Hajdú András „Diszkrét geometriai és fúziós módszerek objektumok detektálásához és döntéstámogatáshoz (Discrete Geometric and Fusion Based Techniques for Object Detection and Decision Support)" című MTA doktori értekezésének bírálata

A disszertáció a képfeldolgozás és alakfelismerés témaköréhez tartozó területeken ért el új eredményeket, egyrészt az alacsony szintű műveleteket tárgyalva, mint a mintavételezés, tömörítés, másrészt pedig a magas szintű alakfelismerési és osztályozási technikákhoz kapcsolódó együttes (ensemble), illetve fúziós módszereket továbbfejlesztve. A fúziós módszerek, amelyek az utóbbi időben mind népszerűbbek képfeldolgozásban, szakítanak a korábbi feldolgozási csővezeték szemlélettel, és több, lehetőleg ortogonális eljárás eredményeit próbálják meg optimálisan kombinálni.

A dolgozat egy bevezető fejezetből és hat, az új eredményeket taglaló fejezetből áll. A bevezető fejezet célirányosan és formálisan összefoglalja az információ fúzió alapjait, a képfeldolgozási eljárások minőségét kifejező mértékeket, a felhasznált adatbázisokat, és a később tárgyalt fúzió építőelemeit. Az olvasónak különösen hasznos a szemvizsgálat célját és terminológiáját leíró a rövid klinikai összefoglaló, amely a dolgozat eredményeinek alapvető alkalmazási területe.

A második fejezet (1.1. tézis), amely az új eredmények tekintetében az első érdemi fejezet, a gyors sablon (template) illesztéssel foglalkozik. A chamfer illesztés két bináris kép különbözőségének (két halmaz Hausdorff távolságának) egy gyakorlati meghatározási módszere, amely úgy gyorsítható, ha a határból csak kevés számú mintapontot hagyunk meg. Az egyéb korlátozás nélküli megoldás ismert, a Voronoi felbontásra alapul. Az 1.1. tézis arra ad megoldást, hogy a mintapontoknak mi az optimális kiválasztási stratégiája, ha az illesztésnek csak egy korlátozott tartományban kell pontosnak lennie. A stratégiát egy Lloyd relaxációs változat valósítja meg. A Jelölt a módszert mind görbékre, mint pedig területekre kiterjeszti, a módszer tulajdonságait elméletileg és kísérletekkel is vizsgálja, jóságát igazolja, és több példát ad a gyakorlati alkalmazásra (humán póz és végtag detektálás, érszegmentálás). Ebben a feladatkörben a mintapontok kiválasztásának specialitása, hogy a cél a távolságmező hű visszaadása. Ennek ellenére megkérdezem, hogy a Jelölt milyen kapcsolódási pontokat lát általánosabb, magasabb dimenziós mintavételi problémákkal, amikor egy ponthalmaz pontjait kell egyenletesen sűrűn, vagy valamilyen valószínűségsűrűséget követve véges számú ponttal mintavételezni, és a cél a diszkrepancia, illetve a Kolgomorov-Szmirnov távolság minimalizálása.

A harmadik fejezet (1.2. tézis) a diszkrét geometriájú, azaz 2D-rácson definiált görbék tömörítésével foglalkozik. A geometriai egyszerű görbéknek (amelyek önmagukat nem metszik és egy darabból állnak), a diszkrét geometriában olyan pixelsorozat felel meg, amelyben minden pixel reguláris, azaz csak két szomszédos pixelben van görbe pont, kivéve a görbe kezdetét és végét, amelynek csak egy-egy szomszédja van. A kezdő, vég és nem reguláris pontokat, reguláris pontokból álló szegmensek kötik össze, így az általános diszkrét görbéhez egy gráf rendelhető, ahol a csomópontok egy-egy nem reguláris pontot, az élek pedig reguláris pontokat tartalmazó szegmensek. Ezzel a gráfelmélet eszköztára felhasználhatóvá válik diszkrét geometriájú görbék kezelésére, amelyből az Euler utak létezése és előállítása kölcsönvehető. Ehhez a Jelölt egyrészt egy geometriai kiegészítést tett, amely célnak tekinti az utak „simaságát" is, ha több Euler út is lehetséges. Másrészt, a fejezet végén a görbék eredeti vastagságát is figyelembe veszi a folytatás kiválasztásánál. A másik eredmény a görbeszegmensek optimális tömörítését éri el. A cél olyan kódkészlet, azaz szomszédos pixelsorozat minták meghatározása, amelyekből a görbeszegmensek minimális biten kódolhatók. Kérdésem, hogy a kódkészlet kialakítása a tömörítendő görbétől függetlenül történik-e, és ha igen, akkor nem volna-e hatékonyabb annak tulajdonságait figyelembe venni a kódkészlet kialakításánál? A másik kérdés a Bresenham algoritmussal kapcsolatos megjegyzésre vonatkozik. Mi a probléma ezzel, amikor a Bresenham és ekvivalens társai, mint pl. az aszimmetrikus DDA az euklideszi egyeneshez legközelebbi egy pixel vastag szakaszt állítja elő?

A 4. fejezet (2.1. tézis) a szem vakfoltjának/látóidegfőjének (optical disc) és sárgafoltjának (macula) detektálásához alkalmazott fúziós stratégiát tárgyalja. A fuzionálandó eljárások az irodalomban ismertek, a cél ezek erősítése a módszerek adott sugarú körbe eső jelöltek száma alapján végzett többségi szavazással, valamint a vakfolt és sárgafolt egyidejű detekciójával, felhasználva az előre ismert távolságot. Megjegyzem a „mutual information” kifejezés használata szerencsétlen, ugyanis az az információelméletben egy valószínűségi mérték, és a viszonylag egyszerű módszer „gráfelméleti” jelzőjét is túlzásnak tartom. Igazán értékes viszont a kiterjesztés, amely a kombinált módszerek több jelöltjét is felhasználja, és a geometriai kényszert a maximális súlyú klikkek meghatározására vezeti vissza. A probléma NP-teljes, de itt egy közelítő eljárás is tökéletesen megfelelő. A középpont pontos helyének kiválasztására több eljárást is tárgyal a dolgozat, amelyek nem robusztusak, így az outlier-ek hibás döntéshez vezethetnek. A szórásból és kovarianciából számított súlyozott átlag eléggé magától értetődő választás. Tudományos újdonságnak inkább a konfidenciák aggregációját tartom, valamint a javasolt Bayes modellt.

Az 5. fejezet (2.2. tézis) a többségi szavazás egy robusztus általánosítása, amely a geometriai kényszereket a szavazási stratégiában is felhasználja - nevezetesen több eredmény akkor szavazhat együtt, ha közel vannak egymáshoz - így akár kisebbségi, de a kényszert kielégítő vélemény is győzhet. A fejezet nagy része elméleti, amikor a Jelölt a fúzióval elérhető pontosságnövekedést vizsgálja a korábbiaknál általánosabb feltételek mellett, és az elméleti eredményeket bizonyítással majd szimulációval is igazolja. A képfeldolgozási cél továbbra is vakfolt és sárgafolt meghatározás. Megjegyzem, hogy véletlenszerűen ejtett pontok adott körlapra kerülésének esélye a geometriai valószínűség eszközeivel is tárgyalható lenne, sőt érdemes volna az integrálgeometria eszköztárát is felhasználni, ugyanis az sokkal általánosabb problémákra is ad megoldásokat. A fejezet végén egy további általánosítás jelenik meg, amely a szavazókhoz súlyozást rendel és figyelembe veszi azok diverzitását is.

A 6. fejezet (2.3. tézis) a mikroaneurizmák fúziós elvű detektálásával foglalkozik, amelyekből több is lehet egyetlen képen. Az alapgondolat az, hogy az előfeldolgozó és jelöltállító algoritmusokból hatékony párokat érdemes formálni. Egy-egy pár jóságát a rendelkezésre álló képadatbázisokon végrehajtott próbával lehet meghatározni. Az alapgondolat jó, azonban a javasolt szimulációs hűtés alapú megvalósítást a feladat szükségtelen elbonyolításának tartom a következő okok miatt. Ezt a feladatot egyszer kell megoldani, így nem időkritikus. Az előfeldolgozó és jelöltállító algoritmusok száma korlátos, tehát minden lehetőség kipróbálásának sincs akadálya, a képezhető együttesek pedig szűkíthetők volnának heurisztikával. Végül, a szimulációs lehűtés előnye akkor jelenne meg, ha egy folytonos halmazon optimalizálunk, ahol a célfüggvény résztartományonként folytonos, tehát ha aktuálisan egy alacsony értékünk van, akkor remélhetjük, hogy az optimum közelében vagyunk. Ezek a feltételek itt nem állnak fenn, így a szimulált lehűtésnek nincs előnye ahhoz képest, hogy a keresési térben véletlenszerűen választott konfigurációk közül vesszük a legjobbat. A fejezetben felvetett kontextus-alapú módszert viszont értékesnek tartom.

A 7. fejezet (3.1. tézis) egyrészt a korábbi eredményeket egy automatikus szemvizsgáló rendszerben foglalja össze, másrészt az ezen a szinten alkalmazandó döntési stratégiára ad javaslatot. A döntési stratégia alapgondolata hasonló 2.3. tézisben alkalmazottnál, de annál kimunkáltabb. Az elv az, hogy egy adatbázison végzett elemzés alapján az együttest lépésenként újabb osztályozó hozzá vételével építjük fel, majd egyenként kivesszük belőle azokat, amelyek szükségtelenek. Végül a 3.2. tézis a Jelölt ebben az integrált rendszerben felhasznált korábbi eredményeit foglalja össze, amelyeket a dolgozat hivatkozik, de nem fejt ki részletesen.

Összefoglalva, a dolgozat számos értékes képfeldolgozási/osztályozási elméleti eredményt közöl, formális matematikai bizonyítással, illetve szimulációkkal igazolja azok helyességét, és megmutatja a gyakorlati alkalmazásukat. A 2. és 3. tézis eredményei előzetes feltevéseikben és mindenekelőtt alkalmazási területükben erősen kötődnek az automatikus szemvizsgálathoz, bár egyes módszerek más területen is használhatók lennének. Az eredményeket a Jelölt a témakör rangos fórumain, legnagyobb impaktú folyóirataiban publikálta. A publikációk rangja, száma és a hivatkozások sokasága is tiszteletre méltó. Az eredményeket tudomány állását előrevivőnek, igazoltan érvényesnek és a gyakorlatban alkalmazhatónak, illetve alkalmazottnak tekintem. A téziseket elfogadom az alábbi megjegyzésekkel: Jó lett volna a 2. és 3. tézisben is hangsúlyozni az elsődleges alkalmazási területet (szemvizsgálat), mert így a tézisek szövege kicsit mást, általánosabbat ígérnek, mint amit a dolgozat konkrétumaival nyújt. A 3.2. tézispontot a Jelölt munkájaként elismerem, de mivel a dolgozatban ezek az eredmények nincsenek részletesen kifejtve, külön tézispontként való szerepeltetését indokolatlannak tartom.

A mű jól szerkesztett, színvonalas munka. **A dolgozatot eredményeit elegendőnek tartom az MTA doktori cím megszerzéséhez, a nyilvános védés kitűzését javaslom, a fokozat odaítélését támogatom.**

Budapest, 2016. augusztus 4.

Szirmay-Kalos László

Dr. Szirmay-Kalos László
MTA doktora, egyetemi tanár