

MATEMATIČKI FAKULTET
UNIVERZITET U BEOGRADU

Marko Makarić 1090/2012

**Prebrojavanje ćelija metodom klasifikacije na snimcima
dobijenim fluorescentnim mikroskopom**

Master rad

Mentor: prof. dr Nenad Mitić

Beograd, 2017.

Mentor:

prof. dr Nenad Mitić

vanredni profesor, Matematički fakultet, Univerzitet u Beogradu

Članovi komisije:

prof. dr Nenad Mitić

vanredni profesor, Matematički fakultet, Univerzitet u Beogradu

prof. dr Saša Malkov

vanredni profesor, Matematički fakultet, Univerzitet u Beogradu

dr Miloš Beljanski

naučni savetnik, Institut za opštu i fizičku hemiju

Datum odbrane:

Sadržaj

Sadržaj.....	3
1. Uvod.....	1
2. Konfokalna mikroskopija i prikaz ćelija na snimcima dobijenim uz pomoć fluorescentnog mikroskopa.....	2
3. Tehnike obrade slika u računarstvu	6
3.1. Komponente koje su uključene u obradu slika u računarstvu.....	7
3.2. Osnovni koraci u procesu obrade slika u računarstvu.....	8
3.3. Tehnike obrade slika korišćene u ovom radu.....	12
4. Tehnike istraživanja podataka.....	18
4.1. Osnovne tehnike istraživanja podataka.....	20
4.2. Klasifikacija	22
5. Cilj rada.....	26
6. Opis, metode i pristup problemu.....	27
7. Rezultati i njihovo poređenje sa drugim programima.....	38
8. Zaključak i ideje za dalji rad.....	48
8.1. Ideje za dalji rad	49
9. Literatura.....	50
9.1. Korišćena literatura	50
9.2. Korisna literatura.....	50

1. Uvod

Prebrojavanje ćelija je procedura koja se u današnje vreme podrazumeva, kako u medicinske, tako i u biološke, naučne i istraživačke svrhe. Rezultati nalaze primenu u medicini kod lečenja bolesti, gde se prebrojavanjem ćelija može zaključiti o zdravstvenom stanju pacijenta ili se može saznati kako se imuni sistem bori sa nekom infekcijom. U biologiji prebrojavanje ćelija nalazi primenu u raznim eksperimentima gde taj broj pruža informaciju o statusu nekog eksperimenta i na taj način može da usmeri dalje istraživanje ili da pomogne u naučnim istraživanjima gde se istražuje kojom brzinom napreduju mikroorganizmi (drugim rečima kojom brzinom se stvaraju nove ćelije). Međutim, problem koji se javlja pre svega na fakultetima, koji su predstavnici istraživačkih i naučnih centara, je u tome da dobri programi koji vrše prebrojavanje ćelija zahtevaju skupe licence da bi omogućili njihovo legalno korišćenje. Pitanje koje se samo nameće je zašto ti programi zahtevaju tako skupe licence?

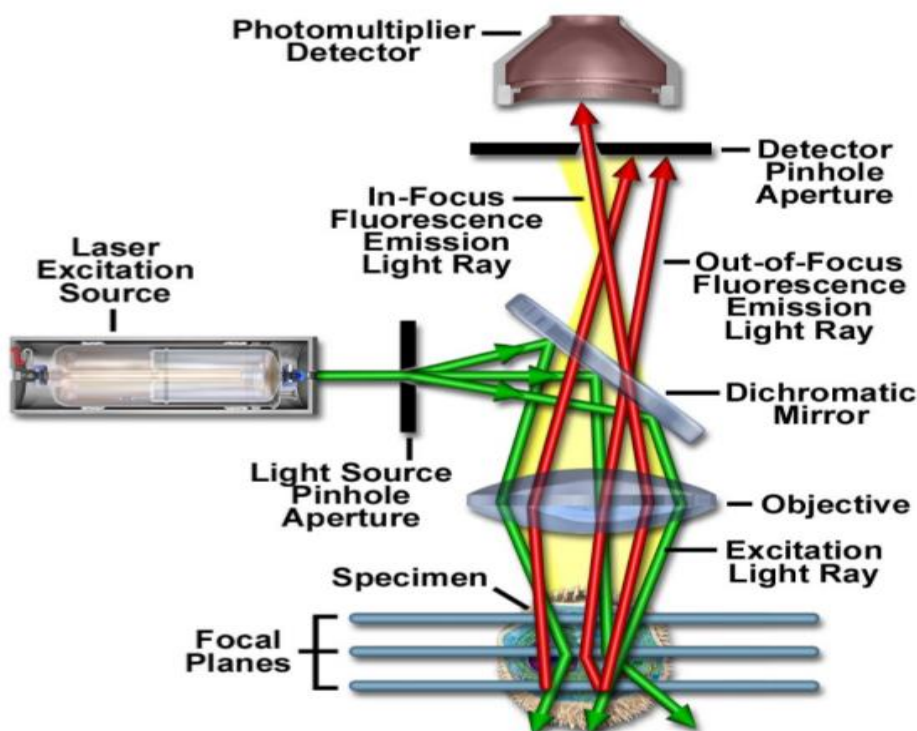
Odgovor leži u tome da prebrojavanje ćelija predstavlja složeniji problem nego što se čini na prvi pogled.

U ovom radu se opisuje problem prebrojavanja ćelija na snimcima dobijenim fluorescentnim mikroskopom i predstavlja se potencijalno rešenje. Rešenje koje će biti predstavljeno kombinuje osnovne metode i tehnike obrade slika u računarstvu uz tehnike istraživanja podataka.

Slike koje su se koristile u izradi ovog rada su slike koje su korišćene u istraživanjima na Biološkom fakultetu u Beogradu. Dobijene su korišćenjem fluorescentnog mikroskopa i obrađene u programu otvorene licence *CellC*. Dostupni su i rezultati prebrojavanja i obrade slika u programu *Imaris* kao predstavniku grupe programa koji zahteva plaćenu licencu za korišćenje. Navedeni programi su uzeti za poređenje rezultata sa rezultatima koje daje rešenje predstavljeno u ovom radu.

2. Konfokalna mikroskopija i prikaz ćelija na snimcima dobijenim uz pomoć fluorescentnog mikroskopa

Konfokalna mikroskopija predstavlja jedan od značajnih koraka napred kada je u pitanju razvoj optičkih mikroskopa i danas je postala neizostavna i nezamenjiva tehnika za molekularna istraživanja i biološka istraživanja nad ćelijama. Konfokalna mikroskopija funkcioniše na principu fluorescencije koja se aktivira uz pomoć jakog snopa svetlosti lasera. Laser osvetljava određene delove posmatranih objekata i na taj način samo ti delovi bivaju osvetljeni i u fokusu posmatraču pri pravljenju snimka. Pri snimanju se posmatra jedna fokalna ravan i kako se svetlo kreće po toj fokalnoj ravni tako je moguće da se formira dvodimenzionalni snimak posmatranih objekata, dok je ostale ravni moguće ignorisati. Ako bi se izvršio niz snimanja na različitim fokalnim ravnima moguće bi bilo i rekonstruisati i trodimenzionalni snimak posmatranih objekata [1]. Bolji uvid o tome kako radi konfokalna mikroskopija je moguće videti na slici 1a [3].

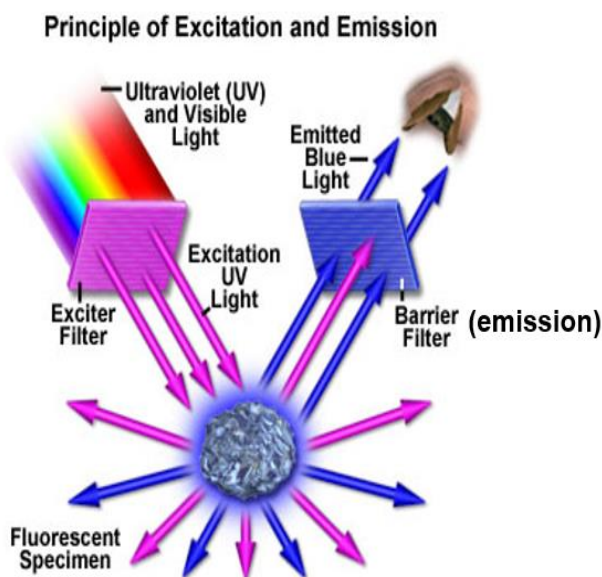


Slika 1a Šematski prikaz delova i načina rada pri snimanju konfokalnim mikroskopom

Ono što je važno u konfokalnoj mikroskopiji je da je istraživanje moguće vršiti i na živim ćelijama što je od izuzetnog značaja za istraživače. U problemu koji je tema ovog rada istraživači vrše različite eksperimente nad ćelijama i prate ponašanje ćelija nakon različitih tretmana prebrojavajući ih. Osnovni zadatak konfokalnih mikroskopa je izrada kvalitetnih i upotrebljivih snimaka (slika). U procesu dobijanja upotrebljivih slika za potrebe istraživanja nailazimo na dva

suštinska izazova. Prvi je da oštećenja na slikama svedemo na najmanju moguću meru, u smislu da kvalitet podataka na slici bude na najvišem nivou (odnos signal-šum da bude na prihvatljivom nivou). Drugi problem je da pri samom snimanju slika ćelije, ili tkiva koja čine ćelije, borave u okruženju koje je za njih prirodno, odnosno koje im omogućava da vrše svoje funkcije i ćelijske procese neometano. Snimanje slika koristeći fluorescentni mikroskop se zasniva na odsjaju fluorescentno obeleženih proteina ili nekih drugih ćelijskih molekula pod određenim svetlom različitih talasnih dužina. Problem je odrediti intezitet i koliko dugo treba primeniti svetlo određene talasne dužine da bi se dobio najbolji efekat i sadržajno upotrebljiva slika. Svetlo jakog inteziteta samo po sebi oštećuje ćelije (posebno je to danas poznato ako se zađe u UV spektar) [4].

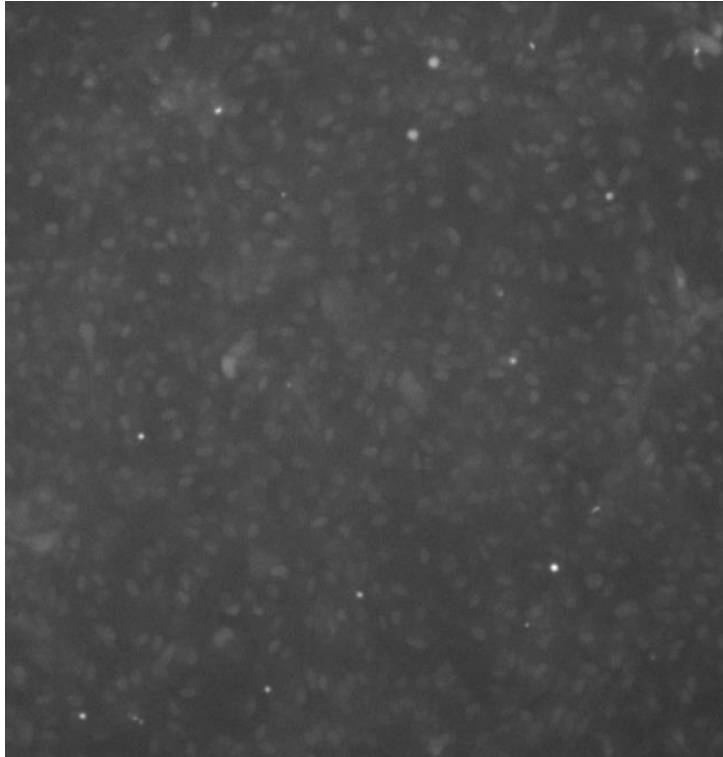
Osnova odsjaja u snimanju ćelija fluorescentnim mikroskopom se zasniva na fluorescentnom pigmentu čiji princip delovanja je prikazan na slici 1b [2].



Nakon svake primene svetlosnog zračenja na ćeliju, ona gubi svoja fluorescentna svojstva, odnosno deo svog fluorescentnog pigmenta koji je nepovratno uništen. Gubitkom fluorescentnog pigmenta, dobijeni snimci mogu sadržati nepravilne podatke i različite anomalije na slici koje mogu dovesti do pogrešnog tumačenja. Jedini siguran način da se gubitak pigmenta svede na najmanju moguću meru, zajedno sa oštećenjima slike, je da se vreme primene i intezitet svetla ograniče koliko je to moguće, naravno uz uslov da je odnos signal-šum na prihvatljivom nivou kao i da su slike sadržajno korisne za potrebe eksperimenta[4].

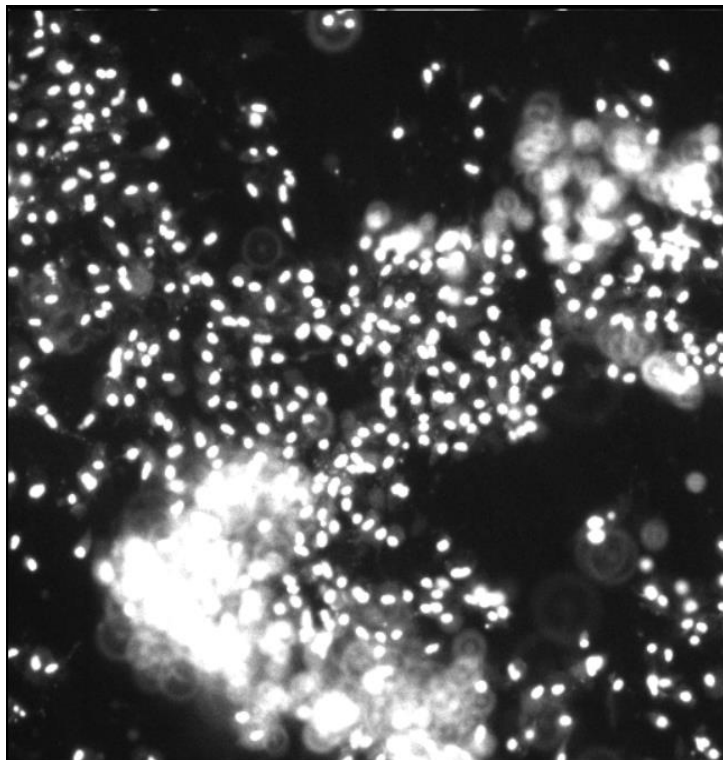
Slika 1b Princip fluorescencije u mikroskopiji

Na slikama 2a, 2b i 2c su predstavljeni neki od snimaka koji su dobijeni kao materijal za rad od istraživača sa Biološkog fakulteta. Na slikama su opisani problemi, ali i dobre strane rada sa konfokalnim mikroskopima.



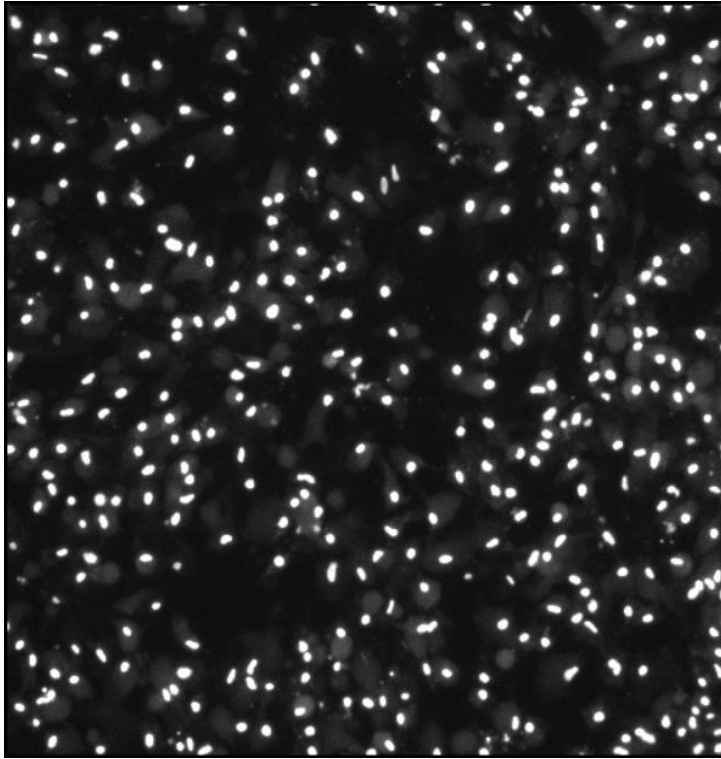
Slika 2a. Gubitak pigmenta

Na slici 2a prikazan je snimak koji je odličan primer ćelija koje su izgubile svoju osobinu fluorescentnog pigmenta gde je njihov odsjaj na niskom nivou što je najverovatnije posledica predugog izlaganja svetlosnom zračenju.



Slika 2b. Prejak intenzitet svetlosti

Slika 2b. prikazuje snimak gde se jasno vidi kako prejak intenzitet svetlosnog zračenja u kombinaciji sa fluorescentnim svojstvima ćelija može da loše utiče na kvalitet sadržaja samog snimka. Posebno je to izraženo u donjem delu snimka gde je jako teško videti broj ćelija koje se nalaze u tom delu.



Slika 2c. Kvalitetan snimak

Na slici 2c je prikazan snimak odličnog kvaliteta na kome se jasno vidi fluorescentni pigment očuvan kod ćelija, intezitet svetlosnog zračenja i odnos signal-šum na prihvatljivom nivou.

Konkretno za potrebe ovog rada i za eksperimente na koje je ovaj rad usmeren bitan je broj ćelija koje se nalaze na snimku, odnosno problem prebrojavanja ćelija. Na taj problem utiče kvalitet dobijenog snimka jer ako ćelije izgube svoj fluorescentni pigment ili intezitet svetlosnog zračenja ne bude odgovarajućeg nivoa prilično se otežava početni problem. U svakom slučaju bez obzira na kvalitet snimka, pre samog prebrojavanja snimak mora proći prvo kroz proces obrade slike koristeći tehnike obrade slika u računarstvu.

3. Tehnike obrade slika u računarstvu

Vid je najnaprednije od svih ljudskih čula i zbog toga ne čudi da slike igraju veoma važnu ulogu u svakodnevnom životu. Obrada slika u računarstvu je jedna od grana koja poslednjih godina izaziva sve više interesovanja među ljudima kojima su računari svakodnevnica. Obrada slika u računarstvu se posmatra iz dva ugla koji predstavljaju ujedno i dva načina primene različitih tehnika obrade. Jedna strana predstavlja unapređenje i bolji prikaz samog sadržaja slike kako bi čovek (posmatrač) bolje video i razumeo sadržaj slike, dok druga strana predstavlja obradu slike sa ciljem da dalju analizu može da izvede i sam računar.

Da bi se shvatilo šta predstavlja obrada slike i koje su tehnike u računarstvu, potrebno je prvo definisati sliku i šta ona predstavlja u računarstvu. Slika može biti definisana kao dvodimenzionalna funkcija $f(x,y)$ gde su x i y prostorne koordinate, a vrednost funkcije f za bilo koji par koordinata (x,y) naziva se intezitet ili nivo boje slike u toj tački. Kada su x , y i vrednosti za intezitet funkcije f konačne diskretne vrednosti takvu sliku nazivamo digitalnom slikom. Tehnike obrade slika u računarstvu se odnose na digitalne slike i njihovu obradu. Iz ove definicije možemo da zaključimo da se slika sastoji od određenog broja elemenata. Te elemente nazivamo elementi slike odnosno kako je to opšte prihvaćeno u računarstvu – pikseli [5]. Izvori dobijanja digitalnih slika su raznovrsni; zahvaljujući razvoju tehnika i mašina postoji mogućnost obrade i dobijanja slika stvari i pojava koje nisu vidljive ljudskom oku. Takav primer su i fluorescentni mikroskopi koji su korišćeni kao izvor za dobijanje digitalnih slika u ovom radu. Ne postoji jasna definicija šta predstavlja obrada slike. Razlog za to je uska povezanost obrade sa analizom slike i ne može se sa sigurnošću reći kada se završava obrada slike i počinje analiza. Ipak u računarstvu je prihvaćena paradigma koja obradu slike deli na 3 nivoa. To su: obrada osnovnog nivoa, obrada srednjeg nivoa i obrada naprednog nivoa [5].

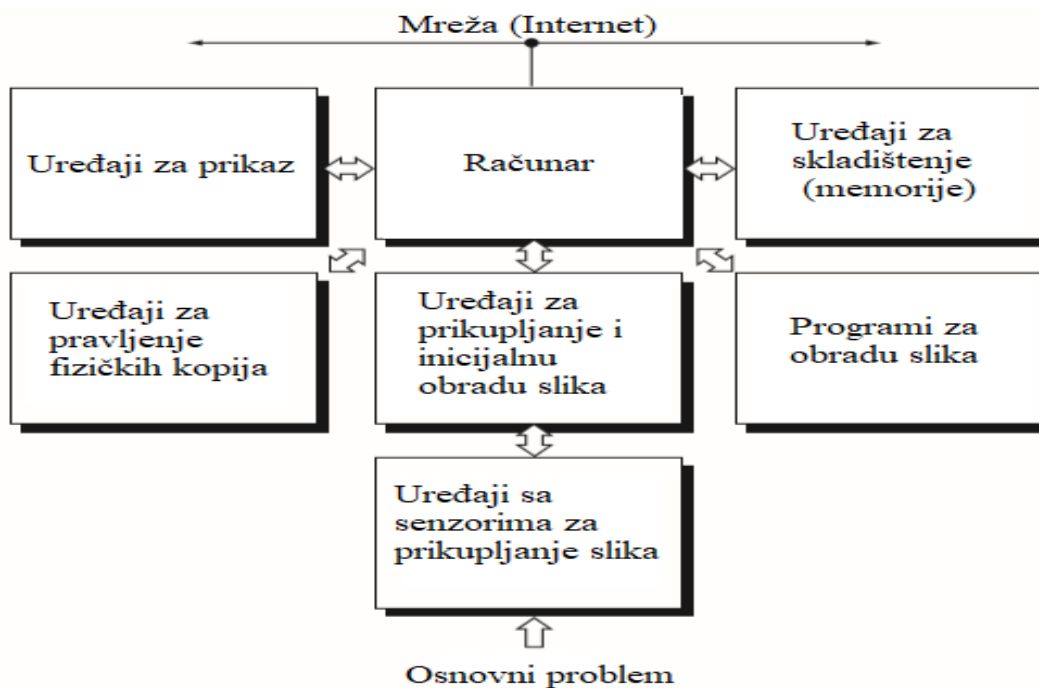
- **Obrada osnovnog nivoa** uključuje osnovne operacije u digitalnoj obradi slika, i najčešće se naziva pretprocesiranje. Uloga pretprocesiranja je u otklanjanju šuma, u poboljšanju kontrasta slike i izoštravanju slike. Obrada osnovnog nivoa se naziva tako jer je i ulazni podatak obrade slika i izlazni (rezultujući) podatak takođe slika.
- **Obrada srednjeg nivoa** uključuje operacije kao što je deljenje slike na regione ili objekte, opis različitih objekata na način da budu prihvatljivi za dalju obradu kao i klasifikaciju (prepoznavanje) određenih objekata. Ono što to je karakteristično za obradu srednjeg nivoa je da je ulazni podatak slika, a izlazni podatak ne mora da bude slika, već može da bude i izvučen skup objekata ili nekih drugih atributa slike (npr. ivice, konture).
- **Obrada naprednog nivoa** podrazumeva davanje značenja prepoznatim objektima, odnosno već uključuje i analizu prethodnih obrada i izvođenje zaključaka koji predstavljaju krajnji rezultat obrade slike.

Da bi se u ovom trenutku bolje razumeli ovi koncepte biće naveden uprošćen primer kako se odvija proces obrade slike u ovom radu. Kasnije će metode i sam način obrade biti detaljnije objašnjeni.

Ako se posmatra problem prebrojavanja broja ćelija na slici dobijenoj fluorescentnim mikroskopom, slika prvo prolazi kroz proces pretprocesiranja (obrada osnovnog nivoa) gde se uklanja šum, izoštravaju bitni delovi i priprema za obradu srednjeg nivoa. U obradi srednjeg nivoa koristeći tehnike obrade slika se prepoznaju ćelije kao posebne jedinice ili grupe ćelija da bi potom u obradi naprednog nivoa došlo do davanja značenja i prebrojavanja prepoznatih ćelija. Nakon toga sledi izvođenje zaključaka oko rezultata prebrojavanja i značaju tog prebrojavanja za određeni eksperiment.

3.1. Komponente koje su uključene u obradu slika u računarstvu

U procesu obrade slika u računarstvu učestvuje niz komponenti, od samih uređaja koji se koriste za proizvodnju slika do računara i drugih mašina specijalizovanih za obradu, do uređaja za prikaz i skladištenje slika i rezultata obrade. Danas postoje i uređaji koji u svom sastavu sadrže više komponenti tako da su sposobni da sami vrše različite obrade slika. Osnovne komponente koje čine proces obrade slika u računarstvu su predstavljene na slici 3.1 [5].



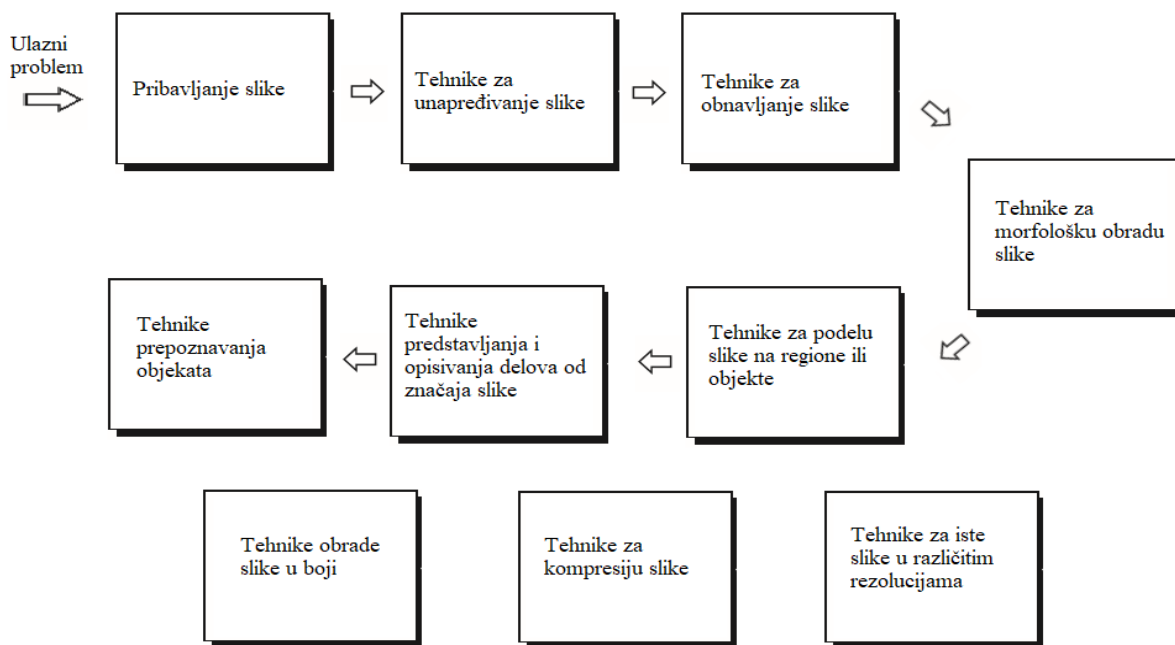
Slika 3.1 Komponente obrade slika u računarstvu

Uređaji sa sensorima za prikupljanje slika predstavljaju fizičke uređaje koji imaju sposobnost prikupljanja odnosno beleženja slika. Da bi se dobila digitalna slika neophodni su **uređaji za**

prikupljanje i inicijalnu obradu slika koji u saradnji sa uređajima sa sa senzorima kao rezultat daju digitalnu sliku. Digitalna slika se dobija inicijalnom obradom kojom se dobijeni signal prevodi u diskretni i proizvodi digitalnu sliku. Ona predstavlja ulaz za obradu **računaru** koji dalje koristi **programe za obradu slika** koji u sebi imaju potrebne funkcije za obradu slike sa ciljem rešavanja **osnovnog problema**. Nakon što je slika obrađena može biti sačuvana na **uređajima za skladištenje**, da se uz pomoć **uređaja za pravljenje fizičkih kopija** napravi fizička kopija, ili da se jednostavno prikaže na **uređajima za prikaz**. Na kraju sve to je moguće i podeliti na mreži što predstavlja deljenje rezultata sa drugima.

3.2. Osnovni koraci u procesu obrade slika u računarstvu

Osnovni koraci u procesu obrade slika u računarstvu se mogu videti na Slici 3.2 [5].



Slika 3.2 Osnovni koraci u procesu obrade slika

Proces obrade slike zavisi od početnog problema. U odnosu na početni problem biraju se koraci koji se obično izvršavaju obeleženim redosledom. Svi navedeni koraci predstavljaju osnovne korake obrade slika u računarstvu i obrada može prestati u bilo kom koraku ukoliko je izlaz iz tog koraka rezultat odnosno rešenje ulaznog problema. Koraci u obradi slike, prema [5] su:

- **Pribavljanje slike** predstavlja osnovni i neizostavni korak u kome se dolazi do digitalne slike koja predstavlja ulazni podatak za obradu. U današnje vreme sve je više uređaja koji su sposobni da proizvedu digitalnu sliku.
- **Tehnike za unapređivanje slike** se koriste da bi se kvalitet sadržaja digitalne slike poboljšao u odnosu na originalnu sliku. Različite tehnike se koriste za različite tipove slika jer ne daju sve tehnike iste rezultate za različite tipove i izvore iz kojih se dobijaju digitalne slike. Za slike dobijene fluorescentnim mikroskopom se koriste drugačije tehnike od onih koje se primenjuju za slike dobijene satelitskim snimanjem zemljine površine. Jedna od vrlo popularnih tehnika za unapređivanje slike je filtriranje koje može da se koristi za izoštravanje ili omekšavanje slike ili nekog njenog dela. Postoji više različitih filtera čija primena i način primene zavise od situacije i ulaznog problema.
- **Tehnike za obnavljanje slike** predstavljaju tehnike koje se koriste najčešće da bi se ispravile greške nastale prilikom procesa pribavljanja slike. Problemi sa kamerom, loša podešavanja mikroskopa ili geometrijski problemi pogrešnog pogleda su neki od prostih primera kako dolazi do grešaka na slici. Za razliku od tehnika za unapređivanje slike koje su zasnovane više na subjektivnom utisku i osećaju kako neka tehnika vizuelno unapređuje sliku, tehnike za obnavljanje slike su zasnovane na matematičkim modelima gde primenu nalazi i teorija verovatnoće u procesu obnavljanja sadržaja slike. Još jedna stvar koja odvaja tehnike unapređivanja od tehnika obnavljanja je da se kod tehnika obnavljanja prepoznaje problem na osnovu poznavanja sadržaja slike, i na osnovu toga se izvodi pretpostavka kako slika treba da izgleda. Kao najčešći problem koji rešavaju tehnike za obnavljanje slike je problem eliminisanja šuma iz originalne slike. Tehnike koje se najčešće primenjuju su različite tehnike filtriranja.
- **Tehnike za morfološku obradu slike** su zasnovane na matematičkom značenju reči morfologija. Ako se u biologiji morfologija odnosi na strukturu i oblik biljaka i životinja tako u slučaju slika morfološka obrada obuhvata izdvajanje ivica, skeleta, konveksnog omotača i različitih drugih geometrijskih oblika. U matematičkoj morfologiji osnovu predstavlja teorija skupova, i za obrade slike ove vrste, sama slika se posmatra kao matematički skup gde elementi slike pripadaju određenim skupovima. Osnovu tehnika i algoritama za morfološku obradu čine dve operacije erozija i dilatacija. Uprošćeno erozija predstavlja operaciju morfološkog filtriranja gde se primenom tehnike u rezultujućoj slici izostavljaju elementi koji su manji od elementa koji se koristi kao strukturni element filtera. Za razliku od erozije, dilatacija predstavlja operaciju morfološkog filtriranja gde se u rezultujućoj slici uvećavaju (podebljavaju) sitniji delovi ili isprekidani delovi slike u odnosu na strukturni element filtera.

- **Tehnike za podelu slike na regione ili objekte** predstavljaju tehnike koje dele sliku na njene sastavne delove ili objekte. Nivo podele i detalja izdvajanja primenom tehnika za podelu zavisi od ulaznog problema koji se rešava. U procesu obrade slika tehnike za podelu slike imaju najteži zadatak, jer od rezultata dobijenih njihovom primenom može da zavisi da li je konačan rezultat procesa obrade slike da li je uspešan ili ne. Imajući to u vidu, tehnikama za podelu je potrebno posvetiti više vremena kako bi rezultati njihove primene bili što bolji. Većina tehnika za podelu slika je zasnovana jednoj od dve osobine vrednosti inteziteta elemenata slike a to su povezanost i sličnost. Povezanost kao osobina primenu nalazi u velikim promenama inteziteta vrednosti što u najvećem broju slučajeva otkriva ivicu nekog objekta ili oblika. Sličnost u vrednosti inteziteta se koristi za podelu na regione koji su slični po vrednosti ili u određivanju praga vrednosti koje elementi moraju da ispunjavaju da bi prošli obradu.
- **Tehnike predavljanja i opisivanja delova od značaja slike** predstavljaju korak u procesu obrade slika koji gotovo uvek ide nakon primene tehnika za podelu slike, jer rezultat primene tih tehnika predavlja delove za koje treba da se utvrde načini predavljanja i opisa kako bi se nastavila dalja obrada. Teorijski gledano, nakon podele slike na regione ili objekte postoje dva moguća izbora za dalji rad. Prvi je predavljanje regiona/objekata korišćenjem njihovih spoljašnjih granica (ivica), a drugi je predavljanje korišćenjem unutrašnjih piksela koji sačinjavaju taj region/objekat. Izbor načina predavljanja je samo deo pripreme podataka za računar. Sledeći korak predavlja opis izabranog načina predavljanja, na primer ako je izabran način predavljanja koristeći ivice regiona/objekata na koji način se opisuju ivice (dužina, širina, broj tačaka, neprekidnost i slične osobine). Tehnike predavljanja koje koriste spoljašnje granice najčešće se koriste u problemima gde se prepoznaju različiti oblici na slici, dok se tehnike predavljanja koje koriste unutrašnje piksele najčešće koriste kod prepoznavanja različitih boja ili geografskih osobina. Ponekad je potrebno koristiti i oba pristupa, ali u svakom slučaju izbor osobina izabranih za opis delova od značaja je od izuzetnog značaja jer opisuje delove koji su glavni predmet obrade i izabrane osobine treba da budu što manje osetljive na različite veličine objekata/regiona, rotacije, translacije i slično.
- **Tehnike prepoznavanja objekata** su tehnike koje kao rezultat primene određuju tip objekta, odnosno na neki način ga svrstavaju u neku grupu objekata koji su poznati od ranije. U procesu obrade slika moguće je prepoznavati različite objekte i uzorke. Da bi prepoznavanje bilo olakšano potrebno je da opis samog objekta ili uzorka bude što je moguće bolji. Prepoznavanje je moguće posmatrati iz dva ugla. Prvi ugao je na osnovu kvantitativnih osobina kao što su dužina, širina ili oblast koju zauzima neki objekat ili region, a drugi ugao je na osnovu kvalitativnih osobina u strukturnom smislu od čega je sačinjen neki objekat koji su delovi i koliko ih ima i izbor načina prepoznavanja u najvećoj meri zavisi od ulaznog problema. Centralnu temu u procesu prepoznavanja čini

koncept učenja na osnovu uzoraka odnosno primera ulaznog problema. U zavisnosti od kvaliteta opisa zavise i rezultati učenja a od njih direktno zavise i rezultati samog prepoznavanja. Neke od najpoznatijih tehnika za prepoznavanje objekata su tehnike podudaranja, tehnike zasnovane na verovatnoći i klasifikacija.

- **Tehnike obrade slike u boji** su tehnike koje imaju sve veću primenu jer broj uređaja koji može da proizvede digitalnu sliku u boji i digitalnih slika koje su dostupne raste velikom brzinom. Tehnike obrade slike u boji su većinom zasnovane na tome da boja sama po sebi predstavlja izuzetnu opisnu osobinu koja može da pomogne u prepoznavanju nečega. Prednost slika u boji je i u tome što ljudsko oko razlikuje veliki spektar različitih boja i lakše primećuje razlike nego kad su u pitanju slike gde je potrebno samo razlikovati nijanse sive boje. Tehnike obrade slike u boji se dele u dve velike celine. Prvu celinu predstavljaju tehnike za obradu u punom spektru boja, dok drugu celinu čine tehnike koje su zasnovane na pseudo bojenju. Tehnike za obradu u punom spektru boja se primenjuju nad slikama koje su proizvedene uz pomoć kamera i drugih senzora koji imaju mogućnosti da proizvedu slike u punom spektru. Danas ove tehnike imaju široku primenu i vrlo su zastupljene u zbog mogućnosti objava i različitih načina deljenja slika na Internetu. Tehnike pseudo bojenja su tehnike koje se zasnivaju na dodeljivanju boja različitim delovima slike u zavisnosti od nekih osobina elemenata slike.
- **Tehnike za kompresiju slike** za razliku od prethodnih opisanih tehnika i koraka ne daju vidljive rezultate u smislu neke vizuelne obrade, već se koriste radi boljeg skladištenja i čuvanja slika u digitalnom obliku kao i njihovog deljenja preko mreže. Da bi se bolje razumela potreba za kompresijom slika može se posmatrati sledeći primer: Koja je potrebna veličina prostora za čuvanje jednog filma u standardnom formatu kvaliteta¹? Ako znamo da film predstavlja niz slika koje se smenjuju brzinom od, u proseku, 30 različitih slika u sekundi, računica bi izgledala **ovako: 30 slika u sekundi * (720 * 480) elemenata slike po jednoj slici * 3 bajta po elementu slike = 31.104.000 bajtova je potrebno za jednu sekundu.**

Ako se pretpostavi da film prosečno traje 2 časa dolazi se do sledeće računice: $31,104,000$ bajtova po sekundi * 7200 sekundi u 2 časa $\approx 2.24 * 10^{11}$ (oko 224 gigabajta) je potrebno za skladištenje filma u trajanju od 2 časa. Ukoliko je potrebno da se takav film sačuva na DVD disku veličine u proseku 8.5 gigabajta, svaka slika mora biti kompresovana približno u proseku 26.3 puta da bi film mogao da bude sačuvan na toj vrsti medijuma. U skladu sa time stepen kompresije za kvalitetnije formate treba da bude još veći. Postoje dva osnovna pristupa za kompresiju slika. Prvi pristup je kompresija bez gubitaka informacija (eng. *lossless*), a drugi pristup je kompresija sa gubitkom

¹ Standardni format kvaliteta (eng. *SD*) podrazumeva predstavljanje slike koristeći niz piksela od $720 * 480 * 24$ bita

informacija (eng. *lossy*). U zavisnosti od ulaznog problema, kao i od mogućnosti mreže ili uređaja za prikaz se bira koji pristup će se koristiti.

- **Tehnike obrade za iste slike u različitim rezolucijama**² kao što se može naslutiti iz samog naziva predstavljaju tehnike za obradu slika koje su prikazane i analizirane u više različitih rezolucija. Potreba za takvom obradom slika i promenama rezolucije leži u tome da neki detalji mogu da prodju neopaženi u određenim rezolucijama dok isti ti detalji u drugim rezolucijama su više nego uočljivi. Ako se posmatraju sitni detalji gde su boje slične i razlike teško uočljive obično se takve slike posmatraju u visokim rezolucijama, dok se u slučaju da je ulazni problem razlikovanje nekih boja ili velikih delova slike gde detalji nisu bitni koriste slike u nižim rezolucijama. Svakako obrada uz pomoć različitih rezolucija predstavlja prednost i može se koristiti za potvrdu rezultata analiza iz drugih rezolucija. Tehnike koje se koriste za obradu slika u različitim rezolucijama su uglavnom tehnike zasnovane na teoriji talasa i transformacijama talasa iz ugla različitih rezolucija.

Važno je napomenuti da opisani koraci kao i tehnike za obradu slika ne moraju nužno da se izvršavaju ovim redosledom kojim su opisani, niti je uvek potrebno da se izvrše svi koraci kako bi se došlo do rezultata obrade. Proces obrade slika se može završiti u bilo kom koraku, ukoliko taj korak daje odgovor na ulazni problem, sa tim u vezi zaključak je da što je složeniji ulazni problem, broj koraka i broj tehnika potrebnih za rešavanje ulaznog problema raste.

3.3. Tehnike obrade slika korišćene u ovom radu

U ovom radu je korišćeno više različitih tehnika za obradu podataka. Tačan pristup biće opisan u kasnijim poglavljima, dok će ovo poglavlje poslužiti za teorijsko objašnjenje tehnika koje su primenjene nad uzorcima. Neke od tehnika su Laplasov filter, prepoznavanje oblika na osnovu rastojanja, određivanje praga dozvoljenog inteziteta signala, normalizacija, izdvajanje ivica, bojenje i prebrojavanje prepoznatih oblika. Opis tehnika će biti dat u skladu sa [5]:

- **Laplasov filter**[5] je jedna od tehnika za unapređivanje kvaliteta slike u odnosu na originalnu sliku. Rezultat primene Laplasovog filtera predstavlja sliku koja je vizualno izoštrena u odnosu na originalnu sliku. Kada se slika posmatra kroz dve dimenzije x i y mogu se praviti različiti filteri koji su zasnovani na dvodimenzionim funkcijama. Laplasov filter je jedan od takvih filtera koji se oslanja na drugi izvod i zasniva se na tome da se definiše diskretan opis drugog izvoda, a zatim se napravi šablon za filter koji će biti primenjen na sliku. Važno je napomenuti da ovakvi filteri nisu osetljivi na rotaciju,

² Rezolucija predstavlja broj elemenata slike koji se koriste za prikaz u odnosu širina * visina

u smislu da ako se prvo rotira slika i primeni filter dobija se isti rezultat kao da se prvo primeni filter pa posle toga izvrši rotacija nad slikom. Može se pokazati da i najprostiji operator zasnovan na izvodu koji nije osetljiv na rotacije predstavlja Laplasov operator. Matematički, ako je slika predstavljenja pomoću funkcije $f(x,y)$ koja zavisi od dve promenljive, Laplasov operator se može predstaviti sledećom formulom:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

Pošto je izvod bilo kog reda linearni operator, Laplasov operator je takođe linearan operator. Da bi se ova jednačina izrazila u diskretnom obliku potrebno je posmatrati ovu jednačinu kroz drugi izvod. U x smeru drugi izvod je predstavljen sledećom formulom

$$\frac{\partial^2 f}{\partial x^2} = f(x + 1, y) + f(x - 1, y) - 2f(x, y)$$

Slično, u y smeru drugi izvod se predstavlja formulom

$$\frac{\partial^2 f}{\partial y^2} = f(x, y + 1) + f(x, y - 1) - 2f(x, y)$$

Iz prethodne tri jednačine sledi da je diskretni Laplasov operator za dve promenljive jednak

$$\nabla^2 f(x, y) = f(x + 1, y) + f(x - 1, y) + f(x, y + 1) + f(x, y - 1) - 4f(x, y)$$

Ova jednačina može da se implementira koristeći šablone za filtriranje. Primena šablona se zasniva na ideji da se vrednost funkcije u svakom elementu slike zameni vrednošću koja zavisi od suseda elementa slike čija se vrednost menja. Različiti filteri imaju različite šablone od kojih zavisi koji susedi su od većeg značaja i koji koeficijenti su im pridruženi. Konkretno u slučaju Laplasovog filtriranja njegova primena ističe razlike u intezitetima na slici i ublažava delove slike gde su promene u intezitetima funkcije elemenata slike male. Ako bi se posmatra slika na kojoj postoje sivkaste linije i slične nepravilnosti izložene na pozadinu slike koja je u crnoj boji, primena ovog filtera bi uzrokovala promene takve da sivkaste linije postanu uočljivije i da se preciznije vide ivice u odnosu na pozadinu. Neki od primera šablona za Laplasov filter predstavljeni su na slici 3.3.

0	1	0
1	-4	1
0	1	0

1	1	1
1	-8	1
1	1	1

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

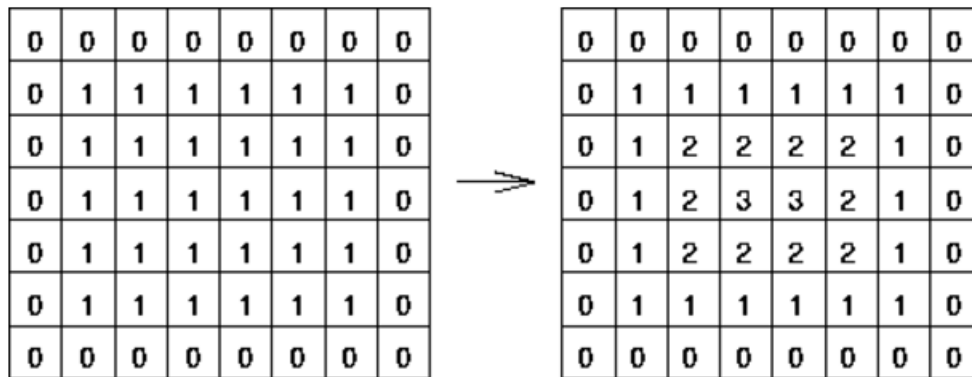
Slika 3.3 Primeri šablona za Laplasov filter

- **Određivanje praga dozvoljenog inteziteta signala** predstavlja jednu od najčešće upotrebljivanih tehnika za obradu slika, posebno među tehnikama za podelu slike na delove ili objekte. Razlog za čestu upotrebu ove tehnike leži u tome da je jako jednostavna za implementaciju i ne zahteva moćne računarske resurse za izvršavanje i obradu. Formalno, neka se slika posmatra kao skup elemenata slike određenih funkcijom $f(x, y)$ i na toj slici neka se razlikuju svetli objekti koji se nalaze na tamnoj pozadini tako da se oni mogu jasno podeliti na dve grupe – objekte i pozadinu. Jedan od načina izvlačenja objekata i razlikovanja od pozadine je da se odredi prag dozvoljenog inteziteta odnosno vrednost T takva da bilo koja tačka (x, y) na slici za koju važi da je vrednost funkcije $f(x, y) > T$ smatra tačkom objekta, a inače ako ovaj uslov nije ispunjen smatra se tačkom pozadine. Rezultat obrade slike ovom tehnikom možemo predstaviti na sledeći način:

$$g(x, y) = \begin{cases} 1, & f(x, y) > T \\ 0, & f(x, y) \leq T \end{cases}$$

Ukoliko se T posmatra kao konstanta kroz obradu cele slike onda se takva obrada naziva određivanje praga dozvoljenog inteziteta signala sa globalnim pragom, a ako se T menja tokom obrade i ima različite vrednosti onda se takva obrada naziva određivanje praga dozvoljenog inteziteta sa promenljivim pragom. Određivanje praga T zavisi od svojstava same slike i sadržaja i u slučaju određivanja globalnog praga najčešće se primenjuje prag dobijen nakon izračunavanja srednje vrednosti inteziteta svih elemenata slike, u odnosu na koju je određen prag. U slučaju određivanja promenljivog praga najčešće se on određuje na osnovu elemenata suseda u određenom delu slike. Primena ove tehnike veliku upotrebljivost ima u uklanjanju šuma sa originalne slike.

- **Prepoznavanje oblika na osnovu rastojanja** predstavlja tehniku koja se koristi za obradu crno belih slika. Rezultat obrade je slika koja sadržajno podseća na originalnu sliku, međutim razlika je u tome što je u intezitet sive boje svakog elemenata slike koji su u prvom planu promenjen na taj način da direktno zavisi od rastojanja do najbliže ivice. Slikovito se ovo može opisati zamišljanjem da su elementi slike koji su u prvom planu takvi da je svaki od njih napravljen od zapaljivog materijala koji sporo gori. Zatim ako se svi elementi na ivici sinhronizovano zapale vatra će se širiti ka unutrašnjosti objekta koji je u prvom planu. Izračunavanjem vremena potrebnog da vatra dođe do nekog elementa se direktno računa tablica rastojanja koja predstavlja ključ ovog algoritma za prepoznavanje oblika.



Slika 3.4 Primena algoritma rastojanja na prost objekat (pravougaonik)

Na slici 3.4 je prikazana primena algoritma rastojanja gde se rastojanje između elemenata slike posmatra kao šahovsko rastojanje³. Rastojanje predstavlja udaljenost između dva elementa slike predstavljena svojim koordinatama (x_1, y_1) i (x_2, y_2) . U tehnikama obrade slika najčešće se koriste sledeći načini računanja rastojanja: Euklidsko rastojanje⁴, Menhetn (gradski blokovi) rastojanje⁵ i šahovsko rastojanje. U zavisnosti od načina računanja rastojanja između elemenata slike postoje različite implementacije algoritma za prepoznavanje oblika na osnovu rastojanja koje se razlikuju po svojoj efikasnosti i algoritamskoj kompleksnosti.

³ Šahovsko rastojanje – ako se elementi slike predstave kao šahovska tabla, kretanje kralja po tabli bi predstavljao opis i osnovni korak šahovskog rastojanja. Matematički predstavljeno $D = \max(|x_2 - x_1|, |y_2 - y_1|)$

⁴ Euklidsko rastojanje – rastojanje koje predstavlja dužinu prave linije između dve tačke. Matematički predstavljeno: $D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

⁵ Menhetn (gradski blokovi) – podrazumeva računanje rastojanja gde je dozvoljeno kretanje po elementima slike samo horizontalno i vertikalno (dijagonalno kretanje nije dozvoljeno). Matematički predstavljeno:

$D = |x_2 - x_1| + |y_2 - y_1|$

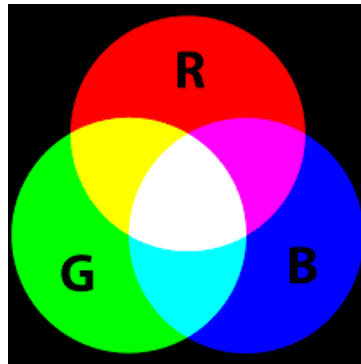
- **Normalizacija** predstavlja tehniku koja se često koristi za promenu vrednosti inteziteta elemenata slike gde se vrednosti inteziteta elemenata slike menjaju tako da se uklape u opseg vrednosti koji je unapred zadat. Normalizacija može biti linearna i nelinearna. Nelinearna normalizacija se koristi kada je nemoguće uspostaviti linearnu vezu između vrednosti inteziteta elemenata originalne slike i opsega inteziteta u koji treba prebaciti vrednosti inteziteta originalne slike. Kod linearne normalizacije, u slučaju n-dimezione crno-bele slike čije se vrednosti inteziteta elemenata nalaze u opsegu opsegu $[Min, Max]$ (matematički zapis: $I: \{X \subseteq \mathbb{R}^n\} \rightarrow \{Min, \dots, Max\}$) normalizacija se predstavlja kao $I_n: \{X \subseteq \mathbb{R}^n\} \rightarrow \{NoviMin, \dots, NoviMax\}$ gde se posle primene normalizacije sve vrednosti nalaze u opsegu $[NoviMin, NoviMax]$. Formula za dobijanje vrednosti inteziteta u nekom elementu slike je

$$\bullet \quad I_n = (I - Min) \frac{NoviMax - NoviMin}{Max - Min} + NoviMin$$

Na primer ako se posmatraju crno-bele slike, gde je intezitet vrednosti elementa slike predstavljen uz pomoć 8 bita odnosno opseg vrednosti za jedan element je: $[0 - 255]$, a na primer u slici koja se posmatra vrednosti inteziteta svih elemenata slike je u opsegu $[50 - 180]$, tada proces normalizacije počinje tako što se od vrednosti inteziteta svih elemenata oduzme 50 i kao rezultat dobije opseg $[0 - 130]$. Zatim se intezitet svakog elementa slike množi sa $\frac{255}{130}$ čime se opseg vrednosti inteziteta svih elemenata slike prevodi u interval $[0 - 255]$.

- **Izdvajanje ivica** predstavlja jednu od najvažnijih tehnika obrade slika koja primenu nalazi u skoro svim programima koji vrše neko prepoznavanje oblika ili analizu različitih objekata / oblika. Ivica sama po sebi u tehnikama obrade slika predstavlja liniju koja povezuje sve elemente slike koji su istog inteziteta i boje, a da čine granicu nekog objekta. Osnovna ideja u pronalazhenju ivica je u tome da se razlikuju unutrašnje ivice objekta od njegovih spoljašnjih ivica i na taj način se određuju granice tog objekta. Originalna slika može biti promenjena u smislu da njeno predstavljanje bude samo uz pomoć ivica koje sadrži. Algoritam koji se najčešće koristi u implementacijama funkcija za izdvajanje ivica je algoritam predstavljen u radu: Satoshi Suzuki and Keiichi Abe "Topological Structural Analysis of Digitized Binary Images by Border Following", CVGIP 30 1, pp 32-46 (1985). Zbog svoje kompleksnosti algoritam neće biti predstavljen u ovom radu, ali svakako predstavlja zanimljivu temu za neku dalju obradu i istraživanje.
- **Bojenje i prebrojavanje prepoznatih oblika** predstavljaju završne tehnike u obradi slike. Ove tehnike nisu obavezne i eventualno se primenjuju na samom kraju obrade da bi se došlo do konačnih rezultata i mogućih zaključaka u vezi sa problemom koji se rešava. Samo bojenje najčešće barata sa već prepoznatim oblicima ili objektima te u zavisnosti od ulaznog problema i traženog rešenja može da elemente slike koji pripadaju

oboji u određenu boju. U računarstvu slika u boji se najčešće predstavlja uz pomoć nekog od sledećih pristupa a to su: RGB, HSV i YCbCr. Verovatno najzastupljeniji od njih je RGB pristup gde se za svaki element slike čuvaju 3 različita inteziteta, gde prvi predstavlja nivo crvene boje, drugi nivo zelene boje, a treći nivo plave boje. Različitim kombinacijama inteziteta crvene, zelene i plave boje moguće je dobiti i ostale boje iz spektra boja koji je poznat ljudskom oku. Najčešće se za predstavljanje svakog od tri nivoa inteziteta odvaja po 8 bita tako da svaki intezitet ima opseg vrednosti [0 – 255]. U slučaju da su vrednosti sva tri nivoa 0 dobija se crna boja, dok u slučaju da su vrednosti sva tri nivoa maksimalne (255) dobija se bela boja. Uprošćeno spekatar boja koristeći RGB pristup je predstavljen na slici 3.4 gde je grubo vizuelno predstavljeno kojim kombinacijama je moguće dobiti određene boje. Kao i bojenje, prebrojavanje takođe predstavlja prostu tehniku koja vrši brojanje prepoznatih oblika ili objekata radi izvođenja potencijalnih rezultata kompletne obrade.



Slika 3.4 RGB spekatar grafički uprošćen prikaz

4. Tehnike istraživanja podataka

U svetu se svakodnevno proizvodi i prikuplja velika količina podataka. Analiza i obrada tih podataka je postala nužna potreba, posebno u oblastima kao što su medicina, inženjerstvo, nauka pa i na nivou poslovnih odluka preduzeća svih veličina. Osnovni cilj istraživanja podataka je da se iz velike količine podataka za obradu izvuku informacije od značaja. Iako na prvi pogled to zvuči trivijalno, trivijalnost se gubi u trenutku kada se shvati prostorno-vremenska zavisnost i obim količine podataka koju je potrebno obraditi da bi se dobile informacije od značaja.

Čest je slučaj da se istraživanje podataka poistovećuje sa pojmom otkrivanja znanja iz podataka⁶, ali generalno se istraživanje podataka posmatra kao jedan od ključnih koraka u procesu otkrivanja znanja iz podataka. Celokupan proces otkrivanja znanja iz podataka je predstavljen na slici 4.1 [7]. Uopšteno proces se može podeliti u 7 koraka.

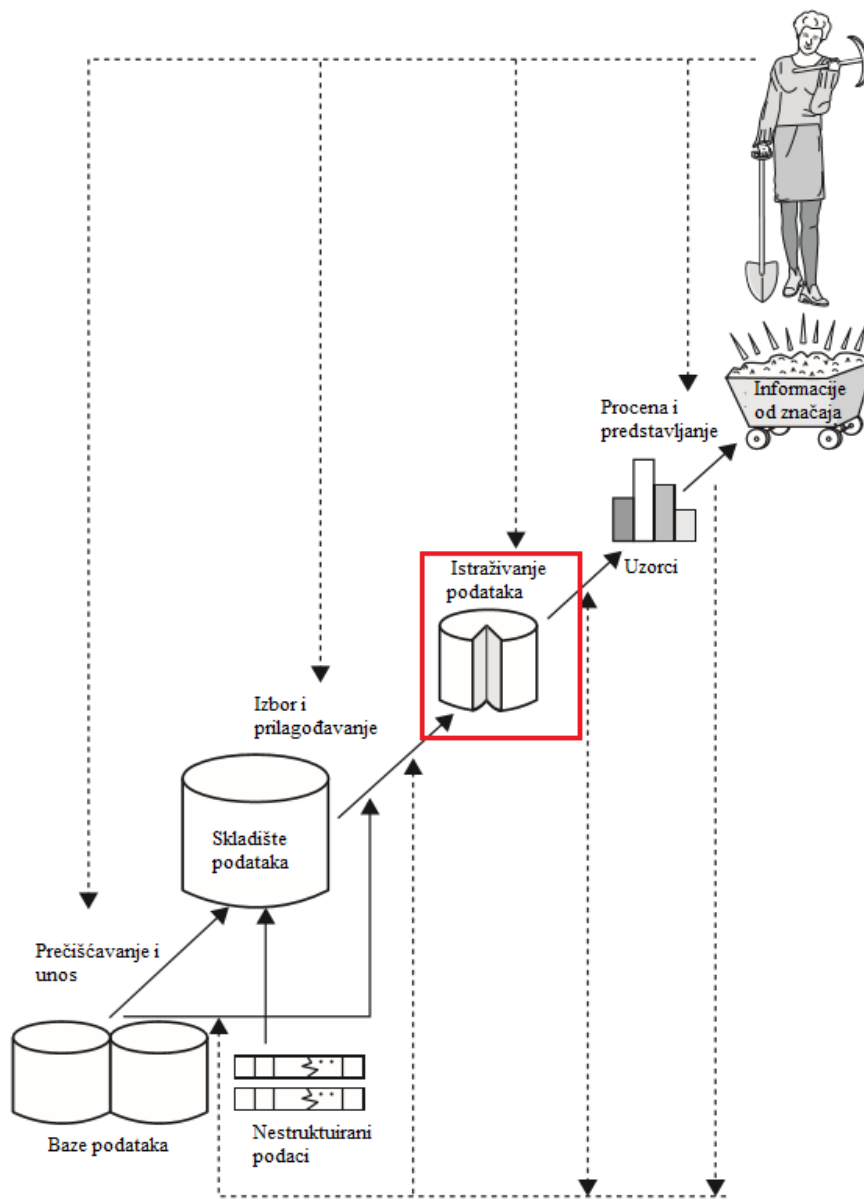
1. **Prečišćavanje podataka** radi eliminacije šuma i nekonzistentnih podataka.
2. **Kombinovanje podataka** u slučaju da postoje podaci iz više različitih izvora.
3. **Izbor podataka** iz skladišta podataka koji su u vezi sa ulaznim problemom.
4. **Prilagodavanje podataka** u oblik koji je pogodan kao ulaz za proces istraživanja podataka.
5. **Istraživanje podataka** proces gde se primenom različitih metoda dobijaju uzorci od potencijalnog značaja.
6. **Procena uzoraka** da se odredi koliko su relevantni dobijeni podaci na osnovu poznatih već informacija od značaja.
7. **Predstavljanje** informacija od značaja kao izlaz obrade ulaznog problema.

Posmatrajući ove korake, prva četiri koraka mogu da se posmatraju kao proces pripreme za istraživanje. Sam proces istraživanja podataka zahteva neku vrste interakcije sa korisnicima ili već poznatim informacijama od značaja. Nakon obrade i pribavljanja rezultata u formi uzoraka, koji predstavljaju potencijalne informacije od značaja i nakon procene rezultata, oni se predstavljaju korisnicima kao izlaz ulaznog problema i u slučaju zadovoljavajuće procene mogu se smatrati novim informacijama od značaja.

Imajući u vidu da istraživanje podataka predstavlja deo procesa otkrivanja znanja iz podataka, uopšteno istraživanje podataka se može definisati kao proces otkrivanja uzoraka od potencijalnog značaja iz velike količine podataka. Nalik na proces otkrivanja znanja iz podataka tako i istraživanje podataka ne može da isključi korake poput čišćenja i pripreme podataka, izbor i izdvajanje uzoraka, predstavljanje uzoraka od značaja itd. Važno je napomenuti da istraživanje podataka nije ograničeno na neki određen domen podataka i izvora, već se tehnike mogu primenjivati na bilo koji tip i izvor podataka u slučaju da ti podaci imaju značaj na ulazni

⁶ Otkrivanje znanja iz podataka (eng. Knowledge Discovery in Data, KDD)

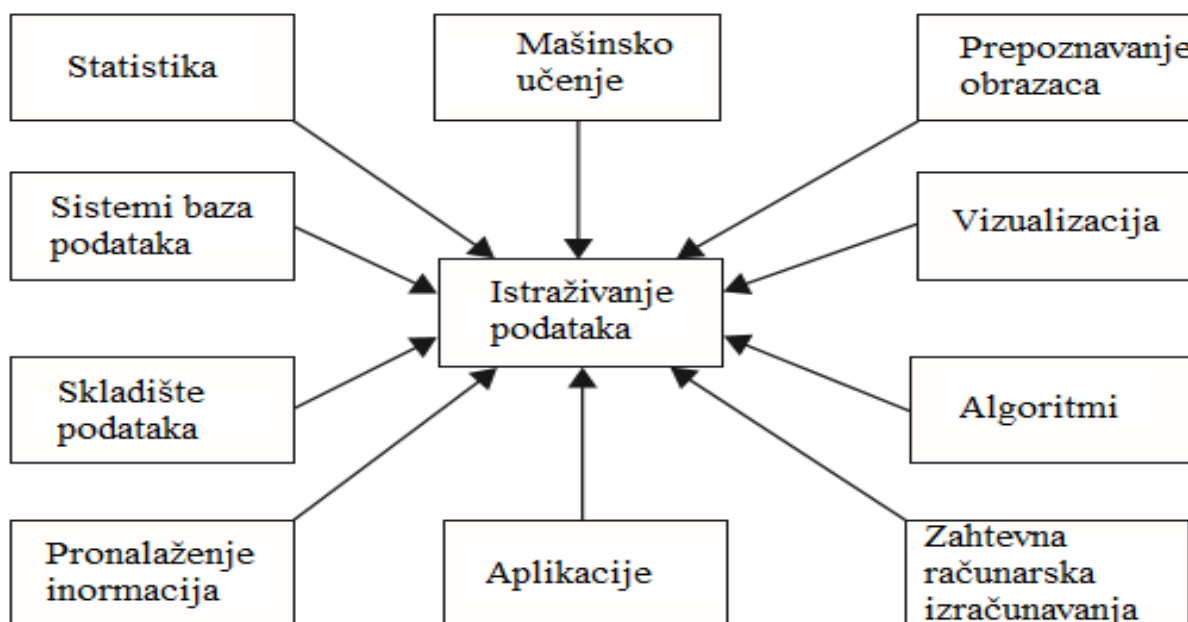
problem ili vezu sa rešenjem izlaznog rezultata. Posmatrano iz tog ugla tehnike istraživanja podatka se primenjuju na podatke iz baza podataka, iz skladišta podataka, transakcione podatke, pa i na podatke složenijeg tipa kao što su tekstualni i multimedijalni podaci, podaci vezani za Internet i računarske mreže i slično [7].



Slika 4.1 Proces otkrivanja znanja iz podataka

Svaka obrada za cilj ima pronalaženje uzoraka od značaja, koji se mogu definisati kao deo podataka za koje važe određene činjenice definisane ulaznim problemom. Naravno uzorci koji se dobijaju kao rezultat ne moraju u potpunosti zadovoljavati uslove i pravila ulaznog problem i zbog toga se koriste različite mere koje definišu koliko je neki uzorak od potencijalnog značaja.

Istraživanje podataka je tokom svog razvijanja usvajalo, a i dan danas koristi metode i tehnike iz različitih oblasti matematike, algoritmike i informatike što je predstavljeno na slici 4.2 [7]



Slika 4.2 Oblasti koje su u direktnoj vezi sa tehnikama istraživanja podataka

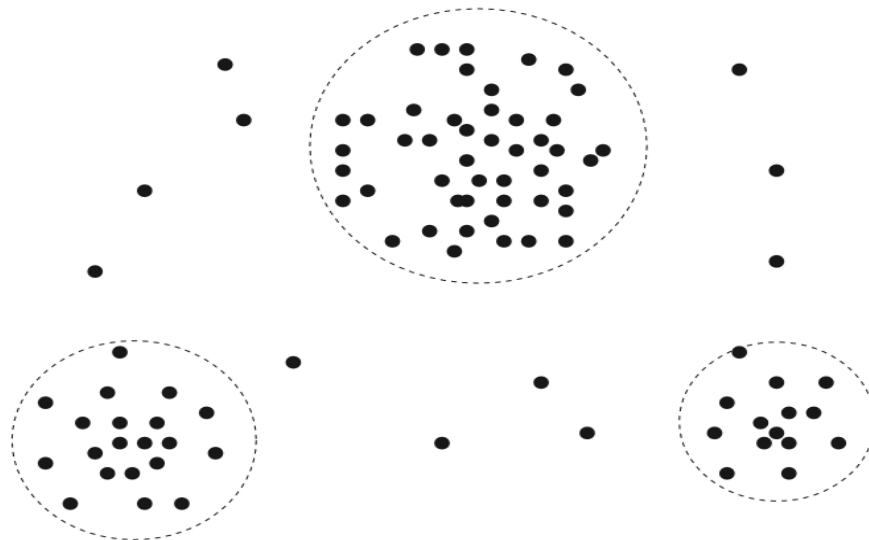
Istraživanje podataka predstavlja oblast punu izazova. Neki od glavnih i najčešćih problema koji se javljaju jesu izbor pravog pristupa, rad sa upotrebljivim podacima, interakcija sa korisnikom, rad sa nestruktuiranim podacima i razni drugi problemi koji mogu da imaju veliki uticaj na krajnji rezultat procesa istraživanja podataka. Ipak i uz različite probleme, tehnike istraživanja podataka su našle veliku primenu u bioinformatici, na internetu, u zdravstvu, informatici, školstvu, bankarstvu i mnogim drugim oblastima svakodnevnog života.

4.1. Osnovne tehnike istraživanja podataka

U zavisnosti od ulaznog problema, od tipa ulaznih podataka i od njihovog stanja mogu se koristiti različite tehnike za obradu u procesu istraživanja podataka. Svrha svih tehnika za obradu je da se na kraju obrade kao rezultat dobijaju informacije od značaja ili uzorci koji mogu da odgovore na pitanje ulaznog problema. Uopšteno, tehnike istraživanja podataka mogu da imaju dve metode, da daju opis ili da daju predviđanje. Metoda u kojoj daje opis se zasniva na određivanju osobina podataka iz ulaznog skupa podataka, dok kod metode predviđanja na osnovu prethodno poznatih informacija formiraju predviđanje o podacima koji su predmet obrade. Neke od najvažnijih tehnika istraživanja podataka su:

- **Klasifikacija** koja će detaljnije biti opisana u delu 4.2

- **Regresija** koja predstavlja statistički pristup koji se najčešće koristi u domenu realnih brojeva gde se na osnovu vrednosti nekih promenljivih predviđa vrednost promenljive koja se obrađuje. Prost primer regresije je da prodavnica želi da predvidi zaradu koju će joj doneti prodaja određenog artikla na popustu na osnovu prethodnih prodaja sa popustom.
- **Klasterovanje** gde se obrada vrši određivanjem klastera koji predstavljaju skupove objekata sa sličnim osobinama. Bitna razlika u odnosu na klasifikaciju i regresiju je da klase (u ovom slučaju klasteri) nisu unapred poznati. Klasterovanje može da se iskoristi za određivanje klasa na ulaznom skupu podataka. Primer određivanja klastera se može videti na slici 4.3 [7] gde se jasno vide 3 određena klastera na osnovu mesta stanovanja kupaca, koji su kupili određen proizvod, koje prodavnica može iskoristiti za svoju reklamnu kampanju.



Slika 4.3 Primer klasterovanja

- **Sumarizacija** predstavlja tehniku koja vrši sumarizaciju (generalizaciju) opštih osobina ili karakteristika nad ulaznim skupom podataka za obradu. Najčešće se obrađuju podaci dobijeni iz baze podataka iz kojih se izdvajaju ili izvode reprezentativne osobine podataka. Izlazni rezultati se najčešće prikazuju u formama različitih dijagrama, višedimenzionih tabela i grafikona. Primer sumarizacije može da predstavlja istraživanje osobina proizvoda čija se prodaja u prethodnoj godini uvećala za neki određeni procenat.
- **Analiza vremenskih serija** predstavlja tehniku gde se istražuju promene vrednosti atributa kroz vreme. Najbolji primer gde se koristi analiza vremenskih serija je praćenje vrednosti akcija na berzi i na osnovu toga odrediti u koje akcije ulagati.
- **Pravila pridruživanja** predstavljaju tehniku koja se koristi za otkrivanje obrazaca koji opisuju međusobno čvrsto povezane osobine podataka. Najčešće kao ulazni podaci se posmatraju transakcioni podaci gde se iz skupa transakcija nalaze različite veze i

pravilnosti između podataka koji učestvuju u transakcijama. Pravila pridruživanja se najčešće predstavljaju uz pomoć skupovnih operacija. Najbolji primer može da bude povezanost artikala koji se kupuju u prodavnici gde se iz skupa različitih transakcija mogu uočiti pravila:

$$\{Laptop, Digitalna\ kamera\} \Rightarrow \{Memorijska\ kartica\}$$

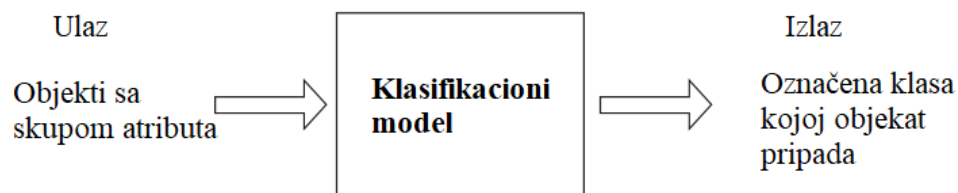
$$\{Memorijska\ kartica\} \Rightarrow \{\text{Čitač memorijskih kartica}\}$$

- **Sekvencijalna analiza** se koristi za određivanje obrazaca u podacima koji zavise od redosleda pojavljivanja. U posmatranom skupu ulaznih podataka koji se posmatraju kao objekti za svaki objekat se dodeljuje njegov istorijat događaja i pronalaze se pravila koja predviđaju strogu zavisnost od redosleda njihovog pojavljivanja.

4.2. Klasifikacija

Osnovni zadatak klasifikacije je da objekte nad kojima se vrši obrada dodeli unapred definisanim klasama. To je danas čest problem i tom zadatku se pristupa u najrazličitijim oblastima. Da bi se bolje razumeo zadatak klasifikacije bitno je opisati celokupan proces obrade kao celinu. Ulazne podatke predstavlja skup objekata koje karakterišu određeni atributi, izlaz, odnosno rezultat, obrade je svrstavanje objekata u unapred određene klase uz pomoć funkcije koja na osnovu skupa atributa objekta svrstava objekat u neku od klasa. Na osnovu ovog opisa klasifikacija grubbo može da se definiše kao **proces konstrukcije funkcije f koja svrstava ulazni objekat sa svojim skupom atributa u jednu od unapred definisanih klasa**. Funkcija koja vrši razvrstavanje u istraživanju podataka poznata je i kao klasifikacioni model. Klasifikacioni model može imati dve uloge:

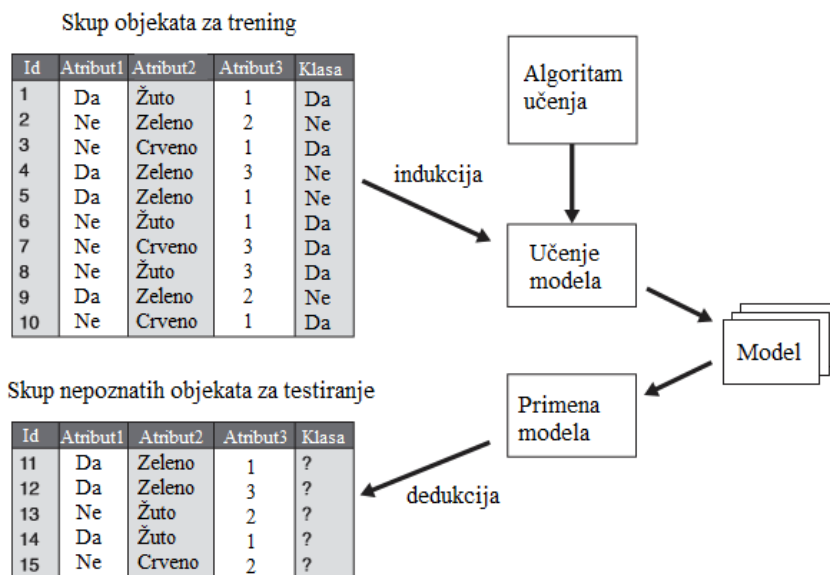
- Opisna – gde klasifikacioni model služi kao objašnjenje odnosno potvrda zašto neki objekti pripadaju određenim klasama (u isto vreme to objašnjava i zašto ne pripadaju nekim drugim klasama).
- Predviđajuća – gde klasifikacioni model služi kao alat za razvrstavanje objekata u klase za one objekte koji još uvek nisu razvrstani.



Slika 4.4 Proces klasifikacije

U opštem slučaju tehnika klasifikacije se zasniva na izvođenju klasifikacionog modela na osnovu ulaznog skupa podataka(objekata). Postoje različiti pristupi i metode za izvođenje klasifikacionog modela kao što su:

- Metode zasnovane na drvetima odlučivanja
- Metode zasnovane na pravilima
- Neuronske mreže
- Statistički zasnovane metode
- Metode zasnovane na podržavajućim vektorima



Slika 4.5 Proces klasifikacije

Ono što je zajedničko za sve metode je da svaka od njih konstruiše algoritam za određivanje klasifikacionog modela koji najbolje opisuje i daje najčvršću vezu između atributa ulaznih objekata i definisanih klasa. Model konstruisan uz pomoć algoritma treba da bude prilagođen i poznatim objektima, ali i da tačno može da predvidi klase objekata koji nisu nikada bili obrađeni. Na Slici 4.5 je predstavljen opšti pristup u rešavanju problema tehnikom klasifikacije. Prvo neophodno je obezbediti skup objekata koji će imati ulogu skupa za trening, gde je važno da su za objekte u skupu za trening poznate klase kojima pripadaju. Skup objekata za trening se koristi za učenje klasifikacionog modela koji kasnije treba da bude primenjiv za skup nepoznatih objekata za testiranje čije klase treba odrediti. Dobijeni klasifikacioni model ne mora da daje u potpunosti precizne rezultate i zbog toga postoje različiti načini za merenje kvaliteta izračunavanja klasifikacionog modela koji se zasnivaju na broju dobro predviđenih i loše predviđenih klasa za neke objekte od strane klasifikacionog modela. Ti brojevi se predstavljaju matricom konfuzije koja je prikazana na slici 4.6 kao primer matrice konfuzije za dve klase. Oznake *a*, *b*, *c*, *d* predstavljaju brojeve predviđenih klasa i stvarnih klasa nakon procesa klasifikacije. Tako je na primer broj *c* broj onih objekata klase 0 koji su pogrešno predviđeni da pripadaju klasi 1.

Matrica konfuzije		Predviđajuća klasa	
		Klasa = 1	Klasa = 0
Stvarna klasa	Klasa = 1	a	b
	Klasa = 0	c	d

Slika 4.6 Matrica konfuzije

Kako broj tačno predviđenih ($a + d$) i broj netačno predviđenih ($b + c$) nisu dovoljni kao mere za poređenje različitih modela, najčešće se, koristeći matricu konfuzije, definišu različite mere. Najčešće korišćena mera je preciznost.

$$\text{Preciznost} = \frac{a + d}{a + b + c + d}$$

Zajedno sa preciznošću, posmatra se i nivo greške kao suprotna mera u odnosu na preciznost.

$$\text{Nivo greške} = \frac{b + c}{a + b + c + d}$$

Klasifikacioni modeli teže da ostvare najveći mogući nivo preciznosti ili ekvivalentno najmanji nivo greške kada se primene na nepoznatom skupu objekata za testiranje. Uz matricu konfuzije često se koristi i matrica cene (Slika 4.7) kao još jedna metrika za određivanje performansi klasifikacionog modela, jer može se desiti da metrika kao što je preciznost nema značaja ako se primeni za određene modele, posebno u slučajevima kada velika većina objekata pripada jednoj klasi dok objekata druge klase ima zanemarljivo malo.

$$\text{Cena} = p * a + q * b + r * c + s * d$$

gde oznake p , q , r , s predstavljaju cene za tačno/netačno predviđene i stvarne klase.

Matrica cena		Predviđajuća klasa	
		Klasa = 1	Klasa = 0
Stvarna klasa	Klasa = 1	p	q
	Klasa = 0	r	s

Slika 4.7 Matrica cena

Postoje i druge mere kao npr. Odziv, lift, F-mera, itd. Ali one će biti korišćene u samom radu tako da njihove formule neće biti detaljno navođene. Više o merama za proveru kvaliteta klasifikacije se može videti u [7].

Razlozi zbog kojih dolazi do grešaka klasifikacionog modela u određivanju klase mogu biti razni, dok greške se uopšteno mogu podeliti u dve kategorije.

- Greške pri treniranju – greške koje nastaju pri klasifikaciji podataka za trening.
- Greške pri uopštavanju – greške klasifikacionog modela u odnosu na nepoznate objekte.

Dobar model mora da ima male greške i pri treniranju i pri uopštavanju. Praktični problemi pri treniranju modela se ispoljavaju tako što isuviše jednostavni modeli mogu da imaju velike greške i pri treniranju i pri uopštavanju (premalno prilagođen model), sa druge strane modeli koji isuviše dobro klasifikuju podatke za trening mogu da imaju veću grešku pri uopštavanju od modela koji ima veću grešku pri treniranju, ovakva pojava dovodi do previše prilagođenog modela (preprilagođen klasifikacioni model). Preprilagođenost se može javiti usled prisustva šuma koji treniranjem može biti pridružen nekoj od klasa i na taj način ugroziti preciznost klasifikacionog modela. Još jedan od načina za preprilagođavanje klasifikacionog modela je treniranje modela sa nedovoljno podataka za trening i na taj način klasifikacioni model može da bude istreniran nedovoljno dobro za primenu na nepoznatim objektima za testiranje.

5. Cilj rada

Cilj ovog rada je implementacija programa, koristeći tehnike obrade slika i tehnike istraživanja podataka, koji će biti sposoban da za ulazne slike dobijene uz pomoć fluorescentnog mikroskopa da izlaznu informaciju u vidu broja prepoznatih ćelija na slici. U isto vreme cilj je da program bude uporediv sa trenutnom ponudom licenciranih i besplatnih programa koji se bave ovim problemom. Svrha izrade ovakvog programa ogleda se u potrebi istraživača sa Biološkog fakulteta Univerziteta u Beogradu kojima ovakav program treba da posluži i olakša istraživačke poslove. Za izradu rešenja biće iskorišćen programski jezik C++, zajedno sa besplatnim *Qt* i *OpenCV* bibliotekama, dok će za sam razvoj biti iskorišćeno besplatna verzija *Microsoft Visual Studio* razvojnog okruženja. Za klasifikacioni model će biti iskorišćeno prosto drvo odlučivanja za čiji trening će biti iskorišćen deo podataka dobijenih od istraživača sa Biološkog fakulteta, dok će drugi deo podataka biti iskorišćen za testiranje i poređenje rezultata sa programima *CellC* i *Imaris* čiji su rezultati prebrojavanja poznati.

Cilj rada je:

- Implementirati i iskoristiti tehnike za obradu slike za poboljšanje stanja i izoštravanje slike, za prepoznavanje oblika i prebrojavanje prepoznatih oblika
- Uz pomoć tehnika za istraživanje podataka odrediti i trenirati klasifikacioni model nad slikama odabranim da budu ulaz za trening klasifikacionog modela
- Iskoristiti klasifikacioni model za primenu na ostatak slika koje će biti iskorišćene kao ulaz za test klasifikacionog modela
- Analizirati i uporediti rezultate prebrojavanja implementiranog programa sa rezultatima prebrojavanja dobijenim programima *CellC* i *Imaris*

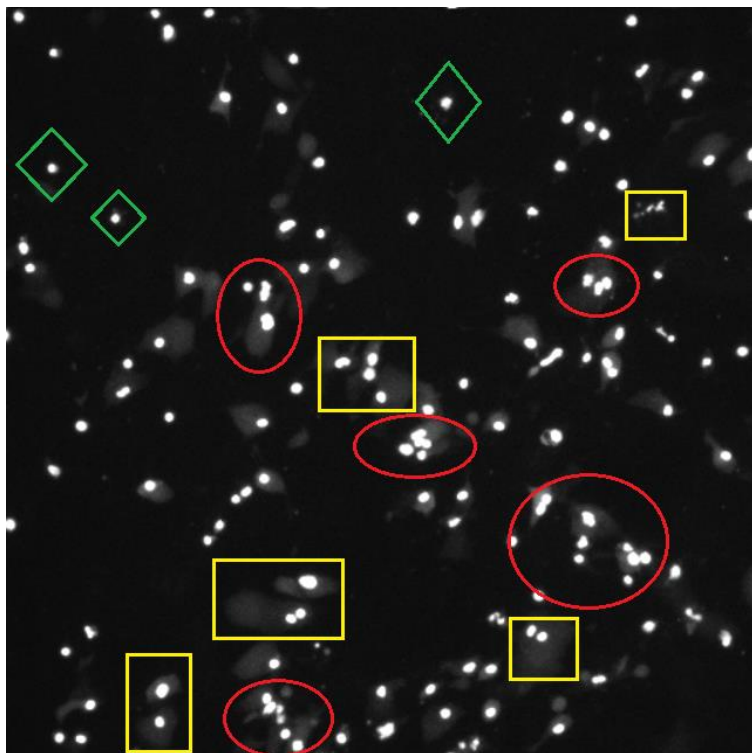
6. Opis, metode i pristup problemu

Za potrebe ovog rada prikupljen je materijal koji je korišćen u eksperimentima na Biološkom fakultetu. Eksperimenti se sprovode kako bi se testirala izdržljivost i preživljavanje ćelija u slučaju primene različitih tretmana nad njima. Mera koja sa sigurnošću može da potvrdi koliko ćelija je preživelo određeni tretman je sam broj ćelija koje su na snimku. Dobijeni snimci odnosno slike su u crno-beloj varijanti gde su ćelije predstavljene belom bojom. Osim problema prisutnosti šuma i prepoznavanja oblika, problem dodatno komplikuju delovi snimaka gde se na malom prostoru preklapa veći broj ćelija što otežava prepoznavanje, a u isto vreme i utiče na prebrojavanje kao izlazni rezultat programa. Neki od primera snimaka koji su korišćeni kao ulazni problem za program su prikazani na slikama 5.1a, 5.1b, 5.1c i 5.1d koji su izmenjeni tako da su jasno označeni delovi snimaka koji predstavljaju izazov ovog rada.

Za potrebe ovog rada prikupljeni su snimci ćelija dobijeni uz pomoć fluorescentnog mikroskopa iz tri različita snimanja u različitim vremenskim periodima kao ulazni problem za program prebrojavanja ćelija. U svakom merenju je prikupljeno oko desetak snimaka za svaki vremenski period što u konačnom zbiru daje više desetina snimaka. Uz snimke prikupljeni su i rezultati prebrojavanja koji su istraživači dobili koristeći program sa besplatnom licencom *CellC* i koristeći program sa *Imaris* koji zahteva plaćenu licencu za korišćenje.

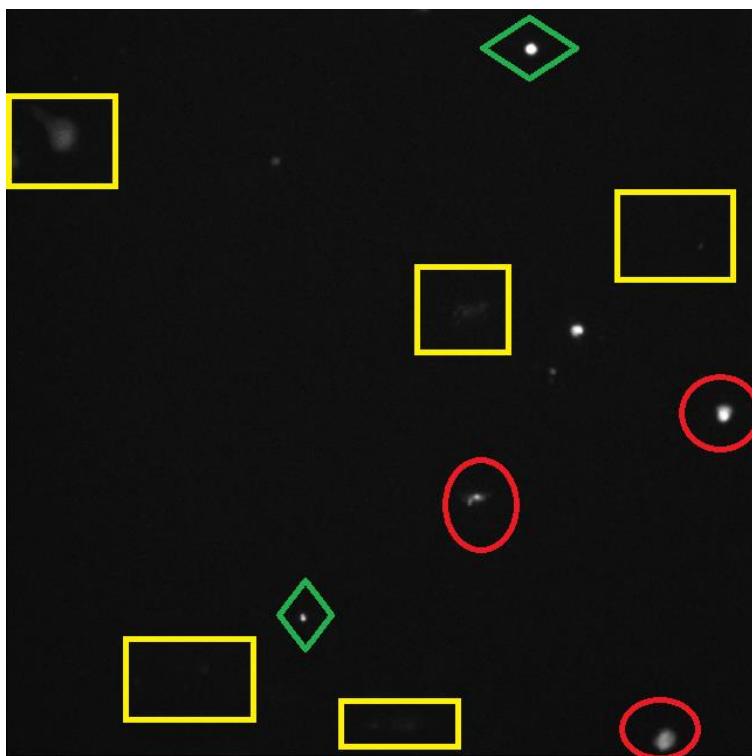
Broj ćelija koje se pojavljuju na snimcima koji su dobijeni kao materijal za rad varira u intervalu od nekoliko desetina do brojke od nekoliko stotina ćelija na jednom snimku. U skladu sa time gustina ćelija na snimcima nije ravnomerna, ali kod snimaka gde je broj ćelija veći, veća je šansa da su pojedine ćelije preklopljene i da su problematični delovi za obradu češći od snimaka gde je ukupan broj ćelija desetak puta manji.

Ono što je zajedničko za skoro svaki snimak je da same ivice ćelija nisu dovoljno jasne i da je potrebno izoštravanje slike, a uz to za skoro svaku sliku postoji i određena doza šuma, koji mora biti eliminisan pre samog procesa prepoznavanja i prebrojavanja i to je bio prvi problem koji je bilo potrebno rešiti u implementaciji programa.



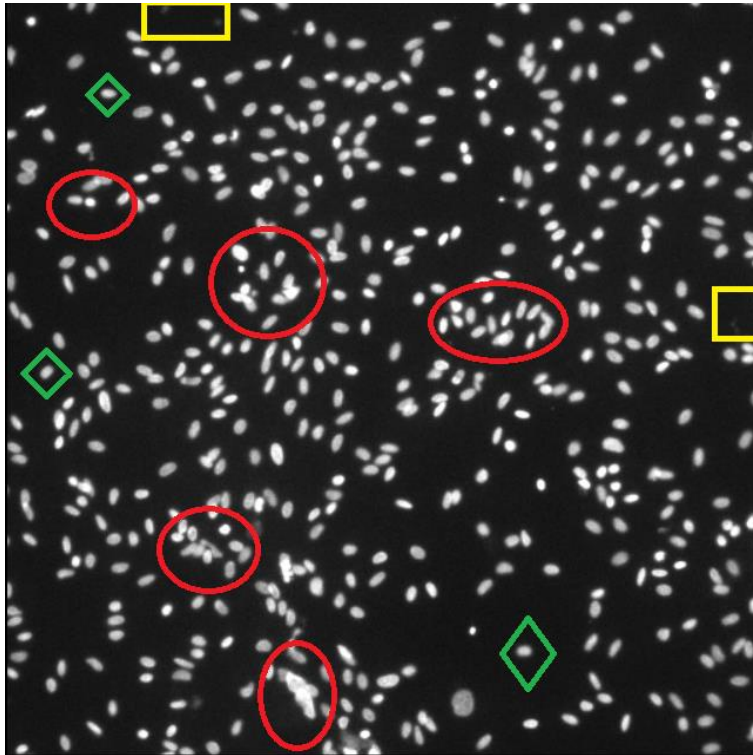
Slika 5.1a Primer ulaznog snimka

Slika 5.1a predstavlja primer ulaznog snimka sa nekim označenim delovima koji su iz ugla posmatrača bitni za razvoj programa. Zelenom bojom su označeni delovi koji predstavljaju ćelije oko kojih nema prisustva šuma i nema drugih ćelija koje bi stvorile problem preklapanja. Žutom bojom su uokvireni delovi koji mogu da budu smatrani šumom (ćelije su uokvirene sivkastim oblicima koji ne predstavljaju druge ćelije već se mogu smatrati šumom). Crvenom bojom su uokvireni delovi gde se jasno vidi problem brojanja kod ćelija koje se preklapaju gde nije najjasnije koliki je njihov broj na malom prostoru.



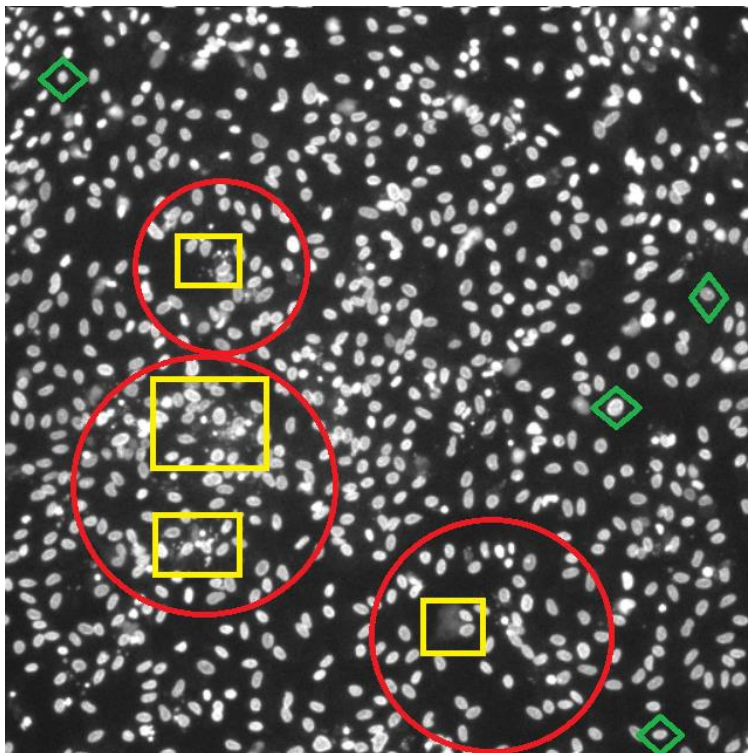
Slika 5.1b Primer ulaznog snimka

Slika 5.1b predstavlja primer ulaznog snimka sa nekim označenim delovima koji su iz ugla posmatrača bitni za razvoj programa. Za označavanje su korišćene iste boje i njihovo značenje kao i na slici 5.1a. Ono što se ističe kod ovog snimka je izuzetno mali broj ćelija koje je potrebno prebrojati.



Slika 5.1c Primer ulaznog snimka

Slika 5.1c predstavlja primer ulaznog snimka sa nekim označenim delovima koji su iz ugla posmatrača bitni za razvoj programa. Za označavanje su korišćene iste boje i njihovo značenje kao i na slici 5.1a. Ovaj snimak je predstavnik snimaka koji imaju i po nekoliko stotina ćelija na slici. Tako su i delovi slike gde se ćelije preklapaju češći i u ovom slučaju su označeni samo oni najuočljiviji gde se golim okom može primetiti i da u takvim grupama ćelija ponekad učestvuje i više od 4-5 ćelija



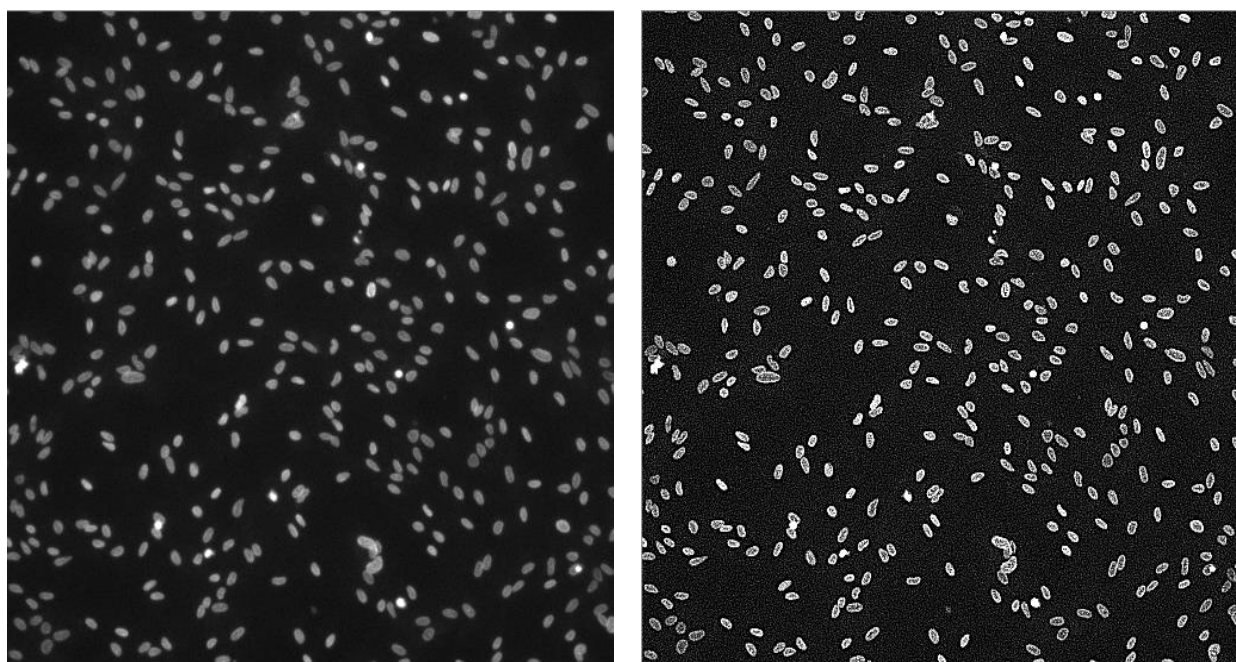
Slika 5.1d Primer ulaznog snimka

Slika 5.1d predstavlja primer ulaznog snimka sa nekim označenim delovima koji su iz ugla pomatrača bitni za razvoj programa. Za označavanje su korišćene iste boje i njihovo značenje kao i na slici 5.1a. Ovaj snimak predstavlja jedan od najtežih snimaka za obradu programa. Granice samih ćelija nisu jasno prepoznatljive, a pored izuzetno velikog broja ćelija i preklapanja ćelija među njima u njihovoj neposrednoj blizini se javlja i prisutnost šuma koja nije zanemarljiva i prilično je teška za obradu.

Nakon upoznavanja i prepoznavanja izazova koje program treba da reši i kao što je već rečeno nakon pregleda snimaka zaključak je bio da je ulazne slike potrebno izoštriti i ukloniti prisustvo šuma. Za izoštravanje je iskorišćen Laplasov filter za koji je primenjen na ulazni snimak. Testirani su različiti šabloni za Laplasov filter i utvrđeno je da najbolje vizuelne rezultate daje šablon predstavljen na slici 5.2. Primena ovog šablona Laplasovog filtera se pokazala kao obavezan prvi korak u radu programa. Rezultati koji su postignuti ovim filterom se mogu videti na slici 5.3 gde se jasno vidi razlika u izoštravanju između originalne slike levo i rezultujuće slike koja nastaje primenom Laplasovog filtera desno.

1	1	1
1	-8	1
1	1	1

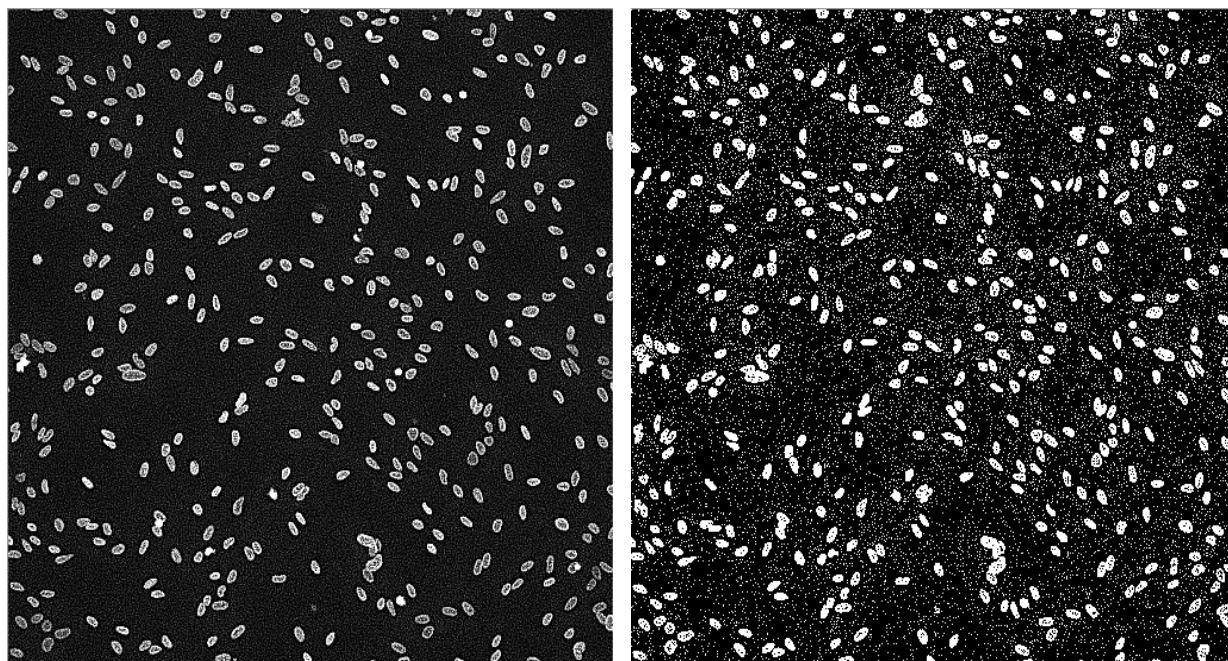
Slika 5.2 Šablon za filter



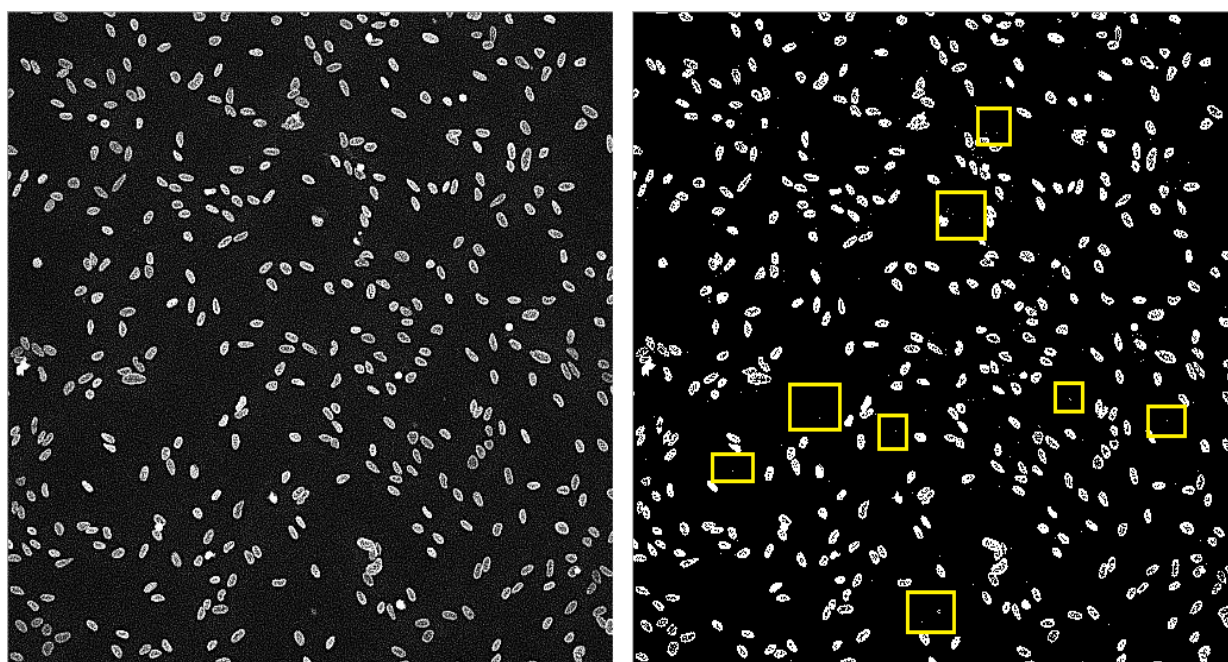
Slika 5.3 Originalna slika (levo) i rezultat primene Laplasovog filtera na originalnu sliku (desno)

Nakon primene Laplasovog filtera i dobijanja izoštrnije slike, gde se jasno raspoznaju okviri ćelija koje je potrebno prebrojati, postalo je jasno da izoštravanje za posledicu ima isticanje i pojavu dodatnog šuma koji je potrebno eliminisati. Pošto su ulazni snimci crno beli, u računarstvu su takve slike predstavljene tako što je svakom elementu slike dodeljena vrednost iz intervala $[0 - 255]$ gde vrednost 0 označava potpuno odsustvo bele boje (crna boja) dok vrednost 255 označava maksimalno prisustvo bele boje. Na izoštranoj slici može da se primeti da su vrednosti kojom su obojene ćelije bliže vrednosti 255 odnosno da je veća prisutnost bele boje, dok je šum bliži pozadini i crnoj boji. Jasno je da u toj raspodeli je potrebno odrediti prag propustljivosti i to je urađeno na taj način da se izabere granična vrednost iz mogućeg intervala $[0 - 255]$, a potom se vrednost svakog elementa slike poredi sa tom vrednošću i ukoliko je vrednost inteziteta piksela veća od granične vrednosti tada se tom elementu slike dodeljuje

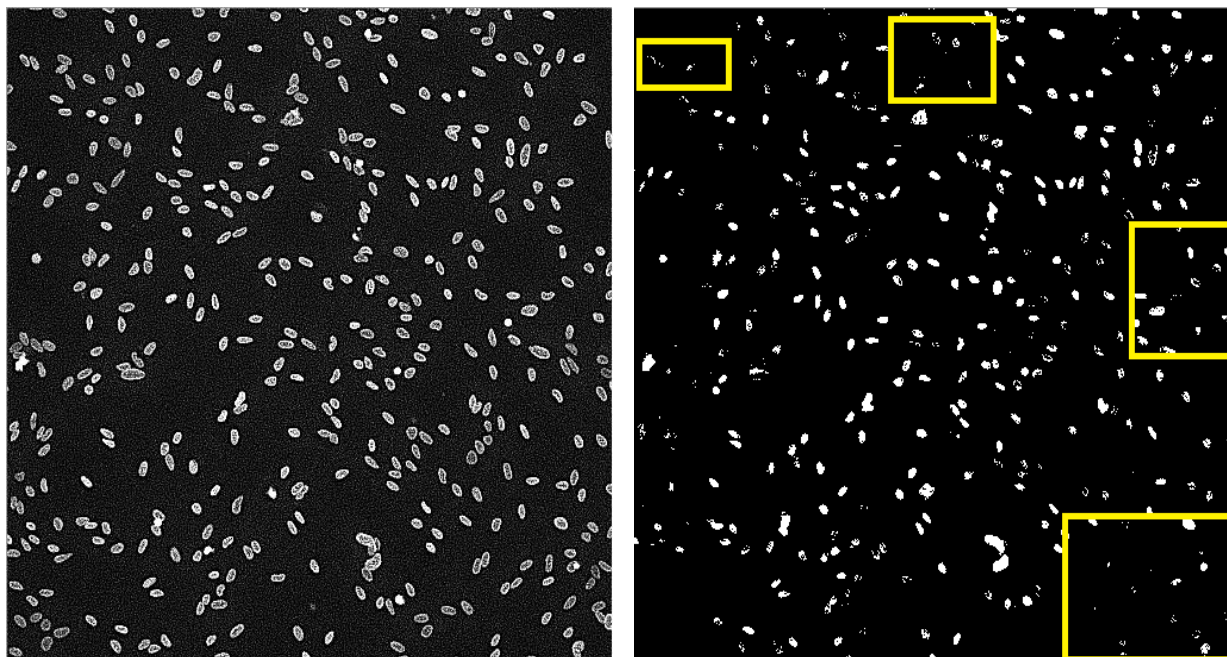
vrednost 255 čime je označeno da je on potencijalno značajan element slike (potencijalno pripada ćeliji). Elementi slike koji imaju vrednost manju od granične se postavljaju na vrednost 0, jer na taj način je moguće smatrati ih pozadinom i nebitnim elementima za dalje istraživanje. Neki od primera primene praga propustljivosti se mogu videti na slikama 5.4a, 5.4b i 5.4c gde je levo prikazana originalna slika (nakon primene Laplasovog filtera) a desno se nalazi slika koja predstavlja rezultat primene praga propustljivosti određene vrednosti.



Slika 5.4a Originalna slika (levo) – Slika nakon primene praga propustljivosti vrednosti 55 (desno)



Slika 5.4b Originalna slika (levo) – Slika nakon primene praga propustljivosti vrednosti 105 (desno)

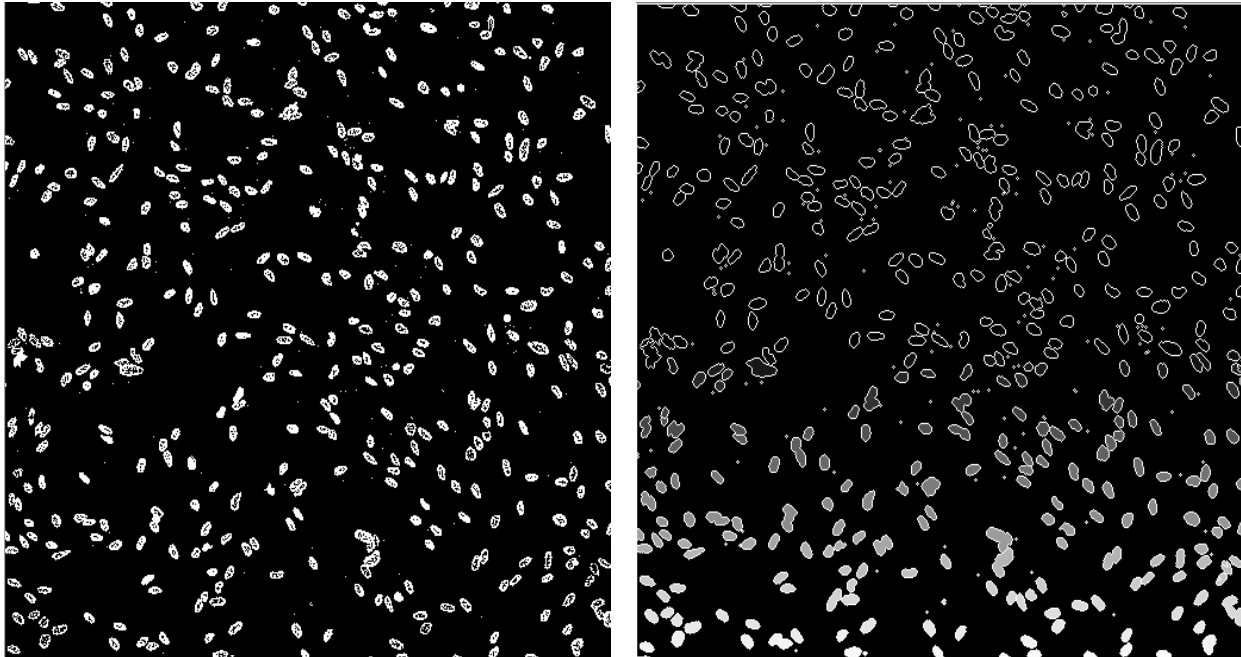


Slika 5.4c Originalna slika (levo) – Slika nakon primene praga propustljivosti vrednosti 125 (desno)

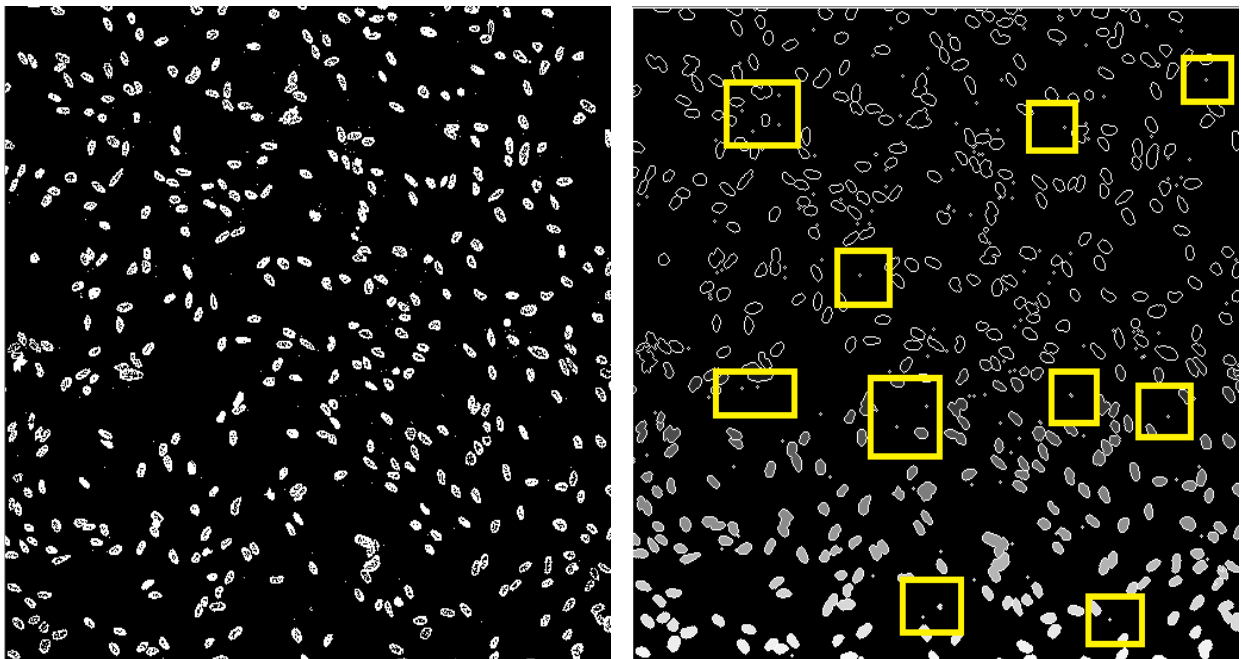
Na predstavljenim primerima na slikama 5.4a, 5.4b i 5.4c jasno se vidi da primena praga propustljivosti ima veliki uticaj na rešavanje ovog problema i da je izbor dobre granične vrednosti od ključne važnosti za dalji rad na ovom problemu. Slika 5.4a prikazuje rezultat kada je granična vrednost premala, uočljivo je da je prisustvo šuma veliko i da može jako da utiče na potencijalno prebrojavanje. Slika 5.4b prikazuje primenu praga propustljivosti sa graničnom vrednošću za koju može da se kaže da je dobro odabrana, žutom bojom je označen šum ali ono što je razlika u odnosu na sliku 5.4a šum je na vrlo niskom nivou. Slika 5.4c prikazuje problem kada je granična vrednost previsoka, šum je eliminisan ali to je dovelo do toga da su i značajni podaci izgubljeni. To se jasno primećuje u označenim delovima gde ako se ti delovi uporede sa originalnom slikom i bez puno truda je uočljivo da su neke ćelije nestale primenom ovog praga. Iz ovih primera može se zaključiti da je pravilan izbor praga propustljivosti odnosno granične vrednosti inteziteta elementa slike jedan od najvažnijih parametara za rad programa i izabran je za jedan od atributa pri građenju klasifikacionog modela za drvo odlučivanja.

Nakon primene Laplasovog filtera za izoštravanje slike i praga dozvoljenog inteziteta elemenata slike može se reći da je završena obrada slike osnovnog nivoa i da je originalna slika obrađena dovoljno da može da se pristupi obradi srednjeg nivoa. U obradi srednjeg nivoa je iskorišćen algoritam izdvajanja ivica koji se na sliku primenjuje nakon primene praga propustljivosti. Algoritam izdvajanja ivica nalazi objekte na slici tako što im određuje spoljašnje granice (pravi razliku između pozadine i objekata) i za rezultat ima niz objekata za koje je poznata njihova pozicija na slici, kao i površina odnosno broj elemenata slike koje objekat zauzima. Posmatrano pojedinačno ćelije su približno slične površine odnosno prostor koji zauzimaju je približno isti, pa se površina prepoznatog objekta može smatrati jako važnom osobinom ćelije za dalji rad. Na

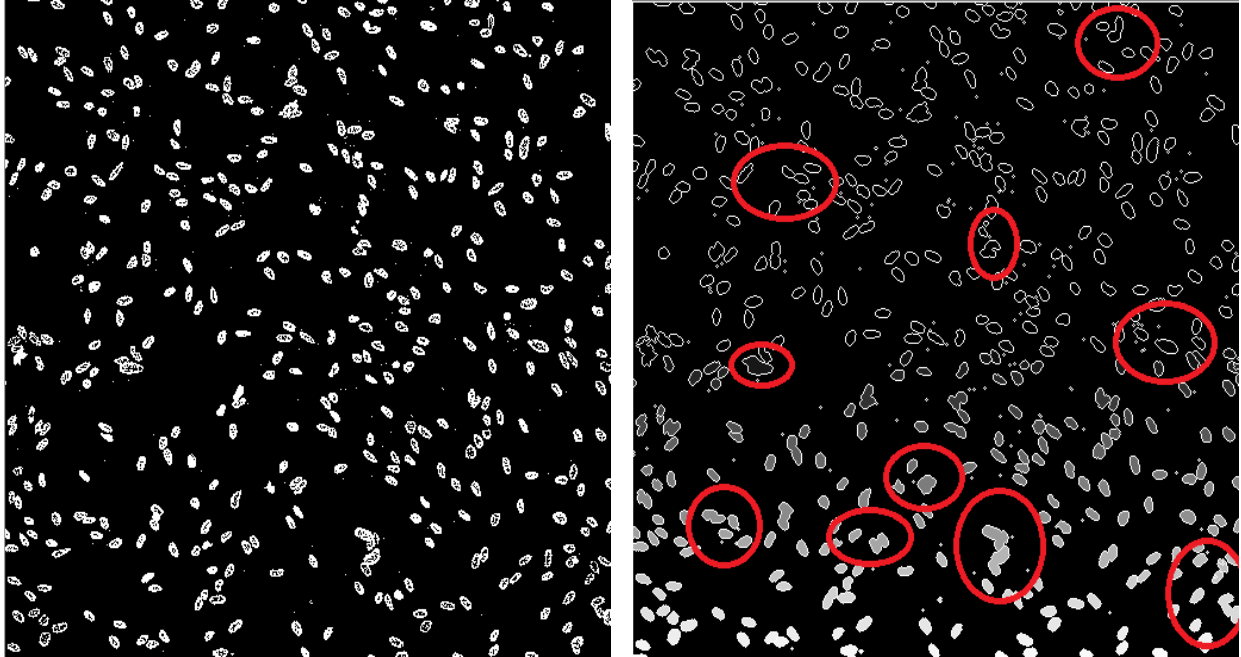
slikama 5.5a, 5.5b i 5.5c su prikazani rezultati primene algoritma za izdvajanje ivica gde se levo nalazi originalna slika (nakon primene praga dozvoljenog inteziteta), a desno se nalazi slika koja je napravljena programski crtajući objekte prepoznate uz pomoć algoritma za izdvajanje ivica.



Slika 5.5a Originalna slika (levo) – Slika dobijena koristeći rezultate algoritma za izdvajanje ivica (desno)

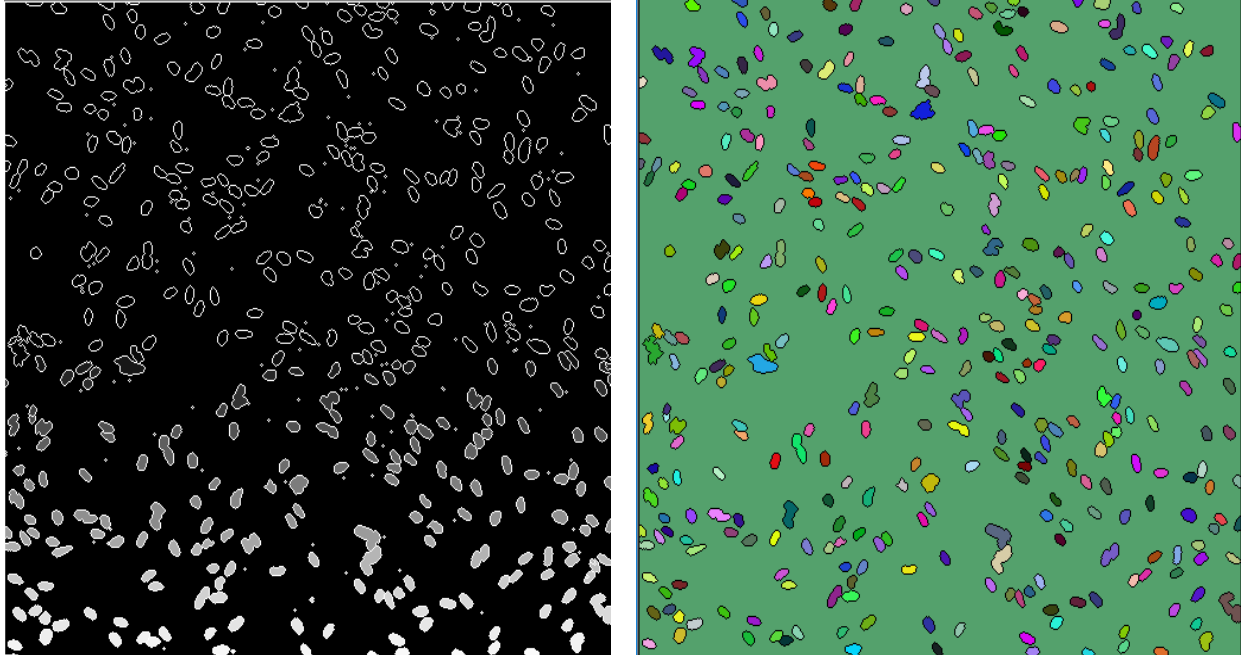


Slika 5.5b Originalna slika (levo) – Slika dobijena koristeći rezultate algoritma za izdvajanje ivica + šum (desno)



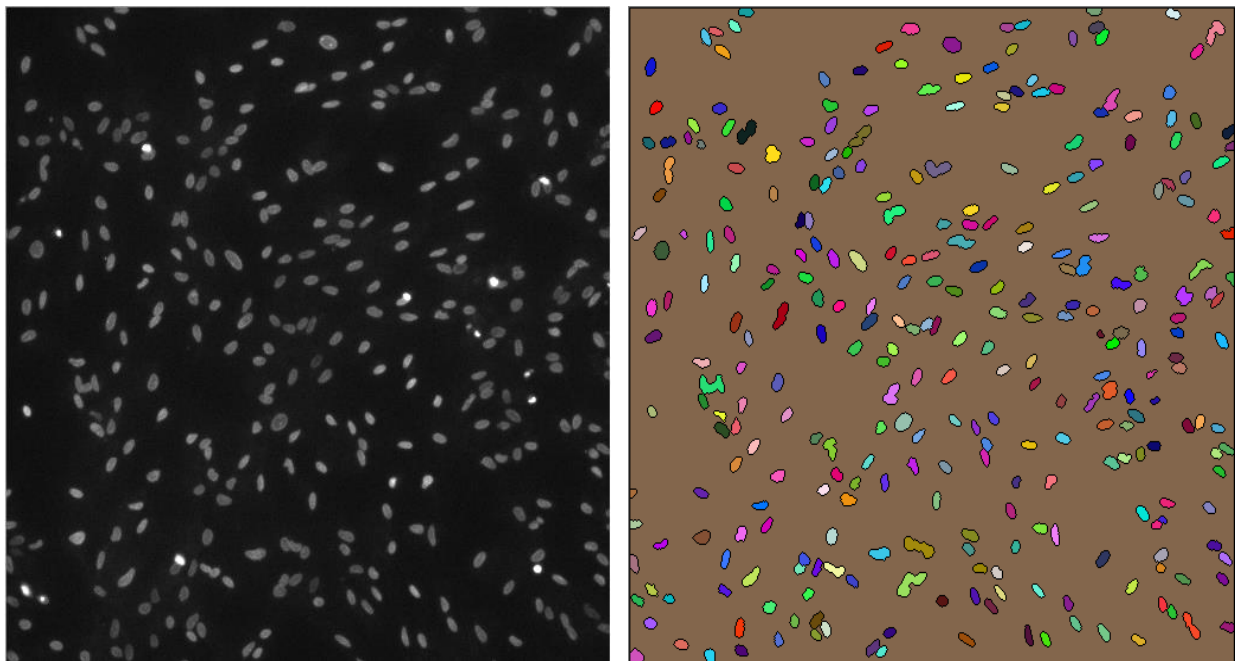
Slika 5.5c Originalna slika (levo) – Slika dobijena koristeći rezultate algoritma za izdvajanje ivica + grupe (desno)

Na slici 5.5b desno može se primetiti prisustvo sitnih tačkica koje su označene i koje su prepoznate kao objekti a predstavljaju šum. Najlakši način za njihovu eliminaciju bi bio da se u krajnjoj obradi prepoznati objekti koji su više puta površinski manji od prosečne površine prepoznatih objekata ne uzimaju u obzir pri daljoj obradi. Isto tako na slici 5.5c desno, označeni su objekti koji su prepoznati kao jedan objekat, a sačinjeni su od potencijalno više ćelija. U krajnjoj obradi se ne može smatrati da je takav prepoznati objekat jedan objekat već na osnovu površine pretpostavlja koliko je objekata učestvovalo u formiranju tog većeg objekta. Na osnovu ove analize površina objekta je uzeta kao drugi atribut za razmatranje za izgradnju klasifikacionog modela. Sa izvojenim objektima može se reći da je obrada slike srednjeg nivoa završena i da je omogućeno da se pređe na obradu naprednog nivoa gde će se izdvojenim objektima dodeliti značenje i izvršiti prebrojavanje kao poslednji korak i rezultat za izlaz iz programa. Iz znanja koje pruža niz izdvojenih objekata pronalazi se prosečna veličina objekta odnosno jedne ćelije na snimku. Svi objekti koji su značajno manji od prosečnog nisu uzeti u razmatranje za brojanje dok su svi objekti koji su nekoliko puta veći od prosečnog smatrani grupom objekata čiji je broj srazmeran veličini objekta u odnosu na veličinu prosečnog prepoznatog objekta. Na taj način formiran je novi niz objekata koji se potom koriste za stvaranje nove slike koja predstavlja vizuelni rezultat rada i obrade programa. Da bi bilo koji korisnik programa mogao lakše da ima uvid u prebrojane objekte kao i u to da li je neki objekat izostavljen u krajnjem rezultatu prepoznati objekti i pozadina su obojeni slučajnim izborom boja. Na taj način je lepo vizuelno predstavljen niz objekata koji su izlaz obrade odnosno oni koji su prepoznati kao jedinstvene ćelije ili grupe ćelija. Rezultat obrade slike izdvojenih ivica je predstavljen na slici 5.6.

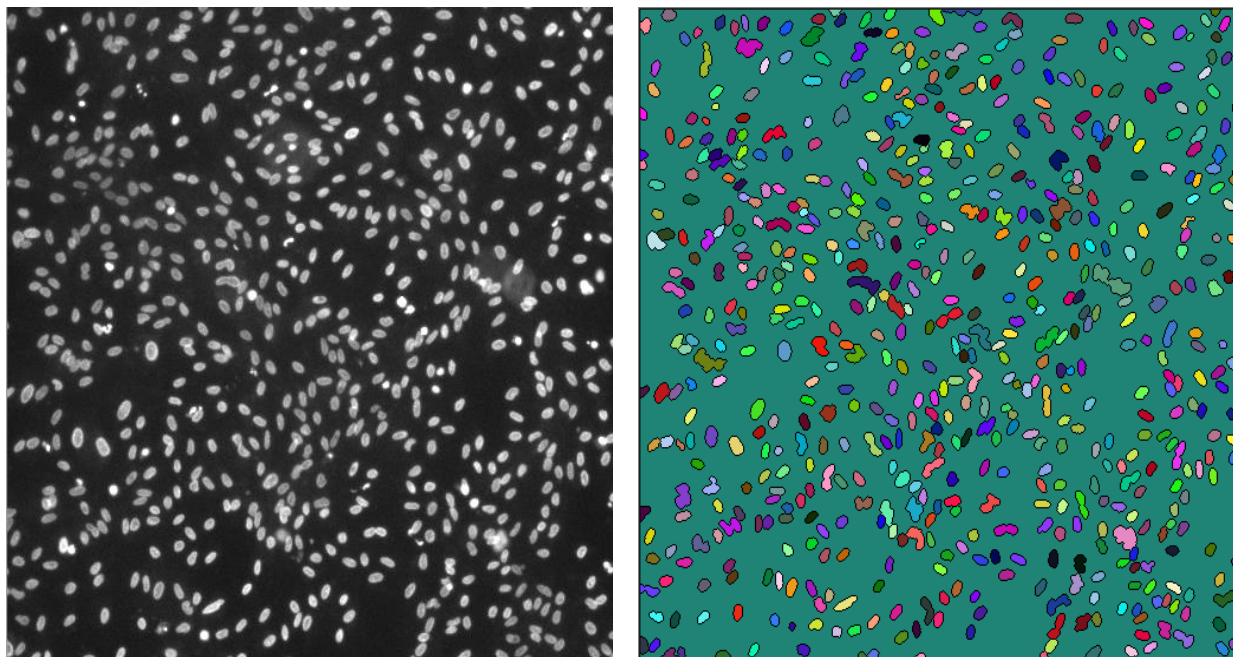


Slika 5.6 Originalna slika izdvojenih ivica (levo) – Rezultujuća slika prepoznatih oblika (desno)

Kao što se može primetiti na slici 5.6 desno, rezultujuća slika je obojena i eliminisan je šum. Prepoznate su ćelije ili grupe ćelija na osnovu kojih se formira krajnji rezultat prebrojavanja bitan za korisnika programa. Još neki primeri rezultata obrade su predstavljeni na slikama 5.7a i 5.7b.

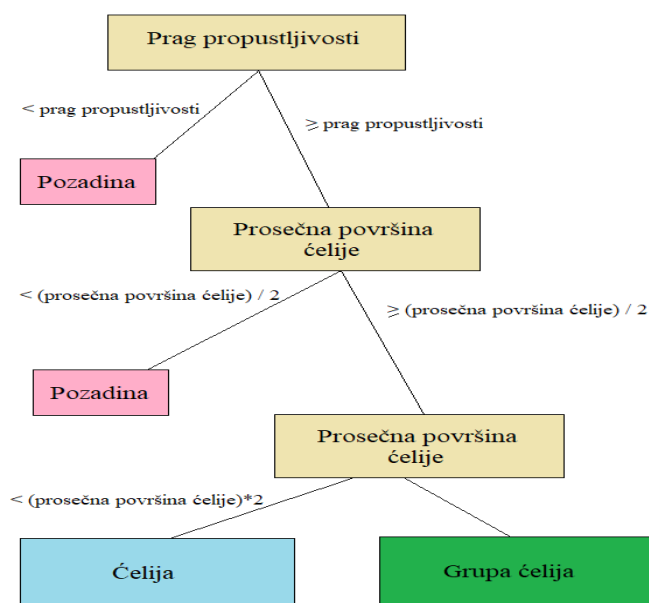


Slika 5.7a Originalna ulazna slika (levo) – Rezultujuća slika prepoznatih oblika (desno)



Slika 5.7b Originalna ulazna slika (levo) – Rezultujuća slika prepoznatih oblika (desno)

Kao što se može videti rezultujuće slike daju jak vizuelni utisak o tome kakva je uspešnost prepoznavanja ćelija bila. Taj utisak je dovoljan za zaključak da li se rezultat može uzeti u razmatranje. Za ozbiljnije analize potreban je ipak pristup i sa strane istraživanja podataka. U tu svrhu izgrađen je klasifikacioni model zasnovan na drvetu odlučivanja. Za svaku grupu tri različita snimanja izabrani su trening podaci na osnovu kojih je određen prag dozvoljenog inteziteta elemenata slike i prosečna površina ćelije. Te vrednosti su direktno učestvovalе u formiranju klasifikacionog modela predstavljenog na slici 5.8.



Slika 5.8 Klasifikacioni model

Klasifikacioni model je formiran na osnovu početnog istraživanja i upoznavanja sa snimcima. Unapred su određene tri klase kojima element slike može pripadati. To su pozadina, ćelija i grupa ćelija. Na osnovu prepoznatih atributa pri izradi programa i početnom testiranju formirano je drvo odlučivanja. Atributi od značaja od kojih zavisi određivanje klasa su prag propustljivosti (dozvoljeni intezitet vrednosti) i prosečna površina ćelije. Vrednost ovih atributa je određena treniranjem modela slučajno odabranim snimcima za trening.

Primenom klasifikacionog modela na snimcima za testiranje dobijeni su rezultati koji će biti predstavljeni u narednom delu zajedno sa poređenjem rezultata koji su postignuti korišćenjem poznatih programa.

7. Rezultati i njihovo poređenje sa drugim programima

Pre samog predstavljanja rezultata i poređenja sa rezultatima drugih programa važno je napomenuti da ni rezultati dobijeni korišćenjem drugih programa nisu u potpunosti precizni i postoje razlike u rezultatima. Razlog toga je kompleksnost problema, u ovom konkretnom slučaju u pitanju su situacije kada se radi o rešavanju prebrojavanja grupa ćelija, stoga nije bilo moguće koristiti uobičajene mere za određivanje preciznosti klasifikacionog modela, pa je upoređivanje rezultata razvijenog programa vršeno u odnosu na konačne rezultate prebrojavanja dobijene zajedno sa snimcima. Dostupni rezultati za poređenje su dobijeni korišćenjem programa *CellC* i *Imaris* i opšti utisak istraživača sa Biološkog fakulteta je da je brojanje pomoću programa *Imaris* dosta preciznije, jer program *CellC* daje rezultate brojanja koji konstantno potcenjuju broj ćelija na snimcima. Poređenje je rađeno na 94 snimka iz 3 različita snimanja.

Rezultati obrade slika iz prvog snimanja

Pri prvom snimanju dobijeno je 10 snimaka i u tabeli 7.1a su prikazani rezultati koje su istraživači dobili. Za potrebe treniranja modela slučajnim izborom je određeno 7 snimaka čiji su rezultati prikazani (obojeni plavom bojom) dok su za test podatke određena 3 snimka sa rezultatima (obojeni zelenom bojom). Snimci su sa oznakom GD-TritonX.

Snimak	CellC	Imaris	U proseku
1	330	375	353
2	375	454	415
3	475	507	491
4	429	448	439
5	515	566	541
6	380	408	394
7	521	533	527
8	485	526	506
9	465	501	483
10	515	545	530

Tabela 7.1a Rezultati prvog snimanja, snimci GD-TritonX

Pri treniranju klasifikacionog modela određen je prag dozvoljenog inteziteta u vrednosti od 88 i prosečna površina ćelije od 34 elementa slike. Primenom ovih vrednosti i na trening i na test podatke dobijeni su rezultati predstavljeni u tabeli 7.1b. Kao što se može videti u tabeli, program sa ovako zadatim pragom dozvoljenog inteziteta i prosečne površine ćelije za rezultat ima više prebrojanih ćelija nego što je prosek programa *CellC* i *Imaris*. Rezultati su približniji rezultatima koje daje program *Imaris*.

Snimak	CellC	Imaris	U proseku	Razvijeni program
1	330	375	353	355
2	375	454	415	452
3	475	507	491	496
4	429	448	439	483
5	515	566	541	546
6	380	408	394	418
7	521	533	527	525
8	485	526	506	505
9	465	501	483	488
10	515	545	530	546

Tabela 7.1b Rezultati prvog snimanja zajedno sa rezultatima razvijenog programa, snimci GD-TritonX

Rezultati obrade slika iz drugog snimanja

Pri drugom snimanju snimljeno je više snimaka u više različitih grupa gde su ćelije bile podvrgnute različitim tretmanima. Za svaku grupu snimaka je rađeno treniranje modela i iskorišćene su oznake dobijene od istraživača.

Prva grupa su snimci GD-TritonX čiji su rezultati predstavljeni u tabeli 7.2a (plavom bojom su obojeni snimci za trening, a zelenom za snimci za test).

Snimak	CellC	Imaris	U proseku
GD-TritonX-01-tritc	738	806	772
GD-TritonX-02-tritc	680	796	738
GD-TritonX-03-tritc	821	844	833
GD-TritonX-04-tritc	770	857	814
GD-TritonX-05-tritc	833	916	875
GD-TritonX-06-tritc	700	779	740
GD-TritonX-07-tritc	762	894	828
GD-TritonX-08-tritc	559	645	602
GD-TritonX-09-tritc	670	785	728
GD-TritonX-10-tritc	591	665	628

Tabela 7.2a Rezultati drugog snimanja snimci GD-TritonX

Treniranjem klasifikacionog modela određene su vrednosti praga dozvoljenog inteziteta od 118 i prosečne površine ćelije 33. Primenom tih vrednosti na trening i test snimke dobijeni su rezultati predstavljeni u tabeli 7.2b. Kao i u radu sa snimcima iz prvog snimanja i sa ovom grupom snimaka se težilo pri treniranju da rezultati budu bliži rezultatima koje daje program *Imaris* tako su i rezultati klasifikacionog modela bliži rezultatima programa *Imaris*.

Snimak	CellC	Imaris	U proseku	Razvijeni program
GD-TritonX-01-tritc	738	806	772	734
GD-TritonX-02-tritc	680	796	738	770
GD-TritonX-03-tritc	821	844	833	826
GD-TritonX-04-tritc	770	857	814	865
GD-TritonX-05-tritc	833	916	875	930
GD-TritonX-06-tritc	700	779	740	727
GD-TritonX-07-tritc	762	894	828	889
GD-TritonX-08-tritc	559	645	602	621
GD-TritonX-09-tritc	670	785	728	812
GD-TritonX-10-tritc	591	665	628	664

Tabela 7.2b Rezultati drugog snimanja zajedno sa rezultatima razvijenog programa (snimci GD-TritonX)

Druga grupa su snimci sa oznakom GD+res. Od dostupnih 8 snimaka 6 je izabrano za trening dok su 2 izabrana za testiranje. Dostupni rezultati su predstavljeni u tabeli 7.3a.

Snimak	CellC	Imaris	U proseku
GD+res-01-tritc	283	275	279
GD+res-02-tritc	266	270	268
GD+res-03-tritc	274	264	269
GD+res-04-tritc	240	229	235
GD+res-05-tritc	223	208	216
GD+res-06-tritc	187	185	186
GD+res-07-tritc	525	531	528
GD+res-08-tritc	462	469	466

Tabela 7.3a Rezultati drugog snimanja, snimci GD+res

Treniranjem modela određene su vrednosti za prag dozvoljenog inteziteta 90 i za vrednost prosečne površine ćelije je određena vrednost 27. Rezultati primene ovog klasifikacionog modela na trening i test podatke su predstavljeni u tabeli 7.3b.

Snimak	CellC	Imaris	U proseku	Razvijeni program
GD+res-01-tritc	283	275	279	275
GD+res-02-tritc	266	270	268	262
GD+res-03-tritc	274	264	269	263
GD+res-04-tritc	240	229	235	231
GD+res-05-tritc	223	208	216	205
GD+res-06-tritc	187	185	186	181
GD+res-07-tritc	525	531	528	536
GD+res-08-tritc	462	469	466	483

Tabela 7.3b Rezultati drugog snimanja zajedno sa rezultatima razvijenog programa (snimci GD+res)

Treća grupa su snimci sa oznakom res+glu+FBS-TritonX. Od ukupno 9 snimaka iz ove grupe 6 je odabrano da predstavljaju snimke za treniranje modela dok su 3 snimka odabrana da budu test snimci. Dostupni rezultati su predstavljeni u tabeli 7.4a.

Snimak	CellC	Imaris	U proseku
res+glu+FBS-TritonX-01-tritc	735	791	763
res+glu+FBS-TritonX-02-tritc	717	751	734
res+glu+FBS-TritonX-03-tritc	706	743	725
res+glu+FBS-TritonX-04-tritc	757	803	780
res+glu+FBS-TritonX-05-tritc	735	775	755
res+glu+FBS-TritonX-06-tritc	---	603	603
res+glu+FBS-TritonX-07-tritc	521	559	540
res+glu+FBS-TritonX-08-tritc	647	676	662
res+glu+FBS-TritonX-09-tritc	630	664	647

Tabela 7.4a Rezultati drugog snimanja snimaka res+glu+FBS-TritonX

Treniranjem modela određene su vrednosti za prag dozvoljenog inteziteta 90 i za vrednost prosečne površine ćelije je određena vrednost 31. Rezultati primene ovog klasifikacionog modela na trening i test podatke su predstavljeni u tabeli 7.4b. Ovime je završena obrada snimaka koji su dobijeni u drugom snimanju.

Snimak	CellC	Imaris	U proseku	Razvijeni program
res+glu+FBS-TritonX-01-tritc	735	791	763	790
res+glu+FBS-TritonX-02-tritc	717	751	734	757
res+glu+FBS-TritonX-03-tritc	706	743	725	732
res+glu+FBS-TritonX-04-tritc	757	803	780	813
res+glu+FBS-TritonX-05-tritc	735	775	755	764
res+glu+FBS-TritonX-06-tritc	---	603	603	610
res+glu+FBS-TritonX-07-tritc	521	559	540	572
res+glu+FBS-TritonX-08-tritc	647	676	662	652
res+glu+FBS-TritonX-09-tritc	630	664	647	646

Tabela 7.4b Rezultati drugog snimanja zajedno sa rezultatima razvijenog programa(res+glu+FBS-TritonX)

Rezultati obrade slika iz trećeg snimanja

Treće snimanje takođe sadrži više različitih grupa ćelija, snimci su izvedeni u različitim vremenskim periodima sa primenom različitih tretmana nad ćelijama. Prva grupa su snimci sa oznakom GD-TritonX. U tabeli 7.5a su predstavljeni dostupni rezultati za tu grupu snimaka.

Snimak	CellC	Imaris	U proseku
GD-TritonX-01-tritc	551	580	566
GD-TritonX-02-tritc	441	508	475
GD-TritonX-03-tritc	384	432	408
GD-TritonX-04-tritc	499	556	528
GD-TritonX-05-tritc	613	668	641
GD-TritonX-06-tritc	610	645	628
GD-TritonX-07-tritc	616	672	644
GD-TritonX-08-tritc	592	628	610
GD-TritonX-09-tritc	653	704	679
GD-TritonX-10-tritc	631	652	642
GD-TritonX-11-tritc	642	667	655

Tabela 7.5a Rezultati trećeg snimanja, snimci GD-TritonX

Treniranjem modela određene su vrednosti za prag dozvoljenog inteziteta 118 i za vrednost prosečne površine ćelije je određena vrednost 32. Rezultati primene ovog klasifikacionog modela na trening i test podatke su predstavljeni u tabeli 7.5b. Zanimljivo je da se prosečna površina ćelije smanjila u kasnijim snimcima, tako da je istrenirani model zbog toga za prvih par snimaka davao veće rezultate od onih dobijenih drugim programima, dok je za kasnije snimke dobijao manje rezultate prebrojavanja od onih dobijenih drugim programima.

Snimak	CellC	Imaris	U proseku	Razvijeni program
GD-TritonX-01-tritc	551	580	566	586
GD-TritonX-02-tritc	441	508	475	511
GD-TritonX-03-tritc	384	432	408	437
GD-TritonX-04-tritc	499	556	528	563
GD-TritonX-05-tritc	613	668	641	644
GD-TritonX-06-tritc	610	645	628	637
GD-TritonX-07-tritc	616	672	644	627
GD-TritonX-08-tritc	592	628	610	611
GD-TritonX-09-tritc	653	704	679	677
GD-TritonX-10-tritc	631	652	642	594
GD-TritonX-11-tritc	642	667	655	621

Tabela 7.5b Rezultati trećeg snimanja zajedno sa rezultatima razvijenog programa(GD-TritonX)

Druga grupa su snimci sa oznakom GD-res i dostupni rezultati prebrojavanja za ovu grupu su predstavljeni u tabeli 7.6a.

Snimak	CellC	Imaris	U proseku
GD-res-01-tritc	215	234	225
GD-res-02-tritc	252	267	260
GD-res-03-tritc	377	401	389
GD-res-04-tritc	370	405	388
GD-res-05-tritc	217	269	243
GD-res-06-tritc	248	269	259
GD-res-07-tritc	492	527	510
GD-res-08-tritc	352	381	367
GD-res-09-tritc	359	389	374
GD-res-10-tritc	367	381	374

Tabela 7.6a Rezultati trećeg snimanja, snimci GD-res

Treniranjem modela određene su vrednosti za prag dozvoljenog inteziteta 126 i za vrednost prosečne površine ćelije je određena vrednost 24. Rezultati primene ovog klasifikacionog modela na trening i test podatke su predstavljeni u tabeli 7.6b. Jako dobri rezultati postignuti ovim klasifikacionim modelom. Vrlo mala odstupanja u odnosu na prosečan broj i na broj koji daje program *Imaris*.

Snimak	CellC	Imaris	U proseku	Razvijeni program
GD-res-01-tritc	215	234	225	229
GD-res-02-tritc	252	267	260	264
GD-res-03-tritc	377	401	389	376
GD-res-04-tritc	370	405	388	405
GD-res-05-tritc	217	269	243	213
GD-res-06-tritc	248	269	259	261
GD-res-07-tritc	492	527	510	514
GD-res-08-tritc	352	381	367	372
GD-res-09-tritc	359	389	374	380
GD-res-10-tritc	367	381	374	382

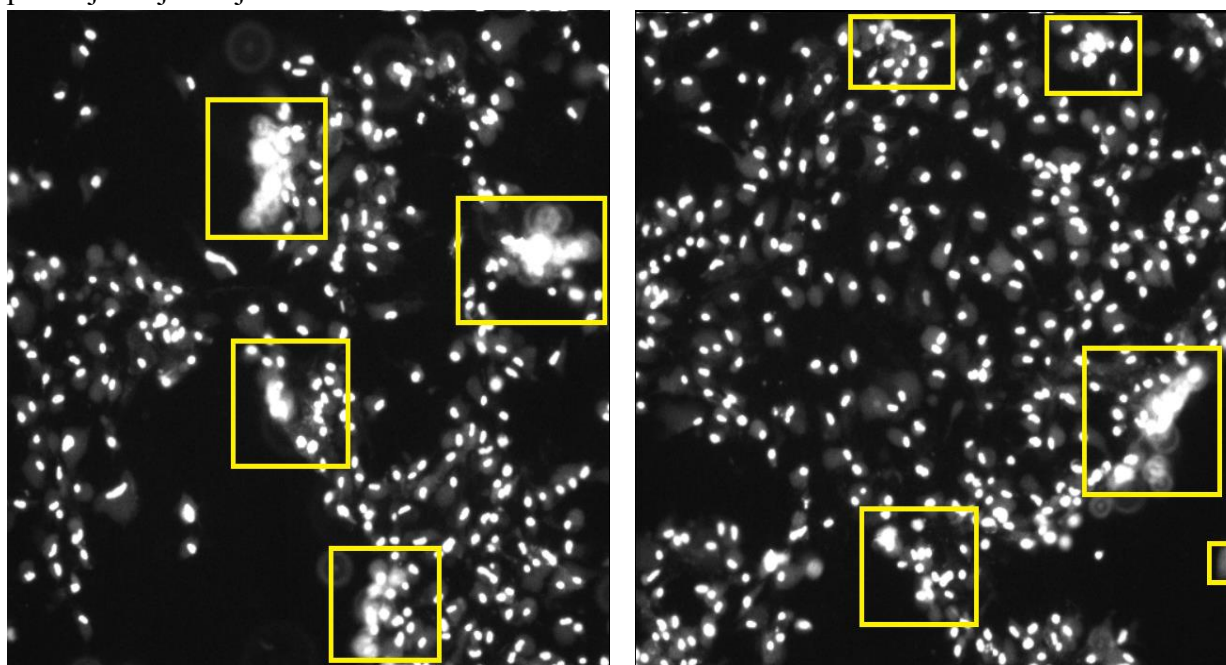
Tabela 7.6b Rezultati trećeg snimanja zajedno sa rezultatima razvijenog programa(GD-res)

Treća grupa su snimci sa oznakom res+glu i dostupni rezultati prebrojavanja za ovu grupu su predstavljeni u tabeli 7.7a.

Snimak	CellC	Imaris	U proseku
res+glu-01-tritc	165	177	171
res+glu-02-tritc	313	332	323
res+glu-03-tritc	288	303	296
res+glu-04-tritc	199	207	203
res+glu-05-tritc	143	150	147
res+glu-06-tritc	124	132	128
res+glu-07-tritc	271	336	304
res+glu-08-tritc	359	421	390
res+glu-09-tritc	171	180	176
res+glu-10-tritc	190	203	197
res+glu-11-tritc	138	150	144
res+glu-12-tritc	172	181	177

Tabela 7.7a Rezultati trećeg snimanja, snimci(res+glu)

Ono što je uticalo na treniranje ovog modela su snimci gde je bilo teško čak i posmatraču da otkrije o kom broju ćelija se radi pošto dolazi do velikog preklapanja. Primer takvih snimaka je predstavljen na slici 7.1 gde je žutom bojom obeležen deo koji predstavlja problem pri radu. Zbog tih snimaka ovaj klasifikacioni model je jedan od nepreciznijih od svih dobijenih u ovom radu kada se uporedi sa rezultatima koje su dali drugi programi. U ovakvim slučajevima predlog je da takvi snimci budu obrađeni jedinstveno bez uticaja na grupu ostalih snimaka. Svakako taj problem predstavlja i osnovu i razmišljanje za dalji rad na problemu prepoznavanja i prebrojavanja ćelija.



Slika 7.1 Snimci koji imaju delove koji predstavljaju problem za prepoznavanje

Treniranjem modela određene su vrednosti za prag dozvoljenog inteziteta 154 i za vrednost prosečne površine ćelije je određena vrednost 32. Rezultati primene ovog klasifikacionog modela na trening i test podatke su predstavljani u tabeli 7.7b.

Snimak	CellC	Imaris	U proseku	Razvijeni program
res+glu-01-tritc	165	177	171	166
res+glu-02-tritc	313	332	323	299
res+glu-03-tritc	288	303	296	277
res+glu-04-tritc	199	207	203	188
res+glu-05-tritc	143	150	147	136
res+glu-06-tritc	124	132	128	120
res+glu-07-tritc	271	336	304	414
res+glu-08-tritc	359	421	390	445
res+glu-09-tritc	171	180	176	380
res+glu-10-tritc	190	203	197	161
res+glu-11-tritc	138	150	144	135
res+glu-12-tritc	172	181	177	179

Tabela 7.7b Rezultati trećeg snimanja zajedno sa rezultatima razvijenog programa(res+glu)

Četvrta grupa su snimci sa oznakom res+glu+FBS i dostupni rezultati prebrojavanja za ovu grupu su predstavljani u tabeli 7.8a.

Snimak	CellC	Imaris	U proseku
res+glu+FBS-01-tritc	6	7	7
res+glu+FBS-02-tritc	22	29	26
res+glu+FBS-03-tritc	149	171	160
res+glu+FBS-04-tritc	52	60	56
res+glu+FBS-05-tritc	90	110	100
res+glu+FBS-06-tritc	50	57	54
res+glu+FBS-07-tritc	56	71	64
res+glu+FBS-08-tritc	96	104	100
res+glu+FBS-09-tritc	302	336	319
res+glu+FBS-10-tritc	108	123	116
res+glu+FBS-11-tritc	75	90	83
res+glu+FBS-12-tritc	91	110	101

Tabela 7.8a Rezultati trećeg snimanja, snimci(res+glu+FBS)

Treniranjem modela određene su vrednosti za prag dozvoljenog inteziteta 96 i za vrednost prosečne površine ćelije je određena vrednost 29. Rezultati primene ovog klasifikacionog modela na trening i test podatke su predstavljani u tabeli 7.8b. Zanimljivo je da broj ćelija u ovoj grupi

snimaka se kreće od jedva nekoliko, do nekoliko stotina ćelija po snimku i da je klasifikacioni model dao korektne rezultate u poređenju sa dostupnim rezultatima drugih programa.

Snimak	CellC	Imaris	U proseku	Razvijeni program
res+glu+FBS-01-tritc	6	7	7	6
res+glu+FBS-02-tritc	22	29	26	26
res+glu+FBS-03-tritc	149	171	160	155
res+glu+FBS-04-tritc	52	60	56	50
res+glu+FBS-05-tritc	90	110	100	103
res+glu+FBS-06-tritc	50	57	54	75
res+glu+FBS-07-tritc	56	71	64	64
res+glu+FBS-08-tritc	96	104	100	121
res+glu+FBS-09-tritc	302	336	319	370
res+glu+FBS-10-tritc	108	123	116	130
res+glu+FBS-11-tritc	75	90	83	75
res+glu+FBS-12-tritc	91	110	101	105

Tabela 7.8b Rezultati trećeg snimanja zajedno sa rezultatima razvijenog programa(res+glu+FBS)

Peta grupa i poslednja iz trećeg snimanja ujedno i poslednja grupa koja će biti obrađena u ovom radu sadrži 12 snimaka sa oznakom res+glu+FBS-TritonX. Dostupni rezultati su predstavljeni u tabeli 7.9a.

Snimak	CellC	Imaris	U proseku
res+glu+FBS-TritonX-01	644	684	664
res+glu+FBS-TritonX-02	601	609	605
res+glu+FBS-TritonX-03	719	720	720
res+glu+FBS-TritonX-04	529	563	546
res+glu+FBS-TritonX-05	606	647	627
res+glu+FBS-TritonX-06	824	772	798
res+glu+FBS-TritonX-07	461	489	475
res+glu+FBS-TritonX-08	698	723	711
res+glu+FBS-TritonX-09	731	727	729
res+glu+FBS-TritonX-10	1094	972	1033
res+glu+FBS-TritonX-11	694	765	730
res+glu+FBS-TritonX-12	669	753	711

Tabela 7.9a Rezultati trećeg snimanja, snimci(res+glu+FBS-TritonX)

Treniranjem modela određene su vrednosti za prag dozvoljenog inteziteta 108 i za vrednost prosečne površine ćelije je određena vrednost 32. Rezultati primene ovog klasifikacionog modela na trening i test podatke su predstavljeni u tabeli 7.9b.

Snimak	CellC	Imaris	U proseku	Razvijeni program
res+glu+FBS-TritonX-01	644	684	664	686
res+glu+FBS-TritonX-02	601	609	605	651
res+glu+FBS-TritonX-03	719	720	720	718
res+glu+FBS-TritonX-04	529	563	546	560
res+glu+FBS-TritonX-05	606	647	627	633
res+glu+FBS-TritonX-06	824	772	798	781
res+glu+FBS-TritonX-07	461	489	475	509
res+glu+FBS-TritonX-08	698	723	711	740
res+glu+FBS-TritonX-09	731	727	729	766
res+glu+FBS-TritonX-10	1094	972	1033	1002
res+glu+FBS-TritonX-11	694	765	730	752
res+glu+FBS-TritonX-12	669	753	711	725

Tabela 7.9b Rezultati trećeg snimanja zajedno sa rezultatima razvijenog programa(res+glu+FBS-TritonX)

Ovim rezultatima je upoređena i poslednja grupa snimaka za koju su bili dostupni i poznati rezultati koji su dobijeni korišćenjem programa *CellC* i *Imaris*.

Kao što je već napomenuto nije bilo moguće izvršiti proveru preciznosti i greške koristeći standardne metode u tehnikama istraživanja podataka jer i dostupni rezultati sadrže greške tako da nije bilo moguće odrediti tačne performanse dobijenih klasifikacionih modela. Ono što je moguće uporediti i kada se pogleda kroz tabele prosečan broj prepoznatih ćelija i broj koji se kao rezultat dobio primenom klasifikacionog modela, ocena je da u odnosu na prosečan broj prebrojanih ćelija, razvijeni program uglavnom daje veći broj a ta razlika je najviše do 9% veći broj u odnosu na prosečan broj dobijen prebrojavanjem korišćenjem drugih programa.

8. Zaključak i ideje za dalji rad

U ovom radu opisan je razvoj programa za prepoznavanje i prebrojavanje broja ćelija na snimcima dobijenim fluorescentnim mikroskopom za potrebe rada istraživača na Biološkom fakultetu. Upoređeni su rezultati koje razvijeni program daje sa drugim poznatim programima koji se koriste za taj problem. Za razvoj programa korišćene su tehnike obrade slika i tehnike istraživanja podataka. Uz pomoć tehnika za obradu slika, snimci su obrađeni tako da je kvalitet snimka za obradu poboljšan, prepoznati su objekti i napravljen način da se razlikuju delovi od značaja u odnosu na šum. Tehnike istraživanja podataka su iskorišćene za formiranje klasifikacionog modela za prebrojavanje ćelija.

U ovom radu obrađeno je 94 snimka iz 3 različita snimanja gde su vršeni različiti tretmani nad ćelijama. Rezultati i zapažanja su sledeći:

1. U prvom snimanju u prvoj grupi pri prebrojavanju prosečan broj prebrojanih ćelija na svim snimcima je 468 uzimajući prosek programa *CellC* i *Imaris*, dok je prosečan broj prebrojanih ćelija razvijenog programa 481, tako da je odstupanje od poznatih programa oko 3%.
2. U drugom snimanju prva grupa snimaka ćelija pri prebrojavanju prosečna dostupna vrednost programa *CellC* i *Imaris* je 756 ćelija na snimcima dok je razvijenog programa 784, odstupanje oko 4%. Kod druge grupe ćelija dostupna prosečna vrednost 306 dok je razvijenog programa 305 odstupanje manje od 1%. Prosečna vrednost broja prebrojanih ćelija kod treće grupe je 689, a razvijenog programa 704 uz odstupanje manje od 3%.
3. U trećem snimanju prva grupa snimaka ćelija je imala u proseku izbrojano 589 ćelija programima *CellC* i *Imaris*, dok je razvijeni program imao rezultat od 591, odstupanje u ovom slučaju manje od 1%. Kod druge grupe ćelija dostupna prosečna vrednost iznosi 338 dok je razvijeni program imao 339, uz takođe odstupanje manje od 1%. Kod treće grupe ćelija dostupna prosečna vrednost prebrojanih ćelija je 221 dok je prosečna vrednost razvijenog programa 241 i odstupanje od oko 9%. Četvrta grupa ćelija je grupa sa najmanjom prosečnom vrednošću prebrojanih ćelija na snimcima i rezultat dostupnih programa je 99 dok je rezultat razvijenog programa 106 za prosečnu vrednost prebrojanih ćelija i odstupanje je oko 7%. Peta i poslednja grupa obrađenih ćelija je za prosečnu vrednost prebrojanih ćelija imala vrednost 695 dok je vrednost razvijenog programa 710 za odstupanje od oko 2%.

Utišak je da je razvijeni program postigao solidne rezultate, odstupanja su u najgorem slučaju bila oko 9% i to za grupu ćelija za koju je bilo karakteristično veliko prisustvo šuma i delova gde snimak nije najjasniji, a kvalitet snimka u tim situacijama nije bilo moguće popraviti. Razvijeni program je uglavnom davao veće rezultate u proseku od proseka rezultate programa *CellC* i *Imaris* što se može smatrati i uspehom jer je opšti utisak istraživača sa Biološkog fakulteta da je

program *Imaris* tačniji, a taj program daje veće rezultate od programa *CellC* za koji je utisak da potcenjuje ćelije pri prebrojavanju.

8.1. Ideje za dalji rad

Najčešći problem koji je razvijeni program imao su bili snimci sa većim grupama ćelija okupljenim na jednom mestu gde posmatrač sam nije u stanju da objasni o potencijalnom broju ćelija u toj grupi. Detaljnije istraživanje i bolje rešavanje tih situacija bi potencijalno donelo bolje rezultate razvijenog programa. Unapređenje tehnika obrade slika za prepoznavanje oblika kao i upošljavanje drugih tehnika istraživanja podataka bi sigurno uticali na poboljšanje performansi razvijenog programa i predstavljaju osnove za dalji rad.

9. Literatura

9.1. Koriščena literatura

- [1] Zhiwei Chen, Dongwu Liu, “*CONFOCAL FLUORESCENCE MICROSCOPY OF MUNG BEAN LEAVES*“, 2007, Natural Science Foundation for Outstanding Young Scholars of Shandong Province, China, Neobjavljen rad
- [2] Chau-Hwang Lee, “*Confocal Microscopy and Related Techniques*“, 2007, Research Center for Applied Sciences, Taiwan, Neobjavljen rad, Strane 13-14
- [3] Nathan S. Claxton, Thomas J. Fellers, and Michael W. Davidson, “*LASER SCANNING CONFOCAL MICROSCOPY*“, 2005, Department of Optical Microscopy and Digital Imaging, National High Magnetic Field Laboratory, The Florida State University, Neobjavljen rad, Strane 1-5
- [4] Andreas Ettinger and Torsten Wittman, “*Fluorescence Live Cell Imaging*“, 2014, Department of Cell and Tissue Biology, University of California, San Francisco, USA, Methods in Cell Biology 123, Strane 77–94
- [5] Rafael C. Gonzalez and Richard E. Woods “*Digital Image Processing (3rd Edition)*“, 2006, Prentice-Hall, Inc. Upper Saddle River, NJ, USA
- [6] Suzuki, S. and Abe, K., “*Topological Structural Analysis of Digitized Binary Images by Border Following*“, 1985, Computer Vision, Graphics, and Image Processing 30 1, Strane 32-46
- [7] Pang-Ning Tan, Michael Steinbach, Vipin Kumar “*Introduction to Data Mining, II ed*“, 2006, Pearson Education

9.2. Korisna literatura

- [1] Web prezentacija programa *CellC* - <https://sites.google.com/site/cellcsoftware/> (stanje 1.7.2017).
- [2] Web prezentacija programa *Imaris* - <http://www.bitplane.com/imaris/imaris> (stanje 1.7.2017)
- [3] Bradbury, S. and Evennett, P., “*Fluorescence microscopy, Contrast Techniques in Light Microscopy*“, 1996, BIOS Scientific Publishers, Ltd., Oxford, United Kingdom
- [4] Dokumentacija biblioteke *OpenCV* - <http://docs.opencv.org/2.4> (stanje 1.7.2017)
- [5] Xindong Wu, Vipin Kumar (eds.): “*The Top Ten Algorithms in Data Mining*“, 2009, CRC Press
- [6] Frank Y. Shih, “*IMAGE PROCESSING AND PATTERN RECOGNITION Fundamentals and Techniques*“, 2010, IEEE Press