

## UREĐIVANJE SKANIRANIH CRNO-BIJELIH TEKSTUALNIH PREDLOŽAKA I PISMA NA KARTI

Stanislav FRANGEŠ, Drago ŠPOLJARIĆ i Paško LOVRIĆ – Zagreb\*

*SAŽETAK. U okviru projekta Hrvatski kartografi skanirani su crno-bijeli tekstualni predlošci naslovnih stranica knjiga, koje najbolje predstavljaju djelatnost određenoga kartografa. Uređivanje takvih predložaka može se izvesti na tri načina: interaktivnom obradom skanirane crno-bijele rasterske slike, ponovnim pisanjem »preko« skanirane rasterske slike i OCR programima. Ukoliko treba urediti veći broj skaniranih crno-bijelih tekstualnih predložaka najprimjerenije je to izvesti OCR programima. Sličan se problem pojavljuje pri uređivanju skanirane karte. Naime, pismo je sastavni dio karte i primjenjuje se s ostalim sredstvima kartografskog izražavanja ili samostalno, pa ga je potrebno urediti nakon skaniranja.*

### 1. UVOD

U okviru projekta *Hrvatski kartografi*, naručitelj kojega je Državna geodetska uprava, pojavila se potreba skaniranja predložaka, točnije naslovnih stranica knjiga, koje bi najbolje predstavile djelatnost određenoga kartografa (Lapaine i dr. 1995).

Kako su naslovne stranice knjiga najčešće crno-bijele, ili ih kopiranjem s crno-bijelim kopirnim uređajem možemo takvima učiniti, ograničili smo se na problem skaniranja crno-bijelih tekstualnih predložaka.

Sličan se problem pojavljuje pri skaniranju karte, gdje se pismo smatra sastavnim dijelom i primjenjuje s ostalim sredstvima kartografskog izražavanja ili samostalno (Lovrić 1988).

### 2. SKANIRANJE CRNO-BIJELIH TEKSTUALNIH PREDLOŽAKA

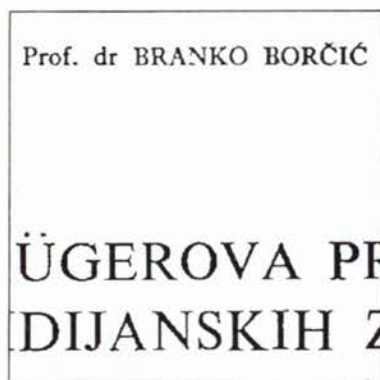
Na raspolaganju smo imali Microtekov stolni skaner ScanMaker II SP i HP Desk Scan II. Skaniranje je izvršeno programom Aldus Photostyler 2.0.

---

\* Mr. sc. Stanislav Frangeš, Drago Špoljarić, dipl. inž. i prof. dr. sc. Paško Lovrić, Geodetski fakultet Sveučilišta u Zagrebu, 10 000 Zagreb, Kačićeva 26.

Kako su tekstualni predlošci za skaniranje (vidi sliku 1a) bili u ovome zadatku crno-bijeli, najoptimalniji je izbor jednobojnog načina skaniranja *line art*, gdje skaner svaki piksel u slikovnoj matrici prepoznaje kao crni ili bijeli.

Nadalje je isti predložak skaniran različitim rezolucijama, a s povećanjem rezolucije rezultati su sve bolji. Pritom treba znati da veličina dokumenta raste s *kvadratom* rezolucije pa na osnovi ispitivanja (Frangješ 1995) predlažemo izbor rezolucije skaniranja od 300 dpi (300 dots per inch – točaka po palcu). Na takvu je odluku utjecalo i vrijeme trajanja skaniranja, koje također raste s povećanjem rezolucije (Vlašić 1995).



Slika 1a. Isječak skaniranog crno-bijelog tekstualnog predloška

### 3.1. Uređivanje interaktivnom obradom rasterske slike

Uređivanje skaniranih crno-bijelih tekstualnih predložaka interaktivnom obradom skanirane rasterske slike može se izvesti svakim programom za obradu rasterske slike. Mi smo koristili program Aldus PhotoStyler 2.0, jer smo njime i skanirali predloške, te jednostavni program Paintbrush koji je sastavni dio svake Windows grupe Accessories.

Uređivanje slova izvodi se pri velikom povećanju, kada je vidljiva pikselna mreža. Zacrnjivanje, odnosno izbjeljivanje željenih piksela postiže se pritiskom na odgovarajuće tipke miša. Također je moguće, nakon selektiranja (izbora) slova, već uređeno slovo kopirati i preseliti na mjesto istog slova koje bi tek trebalo uređivati.

## 3. UREĐIVANJE SKANIRANIH CRNO-BIJELIH TEKSTUALNIH PREDLOŽAKA

Nakon skaniranja trebalo je urediti, tj. grafički popraviti skanirane naslovne stranice knjiga. Pritom mislimo na popravljivanje oštećenih slova, crta i znakova te brisanje nečistoća i mrlja na skaniranim slikama (Frangješ 1995). Postoje tri različita načina uređivanja skaniranih crno-bijelih tekstualnih predložaka:

1. interaktivnom obradom skanirane crno-bijele rasterske slike
2. ponovnim pisanjem »preko« skanirane rasterske slike
3. OCR programima.

Taj je način uređivanja skaniranih crno-bijelih tekstualnih predložaka dugotrajan i zamoran za izvođača pa bi ga trebalo izbjegavati osim ako su potrebni gotovo neznatni ispravci.

### 3.2. Uređivanje ponovnim pisanjem

Drugi je način uređivanja skaniranih crno-bijelih tekstualnih predložaka ponovno pisanje i crtanje »preko« skanirane rasterske slike. Moguće ga je također izvesti svakim programom za obradu rasterske slike koji posjeduje mogućnost pisanja novoga teksta.

Jedan od razloga za izbor programa kojim ćemo izvesti postavljenu zadaću su *fontovi* primijenjeni na predlošcima. Tu se nametnuo CorelDRAW svojim doista velikim izborom fontova, ali i velikim mogućnostima variranja slova i crtanja.

Da bi ponovno napisan tekst i nacrtan crtež bili što sličniji predlošku može se skanirana rasterska slika staviti kao podloga, a nakon završenog pisanja i crtanja izbrisati.

Prednost je tog načina u bržem rješavanju postavljene zadaće, osobito ako je mnogo oštećenja i nečistoća na skaniranoj rasterskoj slici, dokumenti su manji a predlošci kvalitetnije uređeni. Nedostatak je moguća nepodudarnost novonapisanog dokumenta prema predlošku, koja može nastati zbog razlika u fontovima kojima raspolažemo i zbog proizvoljnog smještanja teksta.

### 3.3. Uređivanje OCR programima

*Optical Character Recognition (OCR)* je tehnika prepoznavanja pisma iz skaniranog rasterskog predloška i postupak konvertiranja u znakove – slova, brojeve – te pohranjivanje u neki od raspoloživih tekstualnih formata.

Na našem tržištu postoji nekoliko OCR programa. Testirali smo tri najpoznatija: OmniPage Professional 5.02, Recognita 2.0A (profesionalna inačica koja u sebi uključuje samo drivere za HP skanere) i CorelOCR-Trace 6.0 kao dio programskog paketa CorelDraw.

Prije početka prepoznavanja teksta potrebno je postaviti neke dodatne parametre. Osim određivanja područja (zone) prepoznavanja teksta, dobro je definirati kvalitetu teksta; npr. tekst otisnut na laserskom pisaču, ili na 9-igličnom pisaču, ili faksirani dokument itd. Također je potrebno postaviti veličinu i orijentaciju stranice, jezik na kojem je pisan dokument, ispravnu kodnu stranicu (ovisno o tekstu), te format u kojem će se pohraniti prepoznani tekst. Valja naglasiti da za DOS tekstne procesore treba primjenjivati *Code page 852*, a za Windows programe *East. Ea. ANSI* (drugo ime za Microsoftovu kodnu stranicu 1250 pod Windows CE (EE)).

Sva tri testirana OCR programa imaju standardnu i profesionalnu verziju. Profesionalna inačica ima sposobnost »uvježbavanja i učenja« nestandardnih fontova.

Prvi predložak, koji je poslužio za određivanje kvalitete prepoznavanja navedenih OCR programa, sadržajno nije posebno složen (vidi sliku 1a). Stoga i ova analiza ne daje detaljne i višenamjenske rezultate.

Sva su tri navedena OCR programa podjednako brza i točna. Velika prednost Recognite pred ostalim OCR programima je podrška prepoznavanja teksta najvećeg broja poznatih jezika (svi europski među kojima je i hrvatski) (vidi sliku 1b).



Slika 1b. Isječak skaniranog crno-bijelog tekstualnog predloška ureden OCR programom Recognita

#### 4. UREĐIVANJE SKANIRANOG PISMA NA KARTI

Nakon skaniranja karte vrlo je značajno uređivanje pisma na karti što se izvodi nakon skaniranja popravljanjem oštećenih slova, brisanjem mrlja itd. To je osobito važno pri izradbi faksimila karata skanerom (Lovrić i dr. 1991).

Za primjer skanirana je karta Ivanec i okolica 1900. u mjerilu 1:25 000 (Lovrić i dr. 1966) (vidi sliku 2a). Kako je već u trećem odjeljku navedeno, uređivanje pisma moguće je izvesti na tri načina, pa je to u ovome primjeru i provedeno.



Slika 2a. Isječak skanirane karte Ivanec i okolica 1900. u mjerilu 1:25 000

Uređivanje interaktivnom obradom rasterske slike vrlo je dugotrajno, a u nekim slučajevima čak nemoguće, jer izvodač ne može prepoznati točan oblik slova i odvojiti ga od ostalih susjednih signatura i linija. Dokumenti su golemi, a kvaliteta urednog pisma na karti neprimjerena zahtjevima današnje kartografije.

Uređivanje skaniranog pisma na karti ponovnim pisanjem »preko« skanirane rasterske slike također je teško izvedivo. Prethodno je nužno vektoriziranje

rasterske slike, zatim selektiranje pisma i formiranje sloja u kojemu će biti samo pismo. Ponovno napisani nazivi moraju biti što sličniji onima na predlošku. Teškoće se javljaju zbog razlike u fontovima kojima raspolažemo i zbog proizvoljnog smještaja naziva, što je vidljivo kada se preko sloja u kojemu je samo pismo preklopi sloj s ostalim sadržajem karte.

Uređivanje skaniranog pisma na karti OCR programima također zahtijeva spomenute predradnje. Naime, primjenom OCR programa na čitav sadržaj karte dobivaju se nesuvisli rezultati. Primjenom OCR programa na pismo na karti koje je u zasebnom sloju, dobiveni su najbolji rezultati. Do izražaja dolaze automatsko smještanje naziva i bitno veća brzina postupka, a predložak je vrlo kvalitetno ureden (vidi sliku 2b).



Slika 2b. Isječak skanirane karte Ivanec i okolica 1900. ureden OCR programom Recognita

## 5. ZAKLJUČAK

Uređivanje skaniranih crno-bijelih tekstualnih predložaka može se izvesti na tri načina: interaktivnom obradom skanirane crno-bijele rasterske slike, ponovnim pisanjem »preko« skanirane rasterske slike i OCR programima.

Prvi navedeni način uređivanja skaniranih crno-bijelih tekstualnih predložaka dugotrajan je i zamoran za izvodača, dokumenti su veliki, a kvaliteta uredenih predložaka nije optimalna. Drugim načinom ostvaruju se, za kraće vrijeme, kvalitetnije uredeni predlošci a dokumenti su manji. Prednosti uređivanja skaniranih tekstualnih predložaka OCR programima u odnosu na prethodna dva načina su automatsko smještanje teksta i bitno veća brzina postupka, pogotovo ako je predložak složen. Otežavajuće su okolnosti visoki hardverski zahtjevi i za naše prilike visoka cijena OCR programa. Prema tome, ukoliko je potrebno urediti veći broj skaniranih crno-bijelih tekstualnih predložaka najprimjerenije je to izvesti OCR programima.

Istu je tehniku najbolje primijeniti pri uređivanju skaniranog pisma na karti. Svaki složeniji kartografski orijentiran program trebao bi u svom paketu podržavati neki OCR program kojime bi se uređivalo skanirano pismo na karti.

## ZAHVALA

Posebno zahvaljujemo Odjelu za hidrotehniku Građevinskog fakulteta Sveučilišta u Zagrebu što nam je omogućio rad sa stolnim skanerom HP Desk Scan II i primjenu OCR programa Recognita 2.0A.

## LITERATURA

- Frangeš, S. (1995): Skaniranje i uređivanje skaniranih crnobijelih tekstualnih predložaka za projekt Hrvatski kartografi – Tehničko izvješće. Sveučilište u Zagrebu, Geodetski fakultet, Zagreb.
- Lapaine, M., Lovrić, P., Frančula, N., Frangeš, S. i Vučetić, N. (1995): Hrvatski kartografi – Idejni projekt. Sveučilište u Zagrebu, Geodetski fakultet, Zagreb.
- Lovrić, P. (1988): Opća kartografija. Sveučilišna naklada Liber, Zagreb.
- Lovrić, P., Frangeš, S. i Babić, B. (1991): Izrada faksimila karata elektronskim odvajanjem boja. Savjetovanje: Nove tehnologije u geodeziji. Zbornik radova, Neum, 263–271.
- Lovrić, P., Putar, E. i Frangeš, S. (urednici) (1996): Faksimil karte Ivanec i okolica 1900. Općina Ivanec 1996.
- Vlašić, K. (1995): Priručnik o skeniranju. Kristal print, Zagreb.

## EDITING OF SCANNED WHITE AND BLACK TEXTUAL IMAGES AND LETTER ON A MAP

Within the scope of the project *Croatian Cartographers*, the black and white images of the cover pages which represent the activity of a certain cartographer in the best possible way, have been scanned. Editing of such images can be done in three ways: by interactive processing of the scanned black and white raster image, by repeated writing »over« the scanned raster image and OCR programmes. If it is necessary to edit a larger number of scanned white and black textual images, it can be done in a most convenient way by means of OCR programmes. Similar problem appears while editing a scanned map. Namely, the letters are an integral part of a map and are applied with other means of cartographic expression or independently, so it must be edited after scanning.

Primljeno: 1960–06–26