

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 5187

# **Ekstrakcija tablica na skeniranim dokumentima**

Kristijan Vulinović

Zagreb, ožujak 2017.

*Umjesto ove stranice umetnite izvornik Vašeg rada.*  
*Kako biste uklonili ovu stranicu, obrišite naredbu \izvornik.*

*Zahvala... to be written...*

# SADRŽAJ

<b>1. Uvod</b>	<b>1</b>
<b>2. Binarizacija slike</b>	<b>2</b>
<b>3. Zaključak</b>	<b>3</b>
<b>Literatura</b>	<b>4</b>

# 1. Uvod

U današnje vrijeme postoje izuzetno velike količine papirnatih dokumenata. Samo u Sjedinjenim Američkim Državama nastaje više od milijarde novih papirnatih dokumenata svakog radnog dana. Mogućnost digitalizacije takvih dokumenata može biti od velike koristi prilikom pohrane, slanja ili pretraživanja istih. [5] Digitalizaciju dokumenata možemo podijeliti u dva dijela: prepoznavanje teksta te prepoznavanje grafičkih objekata. [1] Za prepoznavanje teksta je dostupan velik broj alata koji omogućuju optičko prepoznavanje znakova (engl. *optical character recognition*). Prepoznavanje grafičkih objekata dokumenta mnogo je manje zastupljeno u odnosu na prepoznavanje teksta. U to se svrstavaju linije, oblici, slike, simboli, tablice i razni drugi objekti koji se mogu nalaziti na skeniranim dokumentima. Ovaj rad se fokusira isključivo na prepoznavanje tablica, što je prethodno već opisano u radovima poput [1] i [2]. Taj postupak se dijeli na prepoznavanje položaja tablice u odnosu na ostatak dokumenta, prilikom čega je potrebno u dokumentu izdvojiti tablicu od ostatka teksta i ostalih grafičkih objekata, a što je opisano u radu [3]. Nakon što je tablica pronađena određuje se njezin izgled, odnosno broj redaka i stupaca, odnosno koordinate pojedine ćelije, a što je detaljnije opisano u nastavku rada.

Predstavljeno rješenje počinje od slike u sivim tonovima (engl. *gray-scale*), koja se binarizira kako bi se dobila slika koja se sastoji od isključivo crne i bijele boje. Dobivena crno-bijela slika koristi se u daljnjoj obradi te se provjerava je li slika rotirana, odnosno kut rotacije iste, nakon čega se rotacija ispravlja. Ovako obrađena slika koristi se dalje za detekciju tablica, **što će biti objašnjeno jednom kada odlučim kako to radim.**

## **2. Binarizacija slike**

Knjiga [4] opisuje kako koristiti adaptivni wiener filter prema nekim tamo formulama koje su krive u drugim mjestima.

### **3. Zaključak**

Zaključak.

# LITERATURA

- [1] S. Deivalakshmi, K. Chaitanya, i P. Palanisamy. Detection of table structure and content extraction from scanned documents. U *Communications and Signal Processing (ICCSP)*, stranice 270–274. IEEE, apr 2014.
- [2] Basilios Gatos, Dimitrios Danatsas, Ioannis Pratikakis, i Stavros J. Perantonis. Automatic table detection in document images. U *International Journal of Document Analysis*, svezak 8, stranica 172–182, aug 2005.
- [3] Jianying Hu, Ram Kashi, Daniel Lopresti, i Gordon Wilfong. Medium-independent table detection. *Proceedings of SPIE - The International Society for Optical Engineering*, dec 1999.
- [4] Lim J.S. *Two-Dimensional Signal and Image Processing*. PH, 1989. ISBN 0139353224,9780139353222.
- [5] Peng-Yeng Yin. Skew detection and block classification of printed documents. *Image and Vision Computing*, 19(8):567–579, may 2001.



## **Ekstrakcija tablica na skeniranim dokumentima**

### **Sažetak**

Sažetak na hrvatskom jeziku.

**Ključne riječi:** Ključne riječi, odvojene zarezima.

## **Table Extraction on Scanned Documents**

### **Abstract**

Abstract.

**Keywords:** Keywords.