

# Jedna implementacija SVM u CBIR sistemu

Slobodan Čabarkapa<sup>1</sup>, Nenad Kojić<sup>1</sup>, Vladan Radosavljević<sup>2</sup>, Branimir Reljin<sup>3</sup>, Senior Member, IEEE

**Sadržaj** — U radu je prikazana implementacija Support Vector Machine (SVM) u sistem za pretragu slika (CBIR = Content-Based Image Retrieval). Cilj je bio poređenje induktivnog i transduktivnog metoda učenja u SVM. U ranijem radu smo koristili polinomijalni kernel a ovde je razmatran Gausov kernel. Pokazano je da transduktivni metod, koji koristi kako označene (od strane korisnika) tako i neoznačene slike iz baze, daje bolje rezultate od induktivnog metoda. Opisani sistemi su testirani na bazi slika Corel 1K.

**Ključne reči** — CBIR, pretraga slika, SVM, induktivna i transduktivna metoda učenja.

## I. UVOD

**S**a sve većim brojem korisnika računara i Interneta, multimедijalni sadržaji postaju sastavni deo naših života. Zbog potreba tržišta, sve veći broj kompanija svoje proizvode kreira ili reklamira kroz ovakave sadržaje, pa se ukazuje velika potreba za njihovo klasifikovanje, skladištenje i pretragu. Najkompleksniji zadatak je, svakako, kvalitetno pretraživanje sadržaja, pa je i veliki broj radova direktno usmeren ka ovoj temi. Pored pretrage slika [1-5] kojom se bavi veći broj radova, pretraga audio [6] i video [7] sadržaja je, takođe, predmet intenzivnog proučavanja.

Sistem za pretragu slika treba da pronade slike iz baze koje najviše odgovaraju upitnoj slici. Prvi pokušaji pretrage slika bili su bazirani na tekstu. Svakoj od slika bile bi dodeljene karakteristične reči koje opisuju sadržaj slike, pa se pretraga svodila na traženje reči koje su najsličnije upitnoj. Ovakvi sistemi mogu dati dobre rezultate, ali se pokazuju kao lošiji kod velikih baza podataka, i skoro neupotrebljivi za različita govorna područja. Kao alternative ovakvom pristupu pojavljuju se *Content-Based Image Retrieval* (CBIR) sistemi [1-3]. Kod njih se svaka od slika predstavlja obeležjima niskog nivoa (boja, tekstura, oblik, ...), a pretraga se bazira na traženju najkraćeg rastojanja između vektora obeležja upitne slike i slika iz baze. Međutim, objektivna sličnost slika se, često, razlikuje od subjektivnog utiska. Kvalitet pretrage se može

popraviti asistencijom korisnika. Slike dobijene objektivnom metodom se prikažu korisniku na ocenu: označe se kao relevantne one koje subjektivno odgovaraju upitu, dok će neoznačene biti nerelevantne. Označene slike se koriste za korekciju vektora obeležja saglasno subjektivnoj oceni. Asistencija korisnika (RF = relevance feedback) u kombinaciji sa CBIR sistemom značajno otklanja razlike između objektivnog sadržaja i percepcije čoveka [4-5].

U našim prethodnim radovima kao inteligentna logika za klasifikaciju relevantnih i nerelevantnih slika korišćena je neuralna mreža radikalne osnove [8]. Radi ubrzanja postupka inicijalne pretrage, kao moguće poboljšanje predložen je proces adaptivne klasterizacije [9], kada su slike grupisane u klastere po njihovim zajedničkim karakteristikama, i svaki klasa je opisan jednim karakterističnim vektorom obeležja. U inicijalnoj pretrazi se upit poredi sa predstavnicima klastera. Druga mogućnost ubrzanja pretrage je u redukciji vektora obeležja. U radu [10] je pokazano da se brzina rada može mnogostruko povećati uz minimalnu degradaciju kvaliteta pretrage ukoliko se koristi svega oko 10% komponenata potpunog vektora obeležja. Ujedno, čak je postignuto bolje inicijalno pretraživanje u odnosu na korišćenje punog vektora obeležja, jer je redukcijom ostvaren bolji balans između komponenata koje opisuju boju i teksturu.

Umesto neuralne mreže, pri klasifikaciji se može koristiti SVM. U radu [11] je izvršena implementacija SVM u CBIR sistemu i izvršeno je poređenje induktivne i transduktivne metode učenja, za slučaj kada je kernel bio polinomijalnog tipa. U ovom radu je kao kernel korišćen Gausov tip kornela. Akcenat je na transduktivnom metodu (TSVM) učenja [12-13] jer se kod njega u proces odluke uključuju i slike iz baze koje nisu ocenjene od strane korisnika, a kojih je uvek daleko više nego označenih.

U drugom poglavljju je ukratko opisan SVM i osnovne metode obuke. Treće poglavљje opisuje CBIR sistem sa SVM, dok su u četvrom prikazani dobijeni rezultati. Zaključci su navedeni u petom poglavljju.

## II. KLASIFIKATOR TIPO SUPPORT VECTOR MACHINE

Princip SVM u problemima klasifikacije uveden je pre desetak godina [14] i pokazao je veoma dobre rezultate, koji su slični ili bolji od onih koji se dobijaju neuralnim mrežama. U obučavanju SVM se koriste induktivni i transduktivni metod, koje ćemo ukratko opisati.

### 1. Induktivni metod učenja kod SVM

Princip rada SVM može se prikazati grafički kao na slici 1, ako se zamisli trening skup linearno razdvojivih  $N$

<sup>1</sup>Slobodan K. Čabarkapa i Nenad S. Kojić su sa Visoke škole strukovnih studija za informacione i komunikacione tehnologije, Zdravka Čelara 16, 11000 Beograd, Srbija, E-mail: [slobodan.cabarkapa@ict.edu.rs](mailto:slobodan.cabarkapa@ict.edu.rs), [nenad.kojic@ict.edu.rs](mailto:nenad.kojic@ict.edu.rs).

<sup>2</sup>Vladan B. Radosavljević je sa Temple University, 1801 N. Broad Street, Philadelphia, PA19122, E-mail: [vladan@ist.temple.edu](mailto:vladan@ist.temple.edu).

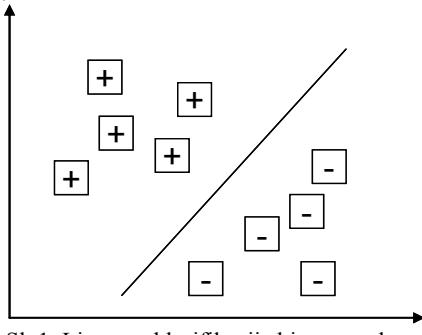
<sup>3</sup>Branimir D. Reljin je sa Elektrotehničkog fakulteta, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11000 Beograd, Srbija, E-mail: [reljinb@etf.bg.ac.yu](mailto:reljinb@etf.bg.ac.yu)

Istraživanja obavljena u ovom radu su delimično podržana od strane Ministarstva nauke i tehnološkog razvoja Republike Srbije, projekat 145096, i EU projekta COST Action 292

elemenata skupa  $Q$ . Neka su elementi skupa  $Q$  označeni binarni elementi  $\pm 1$ , definisani sa

$$\{(\mathbf{x}_i, y_i), y_i = \pm 1, i=1\dots N\}. \quad (1)$$

Ako se ovaj princip primeni na klasifikaciju slika, tada bi  $\mathbf{x}_i$  bio vektor obeležja  $i$ -te slike, a  $y_i$  opisuje da li ta slika pripada jednom (+) ili drugom (-) podskupu skupa  $Q$  (npr. skupu relevantnih ili nerelevantnih slika). Uloga SVM je da pronađe "krivu" kojom se ova dva podskupa mogu razdvojiti, slika 1.

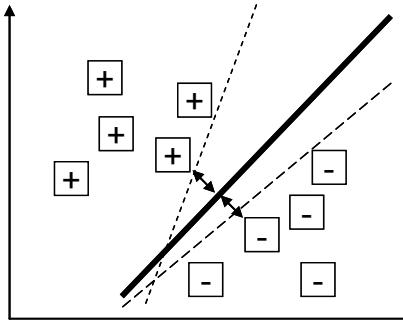


Sl. 1. Linearna klasifikacija binarnog skupa

Jasno je da se ova klasifikacija može uraditi na mnogo načina. To znači da postoje promenljive  $w$  i  $b$  kojima se može izvršiti linearna klasifikacija tipa

$$f(x) = \mathbf{w}^T \mathbf{x} + b \quad (2)$$

što je ilustrovano na slici 2.



Sl. 2. Induktivni metod za SVM: Moguće linearne klasifikacije (isprekidane linije); Klasifikacija sa maksimalnom marginom između elemenata binarnog skupa (puna linija).

Za slučaj na slici 2, potrebno je definisati koja od mogućih klasifikacija (isprekidane linije) je najbolja. U tom cilju definiše se margina koja treba da bude što veća, tj. da rastojanje linije razdvajanja bude maksimalno udaljeno od dva podskupa. Najbolje razdvajanje tada je kroz sredinu razmaka što je, za linearno razdvojiv skup, prikazano punom linijom na slici 2.

Matematički gledano, SVM treba da pronađe parametre vektora  $w$  i promenljive  $b$  tako da optimizuje funkciju

$$\frac{1}{2} \|\mathbf{w}\|^2$$

uz uslov:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i. \quad (3)$$

Ovo rešenje treba da pronađe globalni minimum u problemu koji je definisan redom složenosti  $O(N^3)$  [14]. Međutim, ako elementi skupa sadrže šum, problem neće biti linearno razdvojiv. U tom slučaju mogu se uvesti takozvane "slack" promenljive, čime se "kažnjavaju"

elementi skupa koji nisu ispravno klasifikovani [14].

Formalno, treba minimizovati parametare vektora  $w$  i  $b$  u funkciji

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i^2$$

uz uslov:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \quad (4)$$

gde parameter  $C > 0$  definiše opseg "kažnjavanja".

Da bi se problem rešio, često se pribegava transformaciji originalnog u neki drugi geometrijski prostor uz moguću promenu broja obeležja. Da bi se ove transformacije programski brzo izračunale uvodi se odgovarajući *kernel* kojim se to postiže [14].

## 2. Transduktivni metod učenja kod SVM

Za razliku od induktivnog metoda, kod transduktivne metode se trening skupu dodaje još  $K$  neoznačenih elemenata skupa  $Q$ . Ako se radi o klasifikaciji slika tada je, obično,  $K \gg N$ . Kako su novi elementi neoznačeni, cilj je predvideti njihov status što je bolje moguće. Ovo podrazumeva da greška obuke, u odnosu na induktivni metod, bude što manja. Obuka se sprovodi tako što se izostavi jedan od elemenata skupa, obuka izvrši preostalim elementima, pa se na izostavljenom elementu testira postupak [15]. Da bi se to ostvarilo potrebno je optimizovati parametare  $w$ ,  $b$  i  $y_j^*$  u relaciji

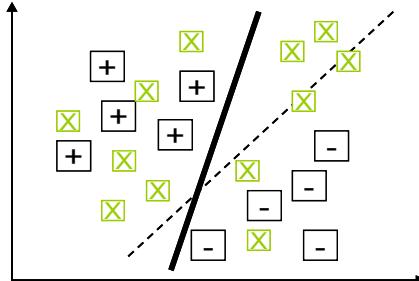
$$\frac{1}{2} \|\mathbf{w}\|^2$$

uz uslov:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i$$

$$y_j^*(\mathbf{w}^T \mathbf{x}_j^* + b) \geq 1, \quad \forall j. \quad (6)$$

Da bi se ovaj problem rešio potrebno je izvršiti označavanje  $y_j^*$  elemenata test skupa, podešavanjem parametara  $w$  i  $b$  tako da se nakon razdvajanja u hiperprostoru postigne maksimalna margina i za trening i za test elemente. Ovo je grafički prikazano na slici 3, gde je razdvajanje sa maksimalnom marginom označeno punom linijom.



Sl. 3. Transduktivni metod za SVM: maksimalna margina kod induktivne metode (isprekidana linija); maksimalna margina nakon uvođenja i neoznačenih elemenata X (puna linija)

Na slici 3 se vidi da je uvođenjem neoznačenih podataka (X) došlo do promene orijentacije krive razdvajanja. I kod transduktivne metode se mogu uvesti "slack" promenljive na sličan način kao i kod induktivne, za slučaj pojave šuma ili za elemente koji su nerazdvojivi, kroz minimizaciju parametara  $w$ ,  $b$  i  $y_j^*$  u funkciji

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i^2 + C^* \sum_j \xi_j^{*2}$$

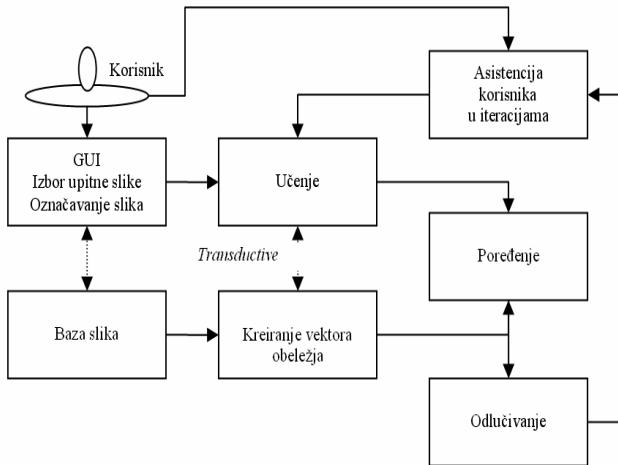
uz uslove:

$$\begin{aligned} y_i(\mathbf{w}^T \mathbf{x}_i + b) &\geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \\ y_j^*(\mathbf{w}^T \mathbf{x}_j^* + b) &\geq 1 - \xi_j^*, \quad \xi_j^* \geq 0, \quad \forall j \end{aligned} \quad (7)$$

gde su  $C$  i  $C^*$  parametri definisani od strane korisnika kojima se balansira između pogrešno klasifikovanih elemenata trening skupa i "isključenih" elemenata za potrebe testa.

### III. CBIR SISTEM SA SVM

Struktura CBIR sistema uz asistenciju korisnika je prikazana na slici 4. Korisnik bira upitnu sliku i očekuje od sistema da mu ponudi grupu slika iz baze koje su joj objektivno (po obeležjima niskog nivoa) najsličnije. Od ponudjenih  $U$  slika korisnik označava njih  $R$  kao relevantne dok je ostalih  $I=U-R$  automatski označeno kao nerelevantne. Na osnovu ovih označavanja SVM se obučava i korisniku u narednoj iteraciji prikazuje grupu slika koje su prethodno označene kao relevantne, i novu grupu dobijenu nakon klasifikacije pomoću SVM. Kako je izabrano da se na ekranu prikazuje  $U=20$  slika, proces se završava kada su svih 20 slika označene kao relevantne.



Sl. 4. CBIR sistem sa asistencijom korisnika.

### IV. REZULTATI

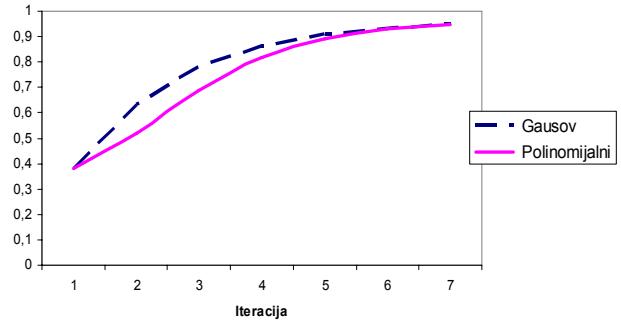
U radu je korišćena Corel baza [16] sa 1000 slika raspoređenih u 10 klasa. Testiranje je izvršeno sa transduktivnom i indiktivnom metodom učenja gde je korišćen kernel Gausovog tipa, za određivanje rastojanja dve slike, sa vektorima obeležja  $\mathbf{x}$  i  $\mathbf{y}$ , definisanim sa:

$$d(\mathbf{x}, \mathbf{y}) = \exp(-(\mathbf{x}-\mathbf{y})^2 / \sigma) \quad (8)$$

Vrednosti standardne devijacije,  $\sigma$ , korišćene u ovom radu su 1, 5 i 9, dok je parametar  $C=1$ . Meren je parametar efikasnosti koji je definisan kao odnos broja dobijenih relevantnih slika i broja prikazanih,  $U=20$ :  $Eff=R/U$ , za svaku od iteracija [11]. Nakon toga, usrednjavanjem se dobija srednja efikasnost,  $Eff_{sr}$ , kojom se opisuju kvalitet i brzina pronalaženja svih  $R=U=20$  relevantnih slika.

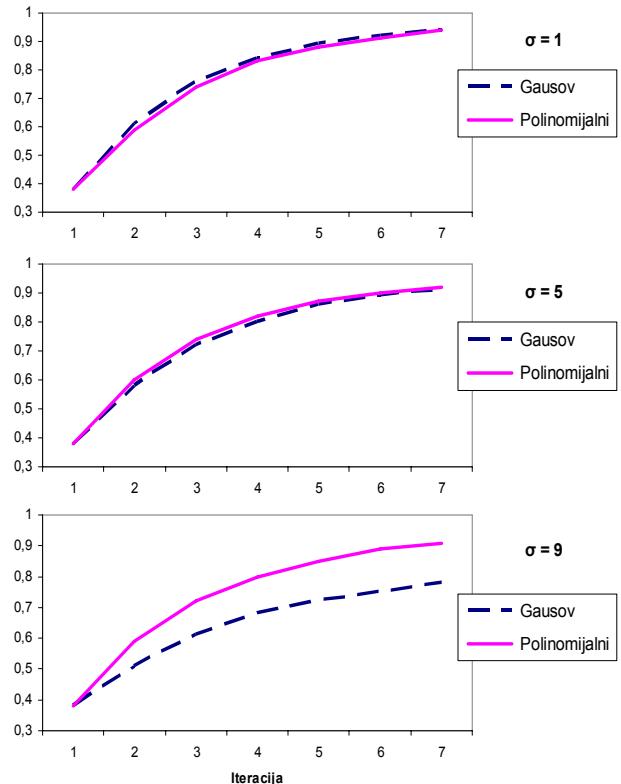
Poređenje srednje efikasnosti po iteraciji za induktivnu

metodu sa kernelom polinomijalnog tipa i redom polinoma  $p=1$  [11], u odnosu na kernel Gausovog tipa kod koga je standardna devijacija  $\sigma=1$ , prikazano je na slici 5.



Sl. 5. Poređenje srednje efikasnosti po iteraciji za induktivnu metodu za različite tipove kernela.

Rezultati pokazuju da induktivna metoda sa kernelom Gausovog tipa daje bolje rezultate za manji broj iteracija. Ukoliko je broj iteracija veći, rezultati ne zavise od tipa upotrebljenog kernela. Ako se parametar  $p=\sigma$  povećava, induktivna metoda sa kernelom polinomijalnog tipa ima veću efikasnost za veći broj iteracija, slika 6.



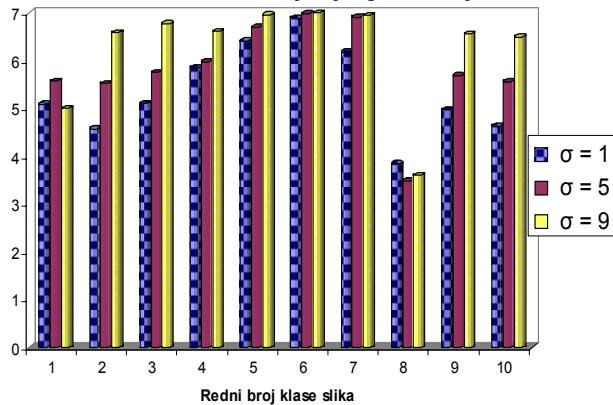
Sl. 6. Poredjenje srednje efikasnosti po iteraciji za transduktivnu metodu, za Gausov i polinomijalni kernel, pri različitim parametrima:  $p=\sigma=1, 5$  i  $9$ .

Rezultati pokazuju da transduktivna metoda sa kernelom Gausovog tipa ima veću srednju efikasnost po iteraciji u odnosu na kernel polinomijalnog tipa, ukoliko je parametar  $p=\sigma=1$ .

Za standardne devijacije  $\sigma=5$  i  $\sigma=9$  kernel Gausovog tipa za svaku iteraciju ima manju srednju efikasnost, tj. manji broj relevantnih slika po iteraciji.

Srednji broj iteracija za transduktivnu SVM, do ostvarenja potpune tačnosti (kada su svih  $U=20$  slika označene kao relevantne), za 10 klasa COREL slika, u

zavisnosti od standardne devijacije, prikazan je na slici 7.



Sl. 7. Srednji broj iteracija do ostvarenja potpune tačnosti, za transduktivnu SVM metodu.

## V. ZAKLJUČAK

U radu su prikazani rezultati implementacije SVM u CBIR sistem uz asistenciju korisnika. Izvršeno je poređenje rezultata nakon induktivne i transduktivne metode učenja, a za dva različita tipa kernela: Gausov i polinomijalni. Pokazano je da je u većini slučajeva transduktivna metoda, koja koristi i označene i neoznačene slike, bolja sa aspekta pronalaženja broja relevantnih slika po iteracijama. Sa aspekta kernela, dobiveni su bolji rezultati za kernel Gausovog tipa, kada standardna devijacija ima malu vrednost ( $\sigma=1$ ). Za veće standarde devijacije ( $\sigma=9$ ) bolje je koristiti kernel polinomijalnog tipa.

## LITERATURA

- [1] A. Pentland, R. Picard and S. Sclaroff, "Photobook: Tools for content-based manipulation of image databases", in *Proc SPIE Conf. Storage and Retrieval of Still Image and Video Databases II*, pp 34-47, San Jose, CA, USA, 1994.
- [2] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang and B. Dom, "Query by image and video content: The QBIC system", *IEEE Computer*, vol. 28, no. 9, pp. 23–32, 1995.
- [3] W. Y. Ma and B. S. Manjunath, "NETRA: A toolbox for navigating large image databases", in *Proc. IEEE Int. Conf. on Image Processing (ICIP '97)*, vol. 1, pp. 568–571, Santa Barbara, CA, USA, October 1997.
- [4] Y. Rui, T. Huang and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS", in *Proc. IEEE Int. Conf. on Image Processing (ICIP '97)*, vol. 2, pp. 815–818, Santa Barbara, CA, USA, October 1997.
- [5] G. Aggarwal, T. Ashwin and S. Ghosal, "An image retrieval system with automatic query modification", *IEEE Trans. Multimedia*, vol. 4, pp. 201-214, June 2002.
- [6] V. Đorđević, N. Reljin, and I. Reljin, "Identifying and retrieving of audio sequences by using wavelet descriptors and neural network with user's assistance", in *Proc. Conf. EUROCON-2005*, pp. 167-170, Belgrade, 21-24 Nov., 2005.
- [7] J. Calic, N. Campbell, A. Calway, M. Mirmehdi, T. Burghardt, S. Hannuna, C. Kong, S. Porter, N. Canagarajah and D. Bull, "Towards intelligent content based retrieval of wildlife videos", in *Proc. 6th Int. Workshop WIAMIS-05*, EPFL, Montreux, Switzerland, April 2005.
- [8] S. Čabarkapa, N. Kojić, V. Radosavljević, G. Zajić, Branimir Reljin, "Adaptive content-based image retrieval with relevance feedback", in *Proc. Conf. EUROCON-2005*, pp. 147-150, Belgrade, 21-24 Nov., 2005.
- [9] V. Radosavljević, N. Kojić, S. Čabarkapa, G. Zajić, I. Reljin and B. Reljin, "An image retrieval system with user's relevance feedback", in *Proc. Conf. WIAMIS-2006*, pp. 9-12, Incheon, Korea, 19-21 April, 2006.
- [10] G. Zajić, N. Kojić, V. Radosavljević, et al., "Accelerating of image retrieval in CBIR system with relevance feedback," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 62678, 13 pages, 2007.
- [11] V. Radosavljević, N. Kojić, G. Zajić, B. Reljin, "The use of unlabeled data in image retrieval with relevance feedback", in *Proc. 9th Symp. NEUREL 2008*, pp. 21-26, Belgrade, 25-27 Sept., 2008..
- [12] Z.-H. Zhou, "Learning with unlabeled data and its application to image retrieval," In: *Proceedings 9th Pacific Rim International Conference on Artificial Intelligence (PRICAI'06)*, pp. 5-10, Guilin, China, LNAI 4099, 2006.
- [13] T. Joachims, "Transductive inference for text classification using support vector machines," In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 200-209, San Francisco, US, 1999.
- [14] V. Vapnik. *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.
- [15] T. Joachims, "Transductive learning via spectral graph partitioning," in *Proc. Int. Conf. on Machine Learning (ICML)*, 2003.
- [16] Corel Gallery Magic 65000 (1999), [www.corel.com](http://www.corel.com).

## ABSTRACT

The paper describes an implementation of Support Vector Machine (SVM) in content-based image retrieval (CBIR) system. The main objective is to compare inductive and transductive learning methods in SVM. In our previous work we explored polynomial kernels while in this paper we consider Gaussian kernel as well. It was shown that transductive method, which uses both labeled (annotated by the user) and unlabeled images available in the database, is better than inductive method. The presented systems are tested over images from the Corel 1K dataset.

## AN IMPLEMENTATION OF SVM IN CBIR SYSTEM

Slobodan Čabarkapa, Nenad Kojić, Vladan Radosavljević, Branimir Reljin