

GYORS, TARTALOMFÜGGŐ KÉPSZEGMENTÁCIÓ

Tézisfüzet a PhD disszertációhoz

Varga Balázs

Témavezető:
Dr. Karacs Kristóf

Konzulens:
Prof. Roska Tamás
az MTA rendes tagja



Pázmány Péter Katolikus Egyetem
Információs Technológiai Kar
Interdiszciplináris Műszaki Tudományok Doktori Iskola

Budapest, 2012

1. Bevezetés

Az emberiség a kezdetek óta komoly munkát fektet azon tevékenységek automatizálásának kutatásába, amelyek lehetőséget nyújtanak az életminőség javítására, az élet időtartamának meghosszabbítására. A modern világ egyre több olyan feladat ellátását követeli meg, amelyek az ember számára túlzottan egysíkúak, monotonak (mint pl. az állandó videokamerás távfelügyelet, vagy egy gyártósor elemeinek a nap huszonnégy órájában minőségbiztosítás céljából történő monitorozása), illetve nem, vagy csak részben végezhetőek el egyszerű, biológiai adottságainkból kifolyólag (pl. a repülésirányítással, gépjármű navigációval kapcsolatos feladatok, vagy az orvosi képzés). Ezen feladatok jelentős részének gépesítése részben, illetve egészben vizuális információ hasznosítására épül, ezért a feldolgozás módszereinek javítása a feladatot ellátó berendezések fejlesztésének elengedhetetlen követelménye. Az emberi tevékenységek összetettsége folytán automatizálásukhoz az lenne a legcélravezetőbb, ha a feladatokat megvalósító eszközöket mind szenzoriális, mind algoritmikus szempontból az emberi érzékekhez és logikus gondolkodáshoz hasonló arzenállal szerelhetnénk fel. A feladat nehézsége abban rejlik, hogy a gépesítendő feladatok többsége nem egyszerűen csak az információ feldolgozását, hanem annak bizonyos fokú „megértését” követeli meg. Egy tárgy, pl. egy teniszlabda azonosítása számunkra, emberek számára teljesen természetes, észrevétlen folyamat. Így még akkor sem okoz gondot

kivenni egy darabot egy labdákkal teli zsákból, ha az egyes labdák határai egy erős esti szürkület miatt nem látszanak rendesen. A feladat számunkra banális, hiszen már *ismerjük* a labda formáját, nagyságrendi méretét.

Ezen példát a képi értés nyelvezetére lefordítva két megfigyelést is tehetünk. Egyrészt kijelenthető, hogy a magas szintű meta adatok rendelkezésre állása (a labda fizikai tulajdonságainak előzetes ismerete) jelentős mértékben segíti a feladat végrehajtását. Másrésztől azonban megállapítható, hogy az ezen, *top-down* rendszerű megközelítés sikere az azt kiegészítő, alacsonyszintű, *bottom-up* elv alapján szerzett információk (fizikálisan látni a labdakupacot, ezen adatok feldolgozását már a retina csatornáiban megkezdeni) nélkül nem lehet sikeres, mert ezek szolgáltatják azokat az alapegységeket, amelyen a magas szintű értelmezés végbermegy.

A fentebb az automatizálás kapcsán felsorolt feladatokhoz a ma elérhető, legpontosabb algoritmusok többségét a *top-down*, illetve *bottom-up* módszerek segítségével szerzett információk hatékony fűzőjét is megvalósító emberi látórendszer inspirálta.

Valószínűleg erre a tényre vezethető vissza, hogy miként a vizuális információk látórendszerben zajló, alacsonyszintű feldolgozásáról már számos ismeret a rendelkezésünkre áll, úgy pixel szintről kiinduló képfeldolgozási feladatok kutatása is előrehaladottabb stádiumban van. A szemantikus beágyazás, illetve a környezeti információk hatékony, absztrakt reprezentációjának felderítése azonban a mai napig alapvető kérdés mind a látórendszert

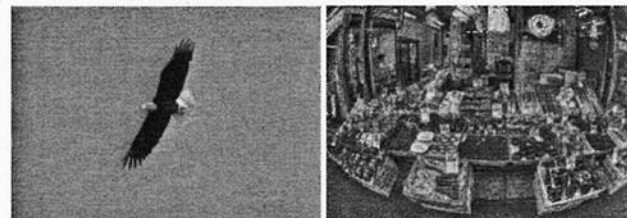
kutató neurobiológusok és pszichológusok, mind pedig a gépi látással foglalkozó kutatók számára.

Céлом, hogy doktori munkámmal előremozdítsam a különböző modalitású információk fúziójára alapuló képi értés tudományterületét. Disszertációmban a hangsúlyt egy olyan bottom-up elvű, moduláris szegmentáció kutatására helyeztem, amely a klaszterezés folyamata során létrejövő osztályhierarchia különböző szintjein és pontjain ad lehetőséget top-down elvű, szemantikus, illetve feladat specifikus információk befecskendezésére.

Elsőként egy olyan eljárást dolgoztam ki, amely gyorsan képes magas felbontású képek szegmentációjára. A magas felbontás használatát az indokolja, hogy segítségével a szegmentáció kimenetéből olyan részleteket is kinyerhetünk, amelyek méretükből adódóan kisebb pixelszámú képeken nem jelennek meg. Ezen többlet információk elősegítik a különböző felismerési-, és klasszifikációs feladatok pontosabb működését, és kiszélesítik a top-down módszerekkel elérhető tudás felhasználásának lehetőségeit. A fizikai képméret növelésének hátránya, hogy a feldolgozandó adatok mennyisége is nő, amely negatív hatást gyakorol a szegmentációs algoritmusok futási idejére. Ezen probléma áthidalására olyan keretrendszert konstruáltam, amely párhuzamosított belső szerkezetéből kifolyólag ki tudja használni a sokprocesszoros számítógép architektúrákban rejlő magas számítási kapacitást. A keretrendszer magját a *mean shift* algoritmus [6] inspirálta, melyet a további sebességnövekedés érdekében kiegészítettem egy kvázi-

véletlen elven működő mintavételezési eljárással, illetve olyan osztályösszevonó lépést alkottam hozzá, amely hatékonyan képes a túlszegmentáció csökkentésére. Mindezekon felül bevezettem egy eljárást, amely csökkenti a párhuzamosításból fakadó többletmunkát.

A második kutatási területem egy olyan, adaptív mintavételezési eljárás megalkotására fókuszált, amely a kép tartalmának (*content*) lokális jellemzői alapján dolgozik. (A képtartalomban található különbségek intuitív szemléltetését az 1. számú ábra tartalmazza.)



1. ábra Egy intuitív példa a kép tartalmának mennyiségében megjelenő különbségek szemléltetésére. Fizikailag mindkét kép 14,5 megapixel felbontású, azonban míg a bal oldalon egyetlen objektum jelenik meg homogén háttér előtt, a jobb oldali kép rengeteg aprólékos részletet tartalmaz.

Kutatásomnak két célja volt: egyfelől a szegmentáció futási idejének minimalizálása a képtartalom függvényében, egyszersmind kimenet minőségének azonos szinten tartása a naiv (tehát az összes képponton dolgozó, mintavételezést nem tartalmazó) eljárással. Ezen célok teljesítéséhez nem csak a mintavételezéshez használt elemek számának, hanem ezek topografikus pozícióinak meghatározása is

szükség van. A problémák megoldására egy autonóm módon működő eljárást alkottam, amely ezen feladat megoldásán túl egy darab paraméter értéke alapján valósítja meg a mintavételezést és ebben tárolja az osztályok és a pixelek között fennálló kötés erősségét is. Ezáltal a rendszer reprezentációja tömör módon képezhető le, így mód nyílik magas pixelszámú képek szegmentációjára is. Ez a stratégia arra is lehetőséget biztosít, hogy a pixeleket a színük és térbeli pozíciójuk alapján hozzájuk leginkább hasonló osztályokhoz rendelhessük hozzá. Ezen adaptív módszert a fent ismertetett párhuzamos keretrendszerbe illesztettem be.

2. Vizsgálati módszerek

Az algoritmus megtervezésében a kernelsűrűség-becslés, a mintavételezés, a Gaussi vegyes modellezés, a szintérelmélet, a hasonlósági metrikák és a párhuzamosított algoritmusfejlesztés területein elérhető szakirodalomra támaszkodtam.

Az algoritmusom vizsgálatát kezdetben azon három leggyakoribb mérési szempont figyelembevételével végeztem, amelyek feltétlenül szükségesek az objektív kiértékeléshez. Ezek a *futási idő* (azon időtartam, amely alatt a szegmentáció végbemegy), a *kimeneti pontosság* (különböző metrikák segítségével számítható, amelyek az adott képhez tartozó *alapigazsághoz* (*ground thruth*-hoz) viszonyítják az algoritmus kimenetét) és a *fizikai felbontás* (a képpontok számával egyezik meg).

A munkám során kiemelten figyeltem arra, hogy a mérések eredményei összevethetőek legyenek az irodalomban publikált

algoritmusok azonos mutatóival, ezért a kimeneti minőséget értékelő tesztek nyilvánosan elérhető adathalmazokon [12][13][14] végeztem el. Ezek előnye, hogy szabványos metrikákat (*Segmentation Covering* [15], *Probabilistic Rand Index* [16], *Variation of Information* [17], *F-measure* [18], *Average Precision* [19], illetve a *Fragmentation* [14]) kínálnak egységesített keretrendszereken belül.

Sajnos azonban ezen adathalmazok aránylag alacsony felbontású képeket tartalmaznak, ezért csak korlátozott módon használhatóak fel a fent felsorolt másik két szempont kiértékeléséhez, az ezen adathalmazokon mért minőségi mutatók pedig nem terjeszthetők ki lineáris módon nagyobb felbontású képekre. A probléma kezelésére két további, valós körülmények között készült, különböző méretű objektumokat megjelenítő, nagyfelbontású képeket tartalmazó képhalmazt állítottam össze. Az első halmaz 15 darab képből áll, amelyeket öt különböző felbontásban vizsgáltam (1. táblázat).

| Az Extended Graphics Array elnevezése | Rövidítés | Felbontás | Felbontás megapixelben (MP) |
|---------------------------------------|-----------|-------------|-----------------------------|
| Wide Quad | WQXGA | 2560 × 1600 | 4,1MP |
| Wide Quad Super | WQSXGA | 3200 × 2048 | 6,6MP |
| Wide Quad Ultra | WQUXGA | 3840 × 2400 | 9,2MP |
| Hexadecatuple | HXGA | 4096 × 3072 | 12,6MP |
| Wide Hexadecatuple | WHXGA | 5120 × 3200 | 16,4MP |

1. táblázat A párhuzamos keretrendszer sebességének kiértékeléséhez használt nagyfelbontású képek névkonvenciói és az ezek által jelölt képméretük.

Az algoritmus skálázódásának vizsgálatához szükséges méréseket öt különböző GPGPU-n (általános felhasználási célú grafikus számítási egység – *general-purpose computing on graphics processing units*) végeztem el (2. táblázat).

| Eszköz neve | Stream processzorok száma | Órajel frekvencia | Memória mérete | Számítási képesség |
|-------------|---------------------------|-------------------|----------------|--------------------|
| 8800GT | 112 | 1500 MHz | 1024 MB | 1.1 |
| GTX280 | 240 | 1296 MHz | 1024 MB | 1.3 |
| S2070SG | 240 | 1440 MHz | 4096 MB | 1.3 |
| C2050 | 448 | 1500 MHz | 3072 MB | 2.0 |
| GTX580 | 512 | 1544 MHz | 1536 MB | 2.0 |

2. táblázat A párhuzamos keretrendszer skálázódásának kiértékeléséhez használt GPGPU-k legfontosabb paraméterei.

A 15 elemű képhalmazon végzett mérések meghőszítették azon hipotézisemet, hogy a mintavételezéssel dolgozó, veszteséges szegmentációs algoritmusok megfelelő kiértékeléséhez szükséges egy negyedik mérési szempont, a *képtartalom* figyelembevétele. Ezen okból kifolyólag összeállítottam egy második, 10 megapixel felbontású képekből álló, 103 db elemet tartalmazó képhalmazt. A korábbi mérésekkel szemben a kiértékelés fő dimenziója immár nem a felbontásban, hanem a képek *tartalmában* jelentkező különbségek által a futási időre gyakorolt hatás volt.

A keretrendszer implementációjához MATLAB-ot [21], illetve az Accelerayes által fejlesztett Jacket [22] eszköztárat használtam. A statisztikai elemzéseket MATLAB illetve Microsoft Excel [23] segítségével végeztem.

3. Új tudományos eredmények

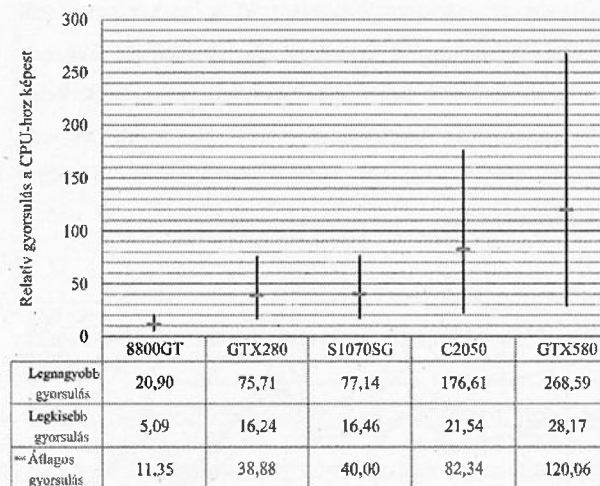
I. Téziscsoport: *Bottom-up* elvű képszegmentációs eljárás párhuzamosítása és megvalósítása sokprocesszoros architektúrán.

Mára a fogyasztói piacokon elterjedt fényképezőgépek és az okostelefonok többsége is sokmegapixeles fényképeket készít. A gépi látásban – következésképp a szegmentációs feladatokhoz is – használt algoritmusok jelentős részére igaz, hogy futási idejüket erőteljesen befolyásolja a bemeneti adathalmaz mérete, azaz a kép felbontása. A képméret növekedése melletti azonos futási idő fenntartásához szükséges az egyre erősebb futtatási környezet, nevezetesen a gyorsabb processzorok használata. A számítási teljesítmény eddig bevett módszere, vagyis az órajel frekvencia emelése a közelmúltban azonban problémákba ütközött. Ezek oka a magas hődisszipáció, illetve azon tény, hogy tranzisztorok gyártástechnológiája mostanra a fizika határait feszegeti. A teljesítmény iránti igény ezzel szemben töretlen, ezért a nagy műveletigényű számítások kiszolgálásához fejlesztett eszközök – így a digitális jelprocesszorok (DSP-k), a felhasználó által programozható kapu-tömbök (FPGA-k), illetve a GPGPU-k – tervezői a processzormagok számának növelésével válaszoltak felmerült kihívásra. Az új rendszerek nyújtotta teljesítmény kihasználásának azonban szükséges előfeltétele a párhuzamosított algoritmikus háttér, melynek megteremtése azonban közelről sem magától értetődő.

A szerző vonatkozó publikációi: [2], [3], [4].

I.1. Párhuzamosítottam a mean shift szegmentációs eljárást, amely ezáltal képessé vált a sokprocesszoros számítógép platformok nyújtotta többletletjesítmény kihasználására. A módszert különböző általános célú számításokat végző grafikus feldolgozó egységekre alkalmazva megmutattam, hogy a párhuzamosítás révén elért sebességnövekedés a processzormagok számával arányosan skálázódik.

Olyan szegmentációs keretrendszert alkottam, amely szimultán módon, több darab kernelen végez mean shift iterációkat. Ezen rendszert implementáltam egy sokprocesszoros architektúrára, és különböző processzorszámú kártyákon végzett mérések segítségével kísérletileg igazoltam, hogy a párhuzamos módon működő keretrendszer jelentősen gyorsabban dolgozik, mint a szekvenciális változat, továbbá, hogy a feldolgozóegységek számának növelésével további gyorsulás érhető el. A 2. ábra a keretrendszer magjának CPU-hoz (Intel Core i7-920 típus, 2.66GHz-es magórajel) viszonyított átlagos gyorsulását mutatja meg az egyes kártyákon.

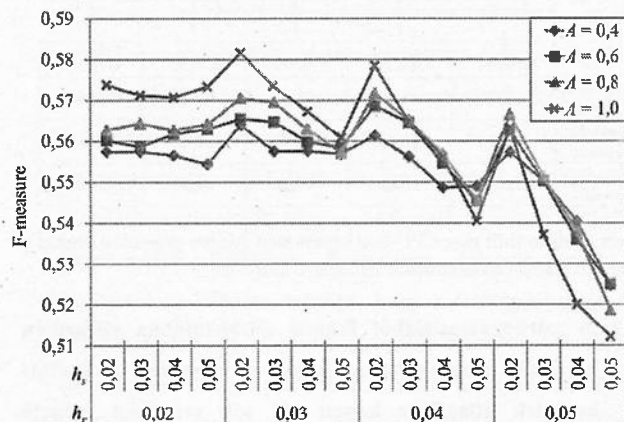


2. ábra A mean shift mag CPU-hoz képest mért átlagos gyorsulási értékei a különböző számú processzorokat tartalmazó kártyákon.

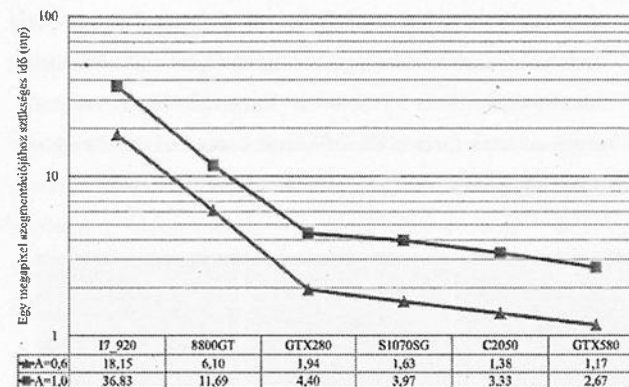
I.2. A párhuzamosításból fakadó többletmunka elemzésén keresztül megmutattam, hogy az egyszerre számított kernelek átlagához képest túl sok számítás igénylő kernelek korai terminálásával úgy lehet jelentős sebességnövekedést elérni, hogy eközben a kimeneti minőség az irodalomban általánosan használt metrikák alapján alig csökken.

Az vizsgálat eredményeként azt találtam, hogy a blokkprocesszalás jellemzőiből adódóan egyesével nem lehetséges a szaturált módusok új kernelekre történő cseréje. Ezért saját (abridgingnek, azaz rövidítésnek, vagy levágásnak

elnevezett) módszert dolgoztam ki a párhuzamosításból fakadó többletmunka csökkentésére. A módszer kimeneti minőségre, illetve futási időre gyakorolt hatását a Berkeley Szegmentációs Adathalmazon [12] (3. ábra), és egy nagyfelbontású képeket tartalmazó halmazon (4. ábra) vizsgáltam meg.



3. ábra A párhuzamosan működő keretrendszer kimeneti minőségének alakulása különböző parametrizációk mellett az F-measure mérőszám alapján. h_i és h_r egyenként a térbeli, illetve szintérbeli sávszélességeket, A az *abridging* paramétert jelöli.



4. ábra Egy megapixel klaszterezéséhez szükséges átlagos futási idők az egyes kártyákon, (illetve a CPU-n) az *abridging* paraméter (A) függvényében. A kimeneti minőség különbsége a két beállítás között 3%.

1.3. Hatékony, párhuzamos osztályösszevonó algoritmust alkottam, amely topografikus elrendezés, illetve a szín alapján képes szegmentált képek túlszegmentációjának csökkentésére.

A kép túlszegmentálása a képfeldolgozásban jól ismert [7][8] és széles körben alkalmazott [9][10][11] eszköz. A módszer előnye, hogy megkönnyíti a mind a magas-, mind pedig az alacsonyszintű feldolgozás segítségével megszerzett többletinformáció rendszerbe fecskenedését, ezáltal a véglegesített osztály struktúra kialakítása az adott feladat szempontjából értelmezett hasonlósági szabályok alapján történhet meg. Olyan, általános felhasználható osztályösszevonó módszert alkottam és implementáltam,

amely párhuzamosított módszerrel határozza meg az osztályok szomszédossági információját és színbeli hasonlóságát. Az 5. ábra a bemeneti képen végzett szegmentációs fázis és az azt követő összevonás kimeneteire mutat egy példát.



5. ábra A bemeneti képet (bal oldal) először homogén tulajdonságokkal rendelkező osztályokba soroljuk (középen), majd azon klasztereket, amelyek szín, illetve szomszédossági információ alapján hasonlóak, összevonjuk (jobb oldal).

II. Téziscsoport: Adaptív, a kép tartalmához idomuló mintavételezési eljárás nemparametrikus képszegmentációhoz.

A kép felbontása magától értetődő módon közvetlenül befolyásolja az adott szegmentációs eljárás futási idejét, illetve kimentének pontosságát. A veszteséges elven működő algoritmusok esetében ezen két érték megváltozásának okai azonban nem kizárólagosan a felbontásban keresendők, mert a valós körülmények között készült képek által hordozott információ mennyisége és topografikus eloszlása erősen heterogén (lásd 1. ábra). Ezen okból

kifolyólag a pusztán a képfelbontás alapján mért eredmények a bemenet tartalmi tulajdonságaitól függően torzulhatnak.

A komplexitás szempontjából az elsődleges prioritás a mintavételezett elemek számának csökkentése, ezzel egyidejűleg azonban szem előtt kell tartanunk, hogy az alul-mintavételezés a részletek elvesztéséhez, a túl-mintavételezés felesleges többlétszámításokhoz vezet. Ezen problémákat a következőkben felsorolt módokon oldottam meg.

A szerző vonatkozó publikációja: [1].

II.1. Definiáltam egy, a mean shift eljárásban implicit módon kiszámított bizonyossági értéket, amely heurisztikául szolgál az adaptív mintavételezéshez, egyszersmind önmagában elégséges támpont a képpontok kimeneti osztályainak meghatározásához.

Olyan, egyetlen paraméter által vezérelt megoldást alkottam, amely térbeli, és szintérbeli távolság alapján regisztrálja egy adott pixel és egy osztályt jellemző módus közötti kötés erősségét. A módszer felhasználásával elérhető, hogy minden pixel a hozzá leginkább hasonló tulajdonságokkal rendelkező osztályba rendelődhesen. Az eljárás központi eleme a *kötésbizonyossági érték (bond confidence value)*, amely a szegmentáció során implicit módon kerül kiszámításra.

II.2. Kidolgoztam egy, a kép tartalmi elemeit figyelembe vevő mintavételezési eljárást, amely a szegmentáció előrehaladtával párhuzamosan, adaptív módon határozza meg a kiválasztandó minták térbeli pozícióját és számát, és egy saját, nagyfelbontású, illetve nyilvános képadatbázisok felhasználásával numerikusan igazoltam, hogy az adaptív eljárás kimeneti minőségének mérőszámai a legelterjedtebb metrikák tekintetében közel azonosak a naiv (összes képponton futó) módszer mutatóival, azonban számításigénye jelentősen alacsonyabb annál.

Olyan adaptív sémát alkottam, amelyben a mintavételezés gyakoriságát a kép lokális tulajdonságai határozzák meg. A séma a homogén képi régiókban kevés, nagy sávszélességű kernelt indít, így a szegmentáció ezeken a régiókban különösen gyorsan végbe megy, a színtérben inhomogén régiók esetében pedig nagyobb számú, kisebb sávszélességű kernelt indít ilyen módon megőrizve az itt található részleteket a kimenetben. A mintavételezésből kifolyólag csökken a számítási-, és memóriaigény, így lehetővé válik a nagyfelbontású képek szegmentációja is.

A kimenet minőségét átfogó módon értékeltem ki a szakirodalomban leggyakrabban használt metrikák [14][15][16][17][18][19] és szegmentációs adatbázisok [12][13][14] segítségével. A kimeneti eredmények alapján kísérleti úton mutattam meg, hogy a felhasznált

mintavételezési séma ellenére a kimenet minősége összemérhető a szakirodalomban megtalálható, mean shiftre épülő módszerekével.

Az algoritmusom futási idejének kiértékelését egy 103 darab nagyfelbontású képet tartalmazó adatbázison végeztem el. Mivel ezen adathalmazhoz nem érhető el kézzel készített referencia, ezért humán alanyok szubjektív értékelése alapján meghatároztam az ezen halmazon legjobb minőségű kimenetet adó parametrizációt a legnépszerűbb, nyilvánosan is elérhető, mean shift alapú szegmentációs eljáráshoz [20], illetve az ezen kimenethez leginkább hasonló kimenetet eredményező beállítást a saját keretrendszeremhez. Az ezen parametrizációk mellett mért futási időket a 3. táblázat tartalmazza.

| | Átlagos futási idő (mp) | | | Átlagos gyorsulás |
|----------------------|--------------------------|------------------------------|-----------------------------|-------------------|
| | Javasolt rendszer | Referencia (Gyors beállítás) | Referencia (Naiv beállítás) | |
| Átlag | 18,01 | 320,01 | 23366,06 | 18,58 |
| Szórás/ (Relatív) | 10,73 (59,59%) | 277,46 (86,71%) | 7385,67 | 14,26 |

3. táblázat A 103 elemet számláló, 10 megapixeles képeket tartalmazó adathalmazon mért futási idők. Az átlagos gyorsulás az általam alkotott keretrendszer és a referenciaként használt algoritmus gyors beállítása közötti különbséget jelöli.

II.3. Nagyfelbontású képeken mért futási idők segítségével, és egy, a képek tartalmi komplexitásának jellemzésére bevezetett, saját mérőszám korrelációanalízisén keresztül igazoltam az algoritmus adaptivitását, tehát, hogy a kevésbé komplex képek szegmentálása rövidebb ideig tart.

Definiáltam egy szubjektív, percepcióra alapuló (*kappa-indexnek* elnevezett) mérőszámot. Értékét egy adott képre humán résztvevők által adott osztályzatok átlagaként határoztam meg, akiket arra kértem, hogy osztályozzák a hasznos képtartalom mennyiségét egy 1-től 5-ig értelmezett skálán úgy, hogy az 1-es érték jelentése „*kevés dolog, illetve sok, nagy homogénnek tekinthető régió, „szellős” kép*”, míg a skála másik vége a „*sok kivehető részlet, gazdag információtartalom, „zsúfolt” kép*” leírással jellemezhető. A kappa-indexet 15 résztvevő által adott osztályzatokat kiátlagolva határoztam meg a 103 elemet tartalmazó, nagyfelbontású képhalmaz minden egyes elemére, majd ezen értékek – tehát az átlagos osztályzatok – alapján három alosztályba soroltam a képeket (4. táblázat – az információ mennyisége az A alosztályba, sorolt képekben a legkevesebb, a C alosztályban a legtöbb).

| | Átlagos futási idő (mp) | | | Átlagos gyorsulás | Kernelek átlagos száma | Képek száma |
|-----------|-------------------------|------------------------------|-----------------------------|-------------------|------------------------|-------------|
| | Javasolt rendszer | Referencia (Gyors beállítás) | Referencia (Naiv beállítás) | | | |
| A osztály | 12,85 | 204,22 | 23273,47 | 15,54 | 599,43 | 30 |
| B osztály | 14,84 | 344,63 | 23371,93 | 22,24 | 674,38 | 45 |
| C osztály | 28,62 | 404,49 | 23455,82 | 15,96 | 1011,18 | 28 |

4. táblázat Az átlagos tartalom mennyiség alapján létrehozott, 10 megapixel-es képeket tartalmazó alosztályokon mért futási idő eredmények. Az átlagos gyorsulás az általam alkotott keretrendszer és a referenciaként használt algoritmus gyors beállítása közötti különbséget jelöli.

A teljes adathalmazon mérve az egyes képek szegmentációjához felhasznált kernelek száma, illetve a képekhez megállapított kappa-indexek közötti korreláció értéke 0,694, amely alátámasztja az erős összefüggés fennállását a humán résztvevők által jelölt tartalom mennyisége, illetve az algoritmusom által feldolgozott tartalom mennyisége között.

4. Az eredmények felhasználási területe

A disszertációban összefoglalt algoritmusok köztes lépésként alkalmazhatók számos különböző mintaillesztési, objektum felismerési, objektum detekciós, illetve magas szintű képi értést megvalósító képfeldolgozási, illetve gépi látási feladatban. A szegmentációs eljárást két alkalommal már sikerrel alkalmaztam különböző valós problémák megoldásához.

Az első alkalommal a rendszer a gyalogos átkelőhely detekcióban [4] került felhasználásra, ami a Bionikus Szemüveg

Projekt [24] egyik kulcs feladata. Ezen kezdeményezés célja személyre szabott segélynyújtás vakoknak és gyengén látóknak a különböző mindennapi feladataik során. A rendszer központi eleme egy hordozható készülék, amely multimodális információ (így pl. GPS koordináták, vizuális adatok, illetve környezeti zajok) alapján képes objektumok (pl. piktogramok, bankjegyek, közlekedési járműszámok, vagy ruhák és azok színének) és különböző situációk, környezetek (pl. a megállóba beálló busz, vagy az otthon/utca/iroda) azonosítására. A különböző szenzoriális bemenetekről származó információk dinamikus módon integrálhatók az algoritmusomba pl. az osztályösszevonást befolyásoló szabályok formájában. Így például amikor a rendszer a GPS koordináták alapján lokalizálja a felhasználó pozícióját, az algoritmusom által használt szabálykönyv automatikus módon kiválaszthatja azokat az elemeket, amelyek a pozícióhoz kapcsolódó komplex, adott (színű, vagy formájú) objektumok (pl. lámpaoszlopok, feliratok) létrehozását detekcióját segítik, továbbá szűrjük a hamis pozitív elemeket.

A jelenleg elérhető okostelefonok bizonyos példányai már nem csak a szenzoriális arzenál területén rendelkeznek komoly potenciállal, hanem mobil GPU-kkal is fel vannak szerelve, melyek párhuzamosításban rejlő képességeit az általam megalkotott keretrendszer képes kihasználni.

A második használati esetben a keretrendszerem a Digitális Holografikus Mikroszkóp Projekt [25] egyik korai prototípusában került felhasználásra. Ezen projekt egy olyan (hardver és szoftver)

környezet megalkotását tűzte ki célul, amely autonóm módon képes valós idejű vízminőség monitorozás megvalósítására. Hogy biztosítsa a vízben található részecskék (pl. por, pollen, algák), azaz az *előtér* robusztus detekcióját, szegmentációját és klasszifikációját, a berendezés végleges verziója több forrásból (így színes-, illetve fluoreszcens mikroszkópiával rögzített képeken, és egy teljesen új módszer, a színes digitális holografikus mikroszkópia segítségével előállított volumetrikus adatokat tartalmazó metszeteken) dolgozik majd. A szegmentációs algoritmusomat ezen, különböző kép karakterisztikát eredményező adatforrások közül sikerrel alkalmaztam színes mikroszkópi, illetve fluoreszcens mikroszkópi felvételek feldolgozására [26]. Mivel a projekt keretében készülő berendezés erre a feladatra épített célhardveren fog megvalósulni, a back-end tartalmaz sokprocesszoros architektúrát, amelynek használata a keretrendszerem alapvető képessége. Mivel az algoritmusom által használt tulajdonságtér csatornái a felhasználó választásának tárgyát képezik, a későbbiekben a pontosság javításának érdekében az egyes szenzorok által szolgáltatott bemenetek közös, kötegelt feldolgozásra is lehetőség van.

Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Karacs Kristófnak, aki bizalommal támogatta a munkámat a doktoranduszi éveim során, és mentorként és oktatóként is kiemelkedő példát mutatott.

Hálás vagyok *Roska Tamás*, *Csurgay Árpád* és *Szolgay Péter* professzor uraknak és *Nyékyné dr. Gaizler Judit* prodékán asszonynak, amiért bölcs tanácsaikkal és gondoskodásukkal fáradhatatlanul segítettek az előrehaladásomat, és biztosították a kutatási lehetőséget a doktori iskola berkein belül.

Köszönöm a *Svájci Hozzájárulás* és a *Tateyama Magyar Laboratórium Kft.* támogatását.

Köszönöm doktorandusz társam, *Szabó Vilmos* önzetlen segítségét, melynek jóvoltából kutatásom első éveiben nagyon sok ötlettel és tudással lettem gazdagabb a gépi tanulás és a gépi látás témaköreiben. Továbbá köszönetemet szeretném kifejezni a Doktori Iskola többi hallgatójának barátságukért és/vagy a megosztott tudásért, illetve a csodálatos időkért, amelyeket együtt átéltünk: *Nemes Csaba*, *Füredi László*, *Balogh Ádám*, *Tornai Kálmán*, *Tisza Dávid*, *Radványi Mihály*, *Kovács Andrea*, *Tibold Róbert*, *Bérci Norbert*, *Feldhoffer Gergely*, *Tornai Gábor*, *Zsedrovits Tamás*, *Rák Ádám*, *Veres József*, *Stubendek Attila*, *Solymár Zóra*, *Vizi Péter*,

Koller Miklós, *Kiss András*, *Horváth András*, *Treplán Gergely*, *Kovács Dániel*, *Tar Ákos*, *Laki László*, *Fülöp Tamás*, *Tuza Zoltán*, *Orosz György*, *Rudan János*, *Horváth Anna*, *Kárász Zoltán*, *Fekete Ádám*, *Gelencsér András*, *Pilissy Tamás*, *Tátrai Antal*, *Karlócai Balázs*, *Szolgay Dániel*, *Bojársky András*, *Bankó Éva*, *Soós Balázs Gergely*, *Kusnyerik Ákos*, *Kozák László*, *Gergelyi Domonkos*, *Reguly István*, *Bihary Dóra*, *Hermann Petra*, *Knakker Balázs*, *Tóth Emília* és *Józsa Csaba*.

Szeretném megköszönni a posztdoktoroknak, hogy munkájukkal életben tartják azt a nagyon különleges szikrát mind a Kar, mind pedig Doktori Iskola életben: *Oláh András*, *Nagy Zoltán*, *Gyöngy Miklós*, *Iván Kristóf* és *Cserey György*.

Külön köszönet illeti *Sifter Viktóriát*, *Adorján Líviát*, *Babiczné Rácz Máriát*, *Körmendyné Érdi Máriát* és *Tholt Pétert*, akik az általam eléjük tárt kisebb-nagyobb problémákra minden alkalommal azonnali megoldásokat kínáltak a bürokrácia útvesztője helyett.

Utolsóként – ám legnagyobb mértékben – a *Családomnak*, a barátomnak, *Dórinak*, a *Barátainak*, és az *RV-NB10-esemnek* szeretném kifejezni a hálámat, akik támogattak a döntéseimben, és osztoztak velem a szép, és a nehéz időkben is.

Publikációk

A szerző folyóirat publikációi

- [1] B. Varga, K. Karacs, "Towards a Balanced Trade-off Between Speed and Accuracy in Unsupervised Data-Driven Image Segmentation," sent to: *Machine Vision and Applications*, Beküldve 2012. május, bíráló alatt
- [2] B. Varga, K. Karacs, "High-resolution Image Segmentation using Fully Parallel Mean Shift," *EURASIP Journal on Advances in Signal Processing*, 2011:111 Springer, 2011

A szerző konferencia publikációi

- [3] B. Varga, K. Karacs, "GPGPU Accelerated Scene Segmentation Using Nonparametric Clustering," *Proceedings of the International Symposium on Nonlinear Theory and its Applications (NOLTA)*, November 2010, Krakow, Poland, pp. 149–152
- [4] M. Radványi, B. Varga, K. Karacs, "Advanced Crosswalk Detection for the Bionic Eyeglass", *Proceedings of the 12th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA)*, January 2010, Berkeley, CA, pp. 1–5

A szerző egyéb publikációi

- [5] H. van Welbergen, Z. Ruttkay, B. Varga, "Informed Use of Motion Synthesis Methods", *Motion in Games*, Springer Verlag, Berlin, ISBN 978-3-540-89219-9, 2008, pp. 132–143

Bibliográfia

Szegmentációval kapcsolatos hivatkozások

- [6] D. Comaniciu, P. Meer, "Mean Shift: a Robust Approach Toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24 no. 5, 2002, pp. 603–619
- [7] A. Duarte, Á. Sánchez, F. Fernández, A. Montemayor, "Improving Image Segmentation Quality Through Effective Region Merging using

- a Hierarchical Social Metaheuristic," *Pattern Recognition Letters* vol. 27 no. 11, 2006, pp. 1239–1251.
- [8] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, R.B. Fisher, "An Experimental Comparison of Range Image Segmentation Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18 no. 7, 1996, pp. 673–689
- [9] S. Beucher, "Watershed, Hierarchical Segmentation and Waterfall Algorithm," *Mathematical Morphology and its Applications to Image Processing*, 1994, pp. 69–76.
- [10] L. Li, R. Socher, L. Fei-Fei, "Towards Total Scene Understanding: Classification, Annotation and Segmentation in an Automatic Framework," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2009, Miami, FL, pp. 2036–2043
- [11] F. Shih, S. Cheng, "Automatic Seeded Region Growing for Color Image Segmentation," *Image and Vision Computing* vol. 23 no. 10, 2005, pp. 877–886.

A minőség kiértékelésével kapcsolatos hivatkozások

- [12] D. R. Martin, C. C. Fowlkes, D. Tal, J. Malik, "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 2001, pp. 416–423
- [13] P. Arbeláez, M. Maire, C. C. Fowlkes, J. Malik, "Contour Detection and Hierarchical Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33 no. 5, 2011, pp. 898–916
- [14] S. Alpert, M. Galun, A. Brandt, R. Basri, "Image Segmentation by Probabilistic Bottom-Up Aggregation and Cue Integration" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, February 2012, pp. 315–327
- [15] M. Everingham, L. van Gool, C. Williams, J. Winn, A. Zisserman, "PASCAL 2008 Results," <http://www.pascal-network.org/challenges/VOC/voc2008/workshop/index.html>, 2008
- [16] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward Objective Evaluation of Image Segmentation Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, 2007, pp. 929–944

- [17] A. Y. Yang, J. Wright, Y. Ma, S. S. Sastry, "Unsupervised Segmentation of Natural Images via Lossy Data Compression," *Computer Vision and Image Understanding*, vol. 110, no. 2, 2008, pp. 212–225
- [18] N. Chinchor, B. Sundheim, "MUC-5 evaluation metrics," *Proceedings of the 5th Conference on Message Understanding*, 1993, pp. 69–78
- [19] B. Bartell, G. Cottrell, R. Belew, "Optimizing parameters in a ranked retrieval system using multi-query relevance feedback," *Proceedings of Symposium on Document Analysis and Information Retrieval (SDAIR)*, 1994
- [20] C.M. Christoudias, B. Georgescu, P. Meer, "Synergism in Low Level Vision," *Proceedings of the 16th International IEEE Conference on Pattern Recognition*, vol. 4, August 2002, pp. 150–155

Az algoritmus fejlesztéssel kapcsolatos hivatkozások

- [21] The MathWorks, Inc. MATLAB. *Addr.: The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098*, www.mathworks.com, accessed June, 2012
- [22] AccelerEyes LLC, *Addr.: 800W Peachtree St NW, Atlanta, GA 30308*, www.accelereyes.com, accessed June, 2012
- [23] Microsoft Corporation, *Addr.: 1 157th Avenue Northeast Redmond, WA 98052*, <http://office.microsoft.com/en-us/excel/>, accessed June, 2012

Az eredmények felhasználásával kapcsolatos hivatkozások

- [24] K. Karacs, A. Lázár, R. Wagner, D. Bálya, T. Roska, and M. Szuhaaj, "Bionic Eyeglass: an Audio Guide for Visually Impaired," *Proc. of the First IEEE Biomedical Circuits and Systems Conference (BIOCAS)*, London, UK, Dec. 2006, pp. 190–193.
- [25] S. Tőkés, V. Szabó, L. Orzó, P. Divós, Z. Krivosija, "Digital Holographic Microscopy and CNN-based Image Processing for Biohazard Detection", *Proceedings of the 11th International Workshop on Neural Networks and Their Applications (CNNA)*, July 2008, Santiago de Compostela, Spain, pp. 8
- [26] B. Varga, "Color-based Object Segmentation for Water Quality Surveillance," *Thesis for Informatics Specialist in Bionic Computing Post-graduate Specialist Training, Pázmány Péter Catholic University*, 2011