



SVEUČILIŠTE U ZAGREBU - GEODETSKI FAKULTET
UNIVERSITY OF ZAGREB - FACULTY OF GEODESY
Zavod za inženjersku geodeziju i upravljanje prostornim informacijama
Institute of Engineering Geodesy and Spatial Information Management
Kačićeva 26; HR-10000 Zagreb, CROATIA
Web: www.igupi.geof.hr; Tel.: (+385 1) 45 61 222; Fax.: (+385 1) 48 28 081

Usmjerenje: Inženjerska geodezija i upravljanje prostornim informacijama

DIPLOMSKI RAD

Primjena OCR-a u vektorizaciji katastarskih planova

Izradio:

Zoran Stipanović

Bana Josipa Jelačića 3

47300 Ogulin

zorans@geodet.geof.hr

Mentor: prof. dr. sc. Miodrag Roić

Zagreb, travanj 2003.

**Zahvala:**

Zahvaljujem se svom mentoru, prof. dr. sc. Miodragu Roiću kao i asistentu mr. sc. Vladi Cetlu na pruženoj pomoći i korisnim savjetima pri izradi diplomskog rada.

Također se zahvaljujem svojim roditeljima na pruženoj podršci i razumijevanju tijekom čitavog školovanja.

Zahvaljujem se i vlasniku privatne geodetske tvrtke Geobiro d.o.o. Ogulin, Željku Kosiću-Vukšiću na pruženoj pomoći i praktičnim savjetima.

I na kraju, zahvaljujem se svim kolegicama i kolegama s fakulteta koji su mi pomogli kad mi je bilo najteže i s kojima sam proveo nezaboravne studentske dane. Ne želim izdvajati nikoga posebno, oni će se sami prepoznati.

Primjena OCR-a u vektorizaciji katastarskih planova

Zoran Stipanović

Sažetak: U ovom diplomskom radu trebalo je ispitati primjenu OCR-a u vektorizaciji planova katastarske općine Grablje. Ova metoda primjenjuje se samo za vektorizaciju tekstualnog sadržaja što uključuje različite nazive i brojeve. Osim toga bilo je potrebno i analizirati dobivene rezultate te ih prikazati u tabličnom i grafičkom obliku. Za ovaj posao upotrebljena je OCR aplikacija koja se nalazi u programu *Microstation Descartes*.

Abstract: The goal of this work was testing OCR on vectorization maps of Cadastral District Grablje. This method applies only for vectorization text which includes different names and numbers. Beside, it was necessary to analyze winning results and present them in the tables and charts. For this work it was used OCR application which is situated in the program *Microstation Descartes*.



Primjena OCR-a u vektorizaciji katastarskih planova

Zoran Stipanović

S A D R Ž A J

1. UVOD.....	6
2. KATASTAR	7
2.1. EVIDENCIJA NA PODRUČJU HRVATSKE	7
2.1.1. <i>Područje Austrijskog katastra.....</i>	8
2.1.2. <i>Područje Mađarskog katastra</i>	10
2.1.3. <i>Područje Jugoslavenskog katastra</i>	11
2.2. KATASTAR NEKRETNINA	12
2.3. ZAKONSKA OSNOVA KATASTRA NEKRETNINA	13
2.4. OPĆE ODREDBE ZAKONA	13
2.5. DJELATNOST KATASTRA NEKRETNINA	13
2.5.1. <i>Određivanje katastarskih prostornih jedinica</i>	14
2.5.2. <i>Katastarska izmjera</i>	14
2.5.3. <i>Izrada i održavanje katastarskih operata</i>	15
3. SKANIRANJE.....	16
3.1. PRIPREMA SKANIRANJA	16
3.2. VRSTE I PRINCIP RADA SKANERA	17
3.2.1. <i>Ručni skaneri</i>	17
3.2.2. <i>Stolni skaneri</i>	18
3.2.3. <i>Rotacijski skaneri</i>	18
3.3. PRIMJENA OCR-A KOD SKANIRANJA.....	19
3.4. METODE SKANIRANJA.....	20
3.5. NEKI NEDOSTACI SKANIRANJA	20
3.5.1. <i>Usporedba ručne i automatske digitalizacije.....</i>	21
4. GEOREFERENCIRANJE	23
4.1. TRANSFORMACIJE.....	23
4.2. TRANSFORMACIJE U KATASTRU.....	24
5. AUTOMATSKA VEKTORIZACIJA KATASTARSKIH PLANOVA I OCR	26
5.1. PODRUČJA PRIMJENE VEKTORIZACIJE I OCR-A	26
5.2. METODE VEKTORIZACIJE	27
5.2.1. <i>Ručne metode.....</i>	27
5.2.2. <i>Automatske metode</i>	27
5.3. METODE OCR-A KOD PREPOZNAVANJA TEKSTA	28
5.4. PRIMJENA OCR-A U OBRADI KATASTASKIH PLANOVA	32
5.4.1. <i>Pregled metoda.....</i>	32
5.4.2. <i>Aplikacije.....</i>	33



6. PRIMJENA OCR-A U VEKTORIZACIJI KATASTARSKIH PLANOVA K.O. GRABLJE	34
6.1. KATASTARSKA OPĆINA GRABLJE	34
6.2. PREUZIMANJE POSTOJEĆIH PODATAKA I OCJENA KVALITETE MATERIJALA	35
6.3. GEOREFERENCIRANJE KATASTARSKIH PLANOVA	36
6.4. OPTIČKO PREPOZNAVANJE ZNAKOVA (OCR).....	41
6.4.1. <i>Binarizacija katastarskih planova</i>	41
6.4.2. <i>Automatska vektorizacija katastarskih planova</i>	43
7. ANALIZA DOBIVENIH REZULTATA	47
7.1. SADRŽAJ PRILOŽENIH MEDIJA.....	52
8. ZAKLJUČAK	53

Literatura

Popis URL-ova

Životopis

Prilozi

1. Uvod

Zemljište i nekretnine dobivaju sve veći značaj, a da bi se njima moglo što učinkovitije upravljati neophodno je njihovo evidentiranje. Osim katastra nekretnina danas postoje i katastar zgrada, katastar šuma, katastar vinograda, katastar vodova, katastar voda itd. Razvojem informacijske tehnologije, koja omogućuje sve bolju međusobnu povezanost i komunikaciju među podacima, stvaraju se sve neophodne pretpostavke za neminovan prijelaz katastarskog sustava na digitalni, informacijski sustav. Prednost ovakvog informacijskog sustava prvenstveno se odnosi na brzinu obrade podataka, pohranjivanju većeg broja podataka na manje mjesta te lakše sortiranje i pretraživanje podataka.

Zato je prevođenje analognih katastarskih planova u digitalni oblik vrlo važan čimbenik u modernizaciji evidencije nekretnina. U tom smislu razvijene su brojne metode i programi koji imaju za cilj ubrzati ovaj postupak. Jedna od tih metoda je i optičko prepoznavanje znakova (engl. Optical Character Recognition – OCR). Razvoj ovih metoda i programa započeo je još 70-tih i 80-tih godina prošlog stoljeća. Danas je na tržištu moguće relativno jeftino nabaviti odgovarajuću OCR aplikaciju, međutim postavlja se pitanje njegove ekonomske isplativosti, s obzirom na pouzdanost i točnost te utrošak vremena.

Jedna od najstarijih metoda prevođenja analognih tekstualnih podataka u digitalni oblik je izravan unos teksta tipkovnicom i pohranjivanje u određenu tekstualnu datoteku. Krajem prošlog stoljeća, kada je digitalizacija planova, karata i GIS sustava postala aktualna, počelo se razmišljati o primjeni OCR-a u tom procesu. Razvitkom računalne tehnologije razvijale su se i različite OCR aplikacije. Sredinom 90-tih, njihova primjena se pokazala opravdanom iako je šira upotreba bila ograničena visokom cijenom softvera (Frangeš i dr. 1996). Do danas, izgrađene metode i algoritmi, zaživjeli su u aplikacijama tako da u novije vrijeme gotovo nema aplikacije koja ne omogućuje OCR, kao i izdvajanje različitih simbola s plana ili karte (Cetl 2001).

Zadatak ovog diplomskog rada je ispitati i analizirati praktičnu primjenu ove metode na konkretnom primjeru katastarske općine Grablje na otoku Hvaru. Za ove potrebe bit će korišten softver Microstation Descartes koji u sebi sadrži OCR aplikaciju koja prepoznaje brojke i slova na rasterskoj podlozi i pretvara ih u digitalni vektorski oblik.

2. Katastar

Katastar nekretnina jest evidencija o česticama zemljišta, zgradama i djelovima zgrada kao i drugim građevinama koje trajno leže na zemljištu ili ispod njegove površine, ako zakonom drukčije nije određeno (NN 128/99).

2.1. Evidencija na području Hrvatske

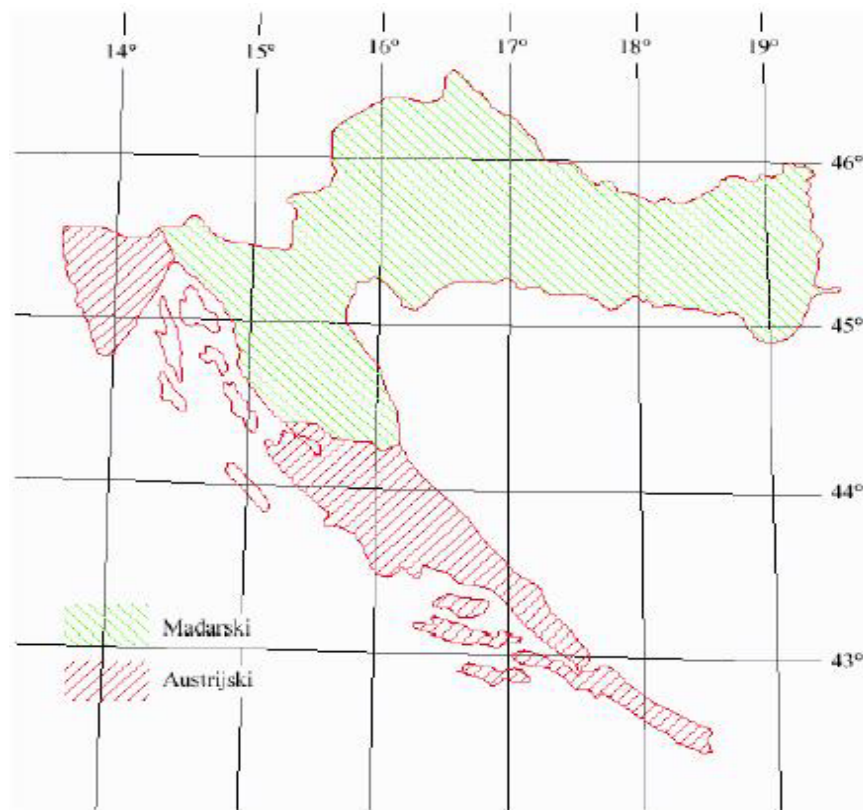
Budući da su se pojedini djelovi Hrvatske kroz povijest nalazili u različitim državama, proces uspostave katastra odvijao se u različitim uvjetima. Posljedica toga je da su danas katastarski podaci za različite djelove Hrvatske razlikuju. Za veći dio zemlje (oko 80%) podaci stare izmjere još uvijek su važeći, dok je za manji dio (oko 20%) provedena «nova izmjera», i to uglavnom za neka naseljena mjesta i veće gradove.

Prvi pokušaj uspostave katastra na području Hrvatske bio je tzv. Jozefinski katastar u 18. st. kad je Hrvatska bila u sastavu Austro-Ugarske monarhije. Glavna jedinica tog katastra bio je bečki hvat (1.896484m), a sam katastar uspostavljen je zbog jednakosti pri plaćanju poreza. Zbog nezadovoljstva stanovništva, 1790. godine stavljen je izvan uporabe.

Sljedeći katastar na našem području bio je tzv. Franciskanski katastar u 19. st. kada je patentom o pristupanju izrade katastra bilo određeno da se pristupi katastarskoj izmjeri te izradi katastarskog operata na području cijele Carevine. Izmjera je izvršena grafičkom metodom (geodetski stol), a tiskane su i prve upute za katastarsku izmjeru sa simbolima.

Uspostava Franciskanskog katastra na našem području obavljena je u periodu od 1818. do 1884. godine. Kako je katastarska izmjera na našem području obavljena u nekoliko vremenskih razdoblja, a katastarski planovi izrađeni u različitim referentnim sustavima, cijeli teritorij Hrvatske može se podijeliti na područja:

1. Austrougarskog katastra (Slika 1),
2. Jugoslavenskog katastra (Slika 6).

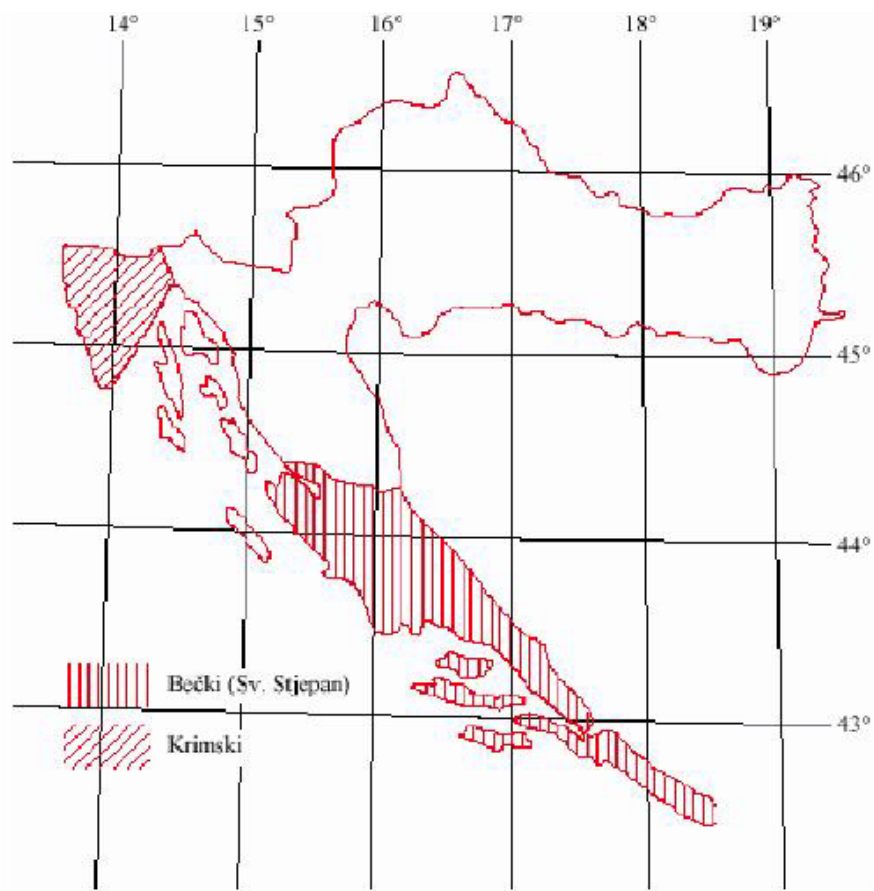


Slika 1. Područje Austrougarskog katastra

2.1.1. Područje Austrijskog katastra

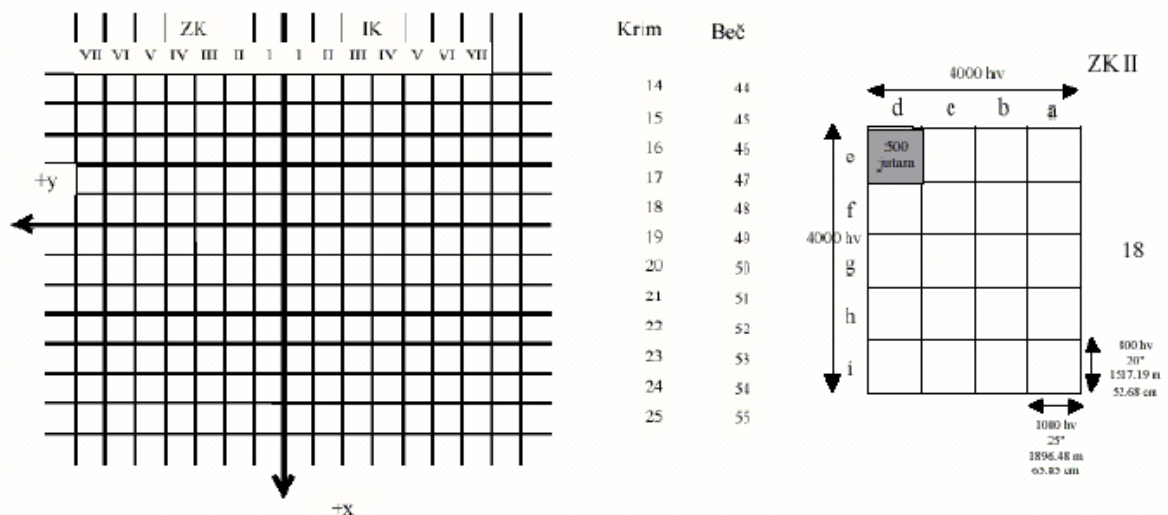
Već spomenutim Carskim patentom od 23. prosinca 1817. godine, određeno je da se krene s izmjerom zemljišta, ustanovljavanjem kultura, klasiranjem zemljišta i sastavljanjem katastarskih operata. Temelj izmjere bila je trigonometrijska mreža (1. – 4. reda) kao mreža trokuta od Beča preko Koruške, Štajerske, sjeverne Hrvatske do Dalmacije, koja je povezana preko Kranjske s Talijansko-francuskom mrežom. Točke 1., 2. i 3. reda određene su numerički, dok su točke 4. reda određene grafički. Veliki nedostatak ove mreže bila je vrlo loša stabilizacija točaka. Naše područje bilo je preslikano u dva koordinatna sustava (Slika 2)

1. Sustav s ishodištem u tornju crkve Sv. Stjepana u Beču ($\varphi = 48^{\circ} 12' 31''54$, $\lambda = 34^{\circ} 02' 27''32$ od Ferra). U ovaj sustav je bio preslikan dio Dalmacije.
2. Sustav s ishodištem u trigonometrijskoj točki Krim kod Ljubljane ($\varphi = 45^{\circ} 55' 43''75$, $\lambda = 32^{\circ} 08' 18''71$ od Ferra). U ovaj sustav je preslikano područje Istre.



Slika 2. Referentni sustavi Austrijskog katastra

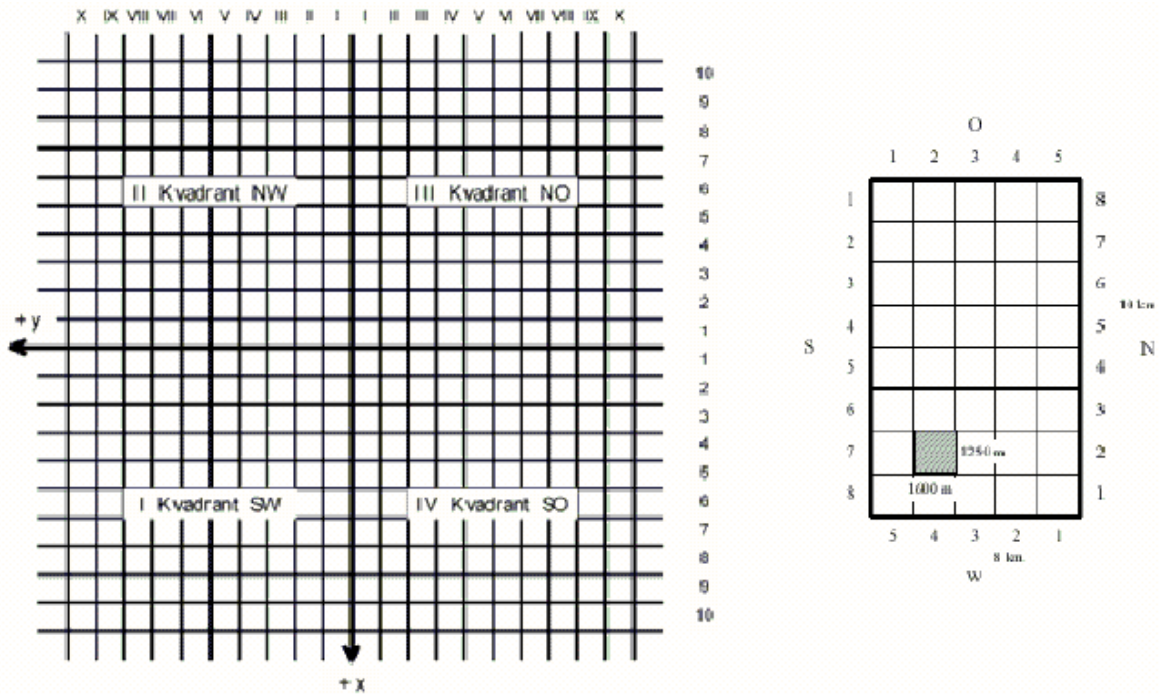
U svim sustavima područje preslikavanja podijeljeno je paralelama s osi x u kolone, a paralelama s osi y u zone (Slika 3).



Slika 3. Podjela i trigonometrijska sekcija u hvatnom sustavu

Jedinica za dužinu u toj izmjeri je 1 hvat, za površine četvorni hvat ($1\text{čhv} = 3.596652\text{m}^2$) te jutro ili ral, koje ima 1600 čhv, odnosno jedno jutro ima

5754.542m², dok je mjerilo planova bilo 1:2880 (1 palac = 40 hvati x 6 stopa x 12 palaca = 2880). 1873. godine, u projekcijskim sustavima se uveo metarski sustav, te je napravljena nova podjela (Slika 4).



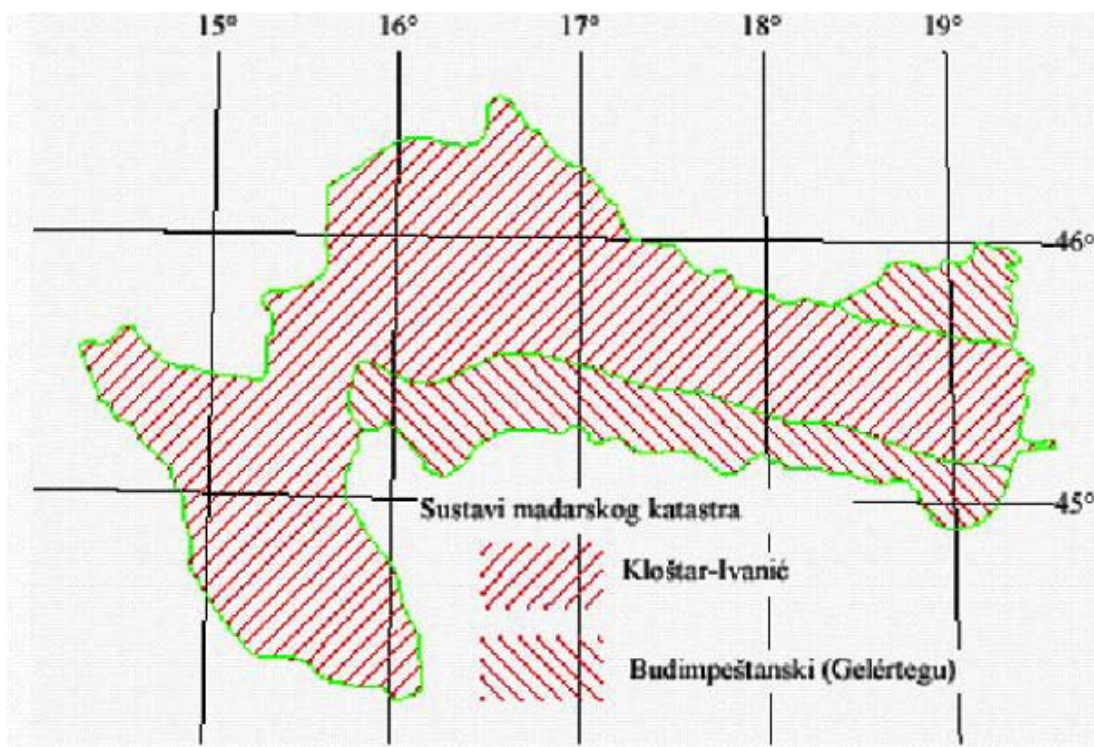
Slika 4. Podjela i trigonometrijska sekcija u metarskom sustavu

Mjerilo planova u ovom slučaju je bilo 1:2500, dok je daljnja podjela bila na listove mjerila 1:1250 i 1:625. Ovakvih planova na području Hrvatske ima malo.

2.1.2. Područje Mađarskog katastra

U sklopu ovog katastra izrađeni su planovi za dio teritorija koji je bio pod Ugarskom upravom, tj. za Hrvatsku bez Istre i Dalmacije. Taj teritorij bio je preslikan u dva sustava (Slika 5).

1. Kloštar-Ivanički koordinatni sustav s ishodištem u franjevačkoj crkvi u Kloštar-Ivaniću ($\varphi = 45^\circ 44' 21''25$, $\lambda = 34^\circ 05' 09''16$ od Ferra)
2. Budimpeštanski sustav s ishodištem u trigonometrijskoj točki Gelérthegu u Budimpešti ($\varphi = 47^\circ 29' 09''64$, $\lambda = 36^\circ 42' 53''57$ od Ferra)

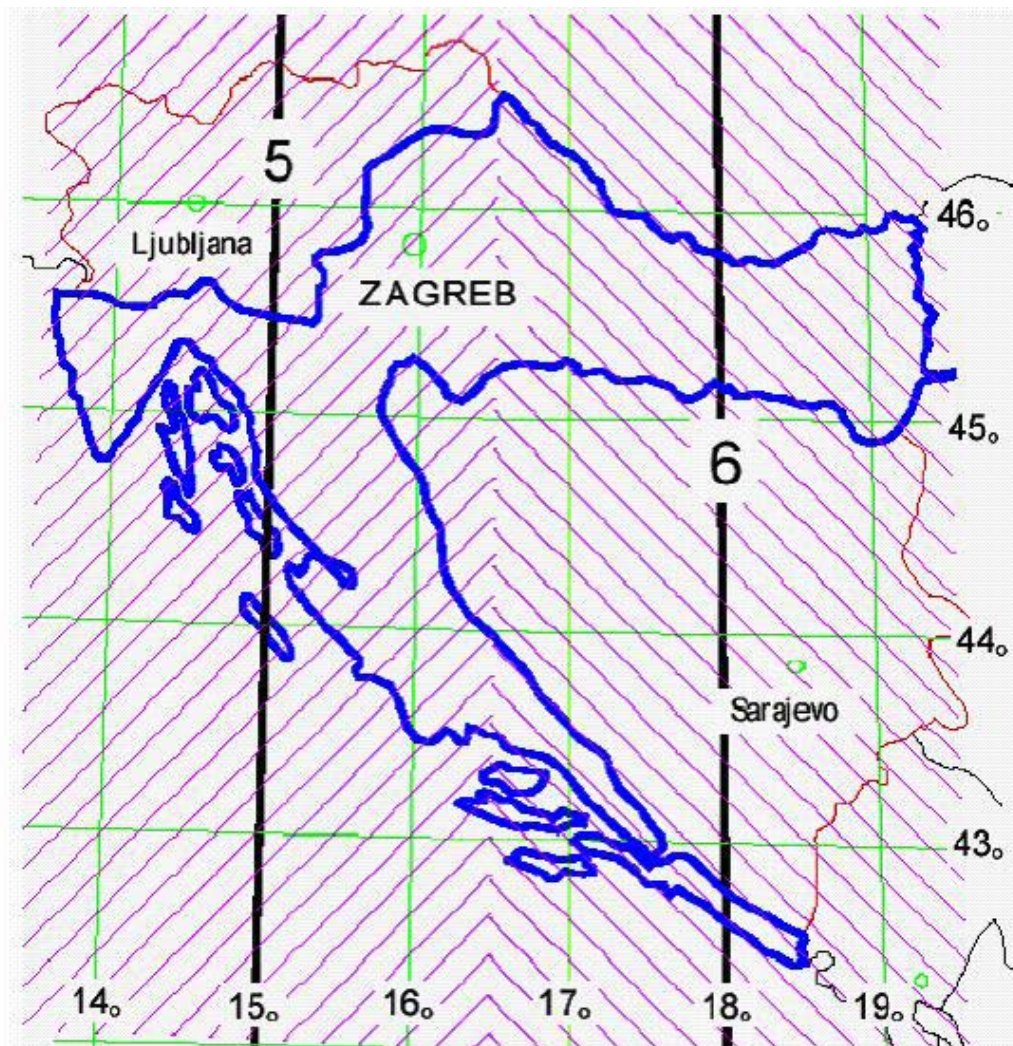


Slika 5. Referentni sustavi Mađarskog katastra

Mađarska izmjera izvedena je na isti način kao i austrijska tako da ima istu podjelu na listove. Cijela katastarska izmjera za ovo područje, obavljena je u razdoblju između 1847. i 1877. godine. Ugarska je mnogo kasnije prihvatila metarski sustav, pa se čak i danas u katastarskim operatima mogu naći podaci s četvornim hvatima i jutrima kao jedinicama.

2.1.3. Područje Jugoslavenskog katastra

1929. godine, za cijelo područje Kraljevine Jugoslavije uvedena je jedinstvena Gauss-Krügerova projekcija. Cijeli teritorij države preslikan je na tri poprečno postavljena cilindra koji dodiruju plovu elipsoida u petnaestom, osamnaestom i dvadesetprvom meridijanu. Područje Hrvatske preslikava se u dva koordinatna sustava, koji se označavaju kao peti i šesti (Slika 6).



Slika 6. Referentni sustavi jugoslavenskog katastra u Hrvatskoj

Paralelama s osi x na udaljenosti od 22.5 km dijeli se područje svakog sustava na kolone, dok se paralelama s osi y na udaljenosti od 15 km područje dijeli na redove. Da ne bi bilo negativnih y koordinata, osi x dodaje se vrijednost od 500 000 m. Pred koordinatu y stavlja se na mjesto milijuna broj sustava u kojem se pojedina točka nalazi. Pa tako osi x imaju ordinatu $y = 5\,500\,000$ m u petom i $y = 6\,500\,000$ m u šestom sustavu.

Ovakvom podjelom na zone i kolone dobiveni su temeljni triangulacijski listovi, dimenzija 22.5 x 15.0 km, koji se dijele na detaljne listove u ovisnosti od mjerila u kojem je izvršeno kartiranje.

2.2. **Katastar nekretnina**

Katastar nekretnina jest evidencija o česticama zemljišta, zgradama i djelovima zgrada kao i drugim građevinama koje trajno leže na zemljištu ili ispod njegove površine (NN 128/99).

2.3. Zakonska osnova katastra nekretnina

Djelatnost, ustrojstvo, način financiranja kao i ostali elementi vezani uz djelovanje katastra nekretnina propisani su Zakonom o državnoj izmjeri i katastru nekretnina. Zakon je donesen na sjednici Zastupničkog doma Hrvatskog državnog sabora 5. studenoga 1999.

Ovim zakonom se uređuje državna izmjera, katastar nekretnina, registar prostornih jedinica, katastar vodova, geodetski radovi za posebne potrebe, nadležnost i obavljanje poslova državne izmjere i katastra nekretnina te ustrojstvo Državne geodetske uprave kao i javne ustanove za obavljanje poslova državne izmjere i katastra nekretnina – «Hrvatski geodetski institut».

2.4. Opće odredbe zakona

Državna izmjera definirana je kao sustav mjernih i opisnih podataka trodimenzionalnog prikaza područja Republike Hrvatske, utemeljen na prikupljanju, obradi i prikazivanju topografskih i zemljišnih podataka geodetskim metodama (fizikalne, matematičke, astronomske, satelitske geodezije, daljinskog pronicanja i dr.).

Podaci državne izmjere i katastra nekretnina su javni i daju se na uvid zainteresiranim strankama, ukoliko posebnim propisima nije drukčije određeno, a to mogu biti razne isprave, ispisi, zračne snimke, dokumenti, kartografski (analogni i digitalni) i drugi materijali koji se smiju koristiti samo u svrhu za koju su izdani. Osim toga ovi su podaci temelj za zemljišne knjige koje vode zemljišnoknjižni sudovi te za vođenje očevidnika u svezi s prostorom, koji se također vode prema ovom zakonu i drugim propisima.

Poslovi državne izmjere i katastra nekretnina obavljaju se na temelju godišnjih i višegodišnjih programa. Godišnje programe donosi Vlada Republike Hrvatske, a višegodišnje programe Hrvatski državni sabor. Ovi se programi objavljuju u Narodnim novinama, a financijska sredstva za izvršenje programa te za izradbu i održavanje katastra nekretnina osiguravaju se u državnom proračunu kao i iz drugih izvora. Pod pojmom «drugi izvori» smatra se da u financiranju mogu sudjelovati i jedinice lokalne samouprave; županije, gradovi, općine te druge zainteresirane fizičke i pravne osobe.

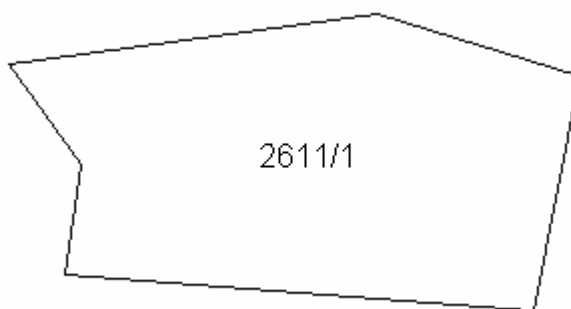
2.5. Djelatnost katastra nekretnina

Poslovi katastra nekretnina obuhvaćaju:

1. Određivanje katastarskih prostornih jedinica
2. Katastarsku izmjeru
3. Izradu i održavanje katastarskih operata

2.5.1. Određivanje katastarskih prostornih jedinica

Osnovna prostorna jedinica katastra nekretnina je katastarska čestica. Prema definiciji, katastarska čestica je dio područja katastarske općine, odnosno katastarskog područja na moru, omeđena granicama (međama i drugim granicama) koje određuju pravni odnosi na zemljištu, te granicama načina uporabe zemljišta (NN 128/99). Ona se u katastarskom operatu prikazuje svojim granicama i označuje posebnim brojem u okviru katastarske općine odnosno katastarskog područja na moru (Slika 7).



Slika 7. Katastarska čestica

Veća prostorna jedinica od katastarske čestice je katastarska općina koja se sastoji od skupa katastarskih čestica. Katastarska općina u pravilu obuhvaća područje jednog naseljenog mjesta s pripadajućim zemljištem. Područje i ime katastarske općine odnosno katastarskog područja na moru određuje Državna geodetska uprava u sporazumu s ministrom nadležnim za poslove pravosuđa.

2.5.2. Katastarska izmjera

Katastarskom izmjerom prikupljaju se i obrađuju podaci o:

1. položaju, obliku, površini, načinu upotrebe, te nositeljima prava na česticama zemljišta;
2. položaju, obliku, načinu upotrebe, te nositeljima prava na zgradama i drugim građevinama;
3. položaju u zgradi, površini, te nositeljima prava na dijelovima zgrada (stanova, te poslovnih i drugih prostora) i drugih građevina, te
4. područjima posebnih pravnih odnosa na zemljištu.

U postupku katastarske izmjere potrebno je prikupiti podatke o nositeljima prava na nekretninama na temelju raspoloživih dokumenata (izvadak iz zemljišne knjige ili katastra zemljišta) kao i izjave zainteresiranih stranaka.

Katastarske čestice prema načinu njihove uporabe svrstavaju se u sljedeće vrste korištenja:

1. poljoprivredno zemljište;
2. šumsko zemljište;
3. unutrašnje vode;
4. površine mora;
5. ostale površine zemljišta.

Odluku o katastarskoj izmjeri katastarske općine donosi Državna geodetska uprava u sporazumu s ministarstvom nadležnim za poslove pravosuđa, u skladu s godišnjim ili višegodišnjim programom.

2.5.3. Izrada i održavanje katastarskih operata

Katastarski operat izrađuje se za svaku katastarsku općinu odnosno katastarsko područje na moru na temelju podataka prikupljenih i obrađenih u katastarskoj izmjeri, a sastoji se od katastarskih planova, preglednih katastarskih karata i drugih grafičkih dokumenata, te pisanih djelova. Podaci o područjima posebnih pravnih odnosa na zemljištu unose se u katastarski operat iz službenih dokumenata kojima su ova područja određena. U okviru katastarskog operata određeni podaci o zgradama, njihovim djelovima i drugim građevinama, zatim podaci o načinu upotrebe zemljišta te podaci o područjima posebnih odnosa na zemljištu, vode se u odvojenim djelovima katastarskog operata. Oblik i sadržaj katastarskih operata propisuje ravnatelj Državne geodetske uprave.

Izvodi, ispisi, preslike i uvjerenja koja se izdaju na temelju katastarskog operata moraju biti ovjereni pečatom i potpisom službene osobe, te predstavljaju javne isprave. Ove isprave nisu dokaz o pravu vlasništva.

Unošenje podataka o nekretninama kojima upravlja Ministarstvo obrane Republike Hrvatske regulirano je posebnim sporazumom nadležnog ministra za poslove obrane i ravnatelja Državne geodetske uprave.

Promjene broja, položaja, oblika, načina uporabe i površina katastarskih čestica (parcelacija), provode se na temelju parcelacijskog elaborata. Za promjene položaja, oblika i načina upotrebe zgrada i drugih građevina potreban je geodetski elaborat. Elaborat mora imati naznačenu osobu koja ga je izradila, datum izrade, ime i potpis te pečat ovlaštenog geodetskog stručnjaka kao i popis suradnika koji su izradili elaborat. Da bi se na temelju parcelacijskog odnosno geodetskog elaborata mogle provesti promjene, on mora biti potvrđen od nadležnog područnog ureda za katastar.

3. Skaniranje

Od samih početaka upotrebe računala pojavila se potreba da se sadržaj dostupan na papiru na neki način prenese u računalo. Za ovu funkciju proizveden je uređaj koji nazivamo skaner. Kao i većina računalnih komponenti u početku je imao iznimno visoku cijenu te je njegova primjena bila ograničena isključivo na profesionalne upotrebe. No kako je rastao broj korisnika računala tako se povećavalo i tržište, što je dovelo do pojave jeftinijih modela, a time i veće upotrebe u kućanstvima.

Rezultat toga je da se danas kvalitetan skaner može dobiti za nekoliko stotina kuna, a za približno tisuću kuna moguće je dobiti uređaje koje mogu koristiti i profesionalci. Ovakav pad cijena povećao je broj korisnika koji su na računalo priključili skaner, a svakodnevno se javljaju i novi korisnici zainteresirani za nabavku ovakvog uređaja.

Osnovna namjena skanera je unos sadržaja dostupnog na papiru u računalo, odnosno pretvaranje analognog sadržaja na papiru u digitalni zapis. Ovakav zahvat skaner može obaviti na samo jedan jedini način. Jednostavno rečeno, dostupan sadržaj u skaneru, uvijek se u računalo prenosi u obliku točkaste slike. Točkasta slika dobivena iz skanera određena je njenom veličinom i razlučivošću (rezolucijom). Veličina podržanog formata skanera obično je vidljiva na samom uređaju, a za većinu kućnih korisnika dovoljan je format A4.

Još jedan podatak koji određuje značajke skanera je podržana paleta boja. Suvremeni skaneri mogu obrađivati slike kao crno-bijele zapise ili zapise u boji. Kod obrade crno-bijelih sadržaja većina uređaja podržava isti format, no pri radu s bojom primjećuju se određene razlike. Kako se boja sastoji od tri komponente (crveno, zeleno, plavo) podatak o boji naveden je kroz informaciju o tome koliko nijansi boje skaner razlikuje u pojedinoj komponenti.

Tako najjeftiniji modeli razlikuju 256 nijansi u svakoj boji što je najčešće navedeno kao podatak da je riječ o 8-bitnoj informaciji o pojedinoj komponenti. Bolji skaneri za upotrebu jedne boje koriste 10, 12 pa i više bita. Tako, na primjer, skaner koji po pojedinoj komponenti koristi 12 bita može razlikovati 4096 nijansi svake pojedine boje, odnosno u stanju je dati sliku s izuzetno preciznim prikazom boje.

Skaneri se razlikuju i prema načinu priključivanja na računalo. Još uvijek se pojavljuju skaneri koji se na računalo priključuju preko paralelnog priključka (za pisač) no ovakvo rješenje valja izbjegavati. Najveću brzinu rada daju skaneri priključeni na SCSI sučelje, no oni zahtijevaju odgovarajući priključak na računalo i imaju nešto veću cijenu. Zapravo, za kućne su korisnike vjerojatno najbolji izbor skaneri koji se priključuju preko USB priključka jer daju zadovoljavajuću brzinu, a izuzetno ih je jednostavno priključiti i instalirati.

3.1. Priprema skaniranja

Prije skaniranja treba napraviti pregled slike. To je skaniranje, ali u nižoj rezoluciji, radi uštede na vremenu. Za stolne skanere to obično znači 20 dpi. Slika dobivena

pregledom važan je izvor informacija. Na njoj se točno vidi smještaj predloška. Najvažniji rezultat pregleda jest histogram, koji omogućava izbor parametara za konačno skaniranje. Krivulju treba modificirati tako da u području sjena piksele posvijetlimo tj. da svakoj ulaznoj vrijednosti pripada veća izlazna vrijednost (Frančula 1999).

3.2. Vrste i princip rada skanera

Razlikujemo tri osnovne vrste skanera. To su ručni, stolni i rotacijski. Za potrebe skaniranja listova k.o. Grablje korišten je rotacijski skaner CHAMELEON 25" poznatog proizvođača Contex koji omogućava širinu skaniranja od 25 inča. S obzirom na dimenzije originalnih listova od max. 75x50 cm, ovaj skaner je bio i više nego dovoljan za ove potrebe.

Sam postupak skaniranja izvodi se tako da se najprije uključi skaner, zatim računalo te se pokreće program za skaniranje. Ime datoteke, širina i rezolucija skaniranja moraju biti postavljeni prije skaniranja ili probe. Ostale pojedinosti mogu se mijenjati u toku skaniranja.

3.2.1. Ručni skaneri

Postupak obrade slike kod ručnog skanera provodi se tako da se glava za skaniranje pomiče iznad predloška i zapravo skanira vrlo tanke retke slike, odnosno po jedan redak točkica na slici. Opisani postupak zahtijeva nešto više vremena jer je potrebno mehanički sklop pomicati ispred uzorka i pri tome obrađivati vrlo veliku količinu podataka. Na kraju je dobivene podatke potrebno prenijeti u računalo, a kako je riječ o većoj količini podataka i nešto sporijoj vezi, ovaj postupak odnosi dosta vremena. Zbog svega opisanog, skaniranje ovom vrstom skanera nije naročito brz postupak te je za unos cijele stranice sadržaja u višoj razlučivosti potrebno oko pola minute kod najskupljih i najbržih kućnih uređaja pa do preko 90 sekundi za nešto jeftinije modele.

(Slika 8) prikazuje model ručnog CCD skanera Meteor MT280. Njegova širina je 80 mm, a pogodan je za uredsku upotrebu malog obima. Uređaj je testiran na pad sa 120 cm visine, a kabel dužine 1,9 m omogućava ugodan rad. Priklijučuje se na port za tipkovnicu.



Slika 8. Ručni skaner

3.2.2. Stolni skaneri

Ova vrsta skanera je najraširenija te je prihvaćena od najvećeg broja profesionalnih korisnika. Princip rada je sljedeći:

Stolni skaneri s gornje strane imaju staklenu ploču na koju treba staviti sliku koju želimo skanirati. Slika se osvjetljava fluorescentnom lampom, a reflektirano svjetlo se prenosi optičkim sustavom na odgovarajući senzor. Obično je to CCD (Charge Coupled Device) poluvodički senzor, kakav se primjenjuje kod video kamera. Zadatak senzora je da pretvori svjetlo u električni signal. Skaner ne pretvara odjednom cijelu sliku u digitalni zapis, već samo jedan red slike. Optički sustav je izveden tako da na senzor odgovarajućeg oblika pada slika jednog reda skaniranog predloška. Nakon što se taj red pretvori u digitalnu informaciju, optički sustav se malo pomakne, tako da na senzor pada slika sljedećeg reda. Čitav proces se ponavlja dok se ne dođe do kraja slike (URL 1).

Postoje crno-bijeli i skaneri u boji. Skaneri u boji imaju CCD senzor koji ima CCD elemente u tri niza, za svaku boju (crvenu, zelenu i plavu) po jedan.

Bitne karakteristike skanera su:

1. Veličina slike koju mogu skanirati, najčešće su to formati A3 i A4.
2. Razlučivost (rezolucija) skanera, koja se obično kreće između 300 i 1200 dpi (točaka po inču).
3. Kvaliteta analogno-digitalne pretvorbe signala iz CCD senzora koju opisuje podatak o broju bita po boji. Tako razlikujemo 24-bitne, 30-bitne, 36-bitne ili čak 48-bitne skanere. Primjerice, ako se radi o 24-bitnom skaneru, to znači da je za svaki piksel slike utrošeno po 8 bita za zapis informacije o svakoj od tri boje (URL 1).

Na sljedećoj slici (Slika 9) prikazan je stolni skaner u boji CanoScan D646ex.



Slika 9. Stolni skaner

3.2.3. Rotacijski skaneri

Ova vrsta skanera se koristi nešto rjeđe, jer zbog visoke cijene često nije dostupna kućnim korisnicima. Ovi skaneri prvenstveno su namjenjeni profesionalnoj upotrebi, a omogućuju kvalitetno skaniranje najzahtjevnijih predložaka. To se prije svega odnosi na predloške velikih formata koje možemo skanirati s razlučivošću od nekoliko tisuća točaka po inču. Zato je bolja kvaliteta reprodukcije, a osim toga i postupak skaniranja je brži. Valja još napomenuti da su rotacijski skaneri mnogo

većih dimenzija od ručnih i stolnih. (Slika 10) prikazuje rotacijski skaner pripremljen za rad.



Slika 10. Rotacijski skaner

3.3. Primjena OCR-a kod skaniranja

Prilikom skaniranja, skaner će u računalo informaciju s papira prenijeti u obliku točkaste slike. Ovakvo rješenje pogodno je za unos fotografija, složenijih ilustracija i slično. No ovaj način zapisivanja informacije sasvim je nepogodan za rad s tekstom. Naime, tekst unesen na ovaj način zapravo predstavlja sliku te ga korisnik ne može ni na koji način mijenjati, koristiti kao dijelove u drugom tekstu i slično. Zbog svega opisanog odavno se pojavila ideja da računalo pokuša "pročitati", odnosno "prepoznati" tekst sa slike.

Zamisao je da računalni program izdvoji manje cjeline za koje pretpostavlja da predstavljaju jedno slovo, a zatim, uspoređujući dobiveni oblik s informacijama o slovu, "prepozna" o kojem je slovu riječ. Tako se prepoznavanjem pojedinih slova postupno kreira zapis koji sadrži tekst sadržan na skaniranoj ilustraciji. Opisani postupak naziva se optičko prepoznavanje slova ili engleski Optical Character Recognition, od čega dolazi kratica OCR.

Međutim, ovaj je postupak prilično nepouzdan pa se može uspješno primjenjivati jedino u slučajevima kada je skanirani tekst obrađen u dovoljno visokoj razlučivosti i s dovoljnim kontrastom. Na kvalitetu prepoznavanja teksta u velikoj mjeri utječe i vrsta slova kojom je tekst napisan. Naime, ukoliko se obrađuje tekst pisan nekom vrstom (fontom) slova koja je već dostupna u računalu, prepoznavanje će biti znatno bolje.

No ni u idealnim uvjetima računalo neće u potpunosti prepoznati sva slova pa se koristi još jedno pomoćno sredstvo. Tako se program za prepoznavanje teksta povezuje s programom za ispravljanje pogrešaka pri pisanju (Spelling Checker). Kad računalo prepoznavanjem teksta dobije riječ koja se ne nalazi u rječniku programa za provjeru ispravnosti pisanja, ono pretpostavi da je riječ o nepravilnom prepoznavanju pa od korisnika zatraži dodatnu intervenciju. Spojem svih opisanih postupaka moguće je relativno uspješno unijeti tekst otisnut na papiru pa se uz utrošak nešto više vremena i veći tekstovi mogu uspješno pretvoriti u tekstualne zapise na računalu.

Za prepoznavanje teksta svakako valja koristiti program koji podržava sve znakove korištene u tekstu. Na žalost, uz većinu skanera korisnik dobiva program

za ovu namjenu bez podrške za hrvatske znakove. Stoga bi bilo korisno osim skanera pribaviti i odgovarajući program jer se tek tada mogu dobiti valjani rezultati obrade.

3.4. Metode skaniranja

S obzirom na mogućnost očitavanja boja predložka skaneri se dijele na one koji mogu čitati jednoboju sliku u dva intenziteta (crno i bijelo), jednoboju sliku u više intenziteta (nijansi sive boje) i sliku u boji. Tako razlikujemo:

Jednoboju skaniranje (engl. line-art) je način rada u kojem skaner za svaku analiziranu točku (piksel) odredi je li crna ili bijela, a dobivena slikovna matrica sastoji se samo od crnih i bijelih točaka. Pohranjivanje slike zahtjeva jedan bit za svaku točku.

Skaniranjem u sivoj skali (engl. grayscale) svakoj se točki dodjeljuje cjelobrojna vrijednost u rasponu od 0 do 255 koja odgovara pojedinoj razini zatamnjenja (sivoj nijansi). Spremanje slike zahtjeva 8 bita (1 bajt) po svakoj točki.

Skaniranje u punoj boji (engl. full color ili true color) znači da se točki dodjeljuje numerička vrijednost od 0 do 255 za svaku od osnovnih svjetlosnih (RGB) komponenti čime se ostvaruje ukupna paleta od 16 777 216 boja. Za spremanje slike potrebna su $3 \times 8 = 24$ bita odnosno 3 bajta po svakoj točki.

Rastersko skaniranje (engl. halftone) služi kao skraćeno postupka rasteriranja slika u sivoj skali. Slika koja ovako nastaje je jednoboju (1bit/piksel), ali se sive nijanse uzorka ne interpretiraju kao crne i bijele, već se tim područjima dodjeljuje rasterski uzorak, čija kvaliteta ovisi o pogonskom programu skanera.

Skaniranje u 256 boja je postupak kojim se šteti memorijski prostor bez gubitaka podataka o boji. On pronalazi svoju primjenu kod relativno jednoličnih uzoraka npr. kod grafikona, omota raznih proizvoda, prozora s ekrana računala i sl. Svaka točka opisana je s 8 bita, a u opis slike ulazi i tzv. paleta tablica upotrijebljenih 256 boja. Za razliku od skaniranja u punoj boji kod koje se svaka boja opisuje RGB komponentama (3×8 bita), kod skaniranja u 256 boja opisuje se rednim brojem položaja (1-256) boje u paleti (8 bita), (Marinović 2004).

3.5. Neki nedostaci skaniranja

Prilikom skaniranja raznih fotografija ili ilustracija otisnutih u časopisima, možemo vrlo brzo uočiti neke osnovne nedostatke skaniranja. Kao prvo, materijal otisnut u novinama, na lecima, prospektima i slično, otisnut je posebnom tehnikom kod koje se koriste sitne točkice čiste boje. Pri tome se najčešće koriste četiri osnovne boje (plavo-zelena, ljubičasta, žuta i crna). Različite nijanse pojedine boje dobivene su promjenom veličine ili gustoće točkica osnovnih tiskarskih boja. Odlučimo li ovakav uzorak unijeti u računalo, pojavit će se određeni problemi. Kao prvo, ovakav zapis boje skaner mora pretvoriti u prikaz sastavljen od tri osnovne komponente boje (crvena, zelena, plava). Osim toga, pri tom postupku, promjene u veličini i gustoći točkica trebalo bi prepoznati kao nijanse, a ne unositi svaku pojedinu točkicu.

U pravilu je ova mogućnost podržana u samom skaneru i najčešće se naziva descreening. Opisano rješenje zasniva se na tome da skaner namjerno malo zamućuje sliku, odnosno promatra veće površine. Takvim širim zahvatom obuhvaća se više sitnih točkica osnovne boje pa se dobivaju informacije o nijansama. Kod većine uređaja moguće je ovo "zamućenje" ručno podešavati (kroz upravljački program) pa korisnik može odabrati najnižu vrijednost koja daje zadovoljavajući rezultat. Neželjeni učinak gubitka oštine može se naknadno dodatno umanjiti izoštravanjem slike kroz programe za obradu točkaste slike.

Drugi problem s kojim će se susresti svaki korisnik posljedica je nepravilnosti na uzorku. Naime, za razliku od ljudskog mozga skaner nije u stanju razlikovati korisne informacije na papiru od različitih nečistoća, tamnijih površina u samom papiru, zaostalih komadića drveta i slično. Zapravo kvalitetan skaner koji radi u visokoj oštini sve će ove neželjene informacije uključiti u sliku, odnosno prenijeti u računalo.

Zbog toga često pri skaniranju dobivamo rezultat koji sadrži puno sitnih, neželjenih točkica, nečistoća i drugih problema. Promjenom parametara kontrasta i oštine i ponavljanjem postupka ove probleme možemo donekle umanjiti, no ako je potreban sasvim čist rezultat, neće se uvijek moći izbjeći potreba za ručnom doradom skaniranog uzorka, odnosno ručnim "čišćenjem" slike. Odlučimo li primijeniti ovaj postupak, najbolje je da uzorak skaniramo u nešto višoj razlučivosti, uklonimo nečistoće, a zatim smanjimo razlučivost slike na željenu vrijednost jer ćemo i pri smanjivanju razlučivosti donekle popraviti uzorak i dodatno ukloniti najsitnije nedostatke.

3.5.1. Usporedba ručne i automatske digitalizacije

Prednosti ručne digitalizacije:

1. Uspostava topoloških odnosa on-line
2. Gotovo nikakva dorada.

Nedostaci ručne digitalizacije:

1. Velik utrošak vremena i mnogo stručnjaka
2. Traži se školovano osoblje, jer se pri digitalizaciji traži interpretacija i strukturiranje sadržaja.

Prednosti automatske digitalizacije:

1. Ukupno manje osoblja i vremena
2. Moguća visoka geometrijska točnost.

Nedostaci automatske digitalizacije:

1. Skupi skaneri

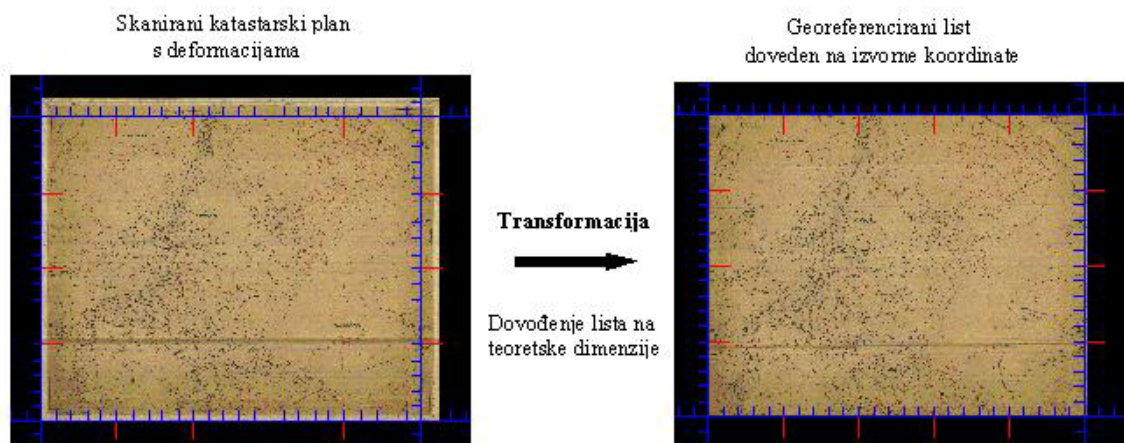


2. Nesavršen softver za vektorizaciju
3. Znatna interaktivna dorada (Frančula 1999).

4. Georeferenciranje

Georeferenciranje je postupak kojim se skanirani list katastarskog plana dovodi na teoretske dimenzije. List koji se georeferencira smješta se u koordinatnu mrežu tako da se uglovima lista i ostalim vidljivim točkama koordinatne mreže (decimetarska ili hvatna mreža) pridružuju njihove teoretske koordinate (Slika 11). Na ovaj način nastoji se smanjiti deformacija listova usljed usuha, rastega ili skaniranja, na najmanju moguću mjeru. Osim toga georeferenciranje je jedan od osnovnih preduvjeta za kvalitetnu vektorizaciju analognih katastarskih planova.

Prilikom georeferenciranja skanirani list katastarskog plana u većoj ili manjoj mjeri mijenja svoj geometrijski oblik odnosno kažemo da se transformira. Pri tome se koriste različite vrste transformacija, a najčešće Helmertova i afina.



Slika 11. Georeferenciranje

4.1. Transformacije

Transformacije kao metode za promjenu geometrijskih podataka koriste se u transformiranju vektorskih i rasterskih podataka. Kod transformacije vektorskih podataka dolazi do promjene položaja karakterističnih točaka vektora dok se kod transformacije rasterskih podataka preračunava vrijednost svakog piksela u slikovnoj matrici. Kao transformacija skaniranih listova katastarskih planova obično se koristi ova druga.

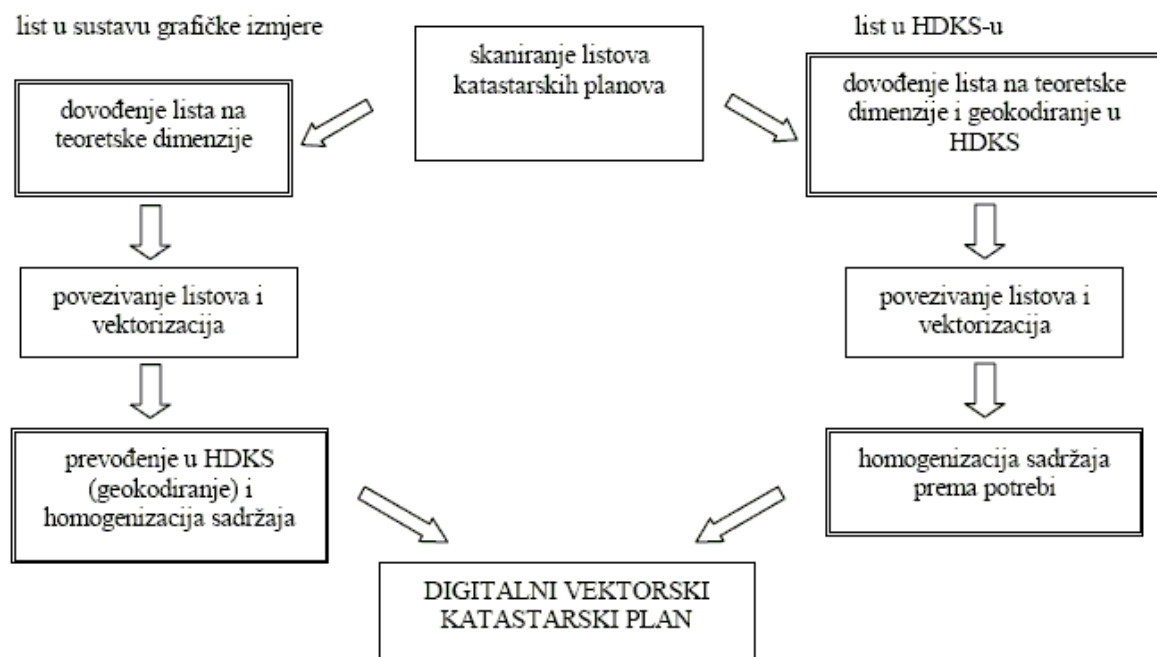
Da bismo mogli transformirati podatke iz jednog sustava u drugi moramo poznavati transformacijske parametre te provesti transformaciju preko odgovarajuće funkcije. Međutim, u praksi su parametri vrlo rijetko poznati. Tada ih moramo izračunati na temelju poznatih koordinata identičnih točaka u oba sustava, a za to su potrebne identične točke. Minimalni broj identičnih točaka ovisi o tipu transformacije. Kod georeferenciranja listova program (npr. MS Descartes) koristi upravo ovu metodu odnosno računa transformacijske parametre na temelju zadanih identičnih točaka. Četiri osnovna tipa transformacija su translacija, rotacija, promjena mjerila i smicanje.

4.2. Transformacije u katastru

Geometrijski podaci u katastru sadržani su na katastarskim planovima, skicama i drugim grafičkim prikazima. Obzirom na sadržaj i vrstu poslova koji se obavljaju u katastru, transformacije nalaze svoju primjenu u svakodnevnim poslovima održavanja katastarskog operata kao i u postupcima vektorizacije katastarskih planova i homogenizacije sadržaja. Preduvjet korištenja bilo kakvih računalom podržanih transformacija u CAD i GIS programima je prevođenje grafičkog izvornika u digitalni oblik.

U postupku prevođenja katastarskog plana u digitalni vektorski oblik i u homogenizaciji njegova sadržaja provode se različite transformacije. Ulazni podaci u postupku vektorizacije su skanirani listovi katastarskih planova u digitalnom rasterskom obliku.

Ovisno o tome da li je katastarski plan u Hrvatskom državnom koordinatnom sustavu ili u nekom starom sustavu grafičke izmjere, razlikujemo dva pristupa u korištenju transformacija koji su shematski prikazani na sljedećoj slici (Slika 12).



Slika 12. Transformacije u postupku izrade digitalnog katastarskog plana

Ako je katastarski plan izrađen u nekom od starih sustava grafičke izmjere tada nam, u pravilu, njegov položaj u HDKS-u nije poznat. Za takve planove moguće je obaviti ispravljanje deformacija skaniranog lista i dovesti ga na njegove teoretske dimenzije. Teoretske dimenzije listova u starim sustavima grafičke izmjere su poznate, a proizlaze iz podjele temeljnih triangulacijskih listova ovisno o mjerilu plana. Za listove plana 1:2880, teoretske dimenzije lista su 1000x800 hvati ($1\text{hv}=1.896484\text{m}$). Ispravljanje deformacija lista i dovođenje na teoretske dimenzije obavlja se nekom od transformacija rasterskih podataka. Zadatak se može obaviti korištenjem Helmertove ili affine transformacije (Tablica 1). Skanirani

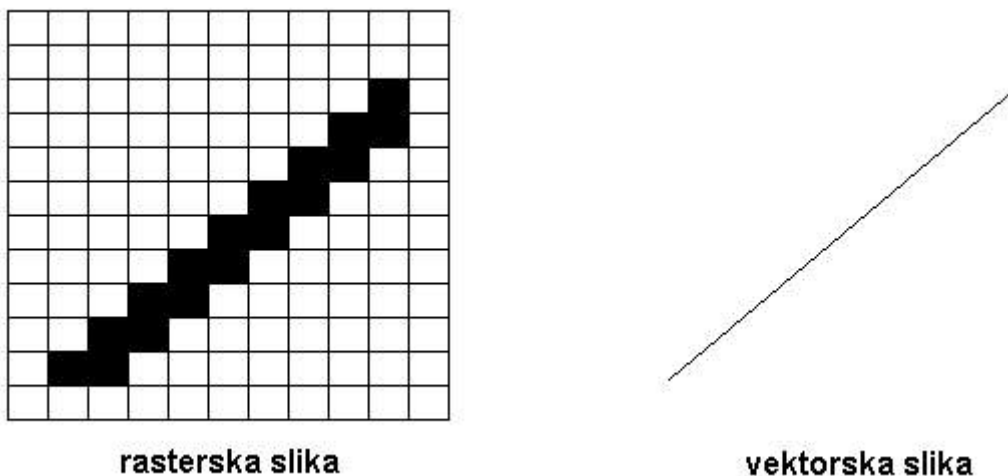
list prevodi se na teoretske dimenzije korištenjem rubnih točaka okvira lista. Ako postoji originalno ucrtana hvatna mreža tada se u transformaciji koriste i te točke. Palčana mreža se ne preporučuje za korištenje obzirom da je najčešće naknadno ucrtavana na listovima pri provođenju promjena. Obzirom na fizičke osobine listova (nejednolike deformacije u smjeru koordinatnih osi) preporučuje se korištenje afine transformacije čije osobine omogućuju postizanje najboljih rezultata (Ceti 2002).

Tablica 1. Transformacijski parametri Helmertove i afine transformacije

Vrsta transformacije	Transformacijski parametri		
Helmertova	2 translacije	1 rotacija	1 promjena mjerila
afina	2 translacije	1 rotacija	2 promjene mjerila

5. Automatska vektorizacija katastarskih planova i OCR

Vektorizacija je postupak kojim se od rasterske slike dobiva vektorska. Rasterska slika je slika prikazana pomoću vrijednosti amplituda svjetlosti (ili boje) točaka, a vektorska je predstavljena matematičkim opisom (najčešće Bezierovim krivuljama). Iako za rasterske slike postoje vrlo kvalitetne metode kompresije, one općenito zauzimaju puno više prostora od vektorskih. Zbog svoje prirode, iz njih se puno teže dobivaju informacije i teže ih se obrađuje. Razlike između vektorske i rasterske slike najbolje se vidi na primjeru (Slika 13). OCR je posebna metoda vektorizacije gdje se iz slike izdvaja tekst. Nakon što se izdvoji, tekst se može obrađivati pomoću bilo kojeg programa za obradu teksta.



Slika 13. Razlika između rasterske i vektorske slike

5.1. Područja primjene vektorizacije i OCR-a

Područja ljudske djelatnosti u kojima se vektorizacija najviše primjenjuje jesu industrija i arhitektura, gdje je vrlo bitno da se planovi i nacrti pohrane ne samo u digitalnom, već i u vektorskom obliku. Na ovaj način omogućava se brži pristup, lakše prenošenje i obrada pohranjene dokumentacije. Kako je vektorizacija relativno zahtjevnja operacija, još uvijek nisu razvijeni komercijalni programi koji bi to radili automatski sa zadovoljavajućom uspješnošću, osim za jednostavnije slike. Zbog toga se još uvijek veći dio posla obavlja ručno koristeći različite CAD/CAM alate. Prednost tog načina je što je većina korisnika već upoznata s tim programima, jer se svi novi nacrti rade pomoću njih.

Osim navedenih primjena, vektorizacija je prisutna u još mnogim područjima i djelatnostima. U građevinarstvu na primjer, koriste se planovi starih kuća i liftova, koji se vektoriziraju da bi se pomoću specijalnih programa odredio najbolji sistem dizala. Tu su još i primjene u vojnoj industriji (posebno američkoj), zatim u tekstilnoj industriji za obrađivanje starih nacrti odjeće iz 18 st. i ranijeg razdoblja. Osim toga vektorizacija se uz obradu slike koristi i za rekonstrukciju narodnih nošnji. Kao jedan od primjera koji pokazuje da je vektorizacija sveprisutna, jest

izrada nadgrobnih spomenika u Americi. Kako većina ljudi donosi nacрте koje žele na nadgrobnom spomeniku, jedna američka pogrebnička firma je izradila poseban program, koji vektorizira donesenu sliku te pomoću matematičkih jednadžbi upravlja strojem koji obrađuje mramor vrlo visokom preciznošću.

Pored svih nabrojениh područja primjene vektorizacije, tu je naravno i geodezija. Vektorizacija se u geodeziji najviše koristi u prevođenju analognih karata i planova u digitalni vektorski oblik. Kod preciznijih karata i planova, automatska vektorizacija rijetko daje željeni rezultat upotrebom različitih komercijalnih programa, pa je u tom slučaju bolje provesti ručnu vektorizaciju. O ovoj temi bit će više riječi u sljedećem poglavlju.

OCR se koristi kako u industriji, tako i u znanosti, poslovanju te općenito svugdje gdje postoji potreba za prenošenjem pisanih materijala i stare dokumentacije u digitalni oblik koji je daleko pogodniji za oblikovanje, ispravljanje, dopunjavanje te pohranjivanje. OCR također omogućava niz automatskih poslova sortiranja, naročito korištenih u pošti (sortiranje pošiljaka prema prepoznatim adresama) te bankama (prepoznavanje čekova i slipova). Posebna grana OCR-a bavi se i prepoznavanjem automobilskih tablica u policijske svrhe (LPR-license plate recognition) (Punek i Kipčić 1999.).

Možemo reći da je OCR daleko najbrži način pretvorbe pisane materije u digitalni oblik, ali još uvijek zahtjeva ljudsku intervenciju i ispravljanje jer ni jedna metoda niti program nisu apsolutno točni.

5.2. Metode vektorizacije

Metode vektorizacije koje se najčešće koriste možemo podijeliti na ručne, poluautomatske i automatske. U ovom potpoglavlju opisane su ručne i automatske metode.

5.2.1. Ručne metode

Kod ručnih metoda prvi korak je nabavljanje analognih planova te njihovo prebacivanje u digitalni oblik, najčešće skaniranjem. Nakon toga se taj predložak (rasterska slika), koristi kao podloga preko koje se crta nova vektorska slika, pomoću CAD/CAM ili specijaliziranog softvera. Pri tome se u određenim slučajevima skaniranje može preskočiti (npr. kod prebacivanja tehničkog crteža u računalo metodom ponovne konstrukcije). Ručnu vektorizaciju nazivamo još i ekranskom vektorizacijom.

5.2.2. Automatske metode

Outline metoda. Ovo je metoda kojom se postižu najbolji rezultati, što znači da vektorska slika najviše slična originalnu rastersku. Temelji se na grupiranju objekata odnosno područja na osnovu sličnosti boja. Broj boja koje se koriste može se unaprijed odrediti ili prepustiti optimizaciju samom algoritmu koja se sastoji u određivanju najpogodnije palete boja.

Centerline metoda se prvenstveno koristi za obradu tehničkih crteža, i to isključivo crno-bijelih predložaka. Osim za tehničke crteže, koristi se za rasterizaciju shema, karata i ostalih crteža koji se sastoje od linija. Boja i debljina dobivenih linija može se podesiti unaprijed.

Woodcut metodom dobiva se vektorska slika koja sadrži objekte varirajuće debljine. Debljina ovisi o intenzitetu određene točke na originalnoj slici.

Sketch metodom dobivamo vektorsku sliku koja se sastoji od odvojenih slojeva (layera) linija, gdje je svako preklapanje slojeva prikazano drugim kutom nagiba crta, kako bi se dobio efekt mreže. Osim toga, moguće je odabrati razmak između linija i obično se koristi za izradu specijalnih efekata.

Mosaic metoda. Rezultat ove metode je slika koja se sastoji od polja simetričnih objekata, a to su obično kvadrati. Boja svakog kvadrata određena je srednjom bojom koju ima originalna slika u tom području. Željena sličnost s originalnom slikom postiže se variranjem ukupnog broja kvadrata. Što je taj broj veći, i slika je bolje kvalitete.

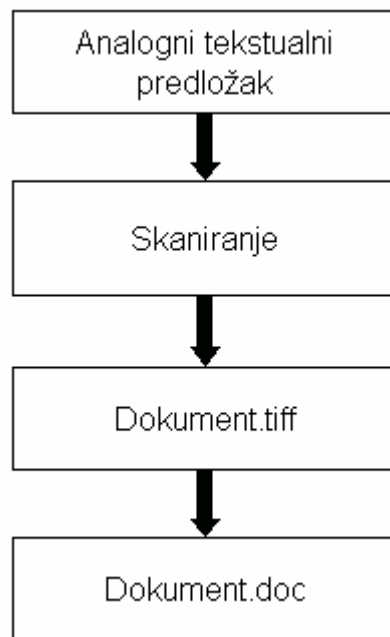
3D Mosaic metoda je zapravo samo modifikacija prethodne. Razlikuju se samo u tome što se može birati vrsta objekta i što se objekti prikazuju tako da ostvaruju 3D efekt.

5.3. Metode OCR-a kod prepoznavanja teksta

Sve metode koje se koriste u OCR-u rade na način da se pojedini znakovi izdvoje sa slike, a zatim prepoznaju jedan po jedan. Osim toga, poželjno je da slike budu binarne (crno-bijele) kako bi se smanjio utjecaj šumova. Ulazna veličina u OCR procesu je digitalna slika u rasterskom formatu. Većina OCR aplikacija omogućuje OCR na već postojećoj rasterskoj slici ili interaktivno, što znači da je slika smještena na skaneru i privremeno se smješta u memoriju računala dok traje sam proces (Slika 14).

Proces optičkog prepoznavanja znakova sastoji se od nekoliko faza:

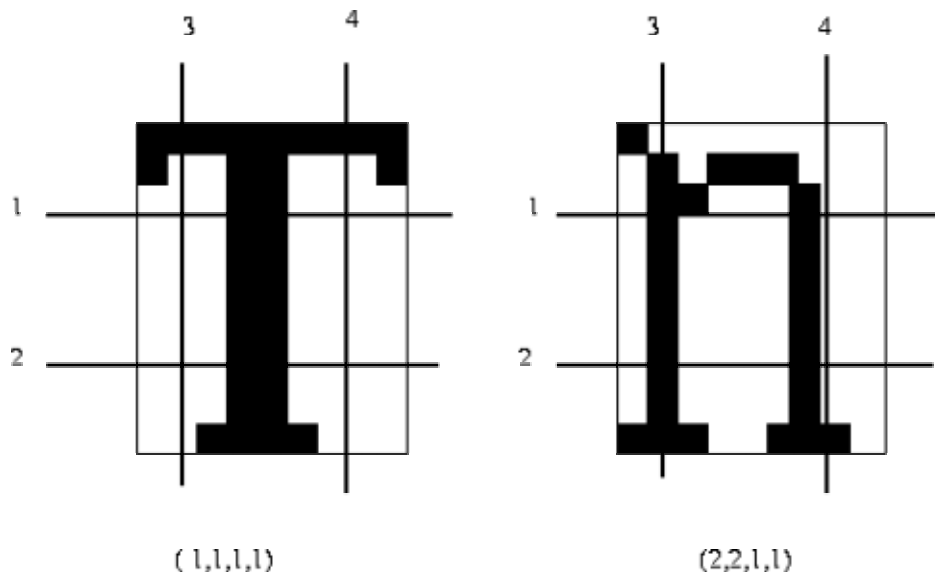
Identifikacija teksta i slikovnih blokova na slici. Većina aplikacija koristi bijeli prostor u pokušaju prepoznavanja teksta u prikladnom redoslijedu. Složenija oblikovanja, kao što su prepoznavanje podebljanih naslova ili tablica, zahtjeva interaktivnu identifikaciju i numeriranje tekstualnih blokova.



Slika 14. OCR proces

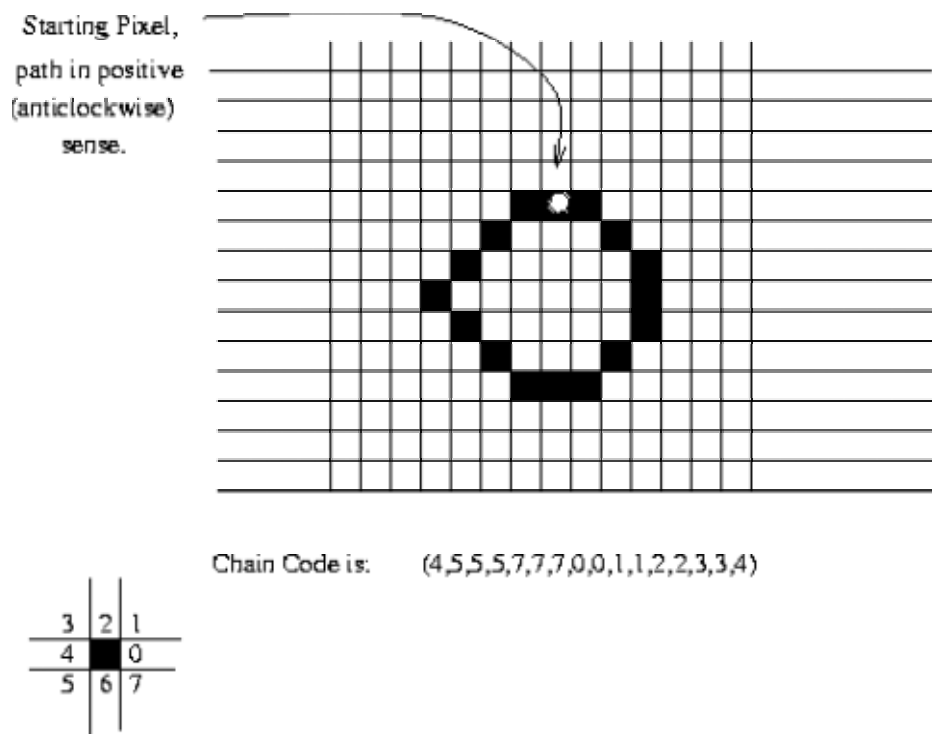
Prepoznavanje znakova. Metode koje se koriste u prepoznavanju znakova podrazumijevaju izdvajanje pojedinih znakova (najčešće brojki i slova), a zatim pojedinačno prepoznavanje. Obzirom na način prepoznavanja postoji nekoliko metoda (Punek i Kipčić 1999):

1. Uspoređivanje s unaprijed definiranom matricom dimenzija $m \times n$. Kod ove metode, znak koji se prepoznaje smješta se u odgovarajuću matricu te se zatim uspoređuje položaj crnih i bijelih polja s unaprijed definiranim matricama. Sam znak se prepoznaje kao vektor u vektorskom prostoru. Ovaj postupak vrlo je osjetljiv na šumove, a također i na translaciju, rotaciju ili neku drugu modifikaciju znaka. To je bila jedna od prvih metoda OCR-a i u novije se vrijeme više ne koristi.
2. Brojanje sjecišta s horizontalnim i vertikalnim linijama. Prvo se odabire određeni broj vertikalnih (m) i horizontalnih (n) linija, a zatim se traži broj sjecišta sa svakom od linija, pomoću čega se definira svaki znak kao vektor u vektorskom ($m \times n$) prostoru (Slika 15). Dobrim odabirom linija bitno se smanjuju dimenzije u odnosu na metodu matrica, a svaki znak se može precizno definirati. Dobra strana ove metode očituje se u određenoj neovisnosti od rotacije i izbora fontova.



Slika 15. Traženje sjecišta s horizontalnim i vertikalnim linijama

3. Metode koje koriste maske. Metoda je slična prethodnoj, s time da se ne mora gledati presjecište s pravcem već s proizvoljnim geometrijskim tijelom. Sličnost znaka s maskom predstavlja udaljenost u nekom n -dimenzionalnom prostoru. Znak se prepoznaje kao ona maska od koje je najmanje udaljen.
4. Metode koje rade na principu računanja momenata. Ova metoda koristi momente prvog i drugog reda za opis i prepoznavanje pojedinih znakova. Pri tome se najčešće se koriste Zernikeovi momenti. Prednost ove metode je neovisnost o translaciji i rotaciji.
5. Rubne metode. Kod ovih metoda pronalaze se granice znakova koje se zatim pretvaraju u funkcije kojima je svaki znak opisan (Slika 16). Iako dobro reagiraju na translaciju i donekle rotaciju, ove metode potpuno su ovisne o promjenama fontova.
6. Pronalaženje hijerarhijskih modela znakova. Ove metode polaze od ideje izgradnje hijerarhijskog stabla prema kojem je svaki znak rađen od djelova, a djelovi od poteza. Ovaj način prepoznavanja znakova još je u fazi razvoja i nije korišten u komercijalne svrhe.



Slika 16. Rubne metode

Prepoznavanje riječi. Kod ove metode najprije se iz teksta izdvajaju odvojeni nizovi znakova (riječi) te se vrši prepoznavanje na temelju unaprijed definiranog riječnika ovisno o odabranom jeziku. Ovdje valja naglasiti da većina OCR aplikacija za tekstualne dokumente podržava hrvatski jezik.

Korektura. U ovoj fazi OCR-a, rezultat se prikazuje u odgovarajućem formatu ovisno o upotrijebljenoj aplikaciji. Softver označava riječi i znakove koji nisu prepoznati, tako da ih korisnik može interaktivno ispraviti.

Izlazni format. Konačni rezultat prepoznavanja pretvara se u izlaznu datoteku za daljnju obradu. To su najčešće datoteke formata ASCII, Word, RTF, PDF i sl.

Danas je na tržištu prisutan velik broj različitih OCR aplikacija namjenjen prepoznavanju tekstualnih dokumenata. Neke od poznatijih su FineReader, Recognita, TextBridge, OmniPage i dr. Sve ove aplikacije mogu se nabaviti na tržištu po relativno prihvatljivim cijenama. Neki proizvođači svoje aplikacije nude kao podršku uz kupovinu skanera, a osim toga, moguće je pronaći demo verzije nekih aplikacija na Internetu.

Točnost i pouzdanost prepoznavanja ovisi o nizu čimbenika kao što su: dimenzija originala, starost originala, kvaliteta papira, različite varijacije teksta (fontovi), lingvistička složenost pojedinih jezika, kvaliteta i rezolucija skanera, ručno ili strojno pisani tekst, brzina prepoznavanja i dr. Za točnost i pouzdanost kažemo da je prihvatljiva pri onoj razini kod koje je OCR prihvatljiviji od ručnog utipkavanja teksta, a obično se uzima da je to 98% i više. Ova razina utvrđuje se prema broju naknadno potrebnih ispravaka u odnosu na ukupan broj znakova.

5.4. Primjena OCR-a u obradi katastarskih planova

Problem prepoznavanja znakova (brojeva, slova, različitih simbola i dr.) na katastarskom planu puno je složeniji nego kod običnih tekstualnih dokumenata. Najveći problem predstavlja nejednakost tekstualnog sadržaja. Brojevi su uglavnom pisani ručno ili pomoću šablona što se očituje u različitom nagibu znakova dok nazivi ulica, vodotoka i ostalih izduženih objekata obično prate smjer njihova protezanja. Tu je još i problem kvalitete analognog predloška, pa na većini starijih planova postoje oštećenja, zamućenja ili promjena boje originala (požutjeli listovi). Ova pojava ometa prepoznavanje znakova jer dolazi do smanjenja kontrasta između teksta i podloge. Poseban problem predstavlja preklapanje teksta s ostalim sadržajem na plan te sjecišta međa, što mnoge OCR aplikacije pokušavaju prepoznati kao tekst. Rješenje navedenih problema zahtijeva pronalaženje i razvoj posebnih metoda prepoznavanja.

Postupak se općenito sastoji od razdvajanja na tekstualni i netekstualni sloj tako da se svaki od njih u postupku vektorizacije posebno obrađuje.

5.4.1. Pregled metoda

Prema Shortis i Trisirisatayawong 1994, proces prepoznavanja i spajanja može se općenito podijeliti u tri faze. Prva faza uključuje skeniranje grafičkog izvornika, predprocesiranje, odvajanje na tekstualni i netekstualni sloj te identifikaciju objekata. Netekstualni sloj, koji sadrži linije, točke i poligone se zasebno vektorizira. U drugoj fazi računaju se momenti za prepoznavanje pojedinačnih znakova identificiranih u prvoj fazi. Grupiranjem prepoznatih znakova formiraju se riječi. U trećoj konačnoj fazi obavlja se spajanje riječi i pojedinih obilježja te se kreira konačna baza podataka.

Za prepoznavanje znakova koristi se metoda računanja momenata. Moment reda $p+q, m_{pq}$, za svaki objekt na rasterskoj slici definira se kao $m_{pq} = \sum \sum x^p y^q$, gdje su x i y koordinate piksela ovisno o položaju ishodišta. Na temelju ove osnovne formule računaju se invarijante momenata za svaki pojedini znak.

Eikvil i dr. 1995, u svom radu koriste statističku klasifikaciju za prepoznavanje simbola i znakova. Slično prethodnoj metodi, postupak se dijeli u tri faze. U prvoj fazi se izdvajaju zasebni znakovi i simboli te se obavlja prepoznavanje. U drugoj fazi prepoznati znakovi i simboli se grupiraju te se u trećoj fazi obavlja sintaksna analiza i ispravak pogrešaka.

Rehier i Li 1995, u svom radu ispituju nekoliko različitih metoda za prepoznavanje simbola i znakova. Za ocjenu kvalitete sustava za prepoznavanje koriste četiri kriterija: područje rada, robustnost, produktivnost i ispravnost. Rezultati njihovih ispitivanja pokazuju da je najbolja metoda prepoznavanja određivanje podudarnosti na osnovu Hausdorff-ovih udaljenosti koja se temelji na određivanju stupnja podudarnosti između objekata, mjerenjem udaljenosti između dva seta točaka.

Amerali i Servigne 1996, u svojoj metodi polaze od traženja ravnih linija na planu. Metoda osim prepoznavanja teksta podržava i prepoznavanje nekih drugih objekata na katastarskom planu. Cjelokupni postupak dijeli se u dvije faze. Prva faza je vektorizacija i izdvajanje segmenata, a druga interpretacija pojedinih segmenata. U fazi vektorizacije iz binarne slike izdvajaju se tri vrste segmenata. Prvu skupinu čine signifikantni segmenti poput parcela, putova i zgrada. Druga skupina su šrafirani segmenti koji se koriste u prepoznavanju zgrada. Treću skupinu čine tekst i mali simboli. Pojedini segmenti pohranjuju se u bazu segmenata. Ključnu kariku u strukturi čini sustav za interpretaciju koji iz segmentne baze stvara objektnu bazu podataka (Cetl 2001).

5.4.2. Aplikacije

Razvoj različitih metoda doveo je i do pojave različitih aplikacija za poluautomatsku i automatsku vektorizaciju koji uključuju OCR i prepoznavanje simbola. Među najznačajnijima su ProVec, R2V (Rastertovector), Microstation Descartes i dr.

Za tekst i ćelije u MS Descartesu moguće je postaviti odgovarajuću temu tj. set parametara koji definiraju simboliku elemenata (font, širina i visina, stil, boja, debljina linije i sl.). Descartes omogućava korisniku odabir između dva načina prepoznavanja. U prvom, aplikacija izabire tekstualne blokove i ćelije za cijelu sliku ili odabrano područje, dok u drugom korisnik interaktivno bira različiti tekst odnosno ćelije. Prepoznavanje obavlja softver, a korisnik interaktivno sudjeluje u tom procesu uz mogućnost ispravljanja.

Prepoznavanje znakova (slova i brojaka), Descartes obavlja dobro, no uspješnost uvelike ovisi o kvaliteti grafičkog izvornika, ali i o kvaliteti teksta obzirom da je većinom pisan rukom. Proces se odvija interaktivno tj. aplikacija prvo odabere tekstualne blokove, a zatim prepoznaje tekst idući od bloka na blok, pri čemu dozvoljava intervenciju korisnika. Pri tome aplikacija često izabere različita sjecišta među te ih pokušava prepoznati kao tekst. To se može izbjeći ako se odabere mod u kojem korisnik sam odabire dio po dio plana. Za jedan list plana 1:2880 dobre kvalitete uz odabir interaktivnog odabira blokova potrebno je oko jedan sat za prevođenje cjelokupnog tekstualnog dijela u vektorsku sliku. Kod planova s velikim oštećenjima i zamrljanošću OCR ne daje najbolje rezultate. U takvim slučajevima bolje je provesti ručnu obradu teksta.

Kod predložaka u boji ili u sivim tonovima, postupak je puno složeniji jer zahtjeva definiranje filtara za boje odnosno nivoe zacrnjenja. Kod katastarskih planova, obzirom da se novo stanje odnosno promjene ucrtavaju crvenom bojom, preporučljivo je plan skanirati u sivim tonovima ili u boji.

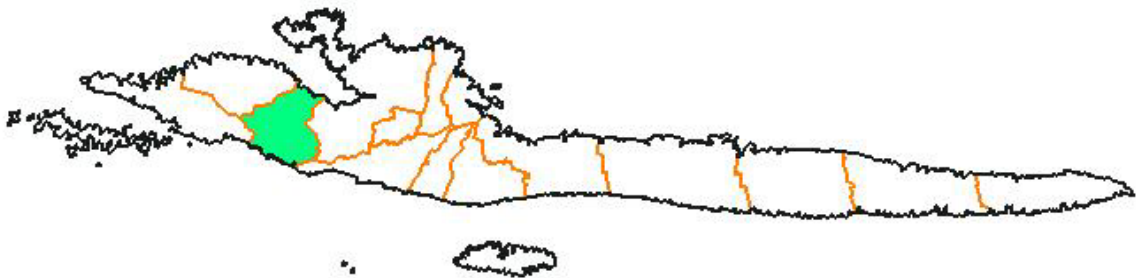
Razina kvalitete algoritma raspoznavanja u Descartesu je vrlo visoka, a posebice u radu s binarnim predlošcima. Slike u boji i slike u sivim tonovima zahtijevaju dodatno vrijeme potrebno za izradu pogodnog filtra. Relativno loša osobina je automatski izbor tekstualnih blokova. U tom slučaju aplikacija se baš i nije pokazala najuspješnijom. Kao nedostatak još se može navesti neprepoznavanje hrvatskih znakova, što nije od presudnog značaja (Cetl 2001).

6. Primjena OCR-a u vektorizaciji katastarskih planova k.o. Grablje

U dosadašnjem dijelu diplomskog rada, OCR je uglavnom teoretski obrađen s kratkim osvrtima na postojeće radove koji se bave ovom problematikom. U ovom poglavlju bit će više riječi o praktičnoj primjeni OCR-a u vektorizaciji te o svim dobrim stranama kao i problemima koji se pri tome javljaju. Ispitivanje je obavljeno na katastarskim planovima k.o. Grablje, a dobiveni rezultati bit će detaljno prikazani i statistički obrađeni.

6.1. Katastarska općina Grablje

K.o. Grablje nalazi se u zapadnom dijelu otoka Hvara, a proteže se od mjesta Zračće na južnoj obali otoka do Starigradskog zaljeva na sjeveru (Slika 17). Najveće naselje je Velo Grablje (Slika 18) koje je smješteno u unutrašnjosti otoka na otprilike pola puta između Starog Grada i grada Hvara. Za obavljanje poslova katastra nekretnina na području ove općine nadležan je Područni ured za katastar Split, ispostava Stari Grad.

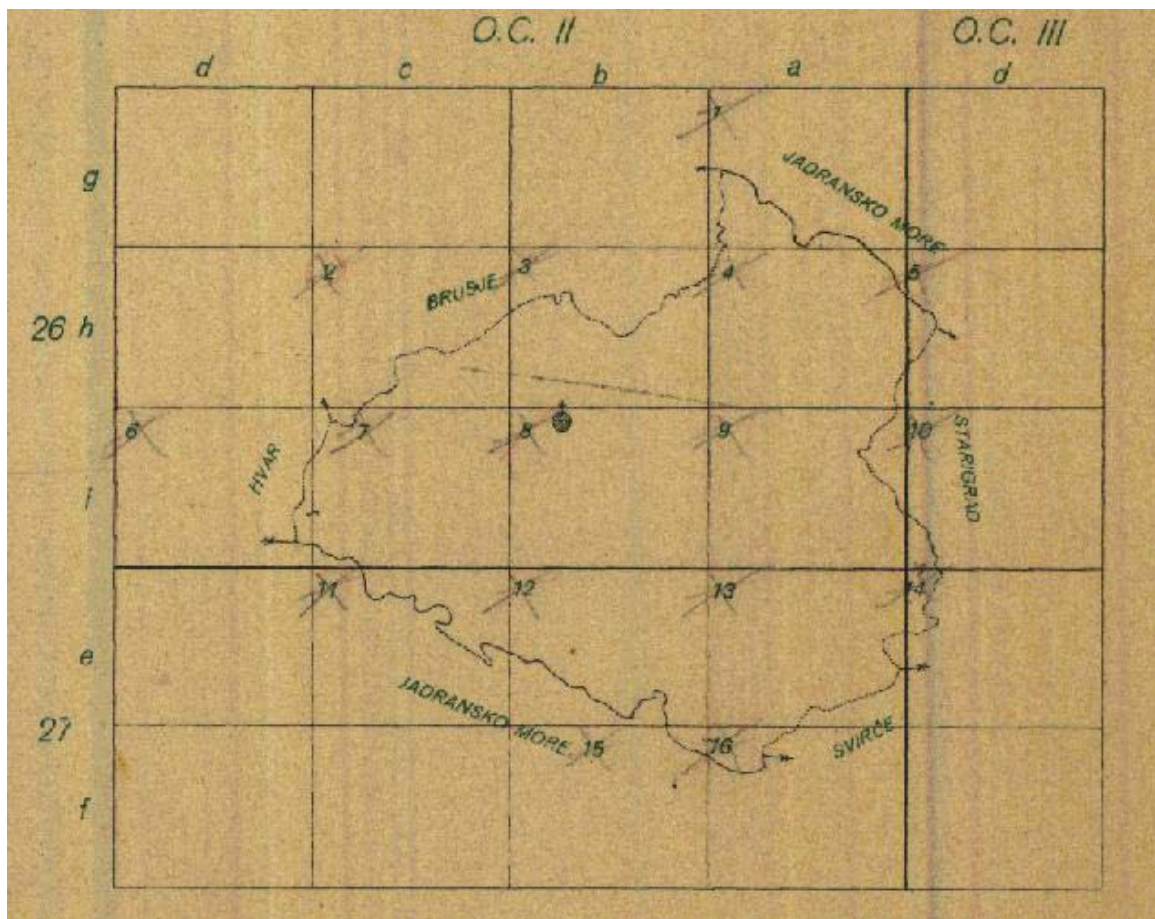


Slika 17. Položaj katastarske općine Grablje



Slika 18. Velo Grablje

Katastarska općina je prikazana na osamnaest katastarskih planova grafičke izmjere, izrađenih u bečkom koordinatnom sustavu za vrijeme Austro-Ugarske monarhije (vidi prilog br. 1). Od toga je šestnaest listova izrađeno u mjerilu 1:2880, dok su dva lista izrađena u mjerilu 1:1440. Na jednom je prikazano područje uvale Zračće, a na drugom se nalazi nekoliko odvojenih planova na kojima su prikazana područja naseljenih mjesta. Mjerna jedinica na tim planovima je hvat. Podjela na listove prikazana je na jednom od skaniranih listova katastarskih planova (Slika 19):

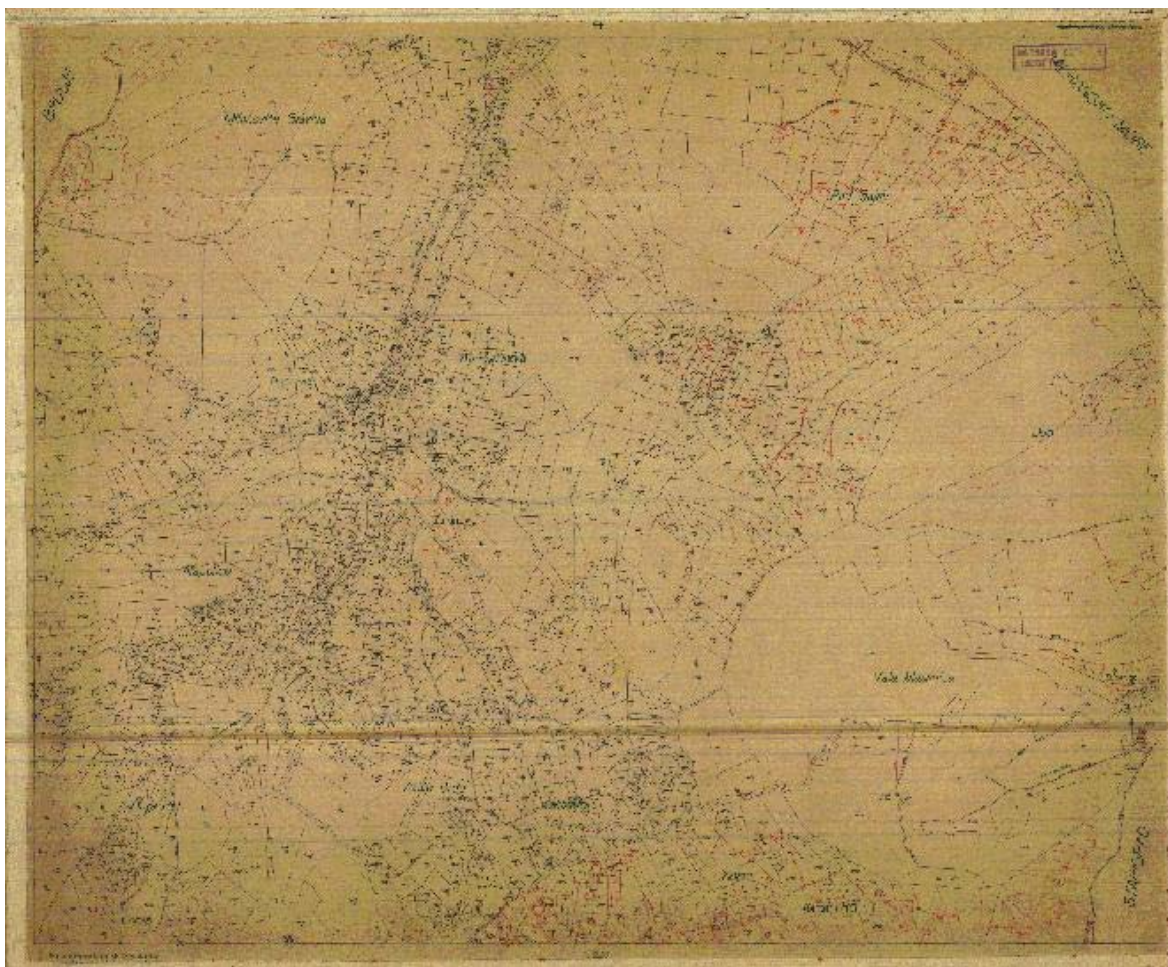


Slika 19. Podjela na listove katastarske općine Grablje

6.2. Preuzimanje postojećih podataka i ocjena kvalitete materijala

Postojeći podaci preuzeti su od strane katastarskog ureda u Starom Gradu, a sastoje se od skaniranih listova katastarskih planova. Podaci su bili pohranjeni na CD-u (311685 Grablje) u slikovnom TIFF formatu i to na način da jedan list plana čini jednu TIFF datoteku. Broj u zagradi je šifra katastarske općine, a ujedno predstavlja i naziv medija. Budući da neki listovi obuhvaćaju manji dio katastarske općine, a i njihova ukupna površina je mala, oni su spojeni sa susjednim većim listom (listovi 6&7) ili su dva lista s manjom prikazanom površinom spojena u jednu datoteku (listovi 10&14 i 15&16). Tako je ukupno preuzeto 15 rasterskih datoteka u LZW komprimiranom TIFF formatu.

Pregledom svake pojedine datoteke ustanovljeno je da su, s obzirom na starost i kvalitetu analognih listova, relativno dobre i pogodne za daljnju obradu. To se prije svega odnosi na koordinatnu mrežu na rubovima listova 1:2880 koja je praktički potpuna i neoštećena. Ovo je vrlo važan podatak obzirom na predstojeće georeferenciranje. Osim toga, tekstualni sadržaj katastarskog plana je čitljiv, a vidljiva je i većina međnih linija. Loše strane su nedostatak koordinatne mreže kod listova mjerila 1:1440, jer nije ni ucrtana na originalne planove te promijenjena boja listova usljed starosti i vanjskih utjecaja. Jedan primjer skaniranog lista katastarskog plana 1:2880 prikazan je na sljedećoj slici (Slika 20):



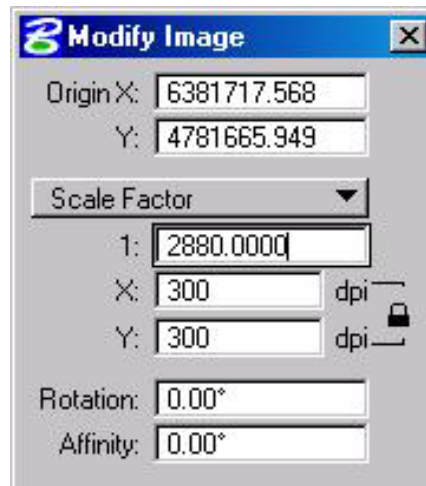
Slika 20. Skanirani list katastarskog plana

6.3. Georeferenciranje katastarskih planova

Jedan od osnovnih preduvjeta za pravilnu vektorizaciju je georeferenciranje listova katastarskih planova. Budući da su katastarski planovi bili u LZW komprimiranom TIFF formatu, a Microstation Descartes ne može otvoriti takve datoteke, trebalo ih je pretvoriti u TIFF datoteke bez sažimanja (kompresije). Taj posao obavljen u programu za obradu slike «Imaging».

Nakon toga je usljedilo učitavanje TIFF datoteka u MS Descartes. Podloga za georeferenciranje bila je datoteka Grablje.dgn. U njoj je prikazana obalna linija

otoka Hvara, granice i nazivi katastarskih općina te koordinatna mreža Bečkog sustava za to područje. TIFF datoteke su učitane interaktivno na način da su dovedene približno na one koordinate gdje bi listovi trebali biti, a zatim je upisano mjerilo plana (Slika 21).



Slika 21. Upisivanje mjerila plana

Da bismo rastersku sliku mogli transformirati odnosno uklopiti u koordinatnu mrežu, potreban nam je odgovarajući alat i identične točke. Alat koji je korišten za ovu namjenu nalazi se unutar prozora Register pod izbornikom Tools (Slika 22), a kao identične točke poslužile su točke koordinatne mreže. Tijekom izbora identičnih točaka, za kontrolu je korištena afina transformacija odnosno reziduali svake pojedine točke prikazani u prozoru «register» (Slika 23).



Slika 22. Register alati

Control Point #	On	Base System		Uncorrected System		Residuals: <u>Uncorrected System</u>		
		X	Y	X	Y	X	Y	XY
1	X	6381717.568	4783183.139	6381756.763	4783213.965	1.115	0.213	1.135
2	X	6383614.048	4783183.139	6383639.985	4783195.187	-2.449	-0.002	2.449
3	X	6383614.048	4781665.949	6383624.827	4781694.066	1.399	-0.874	1.650
4	X	6381717.568	4781665.949	6381747.750	4781712.844	-1.182	-0.659	1.353

On/Total: 90 / 90 Standard Deviation 0.593 0.442 0.523

Slika 23. Prozor register, alat za izradu modela transformacije

Postupak georeferenciranja je sljedeći. Prvo odaberemo alatnu funkciju Place Control Point (prva slijeva na slici 22) za izbor identičnih točaka. Nakon izbora

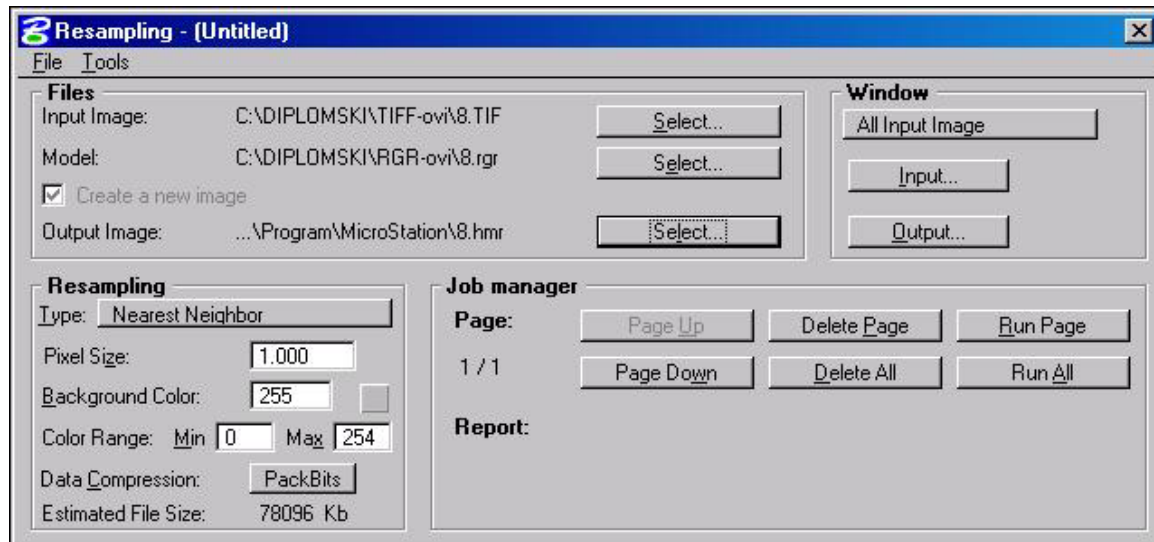
ovog alata, pokazivač (kursor) poprima oblik križića. Ovakav pokazivač omogućuje nam izbor željene točke na digitalnom katastarskom planu. Kod izbora identičnih točaka potrebno je prvo definirati položaj točke sjecišta na koordinatnoj mreži (base control point mark) pritiskom na obje tipke miša (snap). Nakon toga definiramo identične točke koordinatne mreže na rasteru (uncorrected control point mark).

U prozoru Register se automatski registriraju identične točke koje su odabrane «snapiranjem» te se prikazuju njihova odstupanja tj. statističke informacije o ispravnosti transformacijskog modela po kojima se može vidjeti kakva su odstupanja. Po završetku odabira svih identičnih točaka potrebno je prekontrolirati standardno odstupanje ili devijaciju. Taj podatak je prikazan u donjem desnom uglu prozora Register, a trebao bi se okvirno kretati do maksimalno 0.5-1.0 m u smjeru pojedinih koordinatnih osi. Budući da su se ovi podaci za katastarske planove k.o. Grablje manje-više kretali u navedenim okvirima, dobiveni su transformacijski modeli zadovoljavajuće točnosti (Tablica 2). Za svaki list, transformacijski model je pohranjen u registar datoteke u obliku Thin Plate Spline transformacije pod «rgr» ekstenzijom.

Tablica 2. Prikaz standardnih odstupanja transformacijskih modela

Naziv rgr datoteke	Standardno odstupanje (afina transformacija) [m]		
	x	y	xy
1	0.411	0.404	0.430
2	0.631	0.624	0.661
3	0.597	0.492	0.632
4	0.593	0.442	0.523
5	0.336	0.502	0.440
6&7	0.672	0.649	0.706
8	0.911	0.614	0.697
9	1.346	0.772	0.905
10&14	0.405	0.580	0.527
11	0.759	0.672	0.783
12	0.532	0.429	0.557
13	1.647	1.031	1.099
15&16	0.686	0.740	0.767

Nakon definiranja modela transformacije, potrebno je pomoću naredbe Apply/To image otvoriti dijaloški prozor Resampling pomoću kojega se provodi transformacija. Ovaj prozor sadrži mogućnosti za odabir ulazne datoteke, modela transformacije i izlazne (transformirane) datoteke (Slika 24). Unutar ovog prozora također se nalaze i mogućnosti izbora kompresije, tipa budućeg rastera, raspona boja itd. Proces transformacije pokreće se naredbom Run Page.



Slika 24. Dijaloški prozor Resampling

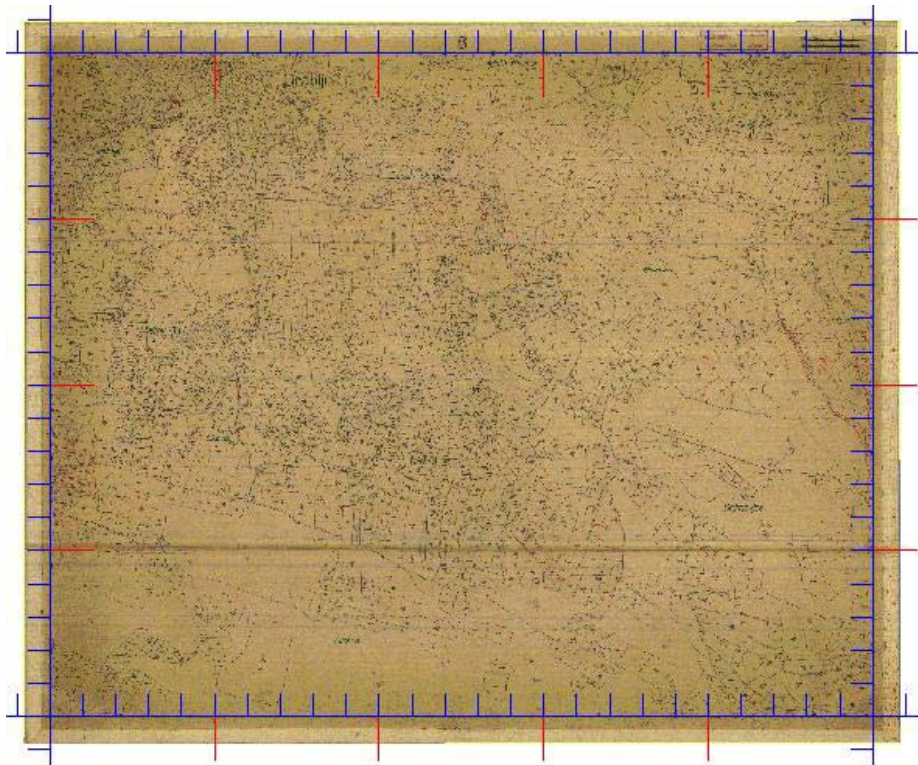
Na ovaj način transformirane su sve preuzete TIFF datoteke osim dvije. To su datoteke u kojima su prikazani katastarski planovi u mjerilu 1:1440. Budući da na njima nije bilo koordinatne mreže niti bilo kakvih drugih točaka koje bi mogle poslužiti kao identične, za njih nije provedeno georeferenciranje.

Kao konačan rezultat transformacije dobiveno je 13 rasterskih HMR datoteka koje prikazuju georeferencirane katastarske planove odnosno listove dovedene na teoretske dimenzije i smještene u koordinatnu mrežu (Slika 26).

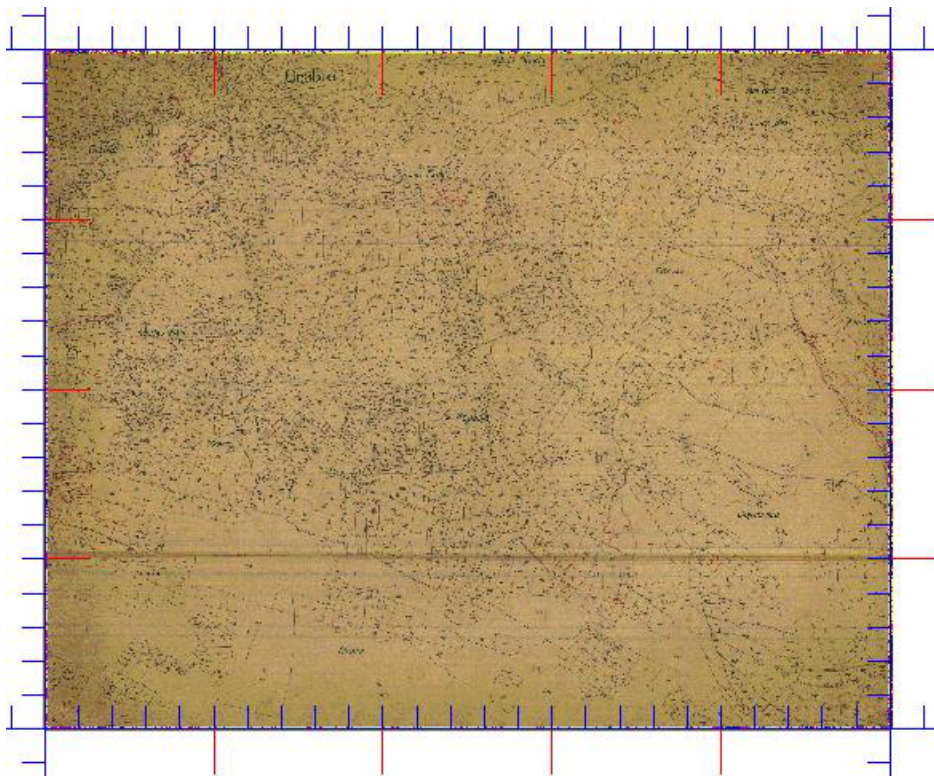
Prije vektorizacije trebalo je još odstraniti određene podatke iz HMR datoteka. To su bili oni djelovi katastarskih planova koji su nakon georeferenciranja ostali izvan okvira koordinatnog područja jednog lista. Njihovo odstranjivanje trebalo je provesti kako ne bi došlo do preklapanja listova (Slika 27). Ovo je učinjeno pomoću alata s glavnog izbornika MS Descartes-a Tools/Image/Image Transform (Slika 25).



Slika 25. Alat Image Transform za obradu slike



Slika 26. Georeferencirani katastarski plan pohranjen u HMR datoteci



Slika 27. Georeferencirani katastarski plan bez odstranjenog dijela

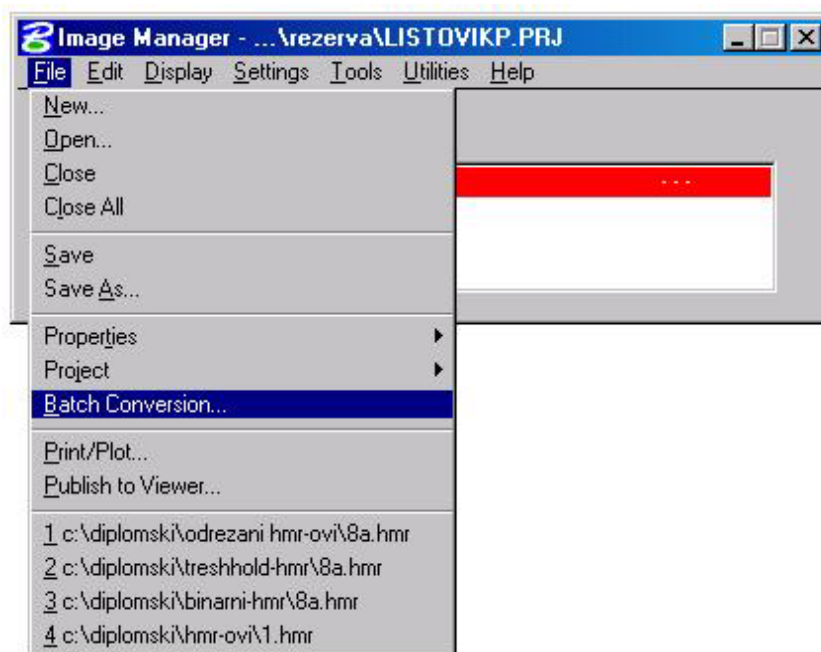
6.4. Optičko prepoznavanje znakova (OCR)

Provođenjem svih opisanih predradnji dobivene su rasterske podloge (HMR datoteke) na kojima se može primjeniti automatska vektorizacija tekstualnog sadržaja ili OCR. Za ove potrebe korišten je programski paket Microstation V7.1 točnije Microstation Descartes koji u sebi sadrži OCR aplikaciju. Način rada ove aplikacije opisan je u prethodnom poglavlju, a za prepoznavanje znakova na katastarskim planovima k.o. Grablje korišten je mod u kojem korisnik sam odabire dio po dio plana pa čak u nekim slučajevima i riječ po riječ. Na ovaj način, naknadni popravci svode se na najmanju moguću mjeru. Iako ovaj mod rada oduzima daleko više vremena, rezultati prepoznavanja su mnogo bolji. U drugom modu rada, korisnik samo odabire rastersku sliku dok aplikacija sama odabire tekstualne blokove. Pri tome često i netekstualni sadržaj plana pokušava prepoznati kao tekst ili dvije odvojene riječi prepoznaje kao jednu i sl.

6.4.1. Binarizacija katastarskih planova

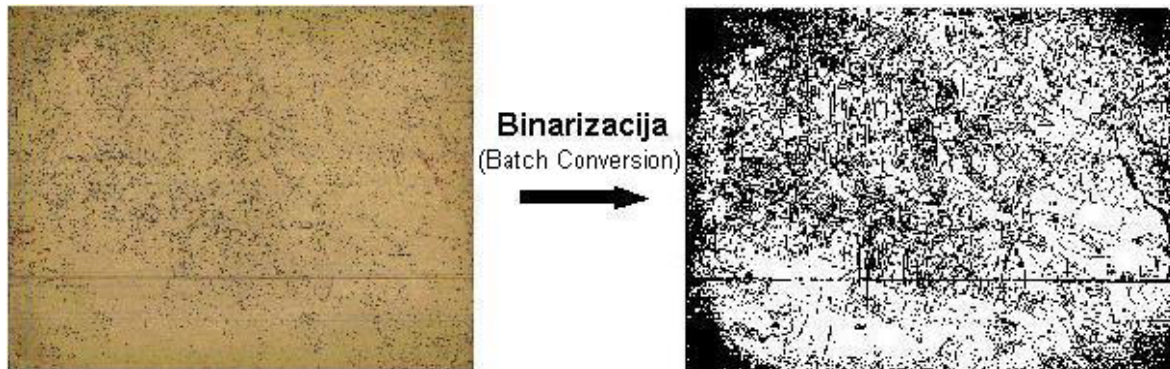
Kao što je već navedeno ranije, kod predložaka u boji ili sivim tonovima postupak je puno složeniji jer zahtjeva definiranje filtera za boje ili razinu zacrnjenja. S obzirom na starost skaniranih analognih katastarskih planova, došlo je do neizbježne promjene boje listova. Tako su listovi požutjeli, a osim toga sve promjene su ucrtavane crvenom bojom. Zato je bilo neophodno provesti binarizaciju HMR datoteka da bi se dobile crno-bijele rasterske slike. Na ovaj način se «olakšava» posao OCR aplikaciji jer su ovakve slike oslobođene šumova, a crni tekst se jasno raspoznaje na bijeloj podlozi.

Kao alat za binarizaciju prvobitno je korištena opcija Batch Conversion (Slika 28) koja se pokazala nepogodnom jer nisu dobivene crno-bijele slike zadovoljavajuće kvalitete.



Slika 28. Alat za obradu rasterskih slika Batch Conversion

Rezultat ovakve binarizacije prikazan je na primjeru datoteke 8.HMR (Slika 29).

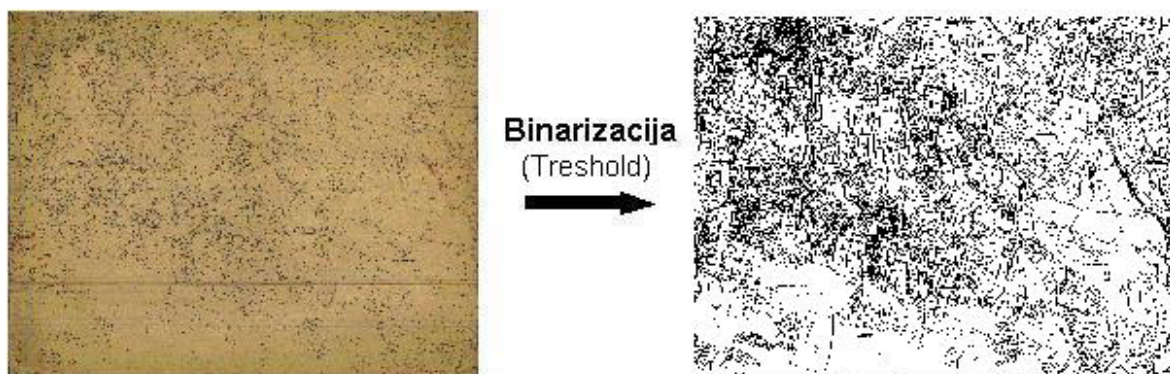


Slika 29. Binarizacija pomoću alata Batch Conversion

Kao što je vidljivo na slici, dijelovi katastarskog plana su zatamnjeni u toj mjeri da se tekst ne može prepoznati niti golim okom, a kamoli OCR aplikacijom.

Zato je binarizacija listova obavljena u programu za obradu slike Corel PhotoPaint. Ovaj program sadrži tzv. Treshold alat koji omogućuje korisniku definiranje razine zatamnjenja i uklanjanje različitih nečistoća. Razina zatamnjenja je prikazana brojčano od 0 do 256. Svi georeferencirani listovi su prije učitavanja u Corel PhotoPaint bili prebačeni u TIFF format jer ovaj program ne podržava format zapisa HMR. Nakon binarizacije georeferencirani listovi su bili vraćeni u HMR format. Za pretvorbu TIFF/HMR korišten je već opisani alat Batch Conversion.

Rezultat ove binarizacije također je prikazan na primjeru datoteke 8.HMR radi lakše usporedbe (Slika 30).



Slika 30. Binarizacija pomoću treshold alata

Usporedbom ove dvije metode jasno je vidljivo da se pomoću treshold alata dobivaju mnogo pregledniji crno-bijeli katastarski planovi. Iz tog razloga, ovom metodom binarizano je svih 13 HMR datoteka. Ovo je bio posljednji korak u pripremi katastarskih planova za automatsku vektorizaciju teksta (OCR).

Potrebna razina zatamnjenja razlikovala se od lista do lista, a ovisila je prije svega o kvaliteti skaniranih originalnih listova. Vrijednosti zatamnjenja prikazuje (Tablica 3).

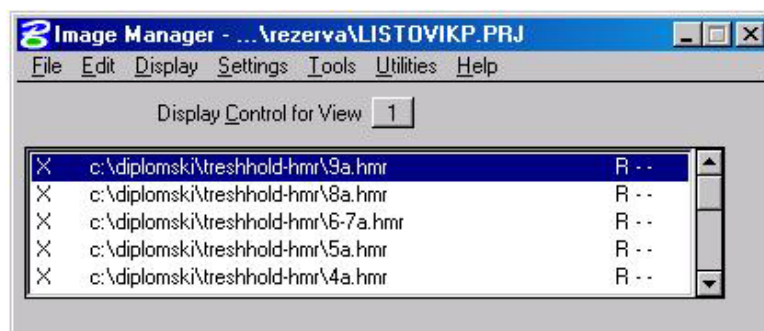
Tablica 3. Korištena razina zatamnjenja listova plana u postupku binarizacije

Naziv HMR datoteke	Razina zatamnjenja
1	100
2	105
3	90
4	95
5	115
6&7	100
8	100
9	105
10&14	105
11	105
12	95
13	95
15&16	110

Iz tablice je vidljivo da su skanirani originalni listovi bili podjednake kvalitete.

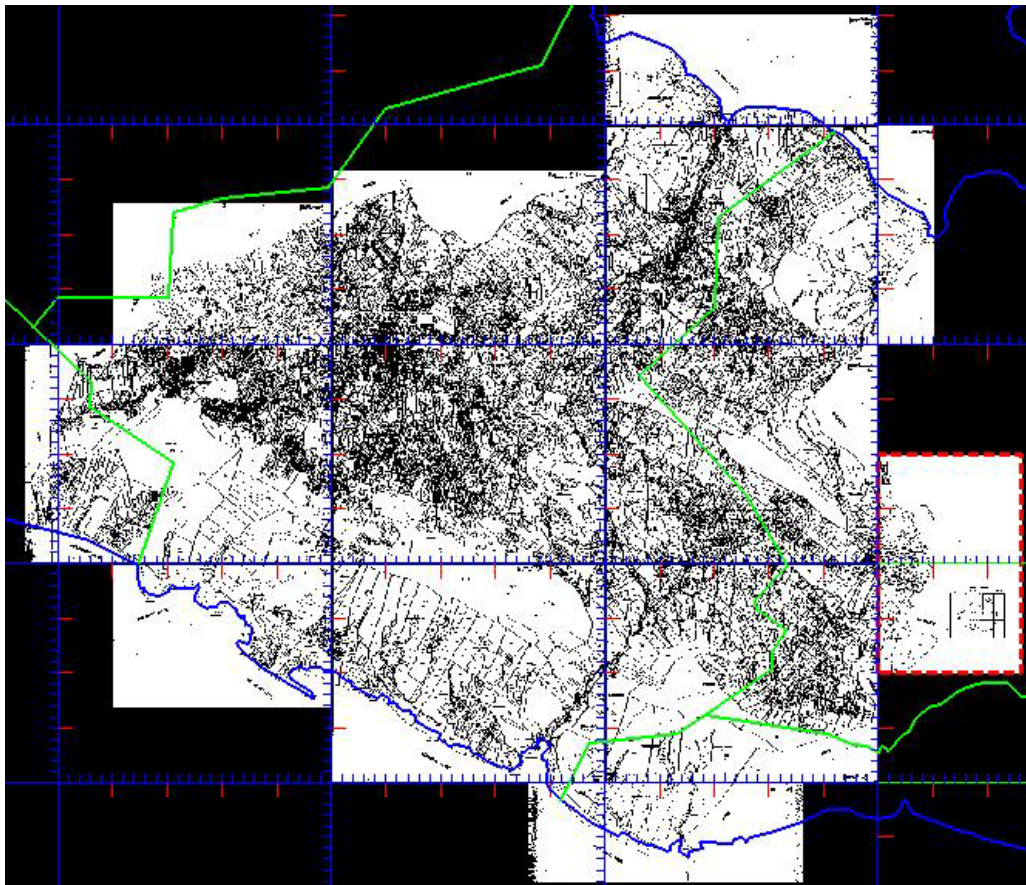
6.4.2. Automatska vektorizacija katastarskih planova

Prethodnim radnjama stvoreni su svi potrebni uvjeti za primjenu OCR-a, a binarne HMR datoteke učitane su u Microstation Descartes pomoću prozora Image Manager (Slika 31). Učitavanje je obavljeno u već spomenutu datoteku Grablje.dgn.



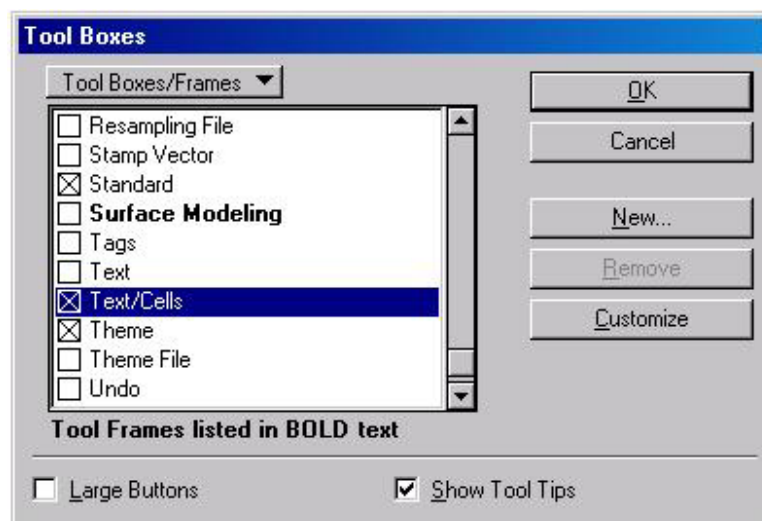
Slika 31. Prozor za učitavanje rasterskih slika Image Manager

Tako je dobiven kontinuirani prikaz cijele katastarske općine bez prekida ili preklapanja listova (Slika 32).



Slika 32. Učitani listovi k.o. Grablje u datoteku Grablje.dgn

Do alata u MS Descartesu koji omogućuje OCR dolazi se preko glavnog izbornika Tools/Tool Boxes, a zatim izborom opcije Text/Cells (Slika 33).



Slika 33. Prozor za izbor alata Tool Boxes

Izborom opcije Text/Cells otvara se novi manji prozor (Slika 34) s alatima koji pružaju korisniku različite mogućnosti.



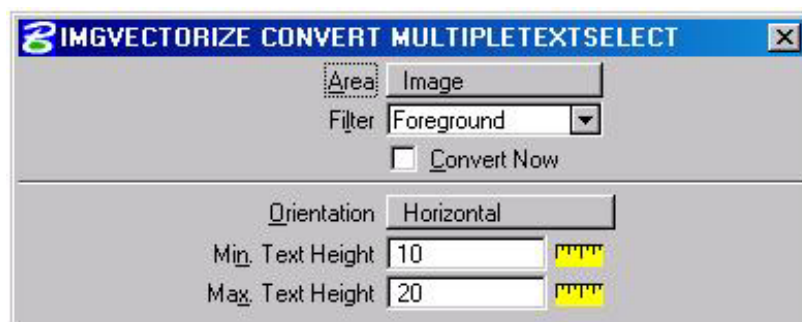
Slika 34. Alati za izbor i pretvaranje teksta u vektorski oblik

Prva opcija slijeva omogućuje izbor horizontalnog teksta, a sadrži još dvije dodatne mogućnosti. To su trenutna pretvorba pojedinog znaka ili riječi (Convert Now) i izbor više znakova ili riječi za naknadno prepoznavanje (Shrink Selection to Text)(Slika 35). Isto vrijedi i za drugi gumb (button) slijeva koji se koristi za odabir kosog teksta. Ovaj način prepoznavanja predstavlja već opisani prvi mod rada.



Slika 35. Izbor između trenutne i naknadne pretvorbe teksta

Treći gumb slijeva služi za odabir tzv. višestrukog teksta (Select Multiple Text), a moguće je odabrati i cijelu sliku. Pri tome korisnik može zadati minimalnu i maksimalnu veličinu teksta koji treba prepoznati kako bi se izbjeglo neželjeno prepoznavanje ostalog sadržaja slike (Slika 36). Ovo je drugi mod rada.



Slika 36. Postavke kod prepoznavanja višestrukog teksta

Četvrti gumb služi za brisanje selekcije odnosno okvira unutar kojeg se nalazi odabrani tekst.

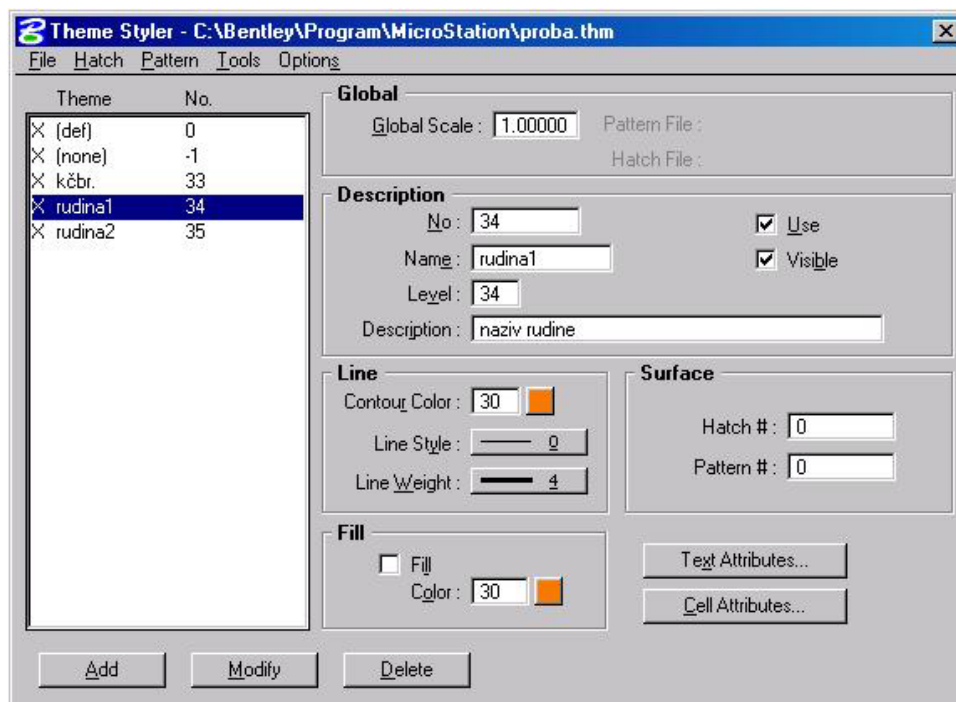
Peti gumb se koristi za prepoznavanje teksta unutar označenog područja dok šesti gumb omogućuje podešavanje različitih postavki teksta poput kvalitete (Poor, Average i Excellent) ili varijacije nagiba znakova (Low, Medium i High). Osim toga moguće je uključiti upotrebu riječnika kod prepoznavanja teksta. OCR aplikacija u MS Descartes-u podržava desetak jezika među kojima nažalost nije i hrvatski.

Posljednja tri gumba služe za selektiranje i prepoznavanje ostalih netekstualnih znakova i simbola (Cells), i nisu korišteni u ovom diplomskom radu.

Kod automatske vektorizacije teksta na katastarskim planovima k.o. Grablje primjenjen je prvi mod prepoznavanja i to na način da su ručno selektirani nazivi i brojevi katastarskih čestica (Shrink Selection to Text). Nakon određenog broja selektiranih naziva i brojeva čestica, za pretvorbu u vektorski oblik *.dgn datoteke izabrana je opcija Convert Now. Ovaj postupak je ponavlján sve dok nisu selektirani i vektorizirani svi brojevi i nazivi na katastarskim planovima k.o. Grablje.

Kako će neki tekst izgledati u vektorskom obliku nakon prepoznavanja, korisnik može definirati pomoću alata Theme Styler dialog box (Slika 37). Tu se mogu kreirati tzv. Theme datoteke s ekstenzijom *.thm. U njima se mogu podesiti gotovo sve postavke teksta (font, širina i visina, stil, boja, debljina linije kao i sloj, naziv sloja, opis i sl.). U jednu Theme datoteku može biti pohranjeno nekoliko tema. Pri tome svaka tema predstavlja jednu vrstu teksta ili simbola.

Budući da na katastarskim planovima k.o. Grablje postoji nekoliko vrsta teksta (brojevi katastarskih čestica, nazivi rudina, potoka, uvala i sl.), formirana je jedna Theme datoteka (Grablje.thm) u kojoj su pohranjene tri teme (kčbr., rudina1 i rudina2). Ovo je učinjeno da bi se različite vrste teksta razlikovale i nakon prepoznavanja u vektorskom obliku.



Slika 37. Prozor za definiranje tema, Theme Styler dialog box

7. Analiza dobivenih rezultata

Vrijeme utrošeno za prepoznavanje svakog pojedinog lista odnosno HMR datoteke ovisilo je veličini samog lista te vrsti, količini i gustoći tekstualnog sadržaja. Kako se ovdje radilo o prepoznavanju u modu rada gdje korisnik sam bira tekstualni sadržaj, utrošak vremena bio je relativno velik. Da bi se dobio jasniji prikaz utrošenog vremena, taj je podatak za svaku pojedinu HMR datoteku prikazan tablično (Tablica 4). U tablici su navedene približne vrijednosti.

Tablica 4. Prikaz utrošenog vremena kod prepoznavanja teksta

Naziv HMR datoteke	Utrošeno vrijeme	
	h	min
1	0	30
2	1	30
3	2	45
4	4	30
5	0	15
6&7	3	30
8	5	45
9	5	15
10&14	0	30
11	0	45
12	1	15
13	2	00
15&16	0	15
Ukupno:	28 h	45 min

Po završetku prepoznavanja, tekstualni sadržaj je razvrstan po slojevima ovisno o tome je li ga OCR aplikacija prepoznala pravilno, pogrešno ili ga uopće nije prepoznala. Pri tome su brojevi i podbrojevi katastarskih čestica tretirani kao odvojeni tekst jer ih aplikacija nije mogla prepoznati kao jedinstvene brojeve. Brojevi i nazivi prema tehničkim uputama (Državna geodetska uprava 2002),

upisani su ručno, kao i oni brojevi i nazivi koje aplikacija uopće nije prepoznala. Slojevi datoteke Grablje.dgn prikazani su u tablici (Tablica 5).

Tablica 5. Slojevi datoteke Grablje.dgn

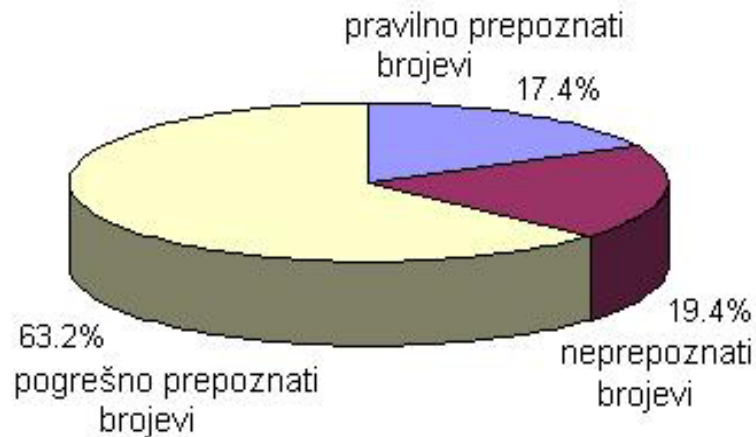
Broj sloja	Ime sloja	Napomena
4	Brojevi	Brojevi čestica (tehničke upute)
10	Listovi	Koordinatna mreža Bečkog sustava
11	Rudine	Nazivi rudina (tehničke upute)
40	Brojevi1	Pravilno prepoznati brojevi
41	Brojevi2	Neprepoznati brojevi
42	Brojevi3	Pogrešno prepoznati brojevi
50	Rudine1	Pravilno prepoznati nazivi rudina
51	Rudine2	Neprepoznati nazivi rudina
52	Rudine3	Pogrešno prepoznati nazivi rudina
60	Hvar	Obalna linija otoka Hvara
61	Granice	Granice katastarskih općina
62	Nazivi	Nazivi katastarskih općina

Iako su preuzeti relativno kvalitetni skanirani listovi katastarskih planova, rezultati prepoznavanja OCR-om nisu zadovoljavajući. To znači da bi kod praktične primjene ove metode u vektorizaciji katastarskih planova stare izmjere bilo mnogo naknadnih ispravaka što predstavlja dodatni utrošak vremena. Ovu tvrdnju potkrepljuju dobiveni rezultati na primjeru katastarske općine Grablje (Tablica 6).

Tablica 6. Rezultati prepoznavanja brojeva katastarskih čestica

Vrsta brojeva	Broj
Pravilno prepoznati brojevi	3927
Pogrešno prepoznati brojevi	14255
Neprepoznati brojevi	4364
Ukupno:	22546

Ovi rezultati se mogu prikazati i grafički (Slika 38):



Slika 38. Grafički prikaz rezultata prepoznavanja brojeva

Glavni razlog za ovako loš rezultat je rukom pisani tekst. Tako su brojevi katastarskih čestica upisani s većim ili manjim nagibom. Iako su lako prepoznatljivi i čitljivi ljudskom oku, to nije bilo dovoljno OCR aplikaciji da ih prepozna u većem postotku. Analizirajući dobivene rezultate u datoteci Grablje.dgn možemo lako uočiti nekoliko pogrešaka u prepoznavanju koje se ponavljaju. Na primjer, aplikacija često broj 1 prepoznaje kao 7 i obrnuto, zatim broj 3 često prepoznaje kao 9. Osim toga može se uočiti da ukoliko aplikacija ne može prepoznati o kojem se broju radi, na njegovo mjesto stavlja znak ~. Još jedan problem javlja se kod brojeva koji su naknadno upisivani prilikom provođenja promjena. Budući da su takvi brojevi upisivani crvenom bojom, nakon binarizacije su izgubili na kvaliteti tako da se jedva raspoznaju i golim okom. Tu su još i poništeni brojevi za koje aplikacija ne može «znati» da ih ne treba prepoznavati. Kod prepoznavanja višeznamenkastih brojeva, broj se smatrao pravilno prepoznat, ako je aplikacija prepoznala sve njegove znamenke.

Što se tiče prepoznavanja naziva rudina, potoka, uvala i ostalog tekstualnog sadržaja, rezultati su još lošiji. Razlozi su uglavnom isti kao i kod prepoznavanja brojeva čestica, međutim tu se još javlja još jedan problem. To je veličina teksta koji se zbog toga u većini slučajeva preklapa s ostalim sadržajem na planu. Rezultati prepoznavanja naziva prikazani su u tablici (Tablica 7).

Tablica 7. Rezultati prepoznavanja naziva

Vrsta naziva	Broj
Pravilno prepoznati nazivi	13
Pogrešno prepoznati nazivi	176
Neprepoznati nazivi	23
Ukupno:	212

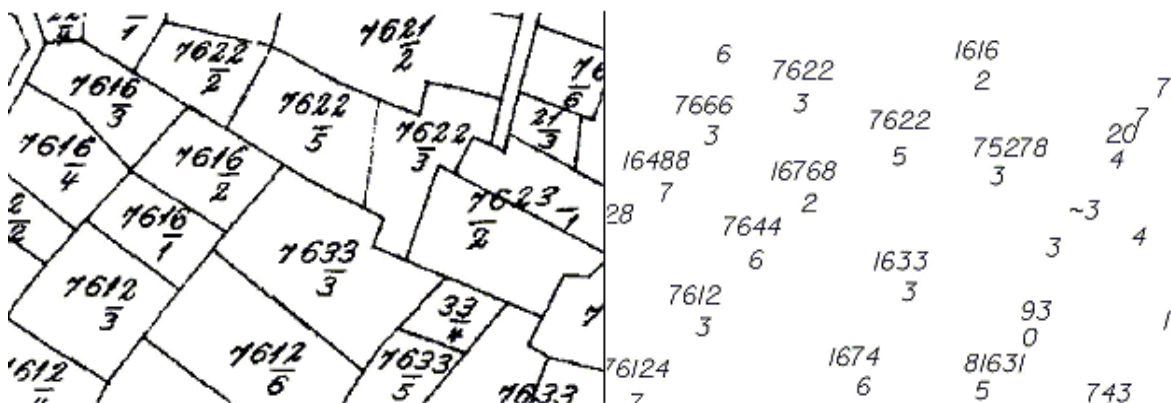
Ove rezultate u grafičkom obliku prikazuje (Slika 39):



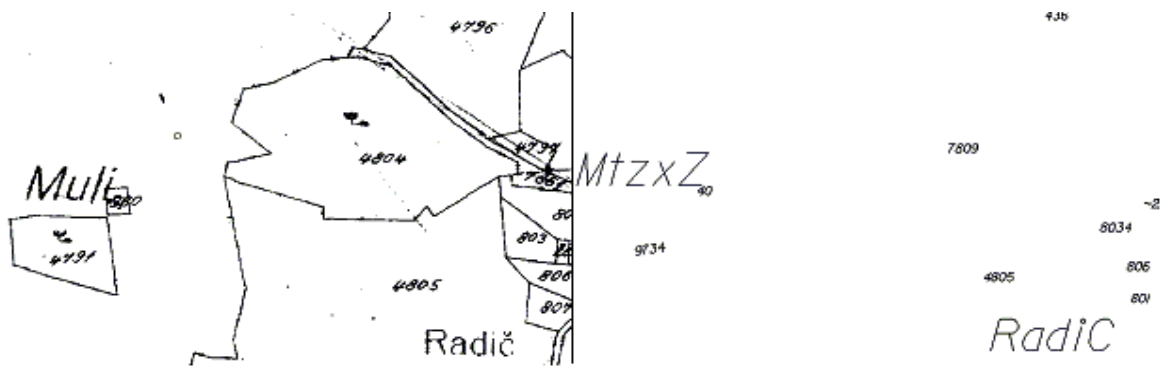
Slika 39. Grafički prikaz rezultata prepoznavanja naziva

Pregledom datoteke Grablje.dgn možemo uočiti da je aplikacija bila uspješnija samo kod naziva koji se ne preklapaju s ostalim sadržajem na planu. Kao još jedan od problema može se navesti neprepoznavanje hrvatskih slova s dijakritičkim znakovima (č, ć, š, ž i sl.). Budući da OCR aplikacija u MS Descartesu ne podržava hrvatski jezik, sva slova s dijakritičkim znakovima tretirana su kao pravilno prepoznata čak i ako ih je aplikacija prepoznala bez dijakritičkih znakova (ž kao z, š kao s i sl.). Kao i kod prepoznavanja brojeva, naziv je tretiran kao pravilno prepoznat samo ako je aplikacija prepoznala sva slova u nazivu.

Na slikama (Slika 40), (Slika 41), (Slika 42), (Slika 43) i (Slika 44) su prikazani isječci binariziranog katastarskog plana k.o. Grablje gdje se može uočiti uspješnost prepoznavanja naziva i brojeva. S lijeve strane su prikazani binarni isječci plana, a s desne strane rezultat prepoznavanja OCR aplikacijom.



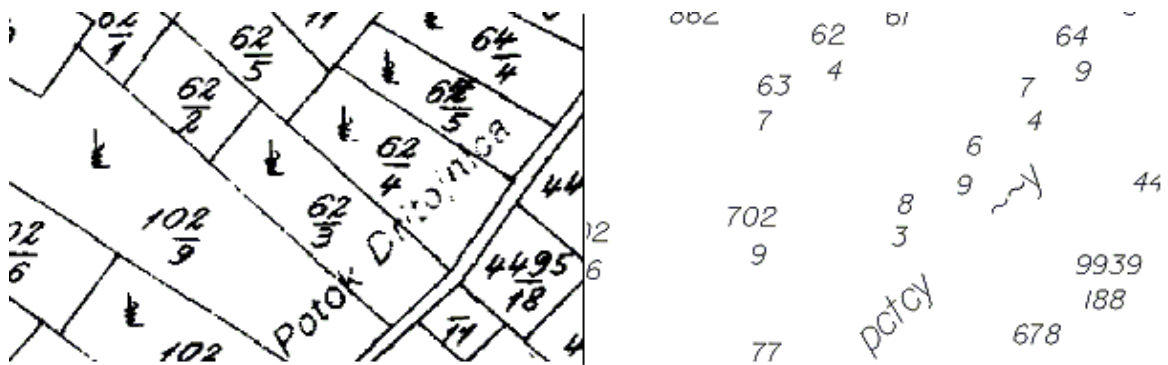
Slika 40. Primjer prepoznavanja relativno kvalitetnog binarnog predloška



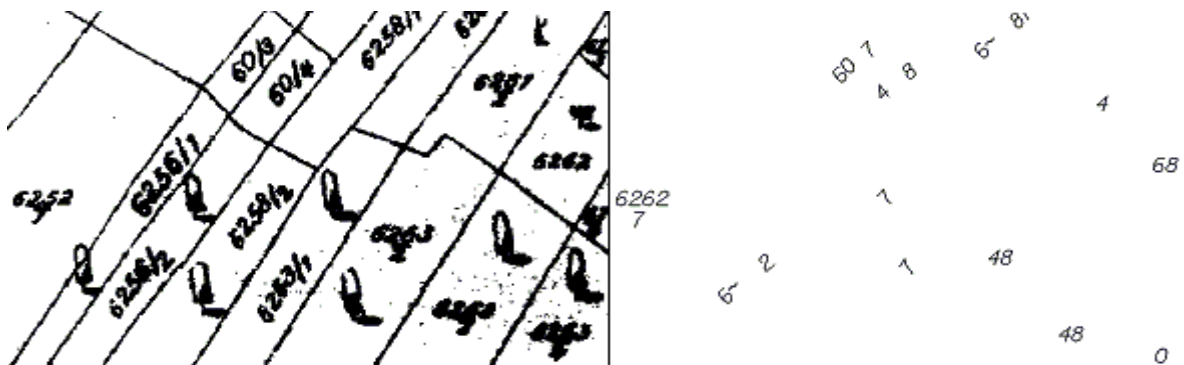
Slika 41. Primjer neprepoznavanja hrvatskih slova



Slika 42. Primjer prepoznavanja kod preklapanja slova s međnim linijama



Slika 43. Primjer prepoznavanja teksta lošije kvalitete



Slika 44. Primjer prepoznavanja lošije binariziranih djelova plana

7.1. Sadržaj priloženih medija

Rezultati obavljenog posla pohranjeni su na dva CD-a. Na jednom CD-u su datoteke pohranjene i složene prema tehničkim uputama (Državna geodetska uprava 2002) dok su na drugom CD-u pohranjeni su svi ostali podaci koji su korišteni u izradi ovog diplomskog rada (Tablica 8).

Tablica 8. Sadržaj drugog priloženog CD-a

RB.	Mapa/ Datoteka	Sadržaj
1	2	3
1.	Diplomski.doc	Tekst diplomskog rada
2.	Grablje.dgn	Vektorski podaci katastarske općine Grablje
3.	Grablje.thm	Različite teme za pojedine vrste teksta
4.	HMR/(1-15&16).HMR	Georeferencirani listovi katastarskog plana (1-15&16)
5.	Binarni/(1-15&16).HMR	Binarizirani listovi katastarskog plana (1-15&16)
6.	RGR/(1-15&16).rgr	Zapis koordinata identičnih točaka kod Thin Plate Spline transformacije

8. Zaključak

Iako je kvaliteta algoritma raspoznavanja u MS Descartesu vrlo visoka, ona dolazi do izražaja tek uz primjenu na vrlo kvalitetnim binarnim predlošcima. To konkretno znači da ova metoda vektorizacije katastarskih planova može doći u obzir uglavnom kod analognih predložaka novije izmjere.

Kad govorimo o upotrebljivosti OCR-a, možemo ocijeniti njegovu uspješnost prema broju prepoznatih znakova u odnosu na njihov ukupan broj. Osim toga vrlo važan je i vremenski čimbenik, što znači da aplikaciju smatramo uspješnom ako se njenom primjenom postiže ušteda vremena u odnosu na ručni unos teksta.

Najveći problem u korištenju OCR-a predstavlja rukom pisani tekst. Za takve predloške trebale bi se koristiti složenije aplikacije koje omogućuju «samoučenje» softvera za prepoznavanje rukom pisanog teksta. Takvi algoritmi su još u fazi razvoja i nisu prisutni u komercijalnim softverima.

Problem rukom pisanog teksta došao je do izražaja i na primjeru katastarske općine Grablje gdje je i kod blagih varijacija teksta bilo poteškoća u prepoznavanju. Međutim, to nažalost nije jedini problem. Dodatni problem predstavlja preklapanje teksta s ostalim sadržajem plana kao i kvaliteta većine analognih planova starije izmjere.

Usprkos problemima koji se javljaju, OCR će zasigurno i dalje biti područje istraživanja u digitalnoj obradi slike. Za očekivati je da će se s vremenom riješiti ili barem umanjiti problem rukom pisanog teksta, a isto tako razviti još točniji i brži algoritmi prepoznavanja. Neka moguća rješenja već se naziru u razvoju sustava umjetne inteligencije.

Literatura:

Amerali, K., Servigne, S. (1996): Methods for vectorization and recognition of some cadastral objects. Second Joint European Conference & Exhibition on Geographical Information, Barcelona, Spain.

Bentley Systems, (1995): Microstation Descartes User's Guide.

Cetl, V. (2001): Optičko prepoznavanje znakova i riječi na slici, Seminarski rad, Geodetski fakultet, Zagreb.

Cetl, V. (2002): Transformacije geometrijskih podataka u katastru, Seminarski rad, Geodetski fakultet, Zagreb.

Državna geodetska uprava (2002): Tehničke upute za PREVOĐENJE KATASTARSKIH PLANOVA IZRAĐENIH U GAUSS-KRUEGEROVOJ PROJEKCIJI U DIGITALNI VEKTORSKI OBLIK, Zagreb.

Eikvil, L., Aas, K., Holden, M. (1995): Tools for Automatic Recognition of Character Strings in maps. CAIP '95, 6th International Conference on COMPUTER ANALYSIS OF IMAGES AND PATTERNS, Prague, Czech Republic.

Frančula, N. (1999): Digitalna Kartografija, 2. prošireno izdanje, Geodetski fakultet, Zagreb.

Frangeš, S., Špoljarić, D., Lovrić, P. (1996): Uređivanje skaniranih crno-bijelih tekstualnih predložaka i pisma na karti, Geodetski list, Zagreb, 3, 279-284.

Marinović, Z. (2004): Georeferenciranje katastarskih planova k.o. Vrbanj, Diplomski rad, Geodetski fakultet, Zagreb.

Narodne novine (1999): Zakon o državnoj izmjeri i katastru nekretnina, 128/99.

Punek, I., Kipčić, D. (1999): Vektorizacija i OCR, Tehnički izvještaj, Zagreb.

Reiher, E., Li, Y. (1995): Map Symbol Recognition. Collection scientifique et technique, Centre de recherche informatique de Montreal, Montreal, Canada

Roić, M. (1997): Digitalni katastar, folije s predavanja, Geodetski fakultet, Zagreb.

Shortis, M., R., Trisirisatayawong, I., (1994): Automatic text recognition and the potential for text to feature association on scanned maps. Geomatics Research Australia, No 61, 85-99.

POPIS URL-ova:

URL 1 http://public.srce.hr/1.tehnicka_skola/rac1/p8_osobna_racunala.pdf

URL 2 <http://www.bentley.com/>

ŽIVOTOPIS

EUROPEAN
CURRICULUM VITAE
FORMAT



OSOBNE OBAVIJESTI

Ime	STIPANOVIĆ, Zoran
Adresa	Bana Josipa Jelačića, 3, 47300, Ogulin, Hrvatska
Telefon	047/525-979
Faks	
E-pošta	zorans@geodet.geof.hr
Državljanstvo	Hrvatsko
Datum rođenja	26. 11. 1979.

RADNO ISKUSTVO

- Datum (od – do) [Navedite posebno za svako zaposlenje, počevši od posljednjeg.]
- Naziv i sjedište tvrtke zaposlenja
- Vrsta posla ili područje
- Zanimanje i položaj koji obnaša
- Osnovne aktivnosti i odgovornosti

ŠKOLOVANJE I IZOBRAZBA

- Datum (od – do) 1994.-1998.
- Naziv i vrsta obrazovne ustanove Obrtnička i Tehnička škola Ogulin
- Osnovni predmet /zanimanje Elektrotehnika
- Naslov postignut obrazovanjem Elektrotehničar
- Stupanj nacionalne kvalifikacije SSS
(ako postoji)

OSOBNJE VJEŠTINE I SPOSOBNOSTI

Stečene radom/životom, karijerom, a koje nisu potkrijepljene potvrdama i diplomama.

MATERINSKI JEZIK

HRVATSKI

DRUGI JEZICI

- sposobnost čitanja
- sposobnost pisanja
- sposobnost usmenog izražavanja

ENGLESKI

IZVRSNO

DOBRO

DOBRO

SOCIJALNE VJEŠTINE I
SPOSOBNOSTI

Življenje i rad s drugim ljudima u višekulturnim okolinama gdje je značajna komunikacija, gdje je timski rad osnova (npr. u kulturnim ili sportskim aktivnostima).

[Opisati te vještine i naznačiti gdje su stečene.]

ORGANIZACIJSKE VJEŠTINE I
SPOSOBNOSTI

Npr. koordinacija i upravljanje osobljem, projektima, financijama; na poslu, u dragovoljnom radu (npr. u kulturi i sportu) i kod kuće, itd.

[Opisati te vještine i naznačiti gdje su stečene.]

TEHNIČKE VJEŠTINE I
SPOSOBNOSTI

S računalima, posebnim vrstama opreme, strojeva, itd.

[Opisati te vještine i naznačiti gdje su stečene.]

UMJETNIČKE VJEŠTINE I
SPOSOBNOSTI

Glazba, pisanje, dizajn, itd.

[Opisati te vještine i naznačiti gdje su stečene.]

DRUGE VJEŠTINE I SPOSOBNOSTI
Sposobnosti koje nisu gore navedene.

[Opisati te vještine i naznačiti gdje su stečene.]

VOZAČKA DOZVOLA

ne

DODATNE OBAVIJESTI

[Ovdje navesti druge obavijesti koje mogu biti značajne, npr. osobe za preporuku, za upućivanje, itd.]

DODATCI

[Popiši sve priložene dodatke.]



PRILOZI

1. Radni originali k.o. Grablje

2. Prepoznati brojevi:

- Ispravno prepoznati brojevi katastarskih čestica (crno)
- Pogrešno prepoznati brojevi katastarskih čestica (crveno)

3. Prepoznato nazivlje:

- Ispravno prepoznato nazivlje (crno)
- Pogrešno prepoznato nazivlje (crveno)