Régiók és alakzatok: Vizuális objektumok kinyerése és illesztése
(From Regions to Shapes: The Extraction and Alignment of Visual Objects)

Válasz Abonyi Jánosnak

az MTA doktori értekezés opponensének

Kató Zoltán

2014. május 16.

Mindenekelőtt köszönöm Prof. Abonyi Jánosnak szakértő, gondos és alapos bírálatát. A bírálat megjegyzéseire és kérdéseire az alábbiakban adok részletes választ.

Kérdésekre, megjegyzésekre adott válaszok

Meglepő, hogy a 2013-ban elkészített értekezés 207 hivatkozása között 2010-nél frissebb dátumú alig szerepel (az önhivatkozások nélkül csak egy hivatkozás származik 2012-ből), a hivatkozások döntő többsége az 2000-es évekből való. Miért nincsenek friss irodalmi hivatkozások, az elmúlt időszak eredményei miként támaszkodtak a javasolt megközelítésmódokra, milyen más kapcsolódó eljárások váltak sikeressé és széles körben alkalmazottá?

Maga az értekezés döntő részben 2012 nyarán készült, hiszen az akkor érvényben lévő eljárásrend szerint az MTMT adataimat már 2012 decemberében zárolni kellett a 2013 marciusi benyújtáshoz. Így csak elvétve kerülhettek be 2011 utáni hivatkozások.

A kérdés második felére válaszolva, talán a legjelentősebb megközelítésmód, ami az utóbbi időben előtérbe került szakterületemen, az a gráfvágáson alapuló módszerek és az ehhez kapcsolódó úgynevezett CRF (*Conditional Random Field*) modellezés, amely az MRF modellek egy speciális esetének tekinthető. Ezeket a módszereket részletesebben tárgyaltuk a 2012-ben megjelent monográfiában [32], így itt csak röviden összefoglalom a terület legújabb kutatási irányait. A gráfvágáson, pontosabban a gráfelméletből jól ismert *max-flow/min-cut* algoritmusokon alapuló energiaminimalizációs módszerek a 2000-es évek elején jelentek meg és terjedtek el az évtized végére. A módszer az utóbbi 5 évben áttörést jelentett a Markov modellek energiafüggvényének minimalizálásában, melynek az addig egyeduralkodó módszere a szimulált hűtés volt. A megközelítés lényege, hogy bizonyos tipusú Markov modellek energiája súlyozott gráfon reprezentálható úgy, hogy az energiafüggvény globális minimuma megegyezik a gráf minimális vágásának költségével, és a vágás eredményeként előálló részgráfok egyértelműen megadják az ehhez tartozó címkézést. A feltételek látszólag erős megkötést jelentenek: kizárólag bináris címkézés és páronkénti potenciálok

használhatóak, ráadásul a potenciál csak a jól ismert homogenitási kritériumhoz hasonló, ú.n. szubmoduláris tulajdonságú lehet. Mégis nagyon sok MRF modell eleget tesz ezeknek a feltételeknek, és így az optimális címkézés garantáltan megtalálható, ráadásul rendkívül alacsony számításigény mellett. Érthető, hogy ennek hatására ma az MRF modellek újra az érdeklődés középpontjába kerültek.

Az utóbbi évek fő kutatási irányai egyrészt a fenti megkötések lazítására irányulnak. Így megjelentek a több címkével is működő energiaminimalizáló módszerek, majd a legutóbbi években az ú.n. magasabb rendű Markov mezőkre (vagyis amelyek nincsnek megszorítva csak páronkénti interakcióra) fókuszáltak a kutatások. Megjegyzem, hogy ezek az új módszerek sokszor már nem is a klasszikus gráfvágáson alapulnak (például az utóbbi években népszerűek lettek az ú.n. duális dekompozíciós módszerek). A másik fő irány a páronkénti interakció megtartása mellett inkább a korábban használt generatív (pl. az értekezésemben is gyakran használt Gauss keverék) modell helyett a több pixel együttes feltételes valószínűségén alapuló CRF modelleket alkalmazza, amelyhez tanulóalgoritmusokkal határozzák meg az eloszlás-paramétereket. Ebben a témakörben több cikk is megjelent az utóbbi 2 évben:

1. Meng Tang, Lena Gorelick, Olga Veksler, Yuri Boykov: GrabCut in One Cut. ICCV 2013: 1769-1776

2. Lubor Ladicky, Christopher Russell, Pushmeet Kohli, Philip H. S. Torr: Inference Methods for CRFs with Co-occurrence Statistics. International Journal of Computer Vision 103(2): 213-225 (2013)

3. Alexander Fix, Thorsten Joachims, Sung Min Park, Ramin Zabih: Structured Learning of Sum-of-Submodular Higher Order Energy Functions. ICCV 2013: 3104-3111

További érdekes fejlemény, hogy megjelentek a gráfok és a Markov modellek a regisztrációban is. Ezek jellemzően nem-parametrikus deformáció meghatározására alkalmasak, ahol a címkék jelentése az adott pozíció elmozdulását jelenti. Ha ezt egy vektorként képzeljük el, akkor a regisztráció egy vektormező segítségével, pontonként megadott transzformáció meghatározását jelenti. A modell hátránya, hogy csak diszkrét elmozdulás-értékeket kezel, és a diszkretizálás finomsága határozza meg a címkék számát, ami viszont a végső minimalizáló algoritmus komplexitását befolyásolja. A korábbi módszerekhez képest azonban még így is nagyon gyors algoritmusokat kaphatunk. Az utóbbi években egyre több ilyen alkalmazás jelent meg, különösen az orvosi képek regisztrációjában:

1. Lorenzo Torresani, Vladimir Kolmogorov, Carsten Rother: A Dual Decomposition Approach to Feature Correspondence. IEEE Trans. Pattern Anal. Mach. Intell. 35(2): 259-271 (2013)

2. Enzo Ferrante, Nikos Paragios: Non-rigid 2D-3D Medical Image Registration Using Markov Random Fields. MICCAI (3) 2013: 163-170

3. Yun Zeng, Chaohui Wang, Xianfeng Gu, Dimitris Samaras, Nikos Paragios: A Generic Deformation Model for Dense Non-rigid Surface Registration: A Higher-Order MRF-Based Approach. ICCV 2013: 3360-3367

Természetesen az értekezésemben bemutatott modellek egy része is átalakítható a gráfvágás követelményeinek megfelelő MRF modellé, és így a számításidők drasztikusan csökkenthetők. Az utóbbi 4 évben két ilyen eredményünk is megjelent:

1. Milan Lesko, Zoltan Kato, Antal Nagy, Imre Gombos, Zsolt Török, Laszlo Vigh Jr., Laszlo Vigh: Live Cell Segmentation in Fluorescence Microscopy via Graph Cut. ICPR 2010: 1485-1488

2. Praveer Singh, Zoltan Kato, and Josiane Zerubia. A Multilayer Markovian Model for Change Detection in Aerial Image Pairs with Large Time Differences. In Proceedings of International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, August 2014. IEEE. (Accepted)

Sztereoszkópikus képalkotás, szuperfelbontás területe rohamosan fejlődött az elmúlt évtizedben. Ezek a feladatok miként támogathatók a létrehozott eszközökkel?

Nagyon örülök ennek a kérdésnek, mert pont az utóbbi egy évben kezdtünk el aktívan foglalkozni folt alapú 3D rekonstrukcióval. Az első eredményeinket bemutató konferenciacikk ugyan még bírálat alatt van, de néhány rekonstrukciós eredményt mellékelek. Az eljárás fontos része az értekezésemben is bemutatott homográfia becslő eljárás, amely a rekonstruálandó 3D sík képei közötti leképezés becslését adja. Ez alapján azután a sík 3D normálvektora és távolsága az 1. kamera koordinátarendszereben már zárt képlettel számolható, így alkalmazható például mobil telefonokon vagy más korlátozott számítási kapacitással rendelkező eszközökön.

A szuperfelbontás lényege, hogy ugyanazon látványról több digitális képet készítünk, amelyek felfoghatók egy végtelen finomságú folytonos kép különböző diszkrét rácsokon vett mintavételezéseiként. Ha sikerül ezeket a rácsokat szubpixeles pontossággal illeszteni, akkor kapunk egy új, potenciálisan ugyan nem reguláris, de mindenképpen nagyobb felbontáson mintavételezett változatot - ez lesz a szuperfelbontás eredményeként előálló nagy felbontású digitális kép. Jól látszik, hogy alapvetően két problémát kell hatékonyan megoldani:

1. Az input képek kezdeti regisztrációja, amely sok esetben hagyományos regisztrációs technikákon alapul és parametrikus transzformációs modellt használ

2. Az előző lépésben kapott kezdeti illesztésből kiindulva a képeket szubpixeles pontossággal kell fedésbe hozni. Ez általában speciális algoritmusokkal történik, szükséges hozzá a képek fotometriai információját felhasználni és jellemzően nem parametrikus transzformációs modellt alkalmaz

A fentiekből jól látható, hogy az 1) lépés megoldásához hatékonyan felhasználhatóak az általunk kifejlesztett regisztrációs eljárások, de a 2) lépés pontos megoldása nehezen képzelhető el tisztán a szegmentált régiók geometriai tulajdonsága alapján.

A gauss keverék modellek végigkísérik a dolgozatot. A szerző is említi, hogy ezek identifikálása gyakran numerikus problémákkal terhelt (pl. 103. oldal). Volt-e szükség valamilyen „trükk" alkalmazására az ilyen esetekben (pl. a kovariancia mátrix

1. ábra. 3D rekonstrukciós eredmények: a képpárokon szín-alapú klaszterezéssel nyertük ki a régiókat és az egymásnak megfelelő régiók közötti síkhomográfiát az értekezésem 6. fejezetében bemutatott módszerrel határoztuk meg.

kezelése kapcsán)?

A Gauss keverék modellek alapvető fontosságúak a képfeldolgozás számos területén. Sokféle módszert dolgoztak ki ezen valószínúségi modell paramétereinek becslésére, melyek közül többet én is alkalmaztam (pl. egyszerű empirikus középérték és kovariancia becslés címkézett minta alapján, EM algoritmus, vagy az összetettebb adaptív szimulált hűtés). Ezen módszereknél ritkán lépnek fel numerikus problémák, de ezekre akkor is ügyelni kell a gyakorlati alkalmazások során. A legkritikusabb hibaforrás a kovariancia mátrix szingularitása lehet, hiszen ekkor nem számolható az inverze és a determinánsával sem tudunk osztani. A szegmentálási modelleknél a Gauss modellt egy-egy régiótipus képjellemzőinek statisztikai jellemzésére használom, tehát ez a probléma akkor léphet fel, ha a Gauss paramétereket egy néhány pixelből álló régióból szeretnénk kiszámolni, hiszen ilyenkor nagy valószínőséggel a pixelek jellemzői sem lehetnek kellően változatosak, vagyis a variancia közel 0 lehet. Ezt az RJMCMC szegmentáló eljárás megfelelően kezeli, és a túlságosan kis méretű régiókat megszünteti (megjegyzendő, hogy ezek kialakulásának esélye kicsi). A bíráló által is említett 103. oldalon található levezetésben a Gauss modellt régiópárok affin illesztésére használjuk. Ebben az esetben a kovariancia mátrix a régió geometriai tulajdonságától függ, így szingularitás a gyakorlatban csak elfajuló alakzatok esetében fordul elő (közel egyenes vagy pontszerű régiók). Az ilyen régiók valós

képelemzési feladatokban általában már fizikai korlátok miatt is ki vannak zárva. Ezen esetek kezelése tehát a konkrét alkalmazásokban elég egyszerű, hiszen ilyen alakzatok megjelenése valamilyen hibára utal, ami akár a feldolgozás megszakítását jelentheti.

Szeged, 2014. május 16. Kató Zoltán