

Óbudai Egyetem

Doktori (PhD) értekezés
tézisfüzete



Adatpárhuzamos sejtmagkeresési eljárás fejlesztése és paramétereinek optimalizálása

Szénási Sándor

Témavezető:

Vámosy Zoltán, PhD

Alkalmazott Informatikai Doktori Iskola

Budapest, 2013. május 30.

A kutatás előzményei

Az orvosi célú digitális képfeldolgozás használata napjainkban egyre elterjedtebb a patológusok körében. A legújabb rendszerekben elérhetővé vált, hogy a szövettani vizsgálatok által igényelt lépések nagy része automatizálható (festések, tárgylemezek továbbítása, digitális felvételek készítése stb.), és várhatóan ez a trend a jövőben is folytatódik [1]. Az újszerű eszközök a szövettanok feldolgozásának közvetlen előnyei mellett (nagyfelbontású, jó minőségű, jól fókuszált képek) számos új lehetőség előtt nyitották meg az utat, a telepatológiai rendszerekben ugyanis nincs szükség többé a tárgylemezek fizikai továbbítására, hanem hálózaton keresztül is van lehetőség a nagyfelbontású felvételek katalógizálására, megosztására, reprodukciójára, távoli elérésére. Az így kiépült rendszerek elterjedése, a hardvereszközök teljesítményének növekedése, és a képfeldolgozási eljárások fejlődése együttesen készítette elő a következő lépést, amely a képfeldolgozó eljárások megjelenését jelenti a napi rutin diagnosztikában.

Ezek egy része pusztán a képalkotási eljárások továbbfejlesztésének tekinthető (különböző képjavitási algoritmusok [2], CT adatok értékelése [3], 3D rekonstrukciók), azonban megjelentek azok a speciális diagnosztikai szoftverek is, amelyek a szövettanok tartalmi analízisére is vállalkoznak. Ez utóbbi esetben az első lépés egy diagnózis felállítása felé általában a különböző szöveti komponensek detektálása. Erre a célra egyelőre nem találhatunk mindenre kiterjedő, általánosan használható módszereket, például az általánosan részletebben vizsgált vastagbél szövettanok esetén is számos, egymástól jelentősen módszert találhatunk a sejtmagok keresésére [4] [5] [6] [7] [8] [9], illetve a további szöveti komponensek elkülönítésére (mirigyek, felszíni hám). A különböző szegmentálási eredmények felhasználhatók akár közvetlenül, a képernyőn megjelenítve különböző morfológiai és morfometriai paramétereket, vagy akár a közeljövőben lehetőség nyílt ezen keresztül egy részben, vagy akár teljesen automatikus diagnosztizáló rendszerek fejlesztése felé [10] [11].

A kidolgozott különböző eljárások egymástól jelentősen eltérő alapelvek alapján működnek. Csak a sejtmagok detektálását végző algoritmusok között is találhatunk olyan, a K-közép [12] eljárásokon alapuló módszert, amely pusztán a képernyőn elhelyezkedő pixelek színeire, illetve azok eloszlására épít, így próbálva meg különböző osztályokat kialakítani, amelyekből az egyikbe remélhetőleg csak a sejtmagokat alkotó pixelek kerülnek. De emellett találhatunk jóval összetettebb, a pixelek színeit és elhelyezkedését is figyelembe vevő régió növelési [13] módszereket, amelyek ugyan jóval nagyobb futásidő mellett, de valamivel pontosabban tudják meghatározni az egyes sejtmagok pontos elhelyezkedését.

A szoftveres környezet változásán túlmenően az utóbbi években jelentősen megváltozott a hardver lehetőségek tárháza is. Az informatikában egészen a 2000-es évekig megszokottá vált, hogy a különböző processzorok teljesítménye évről-évre folyamatosan növekszik, miként ezt a közkeletű Moore-törvény előre meg is jósolta. Ez a nagyon egyszerű jóslat hosszú éveken keresztül időtállóan bizonyult, napjainkban azonban ez a dinamikus fejlődés megtorpant, vagy legalábbis jelentősen irányt változtatott [14]. A processzorgyártók kénytelenek voltak szembesülni azzal, hogy nem tudják az órajelfrekvenciákat az eredetileg elképzelt ütem szerint növelni tovább, és emiatt új megoldásokat kellett keresni a további fejlesztések számára. Ezek közül az utóbbi években a legszembetűnőbb a többmagos

processzorok megjelenése [15], amelyek az évenkénti teljesítménynövekedést egy meglehetősen kézenfekvő ötlettel biztosítják: egy processzoron belül bizonyos egységeket egyszerűen megdupláztak (négyszereztek, nyolcszoroztak), amelynek segítségével az eszköz elméleti számítási kapacitása továbbra is növekszik, kielégítve a piac igényeit.

A gyakorlatban azonban ez a szoftverfejlesztés számára egy drasztikus szemléletmód váltást igényelt. A szoftverfejlesztők ugyanis megszokhatták, hogy a nagy teljesítményigényű rendszerek esetében az optimalizálás mellett az időre is bátran számíthattak, hiszen a későbbiekben megjelenő hardver eszközökön futtatva az alkalmazásokat, azok gyakorlatilag automatikusan gyorsabban működtek. Napjainkban ez azonban már nem mondható ki ennyire egyértelműen, a többprocesszoros rendszerek maximális teljesítménye ugyanis csak az összes processzor(mag) egyidejű terhelésével, tehát párhuzamos algoritmusokkal aknázható ki. Emiatt nagyon sok hagyományos módszert érdemes optimalizálni, vagy ha ez nem lehetséges, akkor teljes egészében újragondolni, újratervezni a mai kor igényeinek megfelelően.

A többmagos processzorok megjelenésén túlmenően azonban időközben megjelentek egészen újszerű architektúrák is, mint például az általános célú számításokra használható grafikus kártyák (GPGPU-k). Az eredetileg a monitoron látható kép megjelenítésére kialakított hardvereszközök idővel különféle 3D számítások végrehajtására lettek alkalmasak, az ipari (és nem kevésbé az otthoni játékos) felhasználók igényeinek megfelelően egyre több egyszerű végrehajtóegységgel rendelkeztek; majd különféle *shader* generációváltások [16] után felmerült a lehetőség, hogy általános célú programok futtatására is alkalmasak legyenek. Ennek adott nagy lökést az Nvidia 2007-es lépése, amikor kiadott egy szoftverfejlesztési környezetet is a saját kártyáihoz (*CUDA 1.0*).

A szoftverfejlesztők számára egy teljesen új világ nyílt meg a több száz magot tartalmazó, ám meglehetősen egyszerű, és sok korlátot tartalmazó rendszerek programozásának lehetőségével. A nyers teljesítményt tekintve ezek az eszközök már évekkel ezelőtt is sokszorosán meghaladták az aktuális csúcs CPU-k feldolgozási kapacitását, így meglehetősen sok, nagy számításigénnyel rendelkező alkalmazás ügyében folynak kutatások azok GPGPU-n való alkalmazhatóságáról [17] [18]. A fejlesztések azonban meglehetősen nehézkesen haladnak, ugyanis az újszerű eszköz számos korláttal és speciális jellemzővel bír a hagyományos CPU-khoz viszonyítva, továbbá hiányoznak a kiforrott eszközök, és a több éves tapasztalatok is, amelyek a hagyományos fejlesztéseknél már rendelkezésre állnak.

Célkitűzések

Az általam elsődlegesen vizsgált hematoxilin-eozin festésű vastagbél szövetmintákban (1. ábra) a sejtmagok megbízható detektálása tűnik az egyik legkritikusabb pontnak a későbbi automatikus diagnosztikai rendszerek felé vezető úton. Egyrészt ezek a szöveti komponensek már önmagukban is számos információt tartalmazhatnak (elhelyezkedésük, méretük, mennyiségük) a későbbi diagnózishoz, másrészt számos további komponenseket (mirigyek, hámsejtek) azonosító eljárás létezik, amelyek a már előzőleg azonosított sejtmagokra alapozzák a további feldolgozást.



1. ábra: Hematoxinin-eosin festésű vastagbél szövetminta.

Emiatt mindenképpen lényeges, hogy rendelkezésre álljanak olyan módszerek, amelyek mind a pontosság, mind pedig a sebesség tekintetében megfelelnek a gyakorlati felhasználás követelményeinek. A különböző eljárások értékelésekor a sebesség mérése nem jelent különösebb problémát, a pontosság tekintetében azonban már meglehetősen nehéz összehasonlítani két, egymástól teljesen különböző alapelvek szerint működő algoritmust.

Céлом emiatt, hogy kidolgozzak egy mutatót, amely megmutatja, hogy egy adott sejtmagkeresési eljárás milyen pontos eredményt adott egy teszt képen (amennyiben ismert annak a referencia eredménye is). Mivel maga a mutató kiértékelése is meglehetősen számításigényes művelet lehet (a nagyfelbontású képek több ezer sejtmagot is tartalmazhatnak akár), emiatt megtervezek egy olyan módszert, ami a lehetőségekhez képest minél kevesebb lépéssel tudja visszaadni ezt az értéket nagyméretű minták esetében. Ez a mutató alkalmas kell, hogy legyen a már meglévő módszerek összehasonlítására, illetve a későbbiekben kifejlesztésre kerülő új módszerek előnyeinek objektív vizsgálatára is.

Az már az előzetes vizsgálatok alapján is nyilvánvaló, hogy nem lesz egy tökéletes módszer a sejtmagok keresésére, általában találkozhatunk kisebb pontosságot nyújtó gyors eljárásokkal, illetve nagyobb pontosságot nyújtó lassabb módszerekkel. Utóbbinak részben az az oka, hogy a különféle eljárások gyakran még a hagyományos képszegmentálási módszereken alapulnak, amelyek gyakran nem használják ki a napjainkban jellemző többprocesszoros rendszerek lehetőségeit, különösen nem az elmúlt években megjelent, legnagyobb teljesítményt rejtő GPGPU-két.

Célkitűzésem, hogy kidolgozzok egy olyan adatpárhuzamos sejtmagkeresési eljárást, amelynek pontossága legalább azonos a napjainkban rendelkezésre álló egyik legjobbnak tekinthető régiónövelési módszerrel, mindezt az eredményt azonban rövidebb futásidő alatt adja vissza. A régiónövelési eljárás két fő részből áll, magából a régiónövelési folyamatból, illetve az ehhez szükséges kiinduló (*seed*) pontok kereséséből. Céлом egy olyan, a GPGPU-k speciális követelményeihez is alkalmazkodó új, adatpárhuzamos módszer kidolgozása, amely

mindkét területen minél jobban kihasználja az eszközök rendelkezésre álló számítási kapacitását.

A különféle képszegmentálási eljárások egészen másféle paramétereket igényelnek, de általában elmondható, hogy bármilyen jó is a kidolgozott módszer, csak a megfelelő paraméterkészlet segítségével lehet elérni annak maximális pontosságát. A lehetséges paraméterek száma azonban gyakran olyan nagy, hogy az ideális paraméterkészlet megtalálása manuálisan reménytelen feladat.

Céлом, hogy kidolgozzak egy olyan evolúciós alapokon nyugvó módszert, amely alkalmas arra, hogy nagyméretű paramétertér esetén is képes ajánlani egy, a gyakorlatban jól használható paraméterkészletet a különféle sejtmagkeresési eljárások számára. A genetikus algoritmusok meglehetősen számításigényesek, így elkészítem egy elosztott rendszer terveit és implementációját is, hogy annak segítségével találjak egy ideális paraméterkészletet a régiónövelési algoritmus számára.

Vizsgálati módszerek

Szegmentáló algoritmusok pontosságának vizsgálata

Az egyes képfeldolgozási eljárások egészen különböző úton juthatnak el hasonló eredményekhez, így azok pontosságának meghatározását célszerűen nem az algoritmus vizsgálatával, hanem csak az általuk nyújtott eredmény alapján végezhetjük el. A szubjektivitás kiküszöbölése érdekében kidolgoztam egy mérőszámot, ami egy referenciaképhez viszonyítva határozza meg, hogy a teszt eredmény attól milyen mértékben tér el. A módszer egy-egy megfeleltetéssel párosítja a teszt és referencia sejtmagokat, ezek alapján megszámlálja a szokásos terminológia szerinti [19] *true-positive*, *true-negative*, *false-positive*, *false-negative* találatok számát, azzal kiegészítve, hogy a hibás találatok alapján az alábbi súlyozást használja:

$$\text{Súly}_i = \text{Min}(\text{Min}_j(\text{Táv}(T_i, R_j)) / \text{Átmérő}(R) * KT, 1) \quad (1)$$

Ahol:

- $\text{Táv}(T_i, R_j)$: A teszt sejtmag i . pixelének és a referencia sejtmag j . pixelének a távolsága.
- $\text{Átmérő}(R)$: A referencia sejtmag (legnagyobb) átmérője.
- KT : Konstans paraméter

Majd ezek alapján kiszámoltam a pontosságot:

$$\text{Pontosság} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (2)$$

Mivel ez a pontosság szövetmintánként különböző lehet, emiatt a vizsgálatot egymást követően több képre is le kell futtatni, és ezek számtani átlaga adja meg az algoritmus pontosságát.

A pontossági vizsgálatokhoz szükség van referenciaeredményekre is, ezt ebben az esetben képzett patológusok által annotált képek (a szövetmintákon kézzel megjelölték, hogy pontosan hol helyezkednek el a sejtmagok) jelentik.

Szegmentáló algoritmusok futásidejének vizsgálata

Mivel az egyes algoritmusok összehasonlítása során lényeges lehet azok futásideje, ezt az alábbi módszerrel mértem:

1. Az alkalmazás elindítása.
2. Egy mérés lefuttatása időmérés nélkül (bemelegítés). Ezt a futtatást a későbbiekben nem vesszük figyelembe, hiszen itt még előfordulhatnak különféle fordítási műveletek, illetve a gyorsítótárak sincsenek olyan állapotban, mint a későbbi mérések során.
3. N darab mérés lefuttatása, ami az alábbiakból áll:
 - a. Időmérés elindítása.
 - b. Teljes sejtmagkeresési folyamat végrehajtása.
 - c. Időmérés leállítása.
 - d. Futásidők feljegyzése.
4. Az alkalmazás leállítása.
5. Az előző mérések alapján átlagos futásidő kiszámítása.

Mivel a futásidő nagyban függ a feldolgozott szövetmintától (kép mérete, sejtmagok száma), így itt is több kép feldolgozása után határoztam meg egy átlagos futásidőt.

Szegmentáló algoritmusok összehasonlító vizsgálata

A kutatás során gyakran volt szükség két algoritmus hatékonyságának az összehasonlítására. Ennek során mindkét algoritmusra elvégeztem a pontossági és időmérési vizsgálatot, majd ezek eredményét hasonlítottam össze. Két szegmentálási algoritmus összehasonlítása során értelemszerűen mind az idő, mint a pontosság esetében csak az azonos módszerrel és azonos mintákon végrehajtott méréseket hasonlítottam össze.

Az idő és pontossági értékeket egy tetszőlegesen megválasztott súlyozással egy közös mérőszámként is lehetne kezelni, de végül azt a megoldást választottam, hogy ezeket külön-külön értékeltem, hiszen mindig az adott feladat és rendelkezésre álló erőforrások döntik el, hogy ezek közül melyik milyen súllyal számít.

Régiönövelés paramétereinek optimalizálása

A régiönövelési algoritmus lépései (előfeldolgozás, régiönövelés, összenőtt sejtek szétvágása) meglehetősen sok (az általunk használt implementációban 27 darab) paramétert igényelnek, amelyek pontos megválasztása esetén várhatjuk csak el, hogy a módszer a lehető legnagyobb hatékonysággal működjön. Maga az alapprobléma egyszerűen modellezhető, adott 27 darab,

egymástól függetlennek tekinthető paraméter megadott értékészlettel, adott egy algoritmus (és annak implementációja) ami szövetminták alapján a megadott paraméterekkel előállítja a mintában található sejtmagok listáját, illetve adott egy kiértékelő függvény, amely a rendelkezésre álló, orvosok által manuálisan annotált minták alapján megadja a vizsgált paraméterkészlethez tartozó pontosságot.

A rendelkezésre álló optimalizálási módszerek közül én az evolúciós módszereket részesítettem előnyben, egészen konkrétan egy genetikus algoritmust készítettem a feladat megoldására. Figyelembe véve a megoldás nagy számításigényét, mindezt egy elosztott környezetben implementáltam, a master-slave modellnek megfelelően.

Új tudományos eredmények

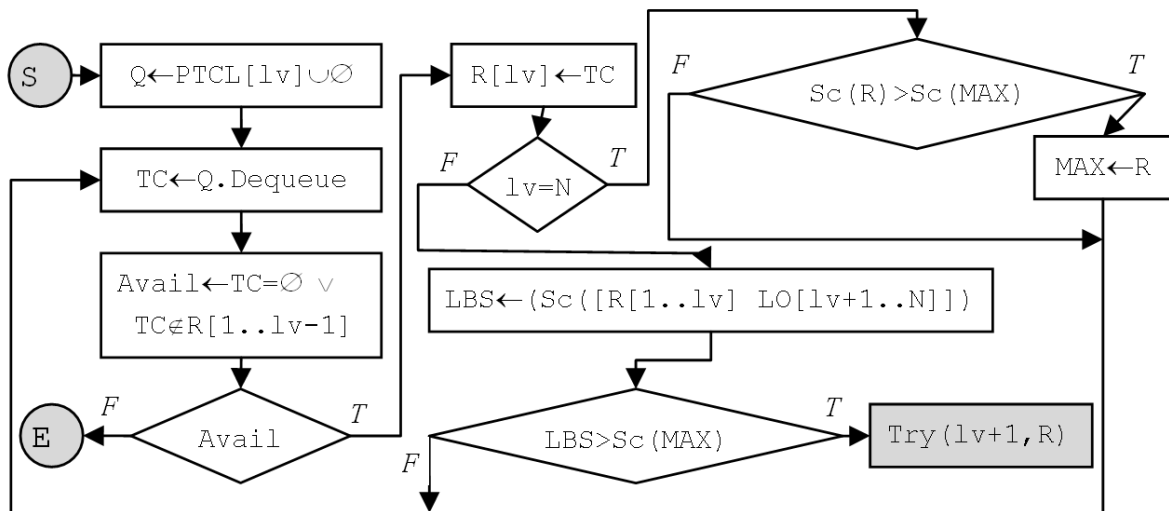
1. téziscsoport

Számos sejtmagkeresési eljárással találkozhatunk, ezek általában jósági vizsgálatokkal is járnak, ezek azonban különféle módszereket alkalmaznak az értékeléshez, így eredményük egymással nem összehasonlítható, másrészt az értékelések csak a hagyományos képfeldolgozó algoritmusokra kidolgozott általános módszereken alapulnak (pl. döntési tábla) és nem veszik figyelembe a sejtmagkeresési eljárások által támasztott speciális követelményeket.

Leggyakrabban a szegmentálási algoritmusok pontosságát egy teszt és egy referencia eredmény pixelenkénti összehasonlításával szokták meghatározni. Ez ugyan egy nagyon jól kezelhető mérőszámot ad arra vonatkozólag, hogy a kép pixeleit egymástól teljesen függetlenül vizsgálva, az egyes pixelekre vonatkozóan az algoritmus jól állapította-e meg, hogy azok egy sejtmag részei vagy sem. A gyakorlatban azonban ez nem ad kielégítő eredményt, hiszen gyakori, hogy több egymáshoz közel álló kisebb sejtmagot a szegmentáló algoritmus egy nagy sejtmagként határoz meg, amit ez az egyszerű ellenőrzés tökéletes eredményként fogadna el. Ehhez hasonlóan, a gyakorlatban nagyobb hibának szeretnék tekinteni, ha egy rendszer sok kisebb sejtmagot nem talált meg, mintha csak egy darab ugyanilyen területű nagy sejtmag maradt ki.

1.1. tézis: Eljárást dolgoztam ki, amely ötvözi a pixel- és objektumszintű értékelések előnyeit, így a meglévő módszereknél jobban alkalmazható szövetmintákban sejtmagokat kereső szegmentáló eljárások objektív összehasonlítására.

Az általam kifejlesztett módszer [20] nem csak a fent említett pixel szintű összehasonlítást végzi el, hanem először megpróbálja egymáshoz rendelni a teszt és a referencia eredményben található sejtmagokat, majd ezekre a sejtmagokra végzi el a részletes összehasonlítást. Ezzel a pontossági mérőszám tartalmazza annak az ellenőrzését is, hogy valóban a referencia eredményben meglévő számú és elhelyezkedésű sejtmagot talált-e a szegmentáló algoritmus, illetve ezen túlmenően a párokra a pixelenkénti összehasonlítást alkalmazva információt ad az egyes megtalált sejtmagok alakjának megfelelőségéről is.



1. algoritmus: Átfedő sejtmagokat vizsgáló visszalépéses keresés vázlata.

1.2. tézis: Kidolgoztam egy keresési algoritmust, amely nagy mennyiségű, egymást átfedő referencia és teszt sejtmagokat tartalmazó szegmentálási eredmények esetén a hagyományos keresésekhez képest kevesebb lépéssel határozza meg a legnagyobb pontossági értéket nyújtó párosításokat.

Az előzőleg felvázolt pontossági mutató a gyakorlatban jól használható, azonban kiszámítása meglehetősen erőforrás-igényes. A pixelenkénti vizsgálat előtt ugyanis szükség van az egyes teszt illetve referencia sejtmagok egymáshoz rendelésére, ez pedig egymást átfedő sejtmagok (illetve így az átfedéseken keresztül kialakuló hosszú láncok) miatt nem egyértelmű. A legnagyobb pontosságot nyújtó párosítás megtalálása több ezer sejtmag esetén használhatatlan futásidőt eredményezne, ezért kidolgoztam egy módszert (1. algoritmus), ami az egymást átfedő sejtmagokat klaszterekbe rendezi [21], majd egy speciális visszalépéses keresésnek köszönhetően a lineáris kereséshez képest kevesebb lépéssel meghatározza a legjobb referencia-teszt sejtmag párosítást [20].

2. téziscsoport

A szöveti mintákon sejtmagok keresésére számos szegmentáló algoritmus létezik, ezek közül az egyik legígéretesebb a különféle régiónövelési eljárások köre. Az algoritmus alapvető lépéseinek (előfeldolgozás, kezdőpont keresés, régiónövelés, utófeldolgozás) egy részét, illetve egyes rész műveleteket (szűrők, konvolúciós műveletek) már sikerült jól párhuzamosítani, a minél rövidebb futásidő érdekében azonban érdemes a kezdőpont keresés és a régiónövelés számára is kidolgozni új, adatkárhuzamos környezetben is működő algoritmusokat.

2.1. tézis: Kidolgoztam egy adatkárhuzamos régiónövelési eljárást, amely alkalmas szöveti mintákon párhuzamosan több kiindulópontból kiindulva sejtmagok detektálására.

Számítási kapacitás tekintetében a GPGPU alapú platformok jelentik az egyik legjobb alternatívát napjainkban [22], emiatt a megvalósított algoritmus alkalmazkodik a rendszer

által megkívánt kötöttségekhez. Mivel a GPGPU esetében több száz végrehajtó egységgel rendelkezünk, és a szálak számát még ennél is jóval nagyobbak érdemes választani azok teljes leterheléséhez, emiatt az algoritmus a lehető legnagyobb mértékű párhuzamosítást használja több szinten is: régiónövelés során az egyes kontúrponatok adatait külön szálak dolgozzák fel, emellett egy magasabb szinten, egyidőben több régiónövelés is futhat párhuzamosan egy kártyán [23].

2.2. tézis: Kidolgoztam egy adatpárhuzamos régiónövelési kiindulóponat keresési eljárást, amely alkalmas a régiónövelési eljárás számára egyidőben több kezdőponat kigyűjtésére.

Az egyes régiónövelések közötti kezdőponat keresés során szintén érdemes kihasználni a GPGPU által nyújtott lehetőségeket, ez pedig új, adatpárhuzamos keresési módszereket igényel, ami a szinkronizációs műveletek nehézsége miatt jelent különösen nagy kihívást. A jelentős sebességnövekedésen túlmenően ez amiatt is szükséges, mivel az előbb említett új régiónövelési módszer egy időben több növelést is tud kezelni párhuzamosan, így a kezdőponat keresést is úgy valósítottam meg, hogy ne csak egy ponatot, hanem egyidőben több lehetséges kiinduló ponatot is keressen, mivel csak így tudja maximálisan táplálni az ezt követő növeléseket [23].

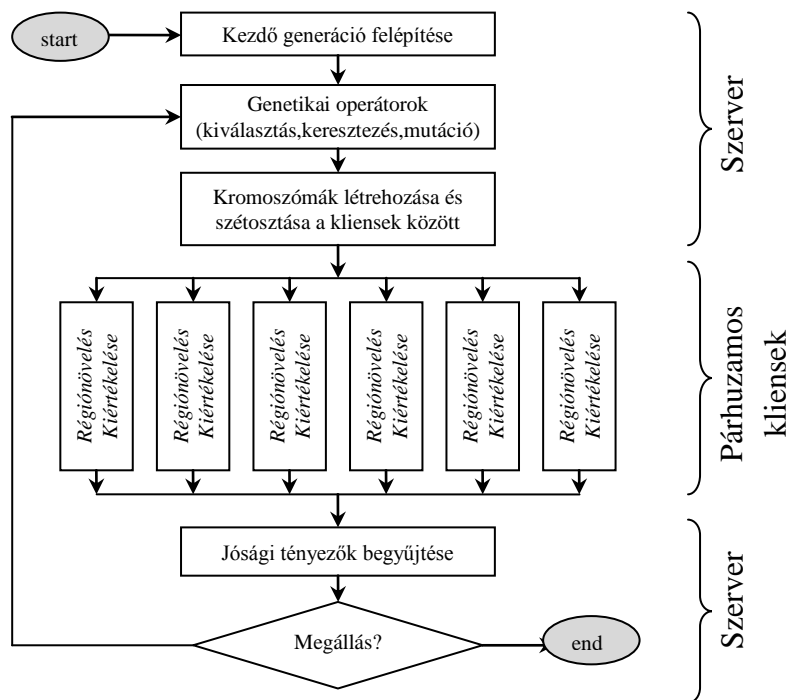
2.3. tézis: Igazoltam, hogy az újonnan kifejlesztett adatpárhuzamos régiónövelési eljárás (kiindulóponat keresés és régiónövelés) a jelenleg meglévő hagyományos régiónövelési eljárásokhoz képest azonos pontosságot ér el jelentősen kisebb futásidő mellett.

Implementáltam CUDA környezetben az adatpárhuzamos régiónövelési algoritmust, majd a fent ismertetett módszerekkel ellenőriztem a módszer pontosságát, illetve futásidejét. A vizsgálat eredményeképpen megállapítottam, hogy a CPU és a GPGPU alapú régiónövelés pontossága gyakorlatilag azonos, míg a GPGPU alapú változat futásideje általában fele/harmada a CPU alapúnak [23] [24].

3. téziscsoport

3.1. tézis: Kidolgoztam egy evolúciós algoritmust, amely alkalmas a régiónövelésen alapuló sejtmagkeresési eljárások nagy mennyiségű, egymástól függetlennek tekinthető paramétereinek optimalizálására. A mérési eredmények igazolják, hogy az általam kidolgozott módszer olyan új paraméterkészletet határoz meg, hogy az a korábban ismert legjobb paraméterkészletnél 4,3%-kal jobb eredményt szolgáltat.

A régiónövelési algoritmus meglehetősen nagy számú paramétert igényel (27 darab), ezek értékkészlete is meglehetősen nagy, illetve ismeretlen az egyes paraméterek egymásra vonatkozó hatása, így az optimális paraméterkészlet manuális megkeresése szinte lehetetlen, de még a hagyományos lineáris kereséssel is valószínűleg évekbe telne. Emiatt kidolgoztam egy genetikai algoritmust [25], amely alkalmas arra, hogy meghatározzon egy, a gyakorlatban is jól használható paraméterkészletet. A gyakorlati tesztek során kiderült, hogy a módszer beváltotta a hozzá fűzött reményeket, 440. generáció után a módszer által ajánlott paraméterkészlet már 4,3%-al nagyobb pontosságot nyújtott, mint az általunk használt régiónövelési implementációhoz tartozó jelenleg ismert legjobb paraméterkészlet [25].



2. algoritmus: Elosztott master-slave alapú genetikai algoritmus felépítése.

3.2. tézis: Statisztikai módszerekkel a rendelkezésre álló manuálisan annotált minták alapján megállapítottam a sejtmagkeresés paramétereit optimalizáló genetikai algoritmus számára szükséges kezdő generáció paramétereinek ideális értéktartományait.

A genetikai algoritmus számára kritikus lehet az első induló generációban található kromoszómák létrehozása. Hogy ezek minél közelebb kerüljenek az elvárt eredményhez, a rendelkezésre álló, manuálisan annotált mintákon található sejtmagok alapján különféle statisztikai módszerekkel megállapítottam néhány paraméterre olyan határértékeket [26] (sejtmag méret, intenzitás, körszerűség stb.), amelyek között célszerű a kezdő generáció géneinek konkrét értékeit megválasztani.

3.3. tézis: Megterveztem és implementáltam egy olyan keretrendszert, amely alkalmas genetikai algoritmusok elosztott futtatására, miközben moduláris felépítésének köszönhetően támogatja a genetikai operátorok és a kommunikációs protokollok szabad megválasztását. A mérési eredmények igazolják, hogy a régiókeresési eljárás paramétereinek optimalizálása során az elosztott rendszer futásideje jelentősen kisebb, mint a hagyományos szekvenciális megoldás időigénye.

A genetikai algoritmusok általában nagy számításigénnyel rendelkeznek, ez ebben az esetben kiemelten igaz, mivel a jóssági függvény kiértékeléséhez az egyes paraméterkészletekkel el kell végezni egy régió növelést illetve egy pontossági kiértékelést egymás után több szövetmintára is. A megoldást ebben az esetben is egy elosztott rendszer jelentette (2. algoritmus), amely a nagy erőforrás-igényre tekintettel egy hálózatokon keresztül

összekapcsolt gépeken futó, több száz kliens egyidejű futtatását is lehetővé teszi. Megterveztem egy ezeknek a feltételeknek megfelelő rendszert [27], mindezt olyan szempontok szerint, hogy mind az egyes genetikai operátorok, mind pedig a kliensek és a szerver közötti kommunikáció formája tetszőlegesen megválasztható legyen. A módszernek köszönhetően az evolúciós algoritmus futásideje jelentősen csökkent.

Az eredmények hasznosítási lehetősége

A képfeldolgozás egy meglehetősen nagy területet lefedő témakör, aminek jelen kutatás csak egy nagyon specializált területét, a szövetmintákon való sejtmagok keresését célozta meg. A vizsgált terület azonban gyakorlati szempontból nagyon fontosnak tekinthető, hiszen napjainkban egyre inkább várhatjuk a különböző automatikus diagnosztikai rendszerek megjelenését. Amikor konkrét betegek diagnosztikájáról beszélhetünk, akkor talán nem is kell részletesen elemezni, hogy milyen sokat jelent akár néhány százalékkal pontosabb kiértékelés, ami az evolúciós algoritmus által nyújtott jobb paraméterkészlet alapján elérhetővé vált.

Az általam kidolgozott GPGPU alapú régiónövelés azonos pontosság mellett kétszer/háromszor gyorsabb futásidőt eredményez. A relatív gyorsulás mellett érdemes figyelembe venni azt is, hogy a hagyományos CPU alapú régiónövelés egy nagyobb szövetmintát közel egy óra futásidővel tud csak feldolgozni, ehhez képest az elérhető 20 perc körüli eredmény nem csak egyszerű gyorsítást jelent, hanem a régiónövelést felveszi a gyakorlatban is használható módszerek közé.

Az általam kidolgozott pontossági mutató pedig a későbbiekben lehetőséget nyújt az újonnan megjelenő sejtmagkeresési eljárások összehasonlítására, ahol külön érdekes lehet az egyes újabb fejlesztések és a megelőző változatok közötti különbségek analízisére. A mutató a gyakorlatban jól használhatónak bizonyult, hiszen ezen a mutatón alapul az általam kidolgozott paraméter optimalizálás, illetve a GPGPU alapú régiónövelés pontosságának igazolása is.

Irodalmi hivatkozások listája

- [1] L. Ficsór, "Digital Microscopy in the Diagnosis of Gastrointestinal Histological Samples," Semmelweis University, 2nd Department of Internal Medicine, Budapest, PhD Thesis 2009.
- [2] A. Nagy, Z. Vámosy, "Super-resolution for Traditional and Omnidirectional Image Sequences," *Acta Polytechnica Hungarica*, vol. 6, no. 1, pp. 117-130, 2009.
- [3] E. Sorantin, E. Balogh, A. Vilanova i Bartrolí, K. Palágyi, L. G. Nyúl, F. Lindbichler, A. Ruppert, "Techniques of Virtual Dissection of the Colon Based on Spiral CT Data," in *Image Processing in Radiology: Current Applications*, 2008, pp. 257-268.
- [4] M. E. Adawy, Z. Shehab, H. Keshk, M. E. El Shourbagy, "A Fast Algorithm for Segmentation of Microscopic Cell Images," in *ITI 4th International Conference on*

Information & Communications Technology (ICICT '06), 10-12 Dec. 2006, pp. 1-1.

- [5] Y. Surut, P. Phukpattaranont, "Overlapping Cell Image Segmentation using Surface Splitting and Surface Merging Algorithms," in *Second APSIPA Annual Summit and Conference*, 2010, pp. 662-666.
- [6] J. Hukkanen, A. Hategan, E. Sabo, I. Tabus, "Segmentation of Cell Nuclei From Histological Images by Ellipse Fitting," in *18th European Signal Processing Conference (EUSIPCO-2010)*, Aalborg, 23-27 Aug. 2010, pp. 1219-1223.
- [7] R. Pohle, K. D. Toennies, "Segmentation of medical images using adaptive region growing," *SPIE*, vol. 4322, pp. 1337-1346, Jul. 2001.
- [8] Pannon University, "Algoritmus- és forráskódléírás a 3DHistech Kft. számára készített sejtmag-szegmentáló eljáráshoz," Veszprém, 2009.
- [9] L. Ficsór, V. S. Varga, A. Tagscherer, Zs. Tulassay, B. Molnár, "Automated classification of inflammation in colon histological sections based on digital microscopy and advanced image analysis," *Cytometry*, vol. 73A, no. 3, pp. 230–237, March 2008.
- [10] L. Krecsák, T. Micsik, G. Kiszler, T. Krenács, D. Szabó, V. Jónás, G. Császár, L. Czuni, P. Gurzó, L. Ficsor, B. Molnár, "Technical note on the validation of a semi-automated image analysis software application for estrogen and progesterone receptor detection in breast cancer," *Diagnostic Pathology*, vol. 6, Jan. 2011.
- [11] K. E. Emblem, F. G. Zoellner, A. Bjornerud, "A fully automated method for predicting glioma patient outcome from DSC imaging. A second reference to histopathology?," in *International Society for Magnetic Resonance in Medicine*, 2013, p. 281.
- [12] A. K. Jain, "Data clustering: 50 years beyond K-means," *Award winning papers from the 19th International Conference on Pattern Recognition* , vol. 31, no. 8, pp. 651-666, June 2010.
- [13] E. C. Smochină, "Image Processing Techniques and Segmentation Evaluation," Technical University "Gheorghe Asachi", Doctoral School of the Faculty of Automatic Control and Computer Engineering,.
- [14] D. Patterson, "The trouble with multi-core," *Spectrum*, vol. 47, no. 7, pp. 28-32, July 2010.
- [15] P. Gepner, M. F. Kowalik, "Multi-Core Processors: New Way to Achieve High System Performance," in *International Symposium on Parallel Computing in Electrical Engineering (PAR ELEC)*, 13-17 Sept. 2006, pp. 9-17.
- [16] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, J. C. Phillips, "GPU Computing," *Proceedings of the IEEE*, vol. 96, no. 5, pp. 879-899, May 2008.
- [17] N. Goswami, R. Shankar, M. Joshi, L. Tao, "Exploring GPGPU workloads: Characterization methodology, analysis and microarchitecture evaluation implications," in *International Symposium on Workload Characterization (IISWC)*, 2-4 Dec. 2010, pp.

1-10.

- [18] L. Hu, X. Che, Z. Xie, "GPGPU cloud: A paradigm for general purpose computing," *Tsinghua Science and Technology*, vol. 18, no. 1, pp. 22-23, Feb. 2013.
- [19] R. Kohavi, F. Provost, "Glossary of Terms," *Machine Learning*, vol. 30, no. 2, pp. 271-274, 1998.

A tézispontokhoz kapcsolódó tudományos közlemények

- [20] S. Szenasi, Z. Vamossy, M. Kozlovsky, "Evaluation and comparison of cell nuclei detection algorithms," in *IEEE 16th International Conference on Intelligent Engineering Systems (INES)*, Lisboa, 13-15 June 2012, pp. 469-475.
- [21] S. Szenasi, D. Jankó, "Black spot treatment system using a „hunting for irregular pattern” process and a safety knowledge-base," in *On safe roads in the XXI. Century*, Budapest, 2006.
- [22] S. Szénási, "GPGPU - Új eszközök az információbiztonság területén," *Hadmérnök*, vol. 4, no. 4, pp. 362-373, Dec. 2009.
- [23] S. Szénási, Z. Vámosy, M. Kozlovsky, "GPGPU-based data parallel region growing algorithm for cell nuclei detection," in *12th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, Budapest, 21-22 Nov. 2011, pp. 493-499.
- [24] A. Reményi, S. Szénási, I. Bándi, Z. Vámosy, G. Valcz, P. Bogdanov, Sz. Sergyán, M. Kozlovsky, "Parallel Biomedical Image Processing with GPGPUs in Cancer Research," in *3rd IEEE International Symposium on Logistics and Industrial Informatics (LINDI 2011)*, Budapest, 25-27 Aug. 2011, pp. 245-248.
- [25] S. Szénási, Z. Vámosy, M. Kozlovsky, "Evolutionary algorithm for optimizing parameters of GPGPU based image segmentation," *Acta Polytechnica hungarica*, 2013.
- [26] S. Szenasi, Z. Vamossy, M. Kozlovsky, "Preparing initial population of genetic algorithm for region growing parameter optimization," in *4th IEEE International Symposium on Logistics and Industrial Informatics (LINDI)*, 5-7 Sept. 2012, pp. 47-54.
- [27] S. Szénási, Z. Vámosy, M. Kozlovsky, "Implementation of distributed genetic algorithm for parameter optimization in cell nuclei detection project," *Acta Polytechnica Hungarica*, 2013.

További tudományos közlemények

- [28] Z. Vámosy, D. Sima, S. Szénási, A. Rövid, P. Kárász, Á. Miklós, S. Sergyán, Á. Tóth, "Párhuzamos számítástechnika modul az új technológiákhoz kapcsolódó megközelítésben", *Informatika a felsőoktatásban 2011 konferencia*, Debrecen, Magyarország, 2011.08.24-2011.08.26., pp. 766-772.

- [29] S. Szénási, D. Jankó, "Közúti baleseti helyazonosítás GPS-koordinátákkal: „POLICE-BAL” (P-BAL) PROGRAM", Közlekedéépítési Szemle, 2010, vol. 60, no. 11, pp. 15-19.
- [30] S. Szénási, "Java programozási nyelv oktatása C# alapokon", Informatika a felsőoktatásban 2008 konferencia, Debrecen, Magyarország, 2008.08.27-2008.08.29., p. 118
- [31] M. Sipos, S. Szénási, "A deklaratív és imperatív programozási paradigma alkalmazása a biztonságos szoftverfejlesztésben", Informatika a felsőoktatásban 2008 konferencia, Debrecen, Magyarország, 2008.08.27-2008.08.29., p. 118
- [32] D. Jankó, F. Kovácsi, S. Szénási, Zs. Tóth-Szabó, "Közúti szolgáltatási szint meghatározása forgalmi mérések adatai alapján", Közúti És Mélyépítési Szemle, 2008, vol. 58, no. 8, pp. 16-22.
- [33] A. György, S. Szénási, I. Vajda, "Szemi-automatikus tudáskiértékelés a vektoralgebrai feladatok példáján", Informatika a felsőoktatásban 2008 konferencia, Debrecen, Magyarország, 2008.08.27-2008.08.29., p. 156
- [34] D. Jankó, S. Szénási, "Internet-based decision-support system in the field of traffic safety on public road networks", 6th European Transport, Budapest, 2007, pp. 131-136.
- [35] D. Kotsis, G. Légrádi, G. Nagy, S. Szénási, "Többnyelvű programozástechnika", Panem Kiadó, 2007
- [36] S. Szénási, Á. Szeghegyi, M. Takács, "Fuzzy Logic Control Problems Simulation Based on Parametrized Operators", 3rd Romanian-Hungarian Joint Symposium on Applied Computational Intelligence (SACI), 2006.05.25-2006.05.26., Temesvár, pp. 25-34.
- [37] S. Szénási, D. Jankó, "Black spot treatment system using a hunting for irregular pattern process and a safety knowledge-base", On safe roads in the XXI. Century., Budapest, Magyarország, 2006, Paper V/7
- [38] S. Preitl, Precup Radu-Emil, J. Gáti, Gy. Kártyás, S. Szénási, "Enhanced Mixed Campus and Distance Higher Education", 3rd Romanian-Hungarian Joint Symposium on Applied Computational Intelligence (SACI), 2006.05.25-2006.05.26., Temesvár, pp. 150-152.
- [39] K. Katona, V. Póserné Oláh, S. Szénási, "A Delphi haladó eszközeinek alkalmazása", Matematika, fizika és számítástechnika oktatók XXX. Konferenciája, 2006.08.23-2006.08.25., Pécs