble of detectors**									***GD***
	OD_1	OD_2	OD_3	OD_4	OD_5	OD_6	OD_7	OD_8	OD_9	
0.9774		X	X	X		X				85.68
0.9765	X	X	X	X	X	X	X			86.35
0.9783	X	X		X	X	X	X	X		86.36
0.9774	X	X	X	X		X		X		89.88
0.9800	X	X	X	X	X	X				89.88

Table 5.12: *Selecting ensembles by using a weighted linear combination of generalized diversity measures.*

Chapter 6

Creating ensembles for the automatic detection of microaneurysms (MAs)

In this chapter, we propose the use of ensembles consisting of individual microaneurysm (MA) detectors for the recognition of these diabetic retinopathy related lesions. Besides collecting some individually robust methods, we also present a procedure to increase divergence among the members with the application of different preprocessing methods before the candidate extractor algorithms. As our experimental results will also show, this approach has proven to be highly efficient in this field. We demonstrate how adding contextual information regarding MAs can further improve their detection and how other lesions can be detected by our approach. Corresponding results have been published in [6–8, 27–32] and are also incorporated in the project DRSCREEN: Developing a computer-based image processing system for diabetic retinopathy screening, TECH08-2 grant of the Hungarian National Office for Research and Technology (NKTH).

Till this point, we analyzed single object detection scenarios. Now, we present some results regarding multiple objects detection, that is, when more objects of the same type may appear in the image. These investigations were motivated by the challenge of detecting microaneurysms (MAs) in color fundus images (see also section 1.4). The recognition of these lesions is of great importance in an automatic screening system, since they are the first signs of diabetic retinopathy (DR) [175].

MA detection is based on the detailed analysis of digital fundus images. State-of-the-art detection approaches usually start with the preprocessing of images, which is followed by candidate extraction. Finally, the extracted candidates are classified as MAs or non-MAs. The reason to separate the latter two steps is that the pixel-wise classification of the whole image would be very time-consuming.

The vast majority of MA detectors can be organized into two categories: the ones based on mathematical morphology, and the others based on shape analysis with non-morphological tools. The largest family of morphology-based candidate extractors are originated from Lay [176] and Baudion [177]. These methods extract the vascular system by taking the maximum of multiple top-hat transformations with rotated linear structuring elements and subtract the resulting image from the original one. The candidates are then extracted by thresholding after applying a Gaussian filter. Oien et al. [178] was the first to apply similar techniques to color images. Spencer et al. [79] proposed a preceding shade correction step to this algorithm, while Frame et al. [179], Mendonca et al. [180], Hipwell et al. [181], Yang et al. [182], Cree et al. [183], Streeter et al. [184] and Fleming et al. [80] proposed modified variants. Mendonca et al. [180], Hipwell et al. [181], Yang et al. [182] and Fleming et al. [80] introduced an extension of this technique to decrease the number of false candidates. Niemeijer et al. [185] also considers a complementary machine learning-based candidate extractor and merge their output. Walter et al. [78] presented a different morphology-based approach, which relies on diameter closing, while in [64], Ravishankar applied morphological filling for establishing a candidate extraction algorithm. Other morphology-based approaches have been proposed in [186], [187] and [188].

The family of the non-morphological shape analysis-based approaches are more diverse. Marino et al. [189] and Bhalerao et al. [190] use a Gaussian mask to match the shape of MAs, while Zhang et al. [82], and Quellec et al. [191] apply multiple Gaussian masks for this purpose. Hahn et al. [192] consider the red/green ratio intensity values and select MA-candidates by applying a shape factor. Other circularity-based operators are used in [193] (double-ring filter), [81] (circular Hough transformation) and [194] (Radon cliff operator).

There are approaches which cannot fit in any of the categories mentioned above, but also worth considering in an ensemble-based system. Grisan et al. [195] extracted candidates based on their densities after a local thresholding. Balasubramanian et al. [196] applied grey-level grouping and proposed an automatic seed generation technique to tackle this problem. Moat operator and region growing are applied in [197] and [198], respectively. Pallawala et al. [199] introduced a normalized cut-based local segmentation technique. Gardner et al. [200] trained an artificial neural network to recognize different retinal features, including MAs. In [37], we have presented an own approach relying on the strength of multiple cross-sectional profiles across the image.

To increase accuracy in this vivid field, we propose the use of ensembles consisting of individual MA detectors. In the preceding chapters, we have already seen that better performance can be expected from an ensemble, if we can reach some independence/divergence among its members. Thus, besides collecting some individually robust methods, we also present a procedure to increase divergence among the members. As our experimental results will also show, this approach has proved to be highly efficient in this field.

To increase divergence, we propose the application of different preprocessing methods before the candidate extractor algorithms. This idea can be easily reasoned by the observation that different preprocessors enhance image content differently. However, this approach is primarily

meaningful in an ensemble-based scenario, where the possible deteriorations of the members can be compensated with increasing divergence among them.

The rest of the chapter is organized as follows. In section 6.1, we explain how to generate ensemble members via combining different preprocessor and MA candidate detector algorithms. Moreover, the determination of the optimal ensemble by a stochastic search algorithm is also explained. The composition of the optimal ensemble may vary according to the selected energy function, so we present our quantitative results regarding this issue, as well. The proposed method is highly competitive among the current MA detectors, which results are presented in details in section 6.2. In this chapter, we also show in what extent our MA detector is able to directly classify images as diseased or not; in other words, the applicability of our ensemble-based detector directly for DR screening. As some complementary results, section 6.3 demonstrates how adding contextual information regarding MAs can further improve their detection.

6.1 Combining preprocessing methods (PPs) and candidate extractors (CEs)

A significant number of recent works focus on the use of multiple algorithms as an ensemble, mostly for classification purposes [56]. In relation to multiple classifier systems (i.e., systems which combine classifiers), we propose the use of a fusion algorithm, where all the ensemble members operate on the whole dataset. Our task is to select the optimal combination of ⟨preprocessing method, candidate extractor⟩ (shortly ⟨PP, CE⟩) pairs on the basis of the MA candidate outputs generated by them. A combination is optimal if it detects as many MAs as possible, while keeping the number of false detections low. Our search framework is based on simulated annealing [201], which is a stochastic search algorithm that does not apply any restrictions to the number of preprocessing methods and candidate extractors.

Simulated annealing is a widely used global optimization method. It is effective for large search space problems by using random sampling to avoid getting stuck in a local minimum. A crucial point in an optimization problem is the choice of the energy function. The selection of the appropriate energy function is specific to the problem, and we will provide a detailed explanation for our choice in section 6.1.2.

In order to minimize the target energy E by simulated annealing, each element of the search space U is a collection of ⟨PP, CE⟩ pairs. A set of MA candidates belongs to each such pair, extracted by the given candidate extraction algorithm on the images with the corresponding preprocessing method applied before. The corresponding energy function value is computed on the union of the candidate sets belonging to the pairs in the collection. The candidates of this collection are compared to a set of MA centroids (ground truth) selected manually by clinical experts. If the Euclidean distance of the centroid of a candidate and a manually selected MA is smaller than a given threshold, then it is considered as a true positive (TP), otherwise it is a false positive (FP) as described in section 1.3. In our experiments, we set this distance threshold to 5 pixels at an image resolution of 768×576 pixels, which value is calculated from the maximal MA diameter ($100\mu m$). The formal description of the proposed combination is as follows.

Algorithm 6.1.1. Optimal combination of preprocessing methods and candidate extractors.

Input:

- *An initial temperature* $T \in \mathbb{R}$.
- *A minimal temperature* $T_{min} \in \mathbb{R}$.

- *A temperature change* $q \in \mathbb{R}$ *with* $0 \leq q \leq 1$.
- *A set* $S = \{\langle PP_i, CE_j \rangle \,|\, i = 1, \ldots, N, \, j = 1, \ldots, M\}$ *containing all* $\langle$*candidate extractor, preprocessing method*$\rangle$ *pairs.*
- *A search space* $U = P(S)$, *where* P *is the power set containing all possible collections of the* $\langle$*candidate extractor, preprocessing method*$\rangle$ *pairs.*
- *A function* $rand(X)$, *which returns a random element* x *from the set* X.
- *A function* $accept : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times [0,1] \rightarrow \{true, false\}$, *which is defined in the following way:*

$$accept(e, e_i, T, r) = \begin{cases} true, & \exp\left(\frac{e - e_i}{T}\right) > r, \\ false, & otherwise. \end{cases}$$

- *An energy function* $E : U \rightarrow \mathbb{R}$.

Output:

- $x_{optimal} \in U$, *where* $E(x_{optimal}) = \min_{x \in U} E(x)$.

---

1. $x \leftarrow r(U)$ {*Initial state.*}
2. $e \leftarrow E(x)$ {*Energy for the initial state.*}
3. $U \leftarrow U - \{x\}$ {*Remove the investigated state from the search space.*}
4. ***while*** $U \neq \emptyset$ *and* $T > T_{min}$ ***do***
5. $\quad x_i \leftarrow r(U)$
6. $\quad e_i \leftarrow E(x_i)$
7. $\quad U \leftarrow U - \{x_i\}$
8. $\quad$ ***if*** $e_i < e$ ***then***
9. $\quad\quad x \leftarrow x_i$ {*If the energy of the investigated state is lower then the current minimum, then it is selected as the current optimum.*}
10. $\quad\quad e \leftarrow e_i$
11. $\quad\quad T \leftarrow T \cdot q$ {*The temperature is decreased with the defined rate.*}
12. $\quad$ ***else***
13. $\quad\quad r \leftarrow rand(\mathbb{R})$ {*A random real number between 0 and 1 is generated for the acceptance function.*}
14. $\quad\quad$ ***if*** $accept(e, e_i, T, r) = true$ ***then***
15. $\quad\quad\quad x \leftarrow x_i$ {*The state can also be accepted via the acceptance function despite its energy value is not lower than the current minimum. The algorithm does not get stuck locally in this way.*}
16. $\quad\quad\quad e \leftarrow e_i$
17. $\quad\quad\quad T \leftarrow T \cdot q$
18. $\quad\quad$ ***end if***
19. $\quad$ ***end if***
20. ***end while***
21. ***return*** x {*The state with the lowest energy is returned.*}

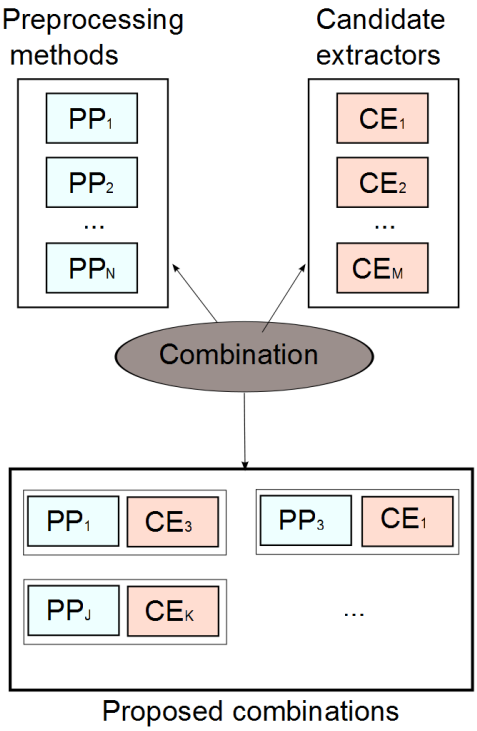


Figure 6.1: *The proposed framework for ensemble selection.*

From the ones described in sections 1.6.1, we consider $N = 4$ preprocessing algorithms PP_1, PP_2, PP_3, PP_4 and $M = 5$ MA candidate extractors MA_1, MA_2, MA_3, MA_4, MA_5 introduced in section 1.6.4, respectively. However, notice that with the use of simulated annealing, we can easily include more methods in the future. In this study, the largest number of ⟨PP, CE⟩ pairs in an ensemble is $5 \times 4 = 20$ in our case, and the search space contains $2^{20} - 1$ possible ensembles.

Depending on the annealing schedule, the algorithm is able to determine the optimal result. However, for large search spaces, it is also possible to select an approximately optimal solution with stopping after a large number of iterations. For a more illustrative representation of the approach see Figure 6.1.

6.1.1 Methodology for ⟨PP, CE⟩ ensembles

We tested our approach on 199 images selected from three databases: the training set of the Retinopathy Online Challenge (ROC) database [58], the DiaretDB1 database [60] and our own dataset, which was provided by the Moorfields Eye Hospital, London, UK (see section 1.5 for their descriptions). When ground truth was not available publicly in the database, we used a consensus regarding the manual MA selection of two ophthalmologists. We have tested our approach in two ways.

- CASE 1: We selected the preprocessing methods for each candidate extractors and measured the change in its sensitivity by using the combination. More formally, we restricted the search space to $\{\langle PP_i, CE\rangle \,|\, i = 1, \ldots, N\}$ in Algorithm 6.1.1 with fixing candidate extractors.
- CASE 2: We measured the effects using multiple candidate extractors in the selection process as it is described in Algorithm 6.1.1.

For both cases, we disclose the results based on 100 runs, splitting the dataset into two disjoint equal sized parts randomly in every run.

CE	E_{SEN}	E_{PPV}	E_{FSCORE}	E_{LN}
MA_1	$\mathbf{0.5523 \pm 0.0479}$	0.1648 ± 0.0247	0.1836 ± 0.0652	0.5386 ± 0.0597
MA_2	$\mathbf{0.2400 \pm 0.0487}$	0.0811 ± 0.0291	0.0116 ± 0.0539	0.2319 ± 0.0422
MA_3	$\mathbf{0.2392 \pm 0.0545}$	0.0934 ± 0.0847	0.1179 ± 0.1011	0.0265 ± 0.0350
MA_4	$\mathbf{0.5583 \pm 0.0387}$	0.4671 ± 0.0296	0.4640 ± 0.0282	0.5546 ± 0.0354
MA_5	$\mathbf{0.5980 \pm 0.0326}$	0.2525 ± 53.0274	0.2563 ± 0.0267	0.3686 ± 0.0460

Table 6.1: *Comparison of the energy functions on the ROC database regarding CASE 1. Best scores are set in bold.*

CE	E_{SEN}	E_{PPV}	E_{FSCORE}	E_{LN}
MA_1	$\mathbf{0.6651 \pm 0.0519}$	0.5174 ± 0.0452	0.5158 ± 0.0461	0.5673 ± 0.0710
MA_2	$\mathbf{0.0349 \pm 0.0121}$	0.0064 ± 0.0060	0.0090 ± 0.0086	0.0262 ± 0.0161
MA_3	$\mathbf{0.1933 \pm 0.0274}$	0.0800 ± 0.0765	0.1164 ± 0.0641	0.1405 ± 0.0833
MA_4	$\mathbf{0.4709 \pm 0.0289}$	0.04029 ± 0.0269	0.4015 ± 0.0289	0.3580 ± 0.1727
MA_5	$\mathbf{0.8066 \pm 0.0379}$	0.4963 ± 0.0385	0.4964 ± 0.0426	0.5283 ± 0.1018

Table 6.2: *Comparison of the energy functions on the DiaretDB1 database regarding CASE 1. Best scores are set in bold.*

6.1.2 Energy functions to optimize ensemble performance

Since it is a crucial point to choose the suitable energy function in an optimization process, we decided to select the most appropriate one by evaluating for multiple possibilities. Namely, we checked the energy functions $E_{SEN} = -SEN/FPI$, $E_{PPV} = -PPV$, $E_{FSCORE} = -\text{F-Score}$ (see also section 1.3.1) and

$$E_{LN} = -\frac{TP}{\ln \frac{FP}{TP}} \tag{6.1}$$

in our search framework. While the first three functions are based on literature recommendations, we also propose the additional one E_{LN}, which is competent with the others. We cannot consider any measure which relies on true negatives (TN) like specificity or Negative Predictive Value, due to the high number of TNs. Namely, every pixels in an image which is not a candidate centroid and not marked as MA by the experts can be considered as TN. For this reason, such evaluation measures are generally excluded for such image processing algorithms [202].

The evaluation of the energy functions is accomplished in the following way: we executed the selection process using all the energy functions listed above simultaneously for CASE 1 and CASE 2 described in section 6.1.1. The results of this evaluation are disclosed in Tables 6.1, 6.2, 6.3 and 6.4 for the ROC, DiaretDB1, Moorfields databases and for all images in the three databases together, respectively, regarding CASE 1 and in Table 6.5 for CASE 2. Each cell in the tables contains the 90% confidence interval of sensitivity values for each case using the given energy function for selection.

From these results, we can conclude that the energy function E_{SEN} fits our goals the most to have a large number of TPs with a rather low FP count. It is also clear that E_{PPV} and E_{FSCORE} are not suitable for this optimization process, while E_{LN} could also be taken into account. However, the variability of E_{LN} is higher than that of E_{SEN}, so we have chosen E_{SEN} in the selection process.

CE	E_{SEN}	E_{PPV}	E_{FSCORE}	E_{LN}
MA_1	$\mathbf{0.7543 \pm 0.0470}$	0.6474 ± 0.0855	0.6497 ± 0.0740	0.7478 ± 0.0410
MA_2	$\mathbf{0.2317 \pm 0.0225}$	0.1169 ± 0.0315	0.1181 ± 0.0326	0.2296 ± 0.0222
MA_3	$\mathbf{0.2307 \pm 0.0183}$	0.2240 ± 0.0204	0.2214 ± 0.0184	0.2297 ± 0.0222
MA_4	$\mathbf{0.5081 \pm 0.0208}$	0.3235 ± 0.0234	0.3278 ± 0.0230	0.4921 ± 0.0269
MA_5	$\mathbf{0.8077 \pm 0.0215}$	0.2723 ± 0.0251	0.2730 ± 0.0242	0.7106 ± 0.1990

Table 6.3: *Comparison of the energy functions on the Moorfields database regarding CASE 1. Best scores are set in bold.*

CE	E_{SEN}	E_{PPV}	E_{FSCORE}	E_{LN}
MA_1	$\mathbf{0.5850 \pm 0.0325}$	0.5312 ± 0.0332	0.5312 ± 0.0312	0.5824 ± 0.0324
MA_2	$\mathbf{0.1755 \pm 0.0229}$	0.0623 ± 0.0170	0.1401 ± 0.0700	0.1749 ± 0.0228
MA_3	$\mathbf{0.0445 \pm 0.0147}$	0.0235 ± 0.0148	0.0250 ± 0.0139	0.0428 ± 0.0167
MA_4	$\mathbf{0.3824 \pm 0.0404}$	0.2698 ± 0.0210	$0.2270 \pm 0.0.0337$	0.3797 ± 0.0385
MA_5	$\mathbf{0.6290 \pm 0.0262}$	0.3370 ± 0.0232	0.3381 ± 0.0271	0.4763 ± 0.1587

Table 6.4: *Comparison of the energy functions on all three databases regarding CASE 1. Best scores are set in bold.*

Database	E_{SEN}	E_{PPV}	E_{FSCORE}	E_{LN}
ROC	$\mathbf{0.7447 \pm 0.0453}$	0.3319 ± 0.0924	0.3061 ± 0.0796	0.7013 ± 0.1446
DiaretDB1	$\mathbf{0.9820 \pm 0.0041}$	0.5346 ± 0.0777	0.5289 ± 0.0757	0.5886 ± 0.2248
Moorfields	$\mathbf{0.9565 \pm 0.0287}$	0.3317 ± 0.0677	0.3314 ± 0.0666	0.8102 ± 0.1886
All	$\mathbf{0.8711 \pm 0.0303}$	0.3785 ± 0.0564	0.3833 ± 0.0616	0.8555 ± 0.0322

Table 6.5: *Comparison of the energy functions on the ROC database regarding CASE 2. Best scores are set in bold.*

CE	SEN	FPI	PPV	DS
MA_1	0.5523 ± 0.0479	572.1612 ± 86.7554	0.0066 ± 0.0015	0.1353 ± 0.0332
MA_2	0.2400 ± 0.0487	151.8692 ± 9.5329	0.0103 ± 0.0030	0.1142 ± 0.0232
MA_3	0.2392 ± 0.0545	512.4052 ± 46.6152	0.0031 ± 0.0009	0.0566 ± 0.0334
MA_4	0.5583 ± 0.0387	540.2896 ± 27.8082	0.0069 ± 0.0014	0.0862 ± 0.0188
MA_5	0.5980 ± 0.0326	569.3936 ± 53.0947	0.0071 ± 0.0016	0.1934 ± 0.0275

Table 6.6: *Performance of the ensembles on the ROC database regarding CASE 1.*

CE	SEN	FPI	PPV	DS
MA_1	0.6651 ± 0.0519	443.9127 ± 29.7674	0.0065 ± 0.0019	0.0712 ± 0.0192
MA_2	0.0349 ± 0.0121	165.7153 ± 13.3925	0.0009 ± 0.0003	0.0162 ± 0.0081
MA_3	0.1933 ± 0.0274	313.5056 ± 18.8335	0.0029 ± 0.0011	0.0428 ± 0.0133
MA_4	0.4709 ± 0.0289	129.0984 ± 14.4546	0.0171 ± 0.0061	0.0703 ± 0.0114
MA_5	0.8066 ± 0.0379	663.5811 ± 16.7470	0.0055 ± 0.0013	0.1778 ± 0.0161

Table 6.7: *Performance of the ensembles on the DiaretDB1 database regarding CASE 1.*

6.1.3 Results and discussion

The performance of the ensembles found by Algorithm 6.1.1 using the energy function E_{SEN} are shown in Tables 6.6, 6.7, 6.8 and 6.9 regarding CASE 1, and Table 6.10 for CASE 2, respectively. Besides sensitivity, we also disclose other measures to describe the performance of the ensemble selection framework: the number of FPs per image FPI, the positive predictive value PPV (see also section 1.3.1), whose negated version is E_{PPV} and the difference DS between the sensitivity of the ensemble and the sensitivity of the best performing individual candidate extractor on the same test dataset.

From the DS values it can be seen that the proposed framework increases sensitivity in a large extent compared to the individual approaches. On the other hand, the number of false positives also increased in a relatively smaller extent. In Figure 6.2, a subimage containing both TPs and FPs is shown. FPs originate from artefacts, image compression errors or thin vessel parts. The proportion of true and false detection can be improved by using classification, voting or other postprocessing techniques, which issue is going to be addressed in section 6.3. In other words, the primary aim of candidate extraction to have a large number of TPs with keeping the number of FPs low is fulfilled.

The proposed approach does not address the grading process of DR directly. Instead, it concentrates on the improvement of the candidate extraction at pixel level. The effect of this method in an actual grading process was needed to be further investigated and addressed in the following sections. The gain in the number of TPs can lead to a higher sensitivity in grading, but the increase of FPs can deteriorate such a system. However, MA detection is only one component

CE	SEN	FPI	PPV	DS
MA_1	0.7543 ± 0.0470	792.4597 ± 32.4551	0.0080 ± 0.0017	0.0888 ± 0.0141
MA_2	0.2317 ± 0.0225	164.6963 ± 6.2115	0.0123 ± 0.0020	0.1092 ± 0.0142
MA_3	0.2307 ± 0.0183	213.3273 ± 11.7408	0.0098 ± 0.0019	0.0025 ± 0.0002
MA_4	0.5081 ± 0.0208	162.2053 ± 8.7651	0.0265 ± 0.0051	0.0983 ± 0.0137
MA_5	0.8077 ± 0.0215	606.1647 ± 15.2953	0.0115 ± 0.0023	0.3089 ± 0.0186

Table 6.8: *Performance of the ensembles on the Moorfields database regarding CASE 1.*

CE	SEN	FPI	PPV	DS
MA_1	0.5850 ± 0.0325	327.4260 ± 24.5482	0.0113 ± 0.0017	0.0556 ± 0.0066
MA_2	0.1755 ± 0.0229	165.4692 ± 2.7000	0.0067 ± 0.0011	0.1077 ± 0.0137
MA_3	0.0445 ± 0.0147	133.8911 ± 11.9872	0.0021 ± 0.0007	0.0114 ± 0.0034
MA_4	0.3824 ± 0.0404	250.7720 ± 14.3805	0.0097 ± 0.0013	0.0723 ± 0.0089
MA_5	0.6290 ± 0.0262	401.1449 ± 17.0726	0.0099 ± 0.0014	0.1279 ± 0.0130

Table 6.9: *Performance of the ensembles on all three databases regarding CASE 1.*

Database	SEN	FPI	PPV	DS
ROC	0.7447 ± 0.0453	808.8792 ± 47.5289	0.0060 ± 0.0012	0.2763 ± 0.0336
DiaretDB1	0.9820 ± 0.0041	1121.8218 ± 70.7609	0.0038 ± 0.0009	0.3516 ± 0.0363
Moorfields	0.9565 ± 0.0287	1135.8917 ± 36.8809	0.0074 ± 0.0014	0.3471 ± 0.0398
All	0.8711 ± 0.0303	985.8611 ± 43.9246	0.0055 ± 0.0008	0.3406 ± 0.0287

Table 6.10: *Performance of the ensembles regarding CASE 2.*

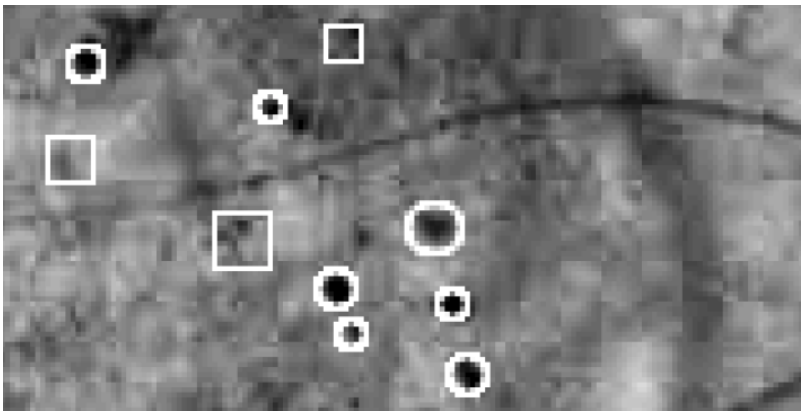


Figure 6.2: *Results of MA detection by an ensemble. Circles depict TPs and squares represent FPs.*

	CASE 1			CASE 2		
Number of pairs	1	2	3	9	10	11
Occurrences	336	962	702	7	389	4

Table 6.11: *Number of ⟨preprocessing method, candidate extractor⟩ pairs selected for the ensembles in different test runs.*

of an automatic screening system, since the proper clinical protocol requires other lesions – like exudates – to be detected, as well.

We have also checked the cardinality of the optimal ensembles for CASE 1 and CASE 2. Table 6.11 shows the relative occurrence of the number of selected pairs in each case. As it can be seen, there is a dominant value for the number of pairs for both cases. In CASE 1, most selections contained 2 pairs, while in CASE 2, in almost every case 10 pairs were selected.

6.2 An ensemble-based MA detector

In this section, an effective MA detector based on the combination of preprocessing methods and candidate extractors is proposed. An exhaustive quantitative analysis is also given to prove the superiority of our approach over individual algorithms.

6.2.1 Ensemble selection for MA detection

We start with describing our ensemble creation approach based on the results presented in section 6.1. Besides the four preprocessing algorithms considered in section 6.1, we have also involved PP_5 (described in section 1.6.1 in details) to increase the divergence of the members further. As for the MA candidate extractors, the same algorithms are considered. As an energy function, we used the competition performance metric CPM [58] (see also section 1.3.1).

The ensemble creation part (see Algorithm 6.1.1) results in a set of ⟨PP, CE⟩ pairs. This ensemble E_{best} then can be used to detect MAs on unknown images. The final ensemble is applied in real detection in the same way as in the training phase. Namely, the final MAs are detected by the fusion of the MA candidates of the individual pairs building up the ensemble E_{best}. Similarly, for every MA candidate $\boldsymbol{c}$, we have a confidence value $conf_{E_{best}}(\boldsymbol{c})$ as defined in section 1.3. Thus, for the final decision on the presence of MAs, the output MA set needs to be thresholded according to the assigned confidence values. The choice of the threshold value is discussed in section 6.2.3 in detail.

6.2.2 Results

We have evaluated the proposed approach for both MA detection and DR grading. In this section, we present the evaluation methodology we used in each case.

MA detection

We have evaluated the MA detection capabilities of the proposed method in the ROC competition for MA detectors [58], as well, as on the publicly available DiaretDB1 (see section 1.5.3) and a private database provided by the Moorfields Eye Hospital (see section 1.5.4). In this section, we provide a brief overview on the methodology we used for the evaluation of MA detection performance of the proposed approach.

Testing. For each database, we provide the Free-response Receiver Operating Characteristic ($FROC$) curves [202], which plots the sensitivity against the average number of false positives per image. To measure the sensitivity at different average false positive per image levels, we thresholded the output set of the MA detector based on the confidence values assigned to each candidate. For the ROC database, we also provide the current ranking of the competition along with the CPM values that serves as the basis for the ranking. In addition, we also calculated a partial area under the curve (AUC) of the algorithms in the same range (between 1/8 and 8) by normalizing the average false positive per image figure by dividing with the maximum value 8 and applying trapezoidal integration. The empirical AUC calculated this way is likely to underestimate the true AUC. However, the uncertainty for the partial AUCs may be quite high due to the low number of images.

DR grading

We have also evaluated our ensemble-based approach to see its grading performance to recognize DR. For this aim, we determined the image-level classification rate of the ensemble on the Messidor database (see section 1.5.5). That is, the presence of any MA means that the image contains signs of DR, while the absence of MAs indicates a healthy case. In other words, a pure yes/no decision of the system has been tested.

Ensemble creation. As there is no training set provided for the Messidor database, we used an independent database (ROC) to train our algorithm. Notice that this is quite a strong handicap in comparison with the usual approach to train on a part of the same database. However, we feel that in this way we can get much closer to measure up the true performance of our system under real circumstances.

Testing. In our evaluation, we classified the retinal images whether they contain signs of DR (R1, R2, R3) or not (R0) (see section 1.5.5 for details). The MA detector classifies an image as diseased if at least one MA is detected, and healthy otherwise. We measured the sensitivity SEN, specificity SPE and accuracy ACC (see section 1.3.1) of the detector at different levels by thresholding the confidence values assigned to the MA candidates. We also measured the percentage of correctly recognized cases for each grade. We provided a fitted Receiver Operating Characteristic (ROC) curve along with the empirical and fitted AUC for the proposed method on the Messidor database. For curve fitting we used JROCFIT [203].

MA detection results

In Table 6.12, we exhibit the ⟨PP, CE⟩ pairs included in the selected ensembles for the three databases, respectively. The rows of the table show the preprocessing methods, while the columns label the candidate extractor algorithms.

Table 6.13 contains the ranked quantitative results of the participants of the ROC competition, with the proposed ensemble DRSCREEN highlighted as the current leader. The performance of the ensemble is also shown in Figure 6.3 in terms of a $FROC$ curve. As we can see from Table 6.13, the proposed ensemble earned both a higher CPM score and a higher partial AUC than the individual algorithms.

Diabetic retinopathy grading results

In Table 6.14, we provide the SEN, SPE and ACC measures of our detector corresponding to different threshold values, respectively. The fitted ROC curve of the detector can be seen in

	MA_1	MA_2	MA_3	MA_4	MA_5
PP_1	R		M	R	R, D
PP_2		M		R	
PP_3	R, D	M		D	R
PP_4	D			R, D	R, D, M
PP_5					R, M

Table 6.12: *$\langle PP, CE \rangle$ pairs selected as members of the MA detector ensembles for the three databases. R, D, M denote whether the pair is selected for the ROC, DiaretDB1 or the Moorfields database, respectively.*

Team	CPM	AUC
DRSCREEN	**0.434**	**0.551**
Niemeijer et al.	0.395	0.469
LaTIM	0.381	0.489
ISMV	0.375	0.435
OKmedical II	0.369	0.465
OKmedical	0.357	0.430
Lázár et al.	0.355	0.449
GIB	0.322	0.399
Fujita	0.310	0.378
IRIA	0.264	0.368
Waikato	0.206	0.273

Table 6.13: *Quantitative results of the Retinopathy Online Challenge for MA detection. For each participating team, the competition performance metric and the partial area under the FROC curve are presented.*

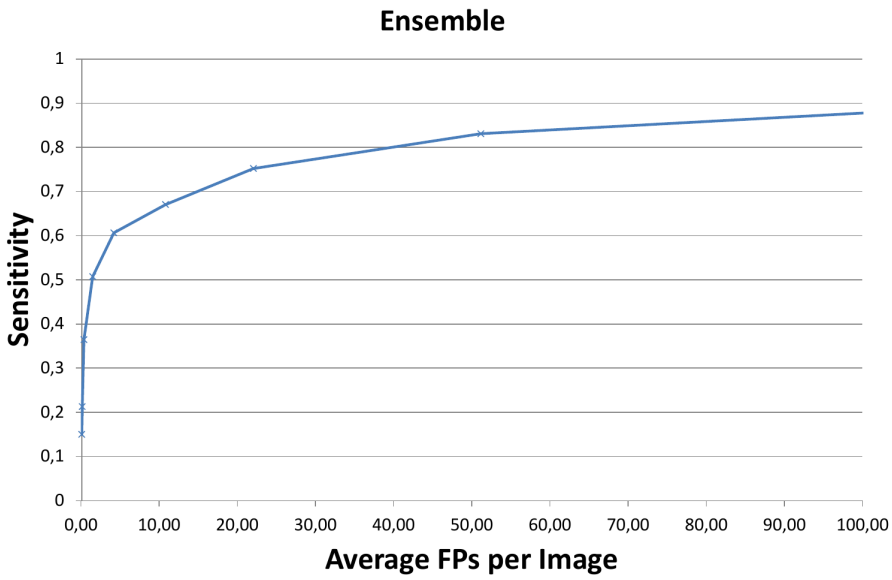


Figure 6.3: *Free-response receiver operating characteristic (FROC) curve of the MA detector ensemble on the Retinopathy Online Challenge.*

Figure 6.4. The empirical area under curve (AUC) is 0.875, while the AUC for the fitted curve is 0.90 ± 0.01. Table 6.14 also contains the percentage of the correctly recognized cases for each

Threshold	0.4	0.5	0.6	0.7	0.8	0.9	1.0
SEN	1	1	1	0.99	0.96	0.76	0.31
SPE	0	0.01	0.03	0.14	0.51	0.88	0.98
ACC	0.53	0.54	0.55	0.59	0.75	0.82	0.62
R0	0.00	0.01	0.03	0.14	0.51	0.88	0.98
R1	1.00	1.00	1.00	0.97	0.92	0.60	0.18
R2	1.00	1.00	1.00	1.00	0.96	0.72	0.29
R3	1.00	1.00	1.00	1.00	0.98	0.92	0.42

Table 6.14: *Results of diabetic retinopathy grading on the Messidor database based on the MA detection results. For each threshold, sensitivity, specificity, accuracy and the percentage of correctly recognized cases for each grade are presented.*

class.

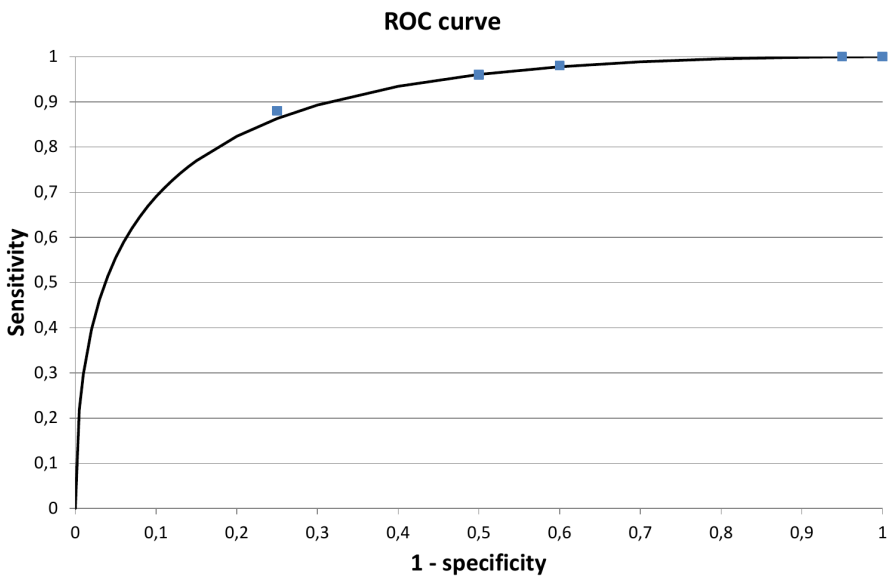
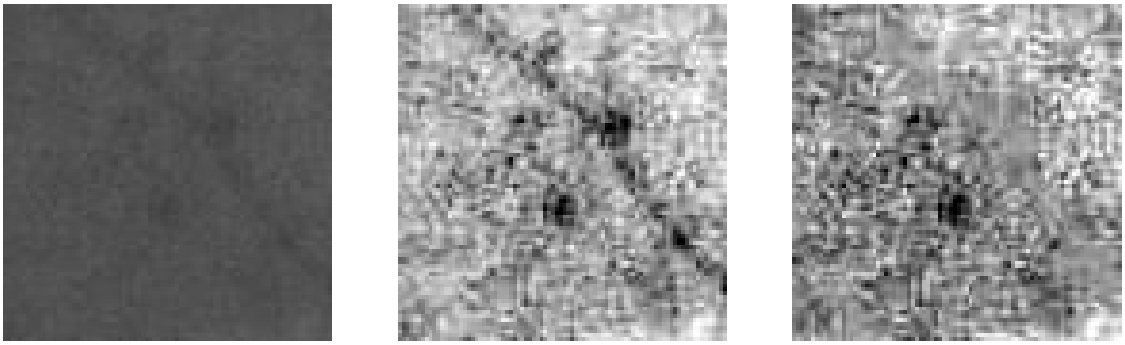


Figure 6.4: *Receiver operating characteristic (ROC) curve of the MA detector ensemble on the Messidor database.*

6.2.3 Discussion on the MA detector

A strong point of the proposed method is that it performs well under difficult circumstances. Figure 6.5 shows an example image, where the application of PP_3 (CLAHE) made it easier to distinguish the MAs from their background. However, the use of the vessel removal and inpainting preprocessing method PP_4 caused the missing of a true MA, while the detection of the remaining MA is easier in the absence of thin retinal vessels. Thus, using different preprocessing methods with candidate extractors creates diversity among the members of the ensemble, which is desired for systems using multiple estimators [56]. This diversity ensures the suppression of false detections, since diverse detectors tend to make different mistakes. Thus, the false detections are likely to receive lower confidence values in the voting procedure.

Our experimental results show that the proposed ensemble-based MA detector outperforms the current individual approaches in MA detection. It has been also proven that the framework has



(a) Original (PP_1) (b) CLAHE (PP_3) (c) Vessel removal (PP_4)

Figure 6.5: *The effect of different preprocessing methods, where MAs are hard to detect.*

high flexibility for different databases. As it can be seen in Table 6.12, the ensemble members may vary, which suggests relatively high variance among databases in this field. Despite this variability, the performance of the ensemble still remained stable. In [58], the FPI rate for a human expert is measured in the ROC database against the consensus of three human experts. This level is approximately 1 FP per image [58] for the ROC database, on which level our ensemble achieved the best score in the competition. Thus, we can recommend to use this level for thresholding at the ensemble creation phase and use it for detecting MAs on unknown images.

As for DR grading, our ensemble also performed well. It is also important to see how the different classes (R0, R1, R2, R3) are recognized at different levels. As it is desired, the severity of DR affects the performance of our detector. At each threshold level, where the sensitivity is less than 1.0, the more severe cases recognized with higher probability.

The selection of the appropriate threshold is also an important issue for our detector to provide sufficient sensitivity and specificity rate. In [204], it has been suggested that sensitivity is more important for a screening system than specificity. In opposition, the British Diabetic Association (BDA) recommends 80% sensitivity and 95% specificity for DR screening [205]. In Table 6.14, we can see that the most accurate result is achieved with the threshold value 0.9. By applying the first idea, we might consider the results corresponding to the threshold value 0.8 as the best in our experiment, where 96% sensitivity and 51% specificity are achieved. That is, we recognized almost all the cases where DR is present, and half of the healthy ones. The closest to the second recommendation is the performance achieved at the 0.9 level: 76% sensitivity and 88% specificity.

It is difficult to compare our method to other screening systems, since those are tested on private databases. Unfortunately, the proportion of non-DR/DR cases are varying in these experiments. Abramoff et al. [204] reported 0.86 AUC on a population where 4.96% of the cases had at least minimum signs of DR. The databases on which Agurto et al. [206] tested, 74.43% and 76.26% of the cases contained signs of DR and they achieved 0.81 and 0.89 AUCs, respectively. The closest to match the requirements of BDA is the system of Jelinek et al. [207] with a 85% sensitivity and 90% specificity, where approximately 30% of patients had DR. Similar proportion (35.88%) of patients having DR are reported by Fleming et al. [208] in their automatic screening system.

Despite the promising results, our system still misclassifies some images, where serious case of DR is present. To improve grading performance, we must take into account the presence or absence of more DR-specific lesions (e.g. exudates), image quality, the recognition of anatomical parts which are essential in a clinical setting. However, our MA detector can serve as a main component of such a system, as we will show it in chapter 7.

The $FROC$ curves of the ensemble for the DiaretDB1 and for the Moorfields database are shown in Figures 6.6 and 6.7, respectively. To the best of our knowledge, no corresponding quantitative results have been published for these databases yet. Thus, we disclose the results of the ensemble-

based method only.

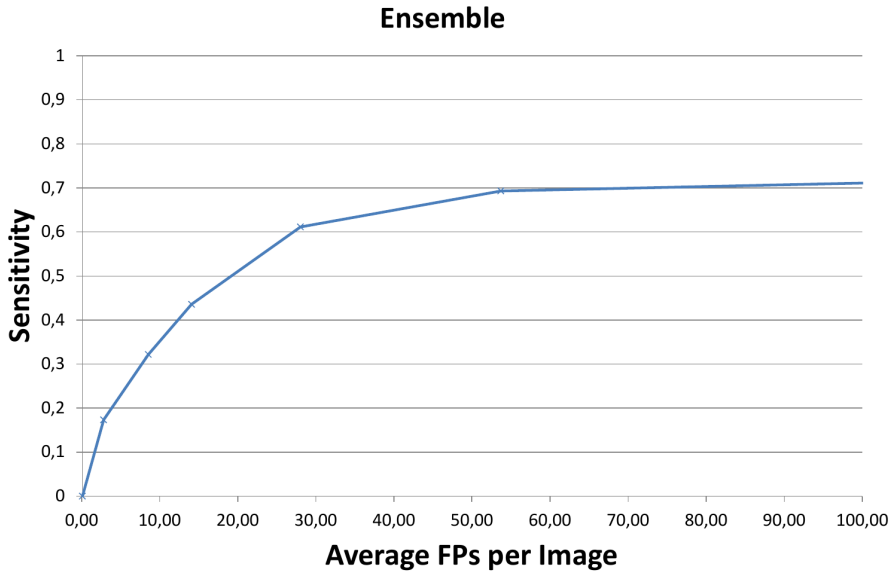


Figure 6.6: *FROC curve of the MA detector ensemble on the DiaretDB1 database.*

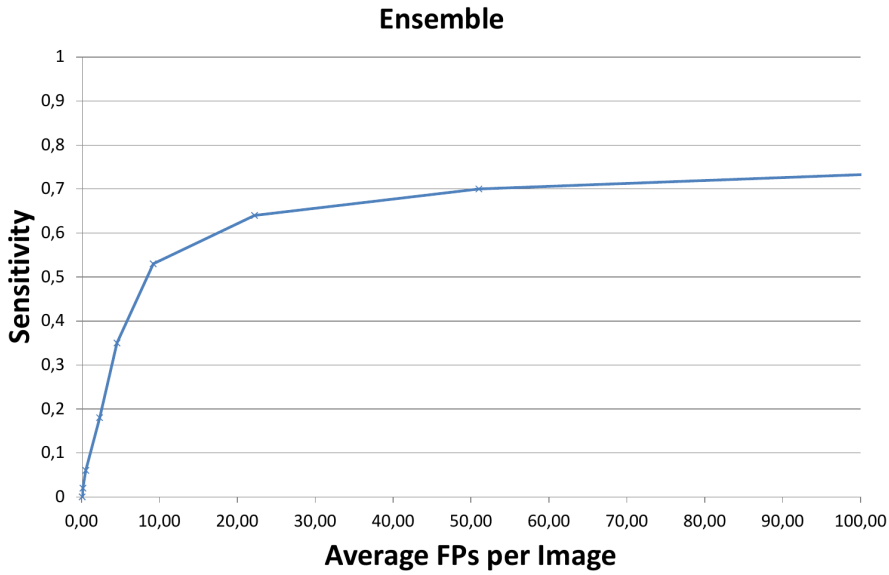


Figure 6.7: *FROC curve of the MA detector ensemble on the Moorfields database.*

6.3 Improving MA detection by a context-aware approach

In this section, we show how MA detection can be further improved by adding some contextual information regarding the location and appearance of these lesions. Namely, the detection of MAs highly depends on the characteristics of the imaging device and other image properties (e.g. type of compression). As a result, some MAs can be easily spotted on the background of the retina, while the recognition of others are more difficult. Besides image characteristics, the spatial location also has influence on the detection of MAs (e.g. proximity of vessel parts, etc.)

In [58], Niemeijer et al. distinguish three categories based on visibility: subtle, regular and obvious. An example for this categorization can be seen in Figures 6.8 and 6.9. In the same

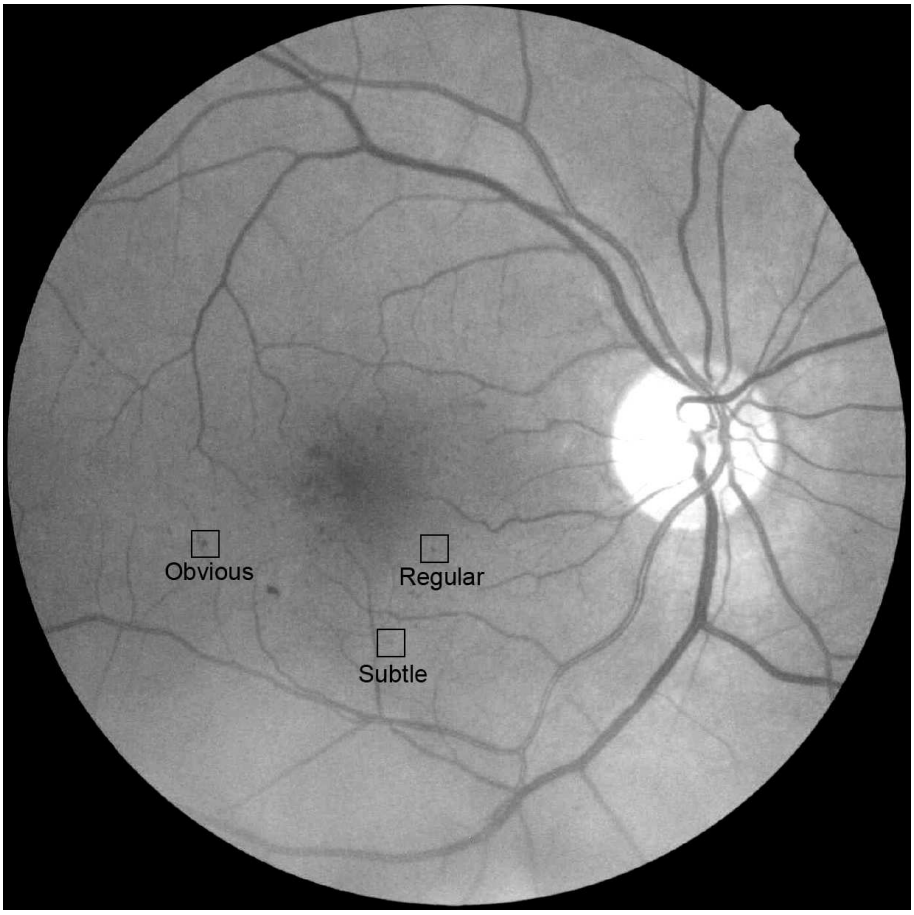
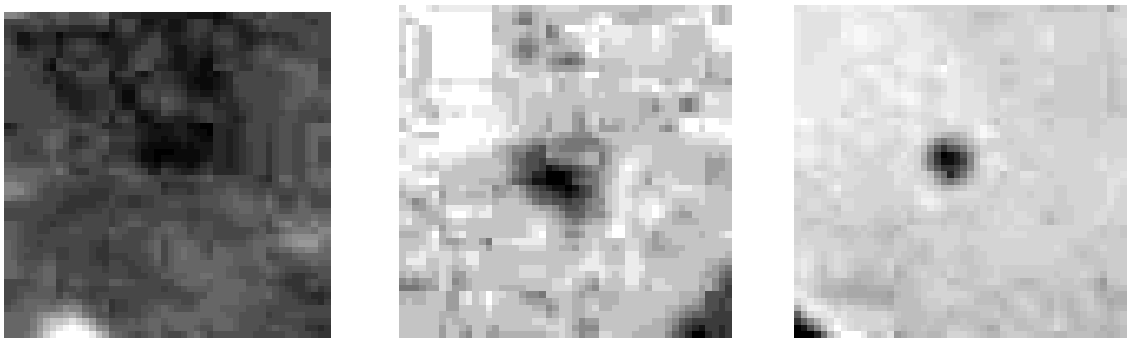


Figure 6.8: *Examples for MAs in different visual categories in a fundus image from the ROC dataset.*

study, they also investigate the detection of MAs near vessel. We extend these categorization with two additional categories by also taking into account the MAs which are detected on the macula and which are on the periphery of the image. Figures 6.10 and 6.11 show examples for the spatial categories. We also provide a computational approach to determine the characteristics of the MAs. In this section, we propose two approaches exploiting this categorization to improve MA detector ensembles.



(a) Subtle (b) Regular (c) Obvious

Figure 6.9: *Microaneurysm categories based on visibility.*

First, to recognize MAs in the different categories, we measure the effect of using different preprocessing methods. As we can see later on, a preprocessing method can enhance the detection

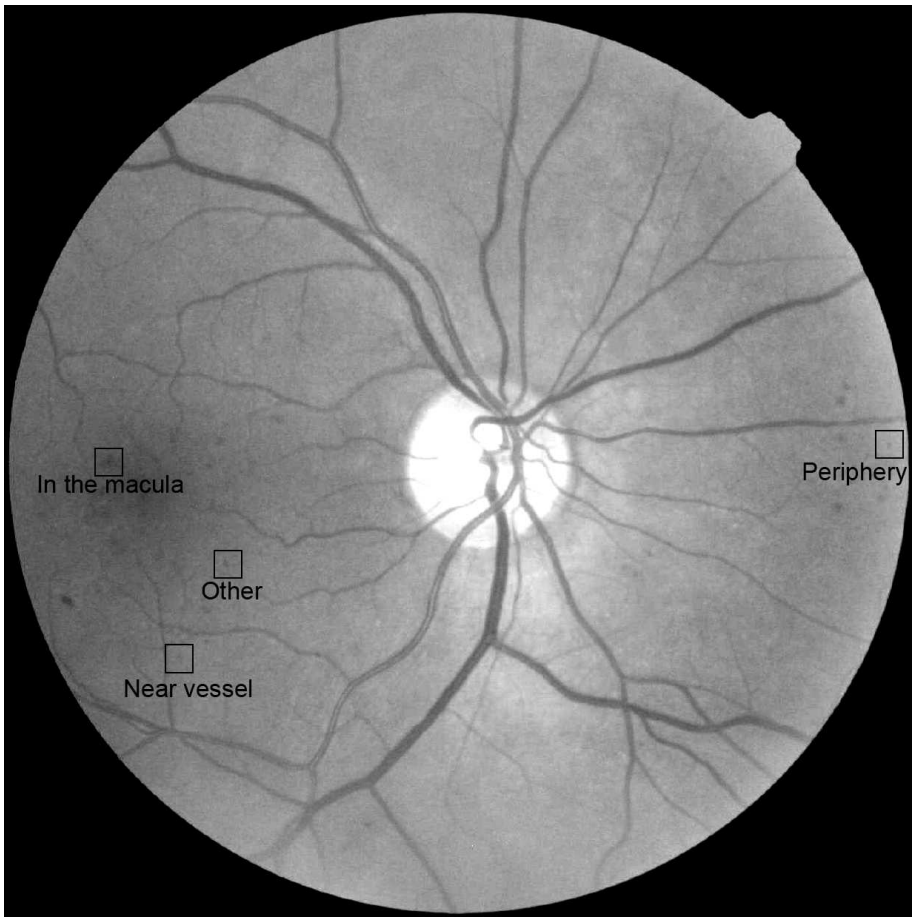
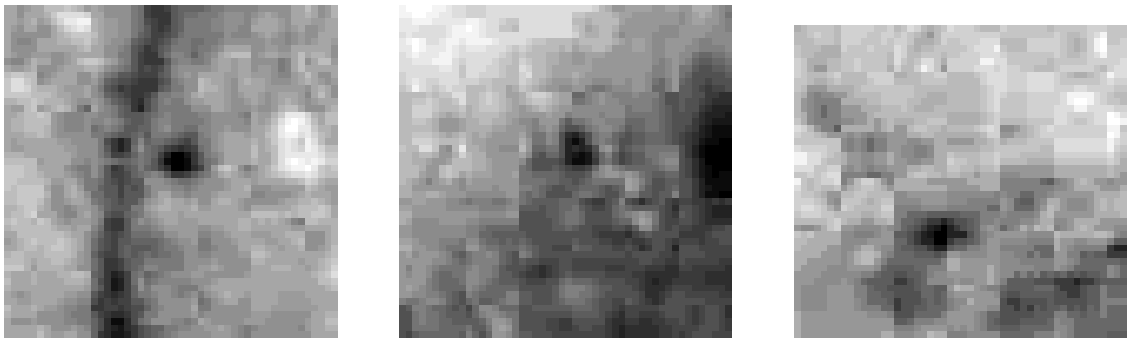


Figure 6.10: *Examples for MAs in different spatial categories in a fundus image from the ROC dataset.*



(a) Near vessel (b) In the macula (c) On the periphery

Figure 6.11: *Microaneurysm categories based on spatial location.*

rate in a few categories, but there is no single best performing preprocessing method for all. To overcome this difficulty, we propose a context-aware selection approach of preprocessing methods for MA candidate extraction.

Moreover, we also present an adaptive weighting approach for ⟨PP, CE⟩ ensembles [32]. This approach assigns an optimal weight to each member of the ensemble based on their performance of detecting MAs having different contrast and spatial locations. The experimental results show that this method is competitive with our ensemble-selection approach [7] described in section 6.2.

6.3.1 Context-aware selection of ⟨PP, CE⟩ pairs

In this section, we describe our context-aware preprocessing method selection approach, which is based on learning. Thus, a training database with manually labelled MAs is needed. Then, the manually labelled MAs of the training database must be categorized. Opposed to the manual way described in [58], we set up an automatic method described next.

Categorization based on visibility. To measure the visibility of an MA, we select an $k \times l$ (e.g. $k = l = 20$ in our case) window centered on the MA centroid and measure the contrast in this region by the formula

$$\sqrt{\frac{1}{kl} \sum_{\boldsymbol{x}(1)=1}^{k} \sum_{\boldsymbol{x}(2)=1}^{l} \left(I\left(\boldsymbol{x}\right) - \mu\right)^2}, \tag{6.2}$$

where $I\left(\boldsymbol{x}\right)$ is the corresponding intensity of the pixel having coordinates $\boldsymbol{x} = (\boldsymbol{x}(1), \boldsymbol{x}(2))$ and μ is the average intensity of the window.

Since we do not have any prior knowledge about the distribution of MAs based on the contrast information, we aimed to divide the MAs into three sets with equal cardinalities. Thus, we categorize an MA as *subtle*, if its contrast is lower than the 33th percentile of the observed contrast values in the training set, *obvious*, if its contrast is higher than the 66th percentile and *regular,* otherwise.

Categorization based on spatial location. We also categorized MAs into three more categories based on their spatial locations: *near to vessel*, *in the macula* and *on the periphery.* For the first category, we must detect the vessel system of the retina, which were solved by our method presented in [35]. We consider an MA as *near to vessel*, if it is closer to a vessel part than the maximal MA diameter. For the second category, we detected macula with the detector proposed in [36]. Then, we collect the MAs falling into the area of the macula to the *in the macula* category.

Finally, MAs *on the periphery* are determined in the following way: first, the radius of the retinal ROI is calculated. Then, each MA having a distance at least 90% of the radius of the ROI from the center of the retina is considered as peripheral MA.

Training

In the training step, each preprocessing method is applied on the training set individually and their effect on the candidate extractor is evaluated. That is, the number of true and false positives, and the number of correctly identified MAs for each category are measured, respectively. The best performing preprocessing method for each category is selected.

Testing

On unknown images, the results of the candidate extractor with the selected preprocessing methods are collected and merged as the union of candidate extractor output as described in section 1.3, as well. The merged MA candidates are the output of the proposed candidate extraction approach.

6.3.2 Adaptive weighting

In this section, we show a way to combine the output of the ⟨PP, CE⟩ pairs by weighting.

⟨PP, CE⟩ ensembles

We combine the output of the ⟨PP, CE⟩ pairs by weighting the ⟨PP, CE⟩ ensembles described in sections 6.1 and 6.2. We assign weights to each candidate with respect to three different pieces of information: which pair detected the candidate, what is the contrast in the neighborhood of the MA and where it is located in the image.

The weighted voting approach

The performance of each ⟨PP, CE⟩ pair is measured for each category in the following way: each extracted candidate is categorized both by visibility and spatial location, then compared to the ground truth whether it is actually an MA or not. Based on this evaluation, for each pair $\mathcal{P}$, visual category v and spatial category s, we calculate the *F-Score* measure as

$$\textit{F-Score}_{v,s}^{\mathcal{P}} = \frac{2 \cdot SEN_{v,s}^{\mathcal{P}} \cdot PPV_{v,s}^{\mathcal{P}}}{SEN_{v,s}^{\mathcal{P}} + PPV_{v,s}^{\mathcal{P}}}, \tag{6.3}$$

where

$$v \in \{\text{subtle}, \text{obvious}, \text{regular}\}, \;\; s \in \{\text{near to vessel}, \text{in the macula}, \text{on the periphery}, \text{other}\} \tag{6.4}$$

and

$$SEN_{v,s}^{\mathcal{P}} = \frac{\text{True MA detections}}{\text{All MAs}}, \;\; PPV_{v,s}^{\mathcal{P}} = \frac{\text{True MA detections}}{\text{All candidates of } \mathcal{P}}. \tag{6.5}$$

Then, we approximate the optimal weights for the ⟨PP, CE⟩ pair $\mathcal{P}$ through the formula [56]

$$w_{v,s}^{\mathcal{P}} = \log \frac{\textit{F-Score}_{v,s}^{\mathcal{P}}}{1 - \textit{F-Score}_{v,s}^{\mathcal{P}}}. \tag{6.6}$$

The weights are normalized for each combination of visual and spatial categories to have sum 1.

Finally, for each candidate in an unknown image, the visual and spatial location categories are determined and the corresponding weight values are summed as the confidence value of the candidate. The final confidence value assigned to an MA candidate is the sum of the weights of the ⟨PP, CE⟩ pairs, which detected this candidate. The selected MA candidates can be filtered by thresholding their confidence values.

6.3.3 Methodology for the context-aware approach

As for the preprocessing methods to set up the ⟨PP, CE⟩ ensembles, we consider PP_1, PP_2, PP_3, PP_4, PP_5, PP_6, PP_7, PP_8 from section 1.6.1. The MA candidate extractors MA_1, MA_2, MA_3, MA_4, MA_5 included in this study are described in section 1.6.4 in details.

Context-aware selection

Our experimental test is performed on the DiaretDB1 database [60] described in section 1.5.3. The database is split into a disjoint training and test set. The selection of the preprocessing method is solely based on the results achieved in the training set. Detailed information about the number of images and MAs can be found in Table 6.15. For each preprocessing method, we used the same parameter setting for the MA candidate extractor MA_5.

Category	Training	Test
MA	126	278
Images	28	61
Subtle	32	101
Regular	48	86
Obvious	46	91
In the macula	4	4
Near vessel	25	43
On the periphery	39	69
Other	58	162

Table 6.15: *The number of all MAs, images, and the MAs belonging to each category for the training and the test databases, respectively.*

Adaptive weighting

Our first experimental test is performed on the DiaretDB0 database [59] described in section 1.5.2. The database is split into a disjoint training and test set. There is also manually marked MAs available for this database as ground truth. The second test is performed on the Retinopathy Online Challenge (ROC) [58] described in section 1.5.1.

6.3.4 Results

Context-aware selection

Table 6.16 contains the results of the training phase with the highest number of correctly recognized MAs are are set in bold with the corresponding preprocessing method. As it can be seen, the preprocessing methods PP_3, PP_4 and PP_7 are selected as each of them performed better than the rest of the algorithms in two categories.

Category	PP_1	PP_2	PP_3	PP_4	PP_5	PP_6	PP_7	PP_8
TP	57	29	61	65	52	16	54	49
FP	1572	1157	1799	1657	859	231	659	919
Subtle	14	8	16	**18**	11	0	6	5
Regular	21	8	**24**	22	17	2	18	15
Obvious	22	13	21	25	25	14	**30**	29
In the macula	2	2	**3**	2	1	0	2	1
Near vessel	4	4	4	4	7	3	**8**	7
On the periphery	12	7	16	**19**	16	2	11	7
Other	39	16	38	**40**	28	11	33	35

Table 6.16: *The number of true and false detections and the number of correctly recognized cases for each MA category in the training database. The preprocessing methods PP_3, PP_4 and PP_7 are selected since they achieved the highest number of correctly identified MAs in at least one category.*

In Table 6.17, the results of the proposed approach and the individual methods can be seen. For fair comparison, we have selected that parameter setting for each method, where the ratio of the true and false detections are the highest. The proposed method provided the highest number of correctly recognized MAs in each category. However, the number of false detections also increased, but it can be lowered further by applying a voting scheme.

	PP_1	PP_2	PP_3	PP_4	PP_5	PP_6	PP_7	PP_8	Proposed
TP	147	54	154	116	60	28	72	67	198
FP	3023	2303	3105	3383	1500	240	765	1625	5521
Subtle	41	12	46	27	11	0	7	9	64
Regular	53	15	55	43	24	22	25	22	67
Obvious	53	27	53	46	25	22	44	36	68
In the macula	2	0	2	2	1	0	1	0	3
Near vessel	7	2	6	8	3	6	9	8	18
On the periphery	23	8	27	18	8	1	7	6	38
Other	115	44	119	88	48	21	55	53	139

Table 6.17: *The number of true and false detections and the number of correctly recognized cases for each MA category in the test database. The proposed combination outperformed the individual preprocessing methods in all categories.*

While these results are reassuring, it should also be noted that the MA detection performance for some categories (e.g. near vessel, periphery) are rather low. However, we can expect that the inclusion of other preprocessing methods and MA candidate extractors can lead to a better performance in these cases.

Adaptive weighting

The results of the weighting approach for the DiaretDB0 [59] and the ROC [58] databases can be seen in Tables 6.18 and 6.19, respectively. The search-based algorithm in the tables for comparison is a former method of ours, which based on the selection of ⟨PP, CE⟩ pairs. As it can be seen from the results, the proposed weighting approach provides better results on the DiaretDB0 database, but not on the ROC one. The reason for the alternating performance of the weighted and the search-based method may lie in the fact that the fundus image databases are rather diverse.

However, both ensemble-based approaches outperformed the individual detectors which shows the strength of the ensembles in this field.

System	1/8	1/4	1/2	1	2	4	8	avg.
DRSCREEN (search)	0.003	0.005	0.011	0.021	0.043	0.087	0.174	0.049
DRSCREEN (weighted)	0.012	0.025	0.037	0.060	0.090	0.129	0.189	0.077

Table 6.18: *Comparison of the search- and the weighting-based MA detector ensembles.*

System	1/8	1/4	1/2	1	2	4	8	avg.
DRSCREEN (search)	0.173	0.275	0.380	0.444	0.526	0.599	0.643	0.434
DRSCREEN (weighted)	0.172	0.201	0.323	0.426	0.478	0.560	0.638	0.399
Niemeijer et al.	0.243	0.297	0.336	0.397	0.454	0.498	0.542	0.395
LaTIM	0.166	0.230	0.318	0.385	0.434	0.534	0.598	0.381
ISMV	0.217	0.270	0.366	0.407	0.440	0.459	0.468	0.375
OKmedical II	0.175	0.242	0.297	0.370	0.437	0.493	0.569	0.369
OKmedical	0.198	0.265	0.315	0.356	0.394	0.466	0.501	0.357
Lázár et al.	0.169	0.248	0.274	0.367	0.385	0.499	0.542	0.355
GIB	0.190	0.216	0.254	0.300	0.364	0.411	0.519	0.322
Fujita	0.181	0.224	0.259	0.289	0.347	0.402	0.466	0.310
IRIA	0.041	0.160	0.192	0.242	0.321	0.397	0.493	0.264
Waikato	0.055	0.111	0.184	0.213	0.251	0.300	0.329	0.206

Table 6.19: *Detailed quantitative results of the Retinopathy Online Challenge (including the weighted ensemble).*

Chapter 7

An ensemble-based system for the automatic screening of diabetic retinopathy (DR)

In this chapter, a framework for the automatic grading of color fundus images regarding diabetic retinopathy (DR) is proposed. The approach classifies images based on characteristic features extracted by lesion detection and anatomical part recognition algorithms. These features are then classified using an ensemble of classifiers. As the results show, the proposed approach is highly efficient for this task. Corresponding results have been published in [9, 10, 33–38] and are also incorporated in the project DRSCREEN: Developing a computer-based image processing system for diabetic retinopathy screening, TECH08-2 grant of the Hungarian National Office for Research and Technology (NKTH).

Many results of this dissertation dedicated to object detection in retinal images were motivated by the intention to build an automatic system for diabetic retinopathy (DR) screening.

Automatic computer-aided screening of DR is a highly investigated field [209]. The motivation for creating reliable automatic DR screening systems is to reduce the manual effort of mass screening [210], which also raises a financial issue [211]. While several studies focus on the recognition of patients having DR [204, 210] and considering the specificity of the screening as a matter of efficiency, we show how both sensitivity and specificity can be kept at high level by combining novel screening features and a decision-making process. Especially, our results are very close to meet the recommendations of the British Diabetic Association (BDA) (sensitivity 80% and specificity 95% [205]).

The basis for an automatic screening system is the analysis of color fundus images [212]. The key to the early recognition of DR is the reliable detection of microaneurysms (MAs) on the retina, which serves as an essential part for most automatic DR screening systems [7, 71, 204, 207]. The role of bright lesions for DR grading has also been investigated with positive [213] and negative outcomes [204] reported. Besides lesions, image quality assessment [208, 214] is also considered to exclude ungradable images. As a new direction, in [206] an image-level DR recognition algorithm is also presented.

The proposed framework extends the state-of-the-art components of an automatic DR screening system by adding prescreening [33] and the distance of the macula (MC) and OD centers (ODC) as novel components. We also use image quality assessment as a feature for classification rather than a tool for excluding images. The comparison of the components used in some recently published automatic DR screening systems are enclosed in Table 7.1.

Screening system	Image quality	Red lesion	Bright lesion	AM/FM	Pre-screening	MC-ODC
[7]		X				
[204]		X				
[206]				X		
[207]		X				
[208]	X	X	X			
[214]	X	X				
Proposed	X	X	X	X	X	X

Table 7.1: *Comparison of components of the automatic screening system.*

Regarding decision making, automatic DR screening systems either partially follow clinical protocols (e.g. MAs indicate presence of DR) [7, 207, 208, 214] or use a machine learning classifier [206, 209, 213]. A common way to improve reliability in machine learning-based applications is to use ensemble-based approaches [56]. For medical decision support, ensemble methods were successfully applied to several fields. In [215], the authors investigated the applicability of ensembles for breast cancer data classification. The prediction of response to certain therapy is improved by the use of a classifier ensemble [216]. In [217], an ensemble of four classifiers was used for cardiovascular disease prediction. Ensemble methods are also provided improvement over single classifiers in a natural language processing environment [218].

Ensemble systems combine the output of multiple learners with a specific fusion strategy. In [7] and [204], the fusion of multiple MA detectors has proven to be more efficient than a single algorithm for DR classification. The proposed system is ensemble-based at more levels: we consider ensemble systems both in image processing tasks and decision making.

In this chapter, a framework for the automatic grading of color fundus images regarding DR is proposed. The approach classifies images based on characteristic features extracted by lesion

detection and anatomical part recognition algorithms. These features are then classified using an ensemble of classifiers. As the results show, the proposed approach is highly efficient for this task. The flow chart of our decision making protocol can be seen in Figure 7.1, as well. We tested our approach on the publicly available dataset Messidor (see section 1.5.5), where it has provided a 0.989 area under the *ROC* curve (*AUC*) value in a disease/no disease setting, which is a relatively high figure compared with other state-of-the-art techniques.

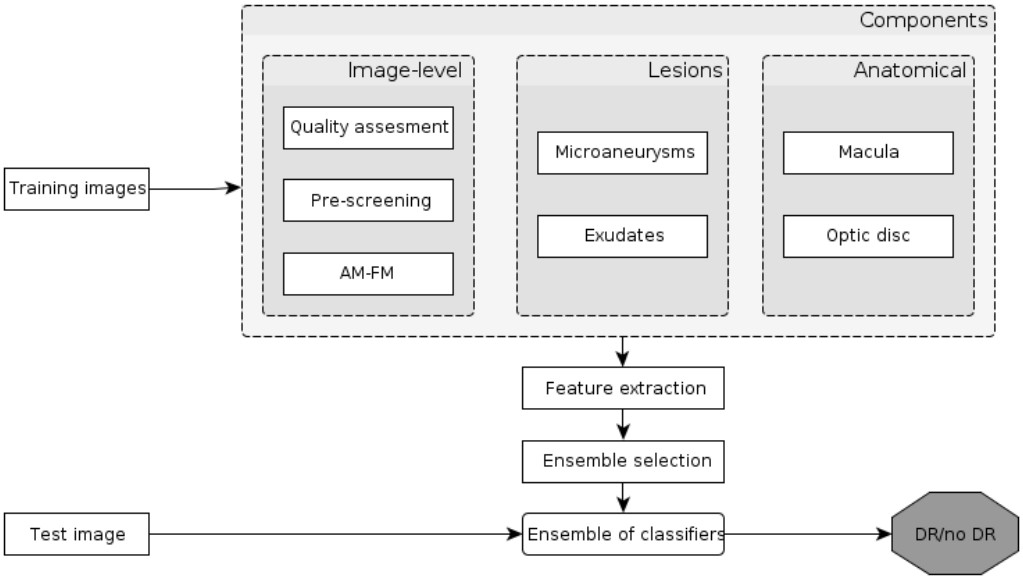


Figure 7.1: *Flow chart of the proposed decision support framework.*

The rest of the chapter is organized as follows. In section 7.1, we present the image processing components of our system. The selection of the classifiers to build up an ensemble for the final decision is described in section 7.2. Our experimental methodology and results can be found in sections 7.3 and 7.4, respectively.

7.1 Components of an automatic system for DR screening

In this section, the components we used for feature extraction are described. They can be classified as image-level, lesion-specific, and anatomical ones.

7.1.1 Image-level components

Quality assessment

We classify the images whether they have sufficient quality for a reliable decision with a supervised classifier, where the box count values of the detected vessel system serve as features [34]. For vessel segmentation, we use an own approach proposed in [35] based on Hidden Markov Random Fields (HMRF). Here, we extend the optimization problem of HMRF models considering the tangent vector field of the image to enhance the connectivity of the vascular system consisting of elongated structures.

Prescreening

During prescreening [33], we classify the images as severely diseased (abnormal) ones or to be forwarded for further processing. Each image is split into disjoint regions and a simple texture descriptor (inhomogeneity measure) is extracted for each region. Then, a machine learning classifier is trained to classify the images based on these features.

Multi-scale AM/FM based feature extraction

The Amplitude-Modulation Frequency-Modulation (AM/FM) [219] method extracts information from an image, decomposing the green channels of the images into different representations, which reflect the intensity, geometry, and texture of the structures with signal processing techniques. The extracted information are then filtered to establish 39 different representations of the image. The images are classified using these features with a supervised learning method. More information on this approach can be found in [219].

7.1.2 Lesion-specific components

MA detection

MAs are normally the earliest signs of DR. They appear as small red dots in the image and their resemblance to vessel fragments make it hard to detect them efficiently. In the proposed system, we apply the own MA detection method described in [7] and chapter 6, which is an efficient approach based on ⟨preprocessing method, candidate extractor⟩ ensembles.

Since robust MA detection is a crucial task in the automatic screening of DR, we put an effort to also develop efficient individual MA detectors to act as members of our ensemble-based system. Namely, we completed our microaneurysm (MA) detector algorithm [37] in our subsequent work [10]. This method realizes MA detection through the analysis of directional cross-section profiles centered on the local maximum pixels of the preprocessed image. Peak detection is applied on each profile, and a set of attributes regarding the size, height, and shape of the peak are calculated subsequently. The statistical measures of these attribute values as the orientation of the cross-section changes constitute the feature set that is used in a Naïve Bayes classification to exclude spurious candidates. A formula is given for the final score of the remaining candidates, which can be thresholded further for a binary output.

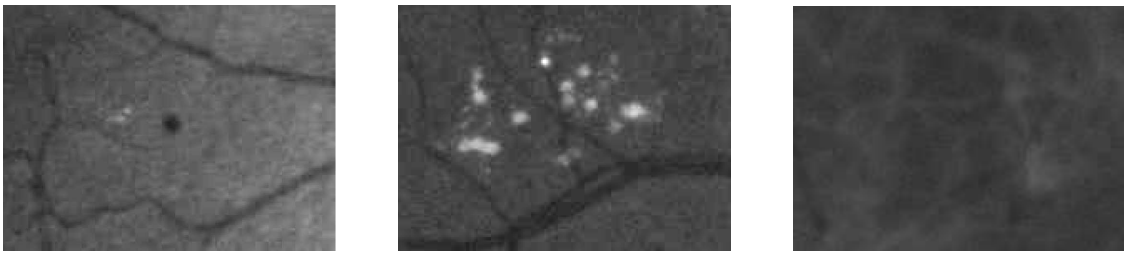
Exudate detection

Exudates are primary signs of DR and occur when lipid or fat leak from blood vessels or aneurysms. Exudates are bright, small spots, which can have irregular shape. Since exudate detection is also a challenging task, we follow the same complex methodology in our own detector as for MA detection described in chapter 6. Thus, we combine preprocessing methods and candidate extractors in the case of exudate detection, as well [38]. In Figure 7.2, we show some examples for the appearance of DR-related symptoms in retinal images.

7.1.3 Anatomical components

Macula detection

The macula is the central region of sharp vision in the human eye with its center referred to as the fovea. Any lesions appearing within the macula can lead to severe loss of vision. Therefore, the efficient detection of the macula is essential in an automatic screening system for DR. The macula



(a) Microaneurysm (b) Exudates (c) Inhomogeneity

Figure 7.2: *Some representative visual features to be extracted from the images.*

is located roughly in the center of the retina, temporal to the optic nerve. In our system, we use the own method described in [36], which extracts the largest component from the image which is darker than its surroundings. The location of the macula together with the OD described below define some features incorporated in our decision framework.

Optic disc detection

The OD is a circular shaped anatomical structure with a bright appearance. It is the area, where the optic nerve enters the eye. If the center and the radius of the OD are detected correctly, they can be used as reference data for locating other anatomical parts e.g. the macula. In our system, we use the own ensemble-based approach described in [3] and section 4.1.2. Recognizing these anatomical parts is important from two aspects: the appearance of certain lesions at specific positions can indicate a more advanced stage of DR and the presence of rare, but serious defects (like retinal detachment) can ruin the detection of the OD and macula.

7.2 Ensemble selection

The most important expectation from a computer-aided medical system is its high reliability. To ensure that, we use ensemble-based decision making [56]. Thus, we train several classifiers to separate DR and non-DR cases and fuse their results. In this section, we describe how we select the ensemble for DR classification based on features extracted from the output of the detectors presented in section 7.1. To combine the decisions of the individual classifiers, we will use the ensemble classifiers $\mathcal{D}_{maj}$, $\mathcal{D}_{wmaj}$, $\mathcal{D}_{avg}$, $\mathcal{D}_{pro}$, $\mathcal{D}_{min}$, $\mathcal{D}_{max}$ described in section 1.3.

To select the optimal ensemble for DR classification, we trained several well-known classifiers that will be described in section 7.3.2. Each ensemble is a subset of these classifiers. Several approaches were tested for selecting the best subset of classifiers $\mathcal{D}'$ for DR grading. The following search methods were investigated based on [174] for a fixed set of classifiers $\{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$ and energy function $E : \mathcal{D}' \subseteq \{\mathcal{D}_1, \ldots, \mathcal{D}_n\} \to \mathbb{R}_{\geq 0}$.

- **Forward search:** First, the best individual classifier is selected. Then, further classifiers are added if the performance of the ensemble increases. The process ends when no further increase of performance is reached by adding more classifiers. Algorithm 7.2.1 gives a formal description of this search method.

Algorithm 7.2.1. Forward search.

---

1. $\mathcal{D}' \leftarrow \operatorname{argmax}\left(E\left(\{\mathcal{D}_1\}\right), \ldots, E\left(\{\mathcal{D}_n\}\right)\right)$
2. $e_{best} \leftarrow E\left(\mathcal{D}'\right)$

3. ***for all*** $\mathcal{D}_i \notin \mathcal{D}', i = 1 \ldots n$ ***do***
4. $\quad e \leftarrow E\left(\mathcal{D}' \cup \{\mathcal{D}_i\}\right)$
5. $\quad$ ***if*** $e > e_{best}$ ***then***
6. $\quad\quad \mathcal{D}' \leftarrow \mathcal{D}' \cup \{\mathcal{D}_i\}$
7. $\quad\quad e_{best} \leftarrow e$
8. $\quad$ ***end if***
9. ***end for***
10. ***return*** $\mathcal{D}'$

- **Backward search:** First, all classifiers are considered as members of the ensemble. Then, classifiers are removed from the ensemble while the performance of the ensemble increases. See Algorithm 7.2.2 for a formal description.

Algorithm 7.2.2. Backward search.

1. $\mathcal{D}' \leftarrow \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$
2. $e_{best} \leftarrow E\left(\mathcal{D}'\right)$
3. ***for all*** $\mathcal{D}_i \in \mathcal{D}'$ ***do***
4. $\quad e \leftarrow E\left(\mathcal{D}' \setminus \{\mathcal{D}_i\}\right)$
5. $\quad$ ***if*** $e > e_{best}$ ***then***
6. $\quad\quad \mathcal{D}' \leftarrow \mathcal{D}' \setminus \{\mathcal{D}_i\}$
7. $\quad\quad e_{best} \leftarrow e$
8. $\quad$ ***end if***
9. ***end for***
10. ***return*** $\mathcal{D}'$

For comparison, we also consider the following two ensembles besides the ones found by the search methods.

- **All:** All classifiers are members of the ensemble.
- **Single best:** The ensemble contains only the best performing classifier.

7.3 Methodology for the automatic screening approach

For experimental studies, we consider the publicly available Messidor database described in section 1.5.5. Some example images corresponding to the different grading scores are shown in Figure 7.3.

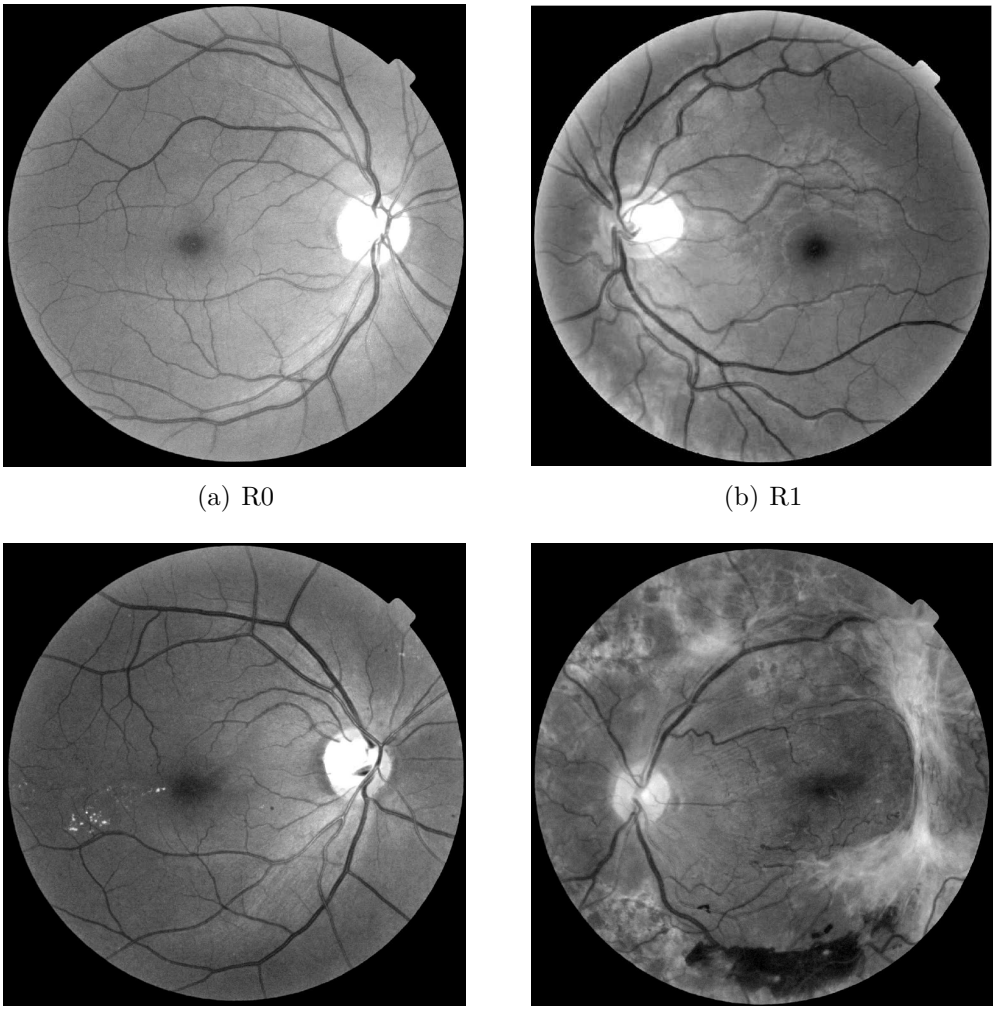
7.3.1 Features to be considered in classification

In this section, we describe the features that were extracted from the output of the image processing algorithms presented in section 7.1. These features are also summarized in Table 7.2.

• χ_0 is the result of quality assessment, which is a real number between 0 (worst) and 1 (best) quality for a color fundus image.

• χ_1 is a binary variable representing the result of prescreening, where 1 indicates severe retinal abnormality and 0 its lack.

• As an essential part of a DR screening system, features $\chi_2 - \chi_7$ describe the results of MA detection. More precisely, χ_i $(i = 2, \ldots, 7)$ stand for the number of MAs found at the confidence levels $\alpha = 0.5, \ldots, 1$, respectively.



(a) R0 (b) R1

(c) R2 (d) R3

Figure 7.3: *Representative images having grades R0, R1, R2, R3 from the Messidor database.*

• $\chi_8 - \chi_{16}$ contain the same information as $\chi_2 - \chi_7$ for exudates. However, as exudates are represented by a set of points rather than the number of pixels constructing the lesions, these features are normalized by dividing the number of lesions with the diameter of the ROI to compensate different image sizes.

• Since abnormalities can make it harder to detect certain anatomical landmarks in an image, χ_{17} represents the Euclidean distance of the center of the macula and the center of the OD to provide important information regarding the patient's condition (see Figure 7.4 for an example). This feature is also normalized with the diameter of the ROI.

• χ_{18} is the result of the AM/FM-based classification, which is a non-negative scalar indicating the confidence of DR detection. The larger χ_{18}, the higher the probability that DR is present.

7.3.2 Classifiers and energy functions

We considered the following classifiers as potential members of ensembles.

- Alternating Decision Tree ($\mathcal{D}_1$),
- kNN ($\mathcal{D}_2$),
- AdaBoost ($\mathcal{D}_3$),

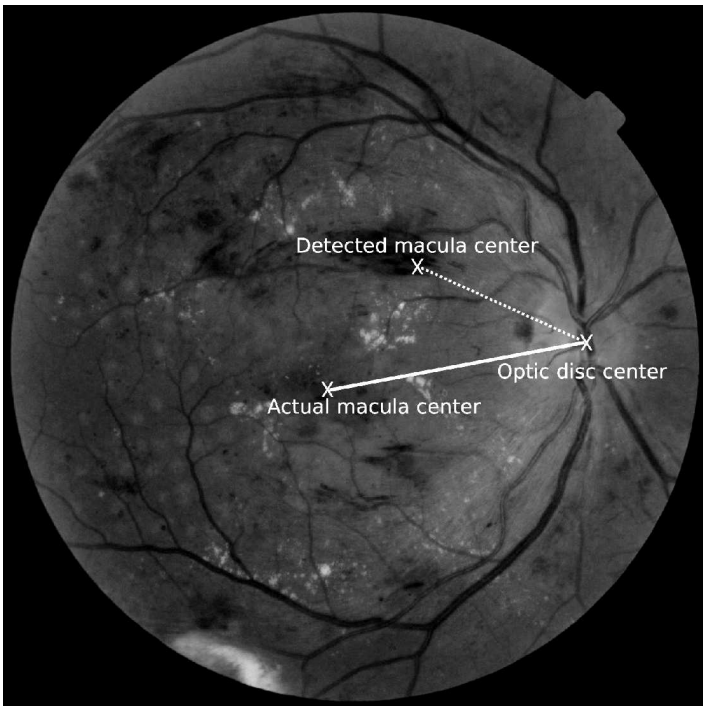


Figure 7.4: *The difference between the actual and the detected OD and macula centers.*

Feature	Description of feature
χ_0	The result of quality assessment.
χ_1	The result of prescreening (non-severe DR / severe DR).
$\chi_2 - \chi_7$	The number of MAs detected at confidence levels $\alpha = 0.5, \ldots, 1.0$, resp.
$\chi_8 - \chi_{16}$	The number of exudate pixels at confidence levels $\alpha = 0.1, \ldots, 1.0$, resp.
χ_{17}	The Euclidean distance of the center of the macula and the center of the optic disc.
χ_{18}	The result of the AM/FM-based classification (No DR/DR).

Table 7.2: *Features for DR grading.*

- Multilayer Perceptron ($\mathcal{D}_4$),
- Naive Bayes ($\mathcal{D}_5$),
- Random Forest ($\mathcal{D}_6$),
- SVM ($\mathcal{D}_7$),
- Pattern classifier [28] ($\mathcal{D}_8$).

For ensemble selection, we considered the energy functions sensitivity SEN, accuracy ACC and *F-Score* defined in section 1.3.1. Notice that to fit this realization to the general framework, the features $\chi_1, \ldots, \chi_{18}$ and classifiers $\mathcal{D}_1, \ldots, \mathcal{D}_8$ should be considered, respectively. Moreover, any of the energy functions SEN, ACC, *F-Score* should be assigned to E in section 7.2.

7.3.3 Training and evaluation

10-fold cross-validation was used for both the training phase and for the evaluation of the ensembles. The figures given in section 7.4 are the average values of the 10-fold cross-validation for the

R0 vs R1 – Forward search			
Energy function / Fusion strategy	*Sensitivity*	*Accuracy*	*F-Score*
$\mathcal{D}_{maj}$	98%/82%/83%	77%/90%/88%	75%/89%/86%
$\mathcal{D}_{wmaj}$	76%/90%/88%	83%/88%/87%	87%/87%/87%
$\mathcal{D}_{avg}$	86%/88%/88%	77%/88%/86%	82%/89%/88%
$\mathcal{D}_{pro}$	74%/90%/86%	80%/88%/87%	79%/89%/87%
$\mathcal{D}_{min}$	74%/90%/87%	74%/91%/87%	85%/88%/88%
$\mathcal{D}_{max}$	77%/90%/87%	71%/91%/86%	81%/88%/87%

Table 7.3: *DR grading results for scenario R0 vs R1 on the Messidor database with forward search method using different fusion strategies and energy functions. Each cell contains the Sensitivity/Specificity/Accuracy of the best ensemble for the corresponding setup.*

respective energy functions in each case on the Messidor database. To measure the performance of the ensembles, we consider the following descriptive values described in section 1.3.1: *Sensitivity*, *Specificity*, *Accuracy*. To compare our results with other approaches, we fitted Receiver Operating Characteristic (*ROC*) curves to the results and calculated the area under these curves (*AUC*) using JROCFIT [203]. We evaluated the ensemble creation strategies in two scenarios.

- R0 vs R1: First, we investigated whether the image contains early signs of retinopathy (R1) or not (R0), that is, $\Omega = \{\text{R0}, \text{R1}\}$ using the framework introduced in section 1.3. Discriminating these two classes are the most challenging task of DR screening, since R1 usually has only minor and visually less distinguishable signs of DR than the advanced stages R2 and R3.

- No DR/DR: Second, we measured the classification performance of the ensembles between all diseased categories (R1, R2, R3) and the normal one (R0), that is, $\Omega = \{\text{R0}, \text{non-R0}\}$ in this case, where non-R0 relates to any of the categories R1, R2, R3.

7.4 Results

7.4.1 Ensemble selection

Tables 7.3 and 7.4 contain the *Sensitivity*, *Specificity* and *Accuracy* values corresponding to the different fusion strategies and search methods for the scenario R0 vs R1, while Tables 7.5 and 7.6 relate to the scenario No DR/DR, respectively. For both scenarios, the table entries corresponding to the most accurate ensembles are set in bold, if they indeed correspond to remarkable better performance than other cases. For better comparison, we also disclose the accuracy values for the ensembles containing all classifiers in Table 7.7.

Regarding the scenario R0 vs R1, from Table 7.4 we can see that the best performing ensemble achieved 94% *Sensitivity*, 90% *Specificity* and 90% *Accuracy* using backward search, output fusion strategy $\mathcal{D}_{avg}$ and energy function *Accuracy*. For the scenario No DR/DR, 90% *Sensitivity*, 91% *Specificity* and 90% *Accuracy* are achieved with the same search method and fusion strategy (see Table 7.6). However, the energy function in this case is *Sensitivity*. For a fair comparison, we also disclose the aggregated results for the energy functions and search methods in Tables 7.8 and 7.10 for the scenario R0 vs R1, and in Tables 7.9 and 7.11 for the scenario No DR/DR, respectively.

R0 vs R1 – Backward search			
Energy function / Fusion strategy	*Sensitivity*	*Accuracy*	*F-Score*
$\mathcal{D}_{maj}$	88%/87%/87%	92%/88%/89%	84%/89%/88%
$\mathcal{D}_{wmaj}$	98%/82%/84%	85%/88%/88%	69%/88%/83%
$\mathcal{D}_{avg}$	85%/89%/88%	**94%/90%/90%**	93%/90%/90%
$\mathcal{D}_{pro}$	0%/78%/78%	0%/79%/80%	0%/78%/80%
$\mathcal{D}_{min}$	81%/90%/88%	83%/89%/88%	64%/96%/85%
$\mathcal{D}_{max}$	98%/81%/82%	98%/81%/83%	76%/89%/86%

Table 7.4: *DR grading results for scenario R0 vs R1 on the Messidor database with backward search method using different fusion strategies and energy functions. Each cell contains the Sensitivity/Specificity/Accuracy of the best ensemble for the corresponding setup.*

No DR/DR – Forward search			
Energy function / Fusion strategy	*Sensitivity*	*Accuracy*	*F-Score*
$\mathcal{D}_{maj}$	88%/79%/86%	91%/76%/88%	88%/84%/88%
$\mathcal{D}_{wmaj}$	88%/84%/87%	88%/88%/87%	91%/68%/85%
$\mathcal{D}_{avg}$	86%/83%/85%	88%/85%/88%	89%/81%/87%
$\mathcal{D}_{pro}$	95%/38%/60%	85%/83%/85%	89%/72%/85%
$\mathcal{D}_{min}$	80%/95%/80%	88%/82%/87%	87%/78%/86%
$\mathcal{D}_{max}$	92%/50%/72%	90%/76%/87%	88%/76%/86%

Table 7.5: *DR grading results for scenario No DR/DR on the Messidor database with forward search method using different fusion strategies and energy functions. Each cell contains the Sensitivity/Specificity/Accuracy of the best ensemble for the corresponding setup.*

No DR/DR – Backward search			
Energy function / Fusion strategy	*Sensitivity*	*Accuracy*	*F-Score*
$\mathcal{D}_{maj}$	89%/78%/86%	90%/80%/89%	90%/88%/90%
$\mathcal{D}_{wmaj}$	88%/93%/85%	86%/83%/85%	89%/90%/88%
$\mathcal{D}_{avg}$	**90%/91%/90%**	87%/80%/86%	89%/92%/90%
$\mathcal{D}_{pro}$	97%/56%/80%	88%/85%/88%	90%/73%/86%
$\mathcal{D}_{min}$	81%/97%/82%	81%/97%/82%	81%/98%/83%
$\mathcal{D}_{max}$	93%/68%/86%	93%/77%/89%	89%/83%/88%

Table 7.6: *DR grading results for scenario No DR/DR on the Messidor database with backward search method using different fusion strategies and energy functions. Each cell contains the Sensitivity/Specificity/Accuracy of the best ensemble for the corresponding setup.*

All classifiers		
Scenario / Fusion strategy	R0 vs R1	No DR/DR
$\mathcal{D}_{maj}$	96%/84%/85%	88%/79%/86%
$\mathcal{D}_{wmaj}$	85%/87%/87%	88%/84%/87%
$\mathcal{D}_{avg}$	80%/88%/87%	86%/83%/85%
$\mathcal{D}_{pro}$	100%/78%/78%	95%/38%/60%
$\mathcal{D}_{min}$	48%/95%/69%	80%/95%/80%
$\mathcal{D}_{max}$	95%/79%/80%	92%/50%/72%

Table 7.7: *DR grading results on the Messidor database with all of the classifiers included in the ensemble. Each cell contains the Sensitivity/Specificity/Accuracy of the best ensemble for the corresponding setup.*

Energy functions

For scenario R0 vs R1 we can state that while the energy functions *Sensitivity* and *Accuracy* have performed similarly, *F-Score* has provided less accurate ensembles. For scenario No DR/DR all the three energy functions performed similarly. The difference in the effectiveness of the measure *F-Score* probably lies in the fact that the dataset for scenario R0 vs R1 is biased to R0, since it contains much more instances belonging to that class. That is, the energy functions *Accuracy* and *Sensitivity* look more robust for less balanced datasets.

Search methods

As for the search methods, the accuracy of the forward and backward search method are similar. However, in both scenarios, the *Sensitivity* and *Specificity* values are more balanced for the backward strategy, which is desired for a grading system.

R0 vs R1			
Energy function	*Sensitivity*	*Specificity*	*Accuracy*
Sensitivity	86%	86%	86%
Accuracy	84%	88%	87%
F-Score	81%	88%	80%

Table 7.8: *Comparison of the energy functions for the scenario R0 vs R1.*

No DR/DR			
Energy function	*Sensitivity*	*Specificity*	*Accuracy*
Sensitivity	90%	79%	86%
Accuracy	88%	84%	87%
F-Score	88%	82%	87%

Table 7.9: *Comparison of the energy functions for the scenario No DR/DR.*

R0 vs R1			
Search method	*Sensitivity*	*Specificity*	*Accuracy*
Forward	80%	89%	87%
Backward	88%	86%	86%
All	84%	85%	81%

Table 7.10: *Comparison of the search methods for the scenario R0 vs R1.*

R0 vs R1			
Search method	*Sensitivity*	*Specificity*	*Accuracy*
Forward	90%	78%	87%
Backward	88%	84%	87%
All	88%	71%	79%

Table 7.11: *Comparison of the search methods for the scenario No DR/DR.*

Classifier output fusion strategies

In Tables 7.12 and 7.13 the comparison of the fusion strategies can be observed. The experimental results indicate that $\mathcal{D}_{avg}$ is the most effective strategy for both scenarios. The aggregated results confirm this observation. However, $\mathcal{D}_{maj}$ and $\mathcal{D}_{wmaj}$ have also provided similar results, suggesting possible alternative choices.

To conclude on the analysis of ensemble selection approaches, it can be stated that backward ensemble search method with energy functions *Sensitivity* or *Accuracy* and fusion strategy $\mathcal{D}_{avg}$ can be recommended for ensemble selection for automatic DR screening with our framework.

7.4.2 Comparison with other automatic DR screening systems

It is a challenging task to compare our approach with other methods. As we can see in Table 7.14, most research groups not only evaluated their approach on private datasets, but the proportion of images showing signs of DR is also completely different. Moreover, the most meaningful measure, the area under the *ROC* curves (*AUC*) is not always disclosed either. However, the proposed approach has provided significantly better performance than the other state-of-the-art techniques regarding the clinically important measures. Also notice that this comparison was able to be made

R0 vs R1			
Fusion strategy	*Sensitivity*	*Specificity*	*Accuracy*
$\mathcal{D}_{maj}$	87%	87%	87%
$\mathcal{D}_{wmaj}$	83%	87%	86%
$\mathcal{D}_{avg}$	85%	89%	88%
$\mathcal{D}_{pro}$	33%	82%	82%
$\mathcal{D}_{min}$	73%	92%	85%
$\mathcal{D}_{max}$	85%	86%	84%

Table 7.12: *Comparison of classifier output fusion strategies for the scenario R0 vs R1.*

No DR/DR			
Fusion strategy	*Sensitivity*	*Specificity*	*Accuracy*
$\mathcal{D}_{maj}$	89%	80%	88%
$\mathcal{D}_{wmaj}$	88%	83%	87%
$\mathcal{D}_{avg}$	88%	84%	88%
$\mathcal{D}_{pro}$	91%	69%	81%
$\mathcal{D}_{min}$	84%	89%	84%
$\mathcal{D}_{max}$	91%	77%	85%

Table 7.13: *Comparison of classifier output fusion strategies for the scenario No DR/DR.*

only for the scenario No DR/DR, because of the lack of data for scenario R0 vs R1 from the other systems.

System	Cases having DR	*Sensitivity*	*Specificity*	*AUC*
[7]	46%	76%	88%	0.90
[204]	4.96%	N/A	N/A	0.86
[206]	74.43%	N/A	N/A	0.81
[206]	76.26%	N/A	N/A	0.89
[207]	30%	85%	90%	N/A
[208]	35.88%	87%	50.4%	N/A
[209]	4.8%	84%	64%	0.84
[214]	37.5%	90.5%	54.7%	N/A
Proposed	46%	90%	91%	0.989

Table 7.14: *Comparison of automatic DR screening systems.*

In [7], we reported grading results for the dataset Messidor based on only MA detection for both scenarios. The comparative results between the proposed system and [7] are given in Table 7.15 for the scenario R0 vs R1, and in Table 7.16 for the scenario No DR/DR, respectively. To highlight the efficiency of the ensemble-based approach, we included the results corresponding to a single classifier-based decision, as well. As we can see, the proposed system outperforms both [7] and the single classifier approach. It is also interesting to notice that the single classifier approach clearly performs better than [7], which is based solely on the detection of MAs. This observation also confirms the necessity of the wide range of components.

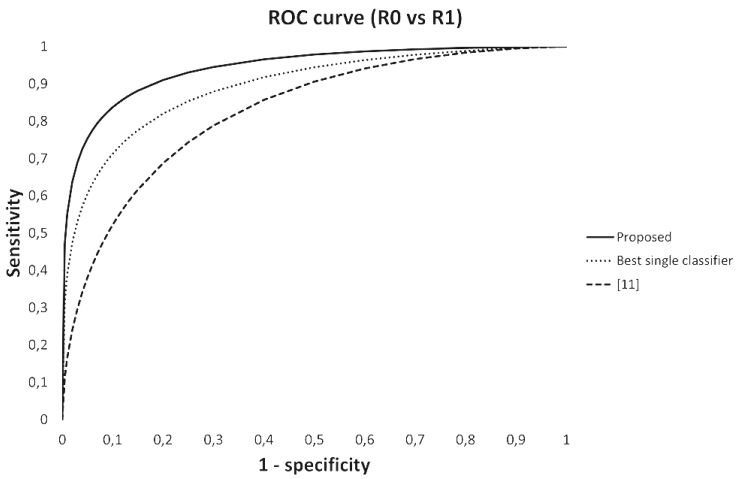


Figure 7.5: *ROC curves of automatic DR screening systems evaluated on the Messidor database for the scenario R0 vs R1.*

R0 vs R1				
System	*Sensitivity*	*Specificity*	*Accuracy*	*AUC*
[7]	97%	14%	32%	0.826
Best single classifier	85%	87%	86%	0.893
Proposed	94%	90%	90%	0.942

Table 7.15: *Comparison of automatic DR screening systems evaluated on the Messidor database for the scenario R0 vs R1.*

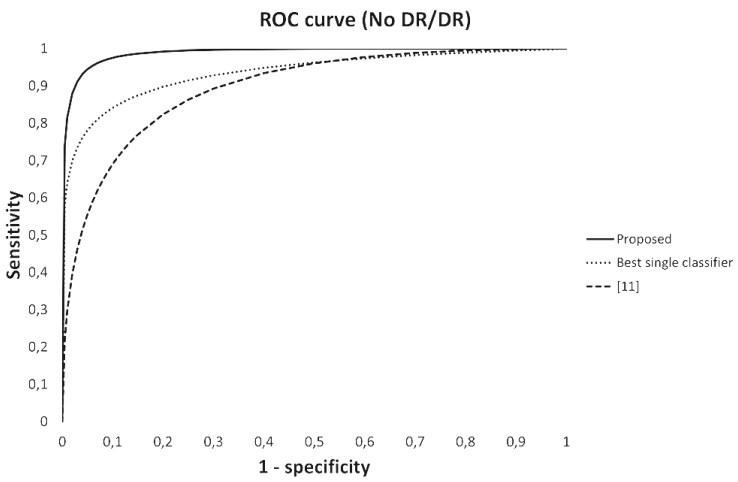


Figure 7.6: *ROC curves of automatic DR screening systems evaluated on the Messidor database for the scenario No DR/DR.*

No DR/DR				
System	*Sensitivity*	*Specificity*	*Accuracy*	*AUC*
[7]	76%	88%	82%	0.90
Best single classifier	90%	81%	86%	0.936
Proposed	90%	91%	90%	0.989

Table 7.16: *Comparison of automatic DR screening systems evaluated on the Messidor database for the scenario No DR/DR.*

Chapter 8

Related results

In this chapter, we briefly present some results which are related to the ones enclosed in the dissertation. The details are omitted to keep up the main focus and also for space reasons. Moreover, we also exhibit results corresponding to some ongoing research on further improvements and such fields, where the results of the dissertation can be exploited.

Within digital geometry, we have devoted remarkable research effort on distance measurement approach in digital grids [11–14, 39]. These works are connected to the content of the dissertation in terms of our results on the approximation of the Euclidean distance which can be applied in object matching and representation problems. We have also applied digital geometric approaches to improve the performance of active contour (snake) models in human body extraction tasks discussed in chapter 2. Namely, we have proposed a speed-up to improve the concavity performance of GVF snakes in [40], extended the GVF snake model to quadtree representation for faster convergence in [41] and generalized a contour tracing method for this representation to address better concavity performance in [42].

As for our clinical application, we have shown that the idea of analyzing cross-section profiles introduced for microaneurysm (MA) detection in [10, 37] can be successfully applied in the segmentation of the retinal vessels, as well [43]. As a more theoretical investigation, we have worked out a new technique to measure dissimilarity in kernel space providing scaling and translation invariance [15]. We have examined the theoretical background of linear invariance in the polynomial kernel space, introduced the centered correlation and centered Euclidean dissimilarity in kernel space, deduced formulas to compute it efficiently and tested the proposed dissimilarity measures with the kNN classifier. The experimental results found for MA detection showed that the presented techniques are highly competitive in similarity or dissimilarity-based classification methods.

Though the dissertation primarily focuses on results regarding the fusion of individual methods, we have devoted effort also for the creation of individual detector algorithms. Besides MA detection we have also developed a competitive method for exudate detection in [16, 44–47]. As our most comprehensive related work for exudate detection, in [16] we apply a grayscale morphology-based method for candidate detection to identify possible regions containing these bright lesions. Then, to extract the precise boundary of the candidates, we introduce a complex active contour-based method. Next, a Naïve Bayes classifier optimized further by an adaptive boosting technique is applied to remove the false exudate candidates.

Retinal image databases are rather inhomogeneous with containing images with different characteristics. Thus, to determine the optimal parameter settings of an individual detector algorithm or an ensemble is a challenging task. To address this issue, we have determined clusters of retinal images coming from different sources. In other words, we have considered individual image characteristics instead of databases in our MA [48] and exudate [49] detection problems. In our study [48] we have selected 19 similarity measures to calculate image differences, and applied k-means clustering to obtain the clusters. For each cluster, an optimal parameter setting is determined for the same MA detector. As for exudate detection addressed in [49], we have extracted also Haralick features and used k-means again for clustering. We have tested our approach on publicly available databases, where the detection performance was successfully increased by the proposed method

for both lesions.

The optimal parameter setting of our macula detector [36] presented in section 1.6.3 is achieved by simulated annealing. A natural idea is to analyze whether the free parameters of the individual algorithms should be adjusted differently from their individually optimal settings if we compose ensembles from them. The main reason for this phenomenon is that besides individually best performance we have to try to reach diversity among the members, as well. Accordingly, we have determined the optimal parameter settings at ensemble level dedicated to MA [50] and exudate [51] detection for the systems presented in [7] and [38], respectively (see also chapter 6). For both detection tasks we could increase the performance further. However, since this search demands massive computational background, we have started to analyze how to make the corresponding simulated annealing procedure computationally efficient with some corresponding results presented in [50, 51].

In the dissertation we have presented image processing related results, however, it is also an important task to store our findings in a standard format. For this aim, we have created metadata schemes based on an XML basis for the content description in both of our systems SHARE [52] and DRSCREEN [53] described in the Introduction. As for SHARE, an XML scheme was designed to provide highly detailed description possibilities of thermal videos for a rescue scene. The scheme description, which is implemented in full compatibility with the MPEG-7 standard, provides fast and simple integration capabilities for database applications, and supports video/image indexing and retrieval for further use. It has been already tested that the proposed schema can be conceptualized in ontology services, which have other input besides video analysis. As for DRSCREEN, the designed XML scheme supports the description of the standard retinal anatomic components and the lesions generally caused by diabetic retinopathy (DR). The scheme also contains schematrons to be able to derive more complex diagnosis based on the amount and spatial distribution of the lesions. The flexible design of the scheme makes any further completions possible to cover acquisitions considering more fundus fields or lesions indicating other diseases.

To increase the performance of our system for DR screening, we have analyzed whether the inclusion of other (non-invasive) modalities may help [54, 55]. Accordingly, we have considered proteomic data gathered from tears. We have found that the joint decision made based on the two modalities image and proteomic data outperformed any of the single modality-based decisions. The main importance of this study lies in the fact that these modalities are rather diverse, thus, their combination in decision making can boost the accuracy of the related screening systems.

Summary of new scientific results

My new scientific results, where my contribution was essential, will be summarized in three thesis points. All of them have been achieved after obtaining my PhD degree. Related chapters and relevant publications of the dissertation are listed at the end of each thesis point.

1. **Optimal sampling and compression of objects**

   A possible way of object detection is to match predefined templates to the processed image. In real time applications, the required time of the matching procedure is a key factor. The compact representation of the matched templates is also an important issue for their efficient storage in devices having limited resources. In this field, I have introduced methods for object sampling and compression based on theoretical results partly obtained by me; I have also verified their efficiency experimentally.

   (i) By the extension on the theory of the centroidal Voronoi tessellation (CVT) [92, 93] I have introduced a new sampling procedure for the simplification of objects in both discrete and Euclidean domains. The sampling error is defined differently as in the classic CVT theory. Namely, it is calculated as the integral of distances between the object and sampled points within a dilated variant of the object. I have proven that this technique is optimal in object detection methods using the chamfer matching framework. I have shown the applicability of the approach for region-based representations of the objects besides their contour-based ones. Especially, I have introduced a weight function to concentrate the sampling of the object to its skeleton. The results have been integrated in the thermal video processing module of the system SHARE to detect human appearances. To summarize, I have shown the efficiency of the proposed sampling method by theoretically proving its optimal behavior and by verifying it experimentally. *See Chapter 2. Related publications: in SCI*[1] *journal:* [1], *other:* [17–19].

   (ii) I have composed a novel technique for the compression of digital curves having arbitrary topology. Opposite to other approaches, the proposed method traces the whole curve, so it is capable of building it up from an alphabet of short line segments. I have given graph theoretical algorithms for the tracing of the curves providing efficient compression. Namely, a curve can be traced by decomposing it to Eulerian graphs or, to avoid decomposition, by adopting the solution of the Chinese Postman Problem, where the weights relate to the compression demands of the specific curve segments. I have introduced a procedure based on a new theoretical model to detect and improve the degenerated digital intersections turning up during the skeletonization of thicker curves. This model interprets the crossing segments as lanes and calculates the volume of deformation in terms of a function of the enclosed angle and widths of these lanes. The curve compression technique has been integrated in the system SHARE to store its template database more efficiently. To summarize, I have shown the efficiency of a

[1] Science Citation Index.

novel curve compression method built upon theoretical derivations; I have performed experimental verification, as well. *See Chapter 3. Related publications: in SCI journals:* [2], *other:* [20, 21].

2. **Detection of objects having single or multiple appearances**

A common image processing task is to reliably recognize and locate the appearances of specific objects. Besides the form of their appearances, like shape and color, we usually have some prior information about the possible number (single or multiple) of appearing objects of certain types. One of the possible approaches to increase the accuracy of object detection is to take as many principles as possible into account from the ones that have been developed for the given problem. In this field, I have introduced novel voting models for the spatial domain; I have proven their validity by theoretical tools and also experimentally.

(i) I have introduced fusion-based techniques for the detection of single objects that can be represented by single pixels. The output of different individual object detector algorithms are aggregated using simple or weighted majority voting-based models. I have made several efforts to improve the detection accuracy. I have composed a graph theoretical procedure for the simultaneous detection of more objects having single appearances with incorporating a constraint regarding their relative spatial locations. I have given also a graph theoretical method that selects the maximum weighted cliques to handle the problem, when the detectors can have more candidates for the location of the object. To take advantage of all the information from the outputs of the individual detectors, I have introduced aggregation models based on axiomatic and Bayesian models. The dependencies between the detectors have been discovered by statistical methods based on the individual variances and pairwise covariances. The dependency issue has been addressed by assigning weights to the outputs, where the weights have been determined via solving a minimization problem for the variance of the final candidate. These approaches have been implemented in the system DRSCREEN to detect the optic disc and macula in retinal images. To summarize, I have experimentally validated the efficiency of a new fusion-based model partly having theoretical foundations. *See Chapter 4. Related publications: in SCI journals:* [3, 4], *other:* [22, 23].

(ii) I have generalized the classic majority and weighted majority voting models to the spatial domain, where the member algorithms vote in terms of pixels for the location of an object. For this purpose, I have introduced a probability term, which can be specifically defined by a geometric constraint according to the shape of the object. I have described the detection accuracy behavior of the new model as a function of the individual accuracies of the member algorithms and the dependencies among them. Namely, I have proven new theoretical results for the minimal and maximal independence of a set of random variables using linear programming. I have determined the change of the accuracy of the ensemble in that practically essential case, when a new member algorithm is added to the system. For disc-like objects, I have derived the behavior of the probability term of the generalized voting model with improving former theoretical results on the diameter of sets of point randomly dropped on a disc. Namely, I have shown that the probability of having the wrong outputs within a disc drops exponentially with the number of detectors in the independent case. I have extended the classic weighted majority voting model to a novel spatial voting one; the appropriate weights to detect disc-like object have been determined by the generalization of the recommendations for the classic case. To discover better the dependencies within the spatial voting model,

I have generalized the classic diversity measures and given a technique to compose efficient ensembles using them. I have demonstrated the applicability of the new models for the detection of the optic disc; the results have been implemented in the system DRSCREEN. To summarize, I have proven the suitability of a novel spatial voting model for the detection of objects that can be represented by single pixels. *See Chapter 5. Related publications: in SCI journal:* [5], *other:* [24–26].

(iii) I have introduced a novel fusion-based technique for the detection of objects having multiple appearances in the image. The method increases the accuracy of the ensemble by making its members more diverse via composing them as pairs of preprocessing algorithms and candidate extractors. To build up an ensemble of such pairs with maximum accuracy, I have composed a stochastic search-based algorithm and tested it exhaustively for energy functions regularly considered to measure the error of detection. I have verified the efficiency of the approach for the detection of microaneurysms in retinal images. To raise further the detection accuracy of the ensemble, I have worked out a context-aware method which takes the appearance features and spatial location of the objects also into account. The results have been implemented in the system DRSCREEN. To summarize, I have shown that the proposed method to make ensembles more diverse is efficient for object detection purposes. *See Chapter 6. Related publications: in SCI journals:* [6–8], *other:* [27–32].

3. **Ensemble-based systems in decision making**

Just in the object detection task, it is worth taking several principles and opinions into consideration to support decisions. The related classic approach is to extract features and to make a decision according to them using a machine learning-based classifier. Besides the high quality features, for the final decision it may be also useful to combine classifiers considering different principles. In this field, I have introduced a new method for the fusion of classifiers; I have validated the approach experimentally within a complex application.

(i) I have introduced a new ensemble-based decision support model, which consists of classic classification methods. The fusion of the classifiers is achieved via axiomatic, majority and weighted majority voting-based models. I have given a method to select the appropriate classifiers and fusion rules to compose ensembles having highest detection accuracies regarding different energy functions. I have verified the efficiency of the approach for both binary and multiclass classification problems according to the requirements of an application field. Namely, the new decision support approach have been implemented in the system DRSCREEN dedicated to the automatic screening of diabetic retinopathy based on retinal images. To summarize, I have shown that the new model considering the fusion of classifiers is efficient for decision support purposes. *See Chapter 7. Related publications: in SCI journal:* [9], *other:* [33, 34].

(ii) I have achieved several results corresponding to the improvement of some components and features of the ensemble-based system discussed in the previous point. Beyond the results presented in the thesis points so far, I have created new methods to exclude obviously diseased or low quality images from further processing based on local inhomogeneity descriptors. I have given a new technique applying mathematical morphology for the detection of the macula, and other ones for the extraction of the retinal vessels using hidden Markov random fields and the detection of microaneurysms through the analysis of intensity profiles. As for the latter field, I have introduced several new features extracted from the functional representations of the intensity profiles for machine

learning-based frameworks. Using a similar principle as for the detection of microaneurysms, I have introduced an ensemble-based system for the detection of region-like objects with verifying the efficiency of this approach for the detection of exudates, which lesions are also specific to diabetic retinopathy. These results have been implemented in the corresponding components of the system DRSCREEN. To summarize, I have shown how the accuracy of a complex decision support system can be raised by improving the efficiency of its components. *See Chapter 7. Related publications: in SCI journal:* [10], *other:* [35–38].

I have achieved several research results related to the ones presented in the dissertation. However, they are not discussed in details and are also excluded from the thesis points partly because of space reasons and to make the dissertation compact. *See Chapter 8. Related publications: in SCI journals:* [11–16], *other:* [39–55].

Appendix – Proof of Theorems

In this chapter, we include the proofs of two theorems from chapter 5.

Theorem 5.1.4. *Suppose that $p \geq 1/2$ and for any k with $0 \leq k \leq n/2$ we have:*

(i) $p_{n,k} + p_{n,n-k} \geq 1$,

(ii) $p_{n,n-k} \geq (n-k)/n$.

Let q be given by (5.7). Then, $q \geq p$, and consequently $\mathbb{E}\xi \geq p$.

Proof. We can write

$$q = \sum_{k=0}^{n} p_{n,k}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{\lfloor n/2 \rfloor}\left(p_{n,k}\binom{n}{k}p^k(1-p)^{n-k} + \right. \tag{8.1}$$

$$\left. p_{n,n-k}\binom{n}{n-k}p^{n-k}(1-p)^k\right) + p_{n,n/2}\binom{n}{n/2}p^{n/2}(1-p)^{n/2}. \tag{8.2}$$

Here if n is odd, the last term should be considered to be zero.

Now by our assumptions $p \geq 1/2$, together with (i) and (ii), using also the identities $\binom{n}{k} = \binom{n}{n-k}$ and $k/n + (n-k)/n = 1$, for any k with $0 \leq k < n/2$ we have:

$$p_{n,k}\binom{n}{k}p^k(1-p)^{n-k} + p_{n,n-k}\binom{n}{n-k}p^{n-k}(1-p)^k \geq \tag{8.3}$$

$$(1-p_{n,n-k})\binom{n}{k}p^k(1-p)^{n-k} + p_{n,n-k}\binom{n}{n-k}p^{n-k}(1-p)^k = \tag{8.4}$$

$$\frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} + \frac{n-k}{n}\binom{n}{n-k}p^k(1-p)^{n-k} + \tag{8.5}$$

$$p_{n,n-k}\binom{n}{n-k}\left(p^{n-k}(1-p)^k - p^k(1-p)^{n-k}\right) \geq \tag{8.6}$$

$$\geq \frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} + \frac{n-k}{n}\binom{n}{n-k}p^k(1-p)^{n-k} + \tag{8.7}$$

$$\frac{n-k}{n}\binom{n}{n-k}\left(p^{n-k}(1-p)^k - p^k(1-p)^{n-k}\right) = \tag{8.8}$$

$$\frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} + \frac{n-k}{n}\binom{n}{n-k}p^{n-k}(1-p)^k. \tag{8.9}$$

In the last inequality, we use (ii) and the fact that $p^{n-k}(1-p)^k - p^k(1-p)^{n-k}$ is non-negative. Furthermore, if n is even, by (ii) we also have

$$p_{n,n/2}\binom{n}{n/2}p^{n/2}(1-p)^{n/2} \geq \frac{n/2}{n}\binom{n}{n/2}p^{n/2}(1-p)^{n/2}. \tag{8.10}$$

Thus, we obtain

$$q \geq \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} p^k (1-p)^{n-k} = p. \tag{8.11}$$

Here, the last equality follows from the proof of Proposition 5.1.2. Since $\mathbb{E}\xi = q$, we have the inequality $\mathbb{E}\xi \geq p$. □

Theorem 5.5.1. *If independent classifiers $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$ are given (conditional independence is considered), then the optimal weight w_i for the classifier $\mathcal{D}_i$ with accuracy p_i can be calculated as*

$$w_i \propto \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)}. \tag{5.56}$$

Proof. Let $s = [s_1, \ldots, s_n]$ denote the vector with the label output of the ensemble, where $s_i \in \Omega$ is the label suggested for $\boldsymbol{\chi}$ by classifier $\mathcal{D}_i$. A Bayes-optimal set of discriminant functions based on the outputs of the n classifiers is

$$g_j(\boldsymbol{\chi}) = \log P(\omega_j) P(s|\omega_j), \quad (j = 1, \ldots, c). \tag{8.12}$$

From the conditional independence, for the discriminant functions $g_j(\boldsymbol{\chi})$ we get

$$\log P(\omega_j) P(s|\omega_j) = \log \left[P(\omega_j) \prod_{i=1}^{n} P(s_i|\omega_j) \right] = \tag{8.13}$$

$$\log P(\omega_j) + \log \left(\prod_{i, s_i = \omega_j} P(s_i|\omega_j) \prod_{i, s_i \neq \omega_j} P(s_i|\omega_j) \right) = \tag{8.14}$$

$$\log P(\omega_j) + \log \left(\prod_{i, s_i = \omega_j} p_i \prod_{i, s_i \neq \omega_j} (1-p_i) r_i \prod_{i, s_i \neq \omega_j} (1-p_i)(1-r_i) \right) = \tag{8.15}$$

$$\log P(\omega_j) + \log \left(\prod_{i, s_i = \omega_j} \frac{p_i (1-p_i)}{1-p_i} \prod_{i, s_i \neq \omega_j} (1-p_i) r_i \prod_{i, s_i \neq \omega_j} (1-p_i)(1-r_i) \right) = \tag{8.16}$$

$$\log P(\omega_j) + \log \left(\prod_{i, s_i = \omega_j} \frac{p_i}{1-p_i} \prod_{i, s_i \neq \omega_j} (1-p_i) r_i (1-r_i) \prod_{i=1}^{n} (1-p_i) \right) = \tag{8.17}$$

$$\log P(\omega_j) + \sum_{i, s_i = \omega_j} \log \frac{p_i}{1-p_i} + \sum_{i, s_i \neq \omega_j} \log((1-p_i) r_i (1-r_i)) + \sum_{i=1}^{n} \log(1-p_i). \tag{8.18}$$

The last term does not depend on the class label j so we can reduce the discriminant function to

$$g_j(\boldsymbol{\chi}) = \log P(\omega_j) + \sum_{i, s_i = \omega_j} \log \frac{p_i}{1-p_i} + \sum_{i, s_i \neq \omega_j} \log((1-p_i) r_i (1-r_i)) = \tag{8.19}$$

$$\log P(\omega_j) + \sum_{i=1}^{n} d_{i,j} \log \frac{p_i}{1-p_i} + \sum_{i=1}^{n} (1-d_{i,j}) \log((1-p_i) r_i (1-r_i)) = \tag{8.20}$$

$$\log P(\omega_j) + \sum_{i=1}^{n} d_{i,j} \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)} + \sum_{i=1}^{n} \log((1-p_i) r_i (1-r_i)). \tag{8.21}$$

The last term of the summation is also independent from the class label j so it can be omitted. If we observe the equations

$$g_j(\boldsymbol{\chi}) = \log P(\omega_j) + \sum_{i=1}^{n} d_{i,j} \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)} \tag{8.22}$$

and

$$g_j(\boldsymbol{\chi}) = \sum_{i=1}^{n} w_i d_{i,j}, \tag{8.23}$$

we get that the weights

$$w_i \propto \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)} \tag{8.24}$$

are supposed to maximize the system accuracy. □

Author's publications

[1] **A. Hajdu** and I. Pitas, "Optimal approach for fast object-template matching," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2048–2057, 2007, IF: 2.462.

[2] **A. Hajdu** and I. Pitas, "Piecewise linear digital curve representation and compression using graph theory and a line segment alphabet," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 126–133, 2008, IF: 3.315.

[3] R. Qureshi, L. Kovács, B. Harangi, B. Nagy, T. Peto, and **A. Hajdu**, "Combining algorithms for automatic detection of optic disc and macula in fundus images," *Computer Vision and Image Understanding*, vol. 116, no. 1, pp. 138–145, 2012, IF: 1.232.

[4] B. Harangi and **A. Hajdu**, "Detection of the optic disc in fundus images by combining probability models," *Computers in Biology and Medicine*, vol. 65, pp. 10–24, 2015, IF: 1.240.

[5] **A. Hajdu**, L. Hajdu, A. Jonas, L. Kovács, and H. Tomán, "Generalizing the majority voting scheme to spatially constrained voting," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4182–4194, 2013, IF: 3.111.

[6] B. Antal and **A. Hajdu**, "Improving microaneurysm detection using an optimally selected subset of candidate extractors and preprocessing methods," *Pattern Recognition*, vol. 45, no. 1, pp. 264–270, 2012, IF: 2.632.

[7] B. Antal and **A. Hajdu**, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1720–1726, 2012, IF: 2.348.

[8] B. Antal and **A. Hajdu**, "Improving microaneurysm detection in color fundus images by using context-aware approaches," *Computerized Medical Imaging and Graphics*, vol. 37, no. 5–6, pp. 403–408, 2013, IF: 1.496.

[9] B. Antal and **A. Hajdu**, "An ensemble-based system for automatic screening of diabetic retinopathy," *Knowledge-Based Systems*, vol. 60, pp. 20–27, 2014, IF: 3.058.

[10] I. Lázár and **A. Hajdu**, "Retinal microaneurysm detection through local rotating cross-section profile analysis," *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 400–407, 2013, IF: 3.799.

[11] **A. Hajdu**, "Geometry of neighbourhood sequences," *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2597–2606, 2003, IF: 0.809.

[12] A. Fazekas, **A. Hajdu**, and L. Hajdu, "Metrical neighborhood sequences in zn," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2022–2032, 2005, IF: 1.138.

[13] **A. Hajdu**, L. Hajdu, and R. Tijdeman, "General neighborhood sequences in zn," *Discrete Applied Mathematics*, vol. 155, no. 18, pp. 2507–2522, 2007, IF: 0.625.

[14] **A. Hajdu** and T. Tóth, "Approximating non-metrical minkowski distances in 2d," *Pattern Recognition Letters*, vol. 29, no. 6, pp. 813–821, 2008, IF: 1.559.

[15] G. Kovács and **A. Hajdu**, "Translation invariance in the polynomial kernel space and its applications in knn classification," *Neural Processing Letters*, vol. 37, no. 2, pp. 207–233, 2013, IF: 1.240.

[16] B. Harangi and **A. Hajdu**, "Automatic exudate detection by fusing multiple active contours and regionwise classification," *Computers in Biology and Medicine*, vol. 54, pp. 156–171, 2014, IF: 1.475.

[17] **A. Hajdu**, A. Roubies, and I. Pitas, "Optimized chamfer matching for snake-based image contour representations," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2006, pp. 1017–1020.

[18] **A. Hajdu**, C. Giamas, and I. Pitas, "Object simplification using a skeleton-based weight function," in *IEEE International Symposium on Signals, Circuits and Systems (ISSCS)*, vol. 2, 2007, pp. 1–4.

[19] R. Harangozó, P. Veres, and **A. Hajdu**, "Subsampling strategies to improve learning-based retina vessel segmentation," in *IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 3349–3352.

[20] **A. Hajdu** and I. Pitas, "Compression optimized tracing of digital curves using graph theory," in *IEEE International Conference on Image Processing (ICIP)*, vol. 6, 2007, pp. 453–456.

[21] H. Tomán, **A. Hajdu**, J. Szakács, D. Hornyik, A. Csutak, and T. Pető, "Thickness-based binary morphological improvement of distorted digital line intersections," in *Fifth Hungarian Conference on Computer Graphics and Geometry*, 2010, pp. 133–139.

[22] B. Harangi, R. Qureshi, A. Csutak, T. Pető, and **A. Hajdu**, "Automatic detection of the optic disc using majority voting in a collection of optic disc detectors," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2010, pp. 1329–1332.

[23] B. Harangi and **A. Hajdu**, "Improving the accuracy of optic disc detection by finding maximal weighted clique of multiple candidates of individual detectors," in *9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012, pp. 602–605.

[24] H. Tomán, L. Kovács, . Jónás, L. Hajdu, and **A. Hajdu**, "A generalization of majority voting scheme for medical image detectors," in *Hybrid Artificial Intelligent Systems*, ser. Lecture Notes in Computer Science, E. Corchado, M. Kurzyński, and M. Woźniak, Eds., vol. 6679, Springer Berlin Heidelberg, 2011, pp. 189–196.

[25] H. Tomán, L. Kovács, A. Jónás, L. Hajdu, and **A. Hajdu**, "Generalized weighted majority voting with an application to algorithms having spatial output," in *Hybrid Artificial Intelligent Systems*, ser. Lecture Notes in Computer Science, E. Corchado, V. Snasel, A. Abraham, M. Wozniak, M. Grana, and S. Cho, Eds., vol. 7209, Springer Berlin Heidelberg, 2012, pp. 56–67.

[26] **A. Hajdu**, L. Hajdu, L. Kovács, and H. Tomán, "Diversity measures for majority voting in the spatial domain," in *Hybrid Artificial Intelligent Systems*, ser. Lecture Notes in Computer Science, J. Pan, M. Polycarpou, M. Woźniak, A. de Carvalho, H. Quintián, and E. Corchado, Eds., vol. 8073, Springer Berlin Heidelberg, 2013, pp. 314–323.

[27] B. Antal and **A. Hajdu**, "An ensemble-based microaneurysm detector for retinal images," in *IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 1621–1624.

[28] B. Antal, I. Lázár, and **A. Hajdu**, "An ensemble approach to improve microaneurysm candidate extraction," in *Communications in Computer and Information Science*, M. Obaidat, G. Tsihrintzis, and J. Filipe, Eds., vol. 222, Springer Verlag, 2012, ch. Signal Processing and Multimedia Applications, pp. 378–394.

[29] B. Antal and **A. Hajdu**, "Evaluation of preprocessing methods for microaneurysm detection," in *International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2013, pp. 723–726.

[30] B. Antal and **A. Hajdu**, "Improving microaneurysm detection in color fundus images by using an optimal combination of preprocessing methods and candidate extractors," in *European Signal Processing Conference (EUSIPCO)*, 2010, pp. 1224–1228.

[31] B. Antal, I. Lázár, **A. Hajdu**, Z. Török, A. Csutak, and T. Pető, "Evaluation of the grading performance of an ensemble-based microaneurysm detector," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2011, pp. 5943–5946.

[32] B. Antal, I. Lázár, and **A. Hajdu**, "An adaptive weighting approach for ensemble-based detection of microaneurysms in color fundus images," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 5955–5958.

[33] B. Antal, **A. Hajdu**, Z. Szabó-Maros, Z. Török, A. Csutak, and T. Pető, "A two-phase decision support framework for the automatic screening of digital fundus images," *Journal of Computational Science*, vol. 3, no. 5, pp. 262–268, 2012.

[34] B. Antal and **A. Hajdu**, "A prefiltering approach for an automatic screening system," in *IEEE International Symposium on Intelligent Signal Processing (WISP)*, 2009, pp. 265–268.

[35] G. Kovács and **A. Hajdu**, "Extraction of vascular system in retina images using averaged one-dependence estimators and orientation estimation in hidden markov random fields," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2011, pp. 693–696.

[36] B. Antal and **A. Hajdu**, "A stochastic approach to improve macula detection in retinal images," *Acta Cybernetica*, vol. 20, pp. 5–15, 2011.

[37] I. Lázár and **A. Hajdu**, "Microaneurysm detection in retinal images using a rotating cross-section based model," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2011, pp. 1405–1409.

[38] B. Nagy, B. Antal, B. Harangi, and **A. Hajdu**, "Ensemble-based exudate detection in color fundus images," in *International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2011, pp. 700–703.

[39] **A. Hajdu**, L. Hajdu, and R. Tijdeman, "Approximation of the euclidean distance by chamfer distances," *Acta Cybernetica*, vol. 20, no. 3, pp. 399–417, 2012.

[40] A. Roubies, **A. Hajdu**, and I. Pitas, "Improving concavity performance of snake algorithms," in *International Symposium on Control, Communications, and Signal Processing (ISCCSP)*, 2006, pp. 1–4.

[41] **A. Hajdu** and I. Pitas, "Content adaptive heterogeneous snakes," in *IEEE International Conference on Image Processing (ICIP)*, vol. 1, 2007, pp. 253–256.

[42] **A. Hajdu** and I. Pitas, "Tracing on heterogeneous grids to improve the concavity performance of snake algorithms," in *International Symposium on Signal Processing and Its Applications (ISSPA)*, 2007, pp. 1–4.

[43] I. Lázár and **A. Hajdu**, "Segmentation of vessels in retinal images based on directional height statistics," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 1458–1461.

[44] B. Harangi and **A. Hajdu**, "Improving automatic exudate detection based on the fusion of the results of multiple active contours," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2013, pp. 45–48.

[45] B. Harangi and **A. Hajdu**, "Detection of exudates in fundus images using a markovian segmentation model," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014, pp. 130–133.

[46] B. Harangi, I. Lázár, and **A. Hajdu**, "Automatic exudate detection using active contour model and regionwise classification," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 5951–5954.

[47] B. Harangi, B. Antal, and **A. Hajdu**, "Automatic exudate detection with improved naive bayes classifier," in *International Symposium on Computer-Based Medical Systems (CBMS)*, 2012, pp. 1–4.

[48] B. Nagy, B. Antal, and **A. Hajdu**, "Image database clustering to improve microaneurysm detection in color fundus images," in *International Symposium on Computer-Based Medical Systems (CBMS)*, 2012, pp. 1–6.

[49] B. Nagy, B. Antal, and **A. Hajdu**, "Image database clustering to improve exudate detection in color fundus images," in *International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2013, pp. 727–731.

[50] J. Tóth, L. Szakács, and **A. Hajdu**, "Finding the optimal parameter setting for an ensemble-based lesion detector," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 32 532–3536.

[51] J. Tóth, H. Tomán, and **A. Hajdu**, "Improving the performance of an ensemble-based exudate detection system using stochastic parameter optimization," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 1–4.

[52] **A. Hajdu**, C. Giamas, N. Vretos, and I. Pitas, "Metadata description of thermal videos for rescue operations," in *International Symposium on Signals, Circuits and Systems (ISSCS)*, vol. 2, 2007, pp. 1–4.

[53] **A. Hajdu**, T. Pető, A. Biró, R. Harangozó, J. Hülvely, Z. Török, and A. Csutak, "Extracting metadata from fundus images for the screening of diabetic retinopathy," in *IEEE International Symposium on Intelligent Signal Processing (WISP)*, 2009, pp. 259–263.

[54] Z. Török, T. Pető, E. Csősz, E. Tukacs, M. Molnár, M. Maros-Szabó, A. Berta, J. Tőzsér, **A. Hajdu**, V. Nagy, B. Domokos, and A. Csutak, "Tear fluid proteomics multimarkers for diabetic retinopathy screening," *BMC Ophthalmology*, vol. 13, no. 40, pp. 1–8, 2013, IF: 1.075.

[55] Z. Török, T. Pető, E. Csősz, E. Tukacs, M. Molnár, A. Berta, J. Tőzsér, **A. Hajdu**, V. Nagy, B. Domokos, and A. Csutak, "Combined methods for diabetic retinopathy screening, using retina photographs and tear fluid proteomics biomarkers," *Journal of Diabetes Research*, vol. Article ID 623619, pp. 1–8, 2014, IF: 3.536.

Bibliography

[56] L. Kuncheva, *Combining pattern classifiers. Methods and algorithms*. Wiley, 2004.

[57] C. van Rijsbergen, *Information Retrieval*, 2nd. London: Butterworths, 1979, p. 208.

[58] M. Niemeijer, B. van Ginneken, M. Cree, A. Mizutani, G. Quellec, C. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M. Abramoff, "Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 185–195, 2010.

[59] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, K. Heikki, and P. Juhani, "Diaretdb0: evaluation database and methodology for diabetic retinopathy algorithms," University of Kuopio, Finland, Tech. Rep., 2006.

[60] T. Kauppi, V. Kalesnykiene, J. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kalviainen, and J. Pietila, "Diaretdb1 diabetic retinopathy database and evaluation protocol," in *IEEE Conference on Medical Image Understanding and Analysis (MIUA)*, 2007, pp. 61–65.

[61] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[62] T. Walter and J. Klein, "Automatic detection of microaneurysms in color fundus images of the human retina by means of the bounding box closing," in *Medical Data Analysis*, ser. Lecture Notes in Computer Science, A. Colosimo, P. Sirabella, and A. Giuliani, Eds., vol. 2526, Springer Berlin Heidelberg, 2002, pp. 210–220.

[63] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, vol. IV, pp. 474–485, 1994.

[64] S. Ravishankar, A. Jain, and A. Mittal, "Automated feature extraction for early detection of diabetic retinopathy in fundus images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 210–217.

[65] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplar-based inpainting," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2003, pp. 721–728.

[66] A. Youssif, A. Ghalwash, and A. Ghoneim, "Comparative study of contrast enhancement and illumination equalization methods for retinal vasculature segmentation," in *Cairo International Biomedical Engineering Conference*, 2006.

[67] R. Gonzalez, R. Woods, and S. Eddins, *Digital image processing using MATLAB*. Gatesmark Publishing, 2009.

[68] T. Lin and Y. Zheng, "Adaptive image enhancement for retinal blood vessel segmentation," *Electronics Letters*, vol. 38, pp. 1090–1091, 2002.

[69] M. Lalonde, M. Beaulieu, and L. Gagnon, "Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching," *IEEE Transactions on Medical Imaging*, vol. 20, no. 11, pp. 1193–1200, 2001.

[70] A. Sopharak, K. Thet New, Y. Aye Moe, M. Dailey, and B. Uyyanonvara, "Automatic exudate detection with a naive bayes classifier," in *International Conference on Embedded Systems and Intelligent Technology*, 2008, pp. 139–142.

[71] M. Niemeijer, M. Abràmoff, and B. van Ginneken, "Fast detection of the optic disc and fovea in color fundus photographs," *Medical Image Analysis*, vol. 13, no. 6, pp. 859–870, 2009.

[72] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 951–958, 2003.

[73] X. Zhu and R. Rangayyan, "Detection of the optic disc in images of the retina using the hough transform," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2008, pp. 3546–3549.

[74] T. Petsatodis, A. Diamantis, and G. Syrcos, "A complete algorithm for automatic human recognition based on retina vascular network characteristics," in *International Scientific Conference*, 2004, pp. 41–46.

[75] S. Sekhar, W. Al-Nuaimy, and A. Nandi, "Automated localisation of retinal optic disk using hough transform," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2008, pp. 1577–1580.

[76] A. Fleming, S. Philip, K. Goatman, J. Olson, and P. Sharp, "Automated assessment of diabetic retinal image quality based on clarity and field definition," *Investigative Ophthalmology & Visual Science*, vol. 47, no. 3, pp. 1120–1125, 2006.

[77] F. Zana, I. Meunier, and J. Klein, "A region merging algorithm using mathematical morphology: application to macula detection," in *International Symposium on Mathematical Morphology and Its Applications to Image and Signal Processing (ISMM)*, ser. ISMM '98, Amsterdam, The Netherlands: Kluwer Academic Publishers, 1998, pp. 423–430.

[78] T. Walter, P. Massin, A. Arginay, R. Ordonez, C. Jeulin, and J. Klein, "Automatic detection of microaneurysms in color fundus images," *Medical Image Analysis*, vol. 11, pp. 555–566, 2007.

[79] T. Spencer, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus," *Computers and Biomedical Research*, vol. 29, no. 4, pp. 284–302, 1996.

[80] A. Fleming, S. Philip, and K. Goatman, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1223–1232, 2006.

[81] S. Abdelazeem, "Microaneurysm detection using vessels removal and circular hough transform," in *National Radio Science Conference*, 2002, pp. 421–426.

[82] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Detection of microaneurysms using multiscale correlation coefficients," *Pattern Recognition*, vol. 43, no. 6, pp. 2237–2248, 2010.

[83] H. Barrow, J. Tenenbaum, R. Bolles, and H. Wolf, "Parametric correspondence and chamfer matching: two new techniques for image matching," in *International Joint Conference on Artificial Intelligence (IJCAI) - Volume 2*, Cambridge, USA: Morgan Kaufmann Publishers Inc., 1977, pp. 659–663.

[84] D. Gavrila and V. Philomin, "Real-time object detection for smart vehicles," in *IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 1999, pp. 87–93.

[85] M. Kumar, P. Torr, and A. Zisserman, "Extending pictorial structures for object recognition," in *British Machine Vision Conference*, London, 2004, pp. 789–798.

[86] P. Felzenszwalb and D. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[87] T. Zhao and R. Nevatia, "Stochastic human segmentation from a static camera," in *Workshop on Motion and Video Computing*, Orlando, Florida, 2002, pp. 9–14.

[88] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 63–71, 2005.

[89] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[90] G. Borgefors, "Hierarchical chamfer matching: a parametric edge matching algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 849–865, 1988.

[91] J. You, W. Zhu, E. Pissaloux, and H. Cohen, "Hierarchical image matching: a chamfer matching algorithm using interesting points," in *Third Australian and New Zealand Conference on Intelligent Information Systems (ANZIIS)*, 1995, pp. 70–75.

[92] Q. Du, V. Faber, and M. Gunzburger, "Centroidal voronoi tessellations: applications and algorithms," *SIAM Review*, vol. 41, no. 4, pp. 637–676, 1999.

[93] Q. Du, M. Gunzburger, and L. Ju, "Constrained centroidal voronoi tessellations for surfaces," *SIAM Journal on Scientific Computing*, vol. 24, no. 5, pp. 1488–1506, 2003.

[94] P. Danielsson, "Euclidean distance mapping," *Computer Graphics and Image Processing*, vol. 14, no. 3, pp. 227–248, 1980.

[95] T. Schouten and E. Van Den Broek, "Fast exact euclidean distance (feed): a new class of adaptable distance transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2159–2172, 2014.

[96] A. Rosenfeld and J. Pfaltz, "Distance functions on digital pictures," *Pattern Recognition*, vol. 1, no. 1, pp. 33–61, 1968.

[97] G. Borgefors, "Distance transformations in digital images," *Computer Vision, Graphics, and Image Processing*, vol. 34, no. 3, pp. 344–371, 1986.

[98] K. Toyama and A. Blake, "Probabilistic tracking with exemplars in a metric space," *International Journal of Computer Vision*, vol. 48, no. 1, pp. 9–19, 2002.

[99] A. Pinz, M. Prantl, and H. Ganster, "A robust affine matching algorithm using an exponentially decreasing distance function," in *Journal of Universal Computer Science*, H. Maurer, C. Calude, and A. Salomaa, Eds., Springer Berlin Heidelberg, 1996, pp. 614–631.

[100] I. Pitas and A. Venetsanopoulos, *Nonlinear digital filters: Principles and applications.* Europe: Kluwer Academic, 1990.

[101] W. Rucklidge, "Locating objects using the hausdorff distance," in *IEEE International Conference on Computer Vision (ICCV)*, 1995, pp. 457–464.

[102] L. Ju, Q. Du, and M. Gunzburger, "Probabilistic methods for centroidal voronoi tessellations and their parallel implementations," *Parallel Computing*, vol. 28, no. 10, pp. 1477–1500, 2002.

[103] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[104] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Berkeley Symposium on Mathematical Statistics and Probability*, L. Cam and J. Neyman, Eds., vol. 1, 1967, pp. 281–297.

[105] Q. Du, M. Emelianenko, and L. Ju, "Convergence of the lloyd algorithm for computing centroidal voronoi tessellations," *SIAM Journal on Numerical Analysis*, vol. 44, no. 1, pp. 102–119, 2006.

[106] D. Pollard, "Strong consistency of k-means clustering," *Annals of Statistics*, vol. 9, no. 1, pp. 135–140, 1981.

[107] I Pitas, *Digital image processing algorithms.* Prentice Hall, 1993.

[108] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[109] C. Xu and J. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.

[110] C. Olson and D. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 103–113, 1997.

[111] C. Metz, "Basic principles of roc analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, 1978.

[112] T. Foures and P. Joly, "Multi-level model for 2d human motion analysis and description," *SPIE Electronic Imaging Science and Technology*, vol. 5018, pp. 61–71, 2003.

[113] Q. Ye, "The signed euclidean distance transform and its applications," in *International Conference on Pattern Recognition (ICPR)*, vol. 1, 1988, pp. 495–499.

[114] J. Serra, *Image analysis and mathematical morphology.* Orlando, FL, USA: Academic Press, Inc., 1983.

[115] L. Nyúl, A. Falcão, and J. Udupa, "Fuzzy-connected 3d image segmentation at interactive speeds," *Graphical Models*, vol. 64, no. 5, pp. 259–281, 2002.

[116] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *SPIE Medical Imaging*, J. Fitzpatrick and M. Sonka, Eds., vol. 5370, SPIE, 2004, pp. 648–656.

[117] A. Rosenfeld, "Arcs and curves in digital pictures," *Journal of ACM*, vol. 20, no. 1, pp. 81–87, 1973.

[118] R. Klette and A. Rosenfeld, "Digital straightness—a review," *Discrete Applied Mathematics*, vol. 139, no. 1–3, pp. 197–230, 2004.

[119] X. Huo and J. Chen, "Jbeam: multiscale curve coding via beamlets," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1665–1677, 2005.

[120] J. Fleischner and H. Fleischner, *Eulerian graphs and related topics. Part 1. Vol. 1.* Amsterdam: North-Holland Publishing Co., 1990.

[121] J. Fleischner and H. Fleischner, *Eulerian graphs and related topics. Part 1. Vol. 2.* Amsterdam: North-Holland Publishing Co., 1991.

[122] R. Klette and A. Rosenfeld, *Digital geometry - geometric methods for digital picture analysis.* Morgan Kaufmann, 2004.

[123] G. Klette, "Branch voxels and junctions in 3d skeletons," in *Combinatorial Image Analysis*, ser. Lecture Notes in Computer Science, R. Reulke, U. Eckardt, B. Flach, U. Knauer, and K. Polthier, Eds., vol. 4040, Springer Berlin Heidelberg, 2006, pp. 34–44.

[124] P. T., *Algorithms for graphics and image processing.* Rockville, MD: Computer Science Press, 1982.

[125] M. Fleury, "Deux problemes de geometrie de situation," *Journal de mathematiques elementaires*, pp. 257–261, 1883.

[126] S. Skiena, *Implementing discrete mathematics: combinatorics and graph theory with mathematica.* Reading, MA: Addison-Wesley, 1990.

[127] I. Debled-Rennesson and J. Reveillès, "A linear algorithm for segmentation of digital curves," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 9, no. 4, pp. 635–662, 1995.

[128] J. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Systems Journal*, vol. 4, no. 1, pp. 25–30, 1965.

[129] A. Katsaggelos, L. Kondi, F. Meier, J. Ostermann, and G. Schuster, "Mpeg-4 and rate-distortion-based shape-coding techniques," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1126–1154, 1998.

[130] J. Ostermann, "Methodologies used for evaluation of video tools and algorithms in mpeg-4," *Signal Processing: Image Communication*, vol. 9, no. 4, pp. 343–365, 1997.

[131] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to algorithms. Second edition.* 2nd revise. MIT Press and McGraw-Hill, 2001.

[132] M. Kwan, "Graphic programming using odd or even points," *Chinese Mathematics*, vol. 1, pp. 273–277, 1962.

[133] M. Guan, "A survey on the chinese postman problem," *Journal of Mathematical Research & Exposition*, vol. 4, no. 1, pp. 113–119, 1984.

[134] R. Floyd, "Algorithm 97: shortest path," *Communication of the ACM*, vol. 5, no. 6, pp. 345–, 1962.

[135] E. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[136] H. Gou and M. Wu, "Fingerprinting curves," in *Digital Watermarking*, ser. Lecture Notes in Computer Science, I. Cox, T. Kalker, and H. Lee, Eds., vol. 3304, Springer Berlin Heidelberg, 2005, pp. 13–28.

[137] V. Solachidis and I. Pitas, "Watermarking polygonal lines using fourier descriptors," *IEEE Computer Graphics and Applications*, vol. 24, no. 3, pp. 44–51, 2004.

[138] E. Saund, "Finding perceptually closed paths in sketches and drawings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 4, pp. 475–491, 2003.

[139] D. Zhong and H. Yan, "Pattern skeletonization using run-length-wise processing for intersection distortion problem," *Pattern Recognition Letters*, vol. 20, no. 8, pp. 833–846, 1999.

[140] R. Al Ajlouni, "The use of digital pattern recognition techniques for virtual reconstruction of eroded and visually complicated archeological geometric patterns," in *ISPRS08*, 2008, B5: 193–198.

[141] E. Deutsch, "Thinning algorithms on rectangular, hexagonal, and triangular arrays," *Communications of the ACM*, vol. 15, no. 9, pp. 827–837, 1972.

[142] C. Sinthanayothin, J. Boyce, T. Williamson, H. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Medicine*, vol. 19, no. 2, pp. 105–112, 2002.

[143] M. Niemeijer, M. Abramoff, and B. van Ginneken, "Segmentation of the optic disc, macula and vascular arch in fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 116–127, 2007.

[144] F. Mendels, C. Heneghan, and J. Thiran, "Identification of the optic disk boundary in retinal images using active contours," in *Irish Machine Vision and Image Processing Conference (IMVIP)*, ser. Cerebrovascular Diseases, IEEE, 1999, pp. 103–115.

[145] M. Abramoff and M. Niemeijer, "The automatic detection of the optic disc location in retinal images using optic disc location regression," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2006, pp. 4432–4435.

[146] W. E., "Smallest enclosing disks (balls and ellipsoids)," in *Results and New Trends in Computer Science*, Springer-Verlag, 1991, pp. 359–370.

[147] C. W.G., "Problems arising in the analysis of a series of similar experiments," *Journal of the Royal Statistical Society*, vol. 4, pp. 102–118, 1937.

[148] D. Kumlander, "A new exact algorithm for the maximum-weight clique problem based on a heuristic vertex-coloring and a backtrack search," in *International Conference on Engineering Computational Technology*, Civil-Comp Press, 2004, pp. 137–138.

[149] R. Clemen and R. Winkler, "Aggregating probability distributions," in *Advances in Decision analysis*, W. Edwards, R. Miles, and D. von Winterfeldt, Eds., Cambridge University Press, 2007, pp. 154–176.

[150] R. Jacobs, "Methods for combining experts' probability assessments," *Neural Computation*, vol. 7, no. 5, pp. 867–888, 1995.

[151] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[152] M. Stone, "The opinion pool," *The Annals of Mathematical Statistics*, vol. 32, no. 4, pp. 1339–1342, 1961.

[153] P. Morris, "Decision analysis expert use," *Management Science*, vol. 20, no. 9, pp. 1233–1241, 1974.

[154] P. Morris, "Combining expert judgments: a bayesian approach," *Management Science*, vol. 23, no. 7, pp. 679–693, 1977.

[155] K. Pearson, "Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.

[156] H. Davies and P. Lam, *Managerial economics: An analysis of business issues*. Financial Times/Prentice Hall, 2001.

[157] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.

[158] D. Chickering, "Learning bayesian networks is np-complete," in *Learning from data*, ser. Lecture Notes in Statistics, D. Fisher and H.-J. Lenz, Eds., vol. 112, Springer New York, 1996, pp. 121–130.

[159] H. Zhang, L. Jiang, and J. Su, "Hidden naive bayes," in *20th National Conference on Artificial Intelligence - Volume 2*, 2005, pp. 919–924.

[160] L. Shijian, "Accurate and efficient optic disc detection and segmentation by a circular transformation," *IEEE Transactions on Medical Imaging*, vol. 30, no. 12, pp. 2126–2133, 2011.

[161] J. Xu, O. Chutatape, and P. Chew, "Automated optic disk boundary detection by modified active contour model," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 3, pp. 473–482, 2007.

[162] H. Li and O. Chutatape, "Boundary detection of optic disk by a modified asm method," *Pattern Recognition*, vol. 36, no. 9, pp. 2093–2104, 2003.

[163] L. Kuncheva, C. Whitaker, C. Shipp, and R. Duin, "Limits on the majority vote accuracy in classifier fusion," *Pattern Analysis & Applications*, vol. 6, no. 1, pp. 22–31, 2003.

[164] L. Lam and C. Suen, "Application of majority voting to pattern recognition: an analysis of its behavior and performance," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 27, no. 5, pp. 553–568, 1997.

[165] D. Gilat, "Monotonicity of a power function: an elementary probabilistic proof," *The American Statistician*, vol. 31, no. 2, pp. 91–93, 1977.

[166] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

[167] H. Altincay, "On naive bayesian fusion of dependent classifiers," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2463–2473, 2005.

[168] L. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.

[169] X. Wang and N. Davidson, "The upper and lower bounds of the prediction accuracies of ensemble methods for binary classification," in *International Conference on Machine Learning and Applications (ICMLA)*, 2010, pp. 373–378.

[170] O. Matan, "On voting ensembles of classifiers," in *Workshop on Integrating Multiple Learned Models (AAAI)*, 1996, pp. 84–88.

[171] M. Appel and R. Russo, "On the h-diameter of a random point set," The University of Iowa, Tech. Rep. 370, 2008.

[172] M. Appel, C. Najim, and R. Russo, "Limit laws for the diameter of a random point set," *Advances in Applied Probability*, vol. 34, no. 1, pp. 1–10, 2002.

[173] M. Niemeijer, M. Abràmoff, and B. van Ginneken, "Fast detection of the optic disc and fovea in color fundus photographs," *Medical Image Analysis*, vol. 13, no. 6, pp. 859–870, 2009.

[174] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Information Fusion*, vol. 6, no. 1, pp. 63–81, 2005.

[175] (). National screening programme for diabetic retinopathy. `http://www.retinalscreening.nhs.uk/`.

[176] B. Lay, "Analyse automatique des images angiofluorographiques au cours de la retinopathie diabetique," PhD thesis, Centre of Mathematical Morphology, Paris School of Mines, 1983.

[177] C. Baudoin, B. Lay, and J. Klein, "Automatic detection of microaneurysms in diabetic fluorescein angiographies," *Revue D'Epidemiologie et de Sante Publique*, vol. 32, pp. 254–261, 1984.

[178] G. Oien and O. P., "Diabetic retinopathy: automatic detection of early symptoms from retinal images," in *Norwegian Signal Processing Symposium*, 1995.

[179] A. Frame, P. Undrill, M. Cree, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms," *Computers in Biology and Medicine*, vol. 28, pp. 225–238, 1998.

[180] A. Mendonca, A. Campilho, and J. Nunes, "Automatic segmentation of microaneurysms in retinal angiograms of diabetic patients," in *International Conference on Image Analysis and Processing*, 1999, pp. 728–733.

[181] J. Hipwell, F. Strachan, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool," *Diabetic Medicine*, vol. 17, pp. 588–594, 2000.

[182] G. Yang, L. Gagnon, S. Wang, and M.-C. Boucher, "Algorithm for detecting micro-aneurysms in low resolution color retinal images," in *Proceedings of the Vision Interface*, 2001, pp. 265–271.

[183] M. Cree, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "A fully automated comparative microaneurysm digital detection system," *Eye*, vol. 11, pp. 622–628, 1997.

[184] L. Streeter and M. Cree, "Microaneurysm detection in colour fundus images," in *Proceedings of the Image and Vision Computing New Zealand*, 2003.

[185] M. Niemeijer, B. van Ginneken, J. Staal, M. Suttorp-Schulten, and M. Abramoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 584–592, 2005.

[186] J. Nayak, P. Bhat, U. Acharya, C. Lim, and M. Kagathi, "Automated identification of diabetic retinopathy stages using digital fundus images," *Journal of Medical Systems*, vol. 32, no. 2, pp. 107–115, 2008.

[187] M. Langroudi and H. Sadjedi, "A new method for automatic detection and diagnosis of retinopathy diseases in colour fundus images based on morphology," in *International Conference on Bioinformatics and Biomedical Technology*, 2010, pp. 134–138.

[188] G. Kande, T. Savithri, and P. Subbaiah, "Automatic detection of microaneurysms and hemorrhages in digital fundus images," *Journal of Digital Imaging*, vol. 23, no. 4, pp. 430–437, 2010.

[189] C. Marino, E. Ares, M. Penedo, M. Ortega, N. Barreira, and F. Gomez-Ulla, "Automated three stage red lesions detection in digital color fundus images," *WSEAS Transactions on Computers*, vol. 7, no. 4, pp. 207–215, 2008.

[190] A. Bhalerao, A. Patanaik, S. Anand, and P. Saravanan, "Robust detection of microaneurysms for sight threatening retinopathy screening," in *Indian Conference on Computer Vision, Graphics, and Image Processing*, 2008, pp. 520–527.

[191] G. Quellec, M. Lamard, P. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1230–1241, 2008.

[192] C. Hann, J. Revie, D. Hewett, J. Chase, and G. Shaw, "Screening for diabetic retinopathy using computer vision and physiological markers," *Journal of Diabetes Science and Technology*, vol. 3, pp. 819–834, 2009.

[193] A. Mizutani, C. Muramatsua, Y. Hatanakab, S. Suemoria, T. Haraa, and H. Fujita, "Automated microaneurysm detection method based on double-ring filter in retinal fundus images," in *Proceedings of the SPIE Medical Imaging*, vol. 7260, 2009, pp. 1–8.

[194] L. Giancardo, F. Meriaudeau, T. Karnowski, K. Tobin, Y. Li, and E. Chaum, "Microaneurysms detection with the radon cliff operator in retinal fundus images," in *Proceedings of the SPIE Medical Imaging*, vol. 7623, 2010.

[195] E. Grisan and A. Ruggeri, "Segmentation of candidate dark lesions in fundus images based on local thresholding and pixel density," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2007, pp. 6735–6738.

[196] S. Balasubramanian, S. Pradhan, and V. Chandrasekaran, "Red lesions detection in digital fundus images," in *IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 2932–2935.

[197] C. Sinthanayothin, J. Boyce, T. Williamson, H. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Medicine*, vol. 19, pp. 105–112, 2002.

[198] D. Usher, M. Dumskyj, M. Himaga, T. Williamson, S. Nussey, and J. Boyce, "Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening," *Diabetic Medicine*, vol. 21, pp. 84–90, 2004.

[199] P. Pallawala, W. Hsu, M. L. Lee, and S. S. Goh, "Automated microaneurysm segmentation and detection using generalized eigenvectors," in *IEEE Workshops on Application of Computer Vision (WACV/MOTIONS) Volume 1*, vol. 1, 2005, pp. 322–327.

[200] G. Gardner, D. Keating, T. Williamson, and A. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool," *British Journal of Ophthalmology*, vol. 80, no. 11, pp. 940–944, 1996.

[201] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[202] D. Chakraborty, "Clinical relevance of the ROC and free-response paradigms for comparing imaging system efficacies," *Radiation Protection Dosimetry*, vol. 139, no. 1-3, pp. 37–41, 2010.

[203] J. Eng. (). Roc analysis: web-based calculator for roc curves. `http://www.jrocfit.org` Downloaded on 07/11/2012., Johns Hopkins University, Baltimore.

[204] M. Abramoff, J. Reinhardt, S. Russell, J. Folk, V. Mahajan, M. Niemeijer, and G. Quellec, "Automated early detection of diabetic retinopathy," *Ophthalmology*, vol. 117, no. 6, pp. 1147–1154, 2010.

[205] @lasthash"Retinal photography screening for diabetic eye disease," British Diabetic Association, Tech. Rep., 1997.

[206] C. Agurto, E. S. Barriga, V. Murray, S. Nemeth, R. Crammer, W. Bauman, G. Zamora, M. S. Pattichis, and P. Soliz, "Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 8, pp. 5862–5871, 2011.

[207] H. Jelinek, M. Cree, D. Worsley, A. Luckie, and P. Nixon, "An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice.," *Clinical and Experimental Optometry*, vol. 89, no. 5, pp. 299–305, 2006.

[208] A. Fleming, K. Goatman, S. Philip, G. Prescott, P. Sharp, and J. Olson, "Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts," *British Journal of Ophthalmology*, vol. 94, no. 12, pp. 1606–1610, 2010.

[209] M. Abramoff, M. Niemeijer, M. Suttorp-Schulten, M. Viergever, S. Russel, and B. van Ginneken, "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes," *Diabetes Care*, vol. 31, pp. 193–198, 2008.

[210] A. Fleming, S. Philip, K. Goatman, G. Prescott, P. Sharp, and J. Olson, "The evidence for automated grading in diabetic retinopathy screening.," *Current Diabetes Reviews*, vol. 7, no. 4, pp. 246–252, 2011.

[211] G. Scotland, P. McNamee, A. Fleming, K. Goatman, S. Philip, G. Prescott, P. Sharp, G. Williams, W. Wykes, G. Leese, J. Olson, and S. D. R. C. R. Network, "Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy.," *British Journal of Ophthalmology*, vol. 94, no. 6, pp. 712–719, 2010.

[212] M. Abramoff, M. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[213] A. Fleming, K. Goatman, S. Philip, G. Williams, G. Prescott, G. Scotland, P. McNamee, G. Leese, W. Wykes, P. Sharp, J. Olson, and S. D. R. C. R. Network, "The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy.," *British Journal of Ophthalmology*, vol. 94, no. 6, pp. 706–711, 2010.

[214] S. Philip, A. Fleming, K. Goatman, S. Fonseca, P. Mcnamee, G. Scotland, G. Prescott, P. Sharp, and J. Olson, "The efficacy of automated disease/no disease grading for diabetic retinopathy in a systematic screening programme," *British Journal of Ophthalmology*, vol. 91, no. 11, pp. 1512–1517, 2007.

[215] D. West, P. Mangiameli, R. Rampal, and V. West, "Ensemble strategies for a medical diagnostic decision support system: a breast cancer diagnosis application," *European Journal of Operational Research*, vol. 162, no. 2, pp. 532–551, 2005.

[216] H. Moon, H. Ahn, R. Kodell, S. Baek, C.-J. Lin, and J. Chen, "Ensemble methods for classification of patients for personalized medicine with high-dimensional data," *Artificial Intelligence in Medicine*, vol. 41, pp. 197–201, 2007.

[217] J.-H. Eom, S.-C. Kim, and B.-T. Zhang, "Aptacdss-e: a classifier ensemble-based clinical decision support system for cardiovascular disease level prediction," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2465–2479, 2008.

[218] S. Doan, N. Collier, H. Xu, P. Duy, and T. Phuong, "Recognition of medication information from discharge summaries using ensembles of classifiers," *BMC Medical Informatics and Decision Making*, vol. 12, no. 1, pp. 1–10, 2012.

[219] C. Agurto, V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, S. Russell, M. Abramoff, and P. Soliz, "Multiscale AM-FM methods for diabetic retinopathy lesion detection," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 502–512, 2010.