

INDEKSIRANJE I PRETRAŽIVANJE MULTIMEDIJALNOG MATERIJALA

Branimir Reljin
Elektrotehnički fakultet Univerziteta u Beogradu

Sadržaj: Nagli razvoj elektronike, telekomunikacija i računara, koji je kulminirao eksplozijom Interneta, doveo je do ogromne količine multimedijalnog materijala. Označavanje i klasifikacija tog materijala, što se naziva indeksiranjem, kao i efikasno pretraživanje audio-vizuelnih sadržaja, predstavlja veliki problem kako profesionalnim organizacijama, tako i korisnicima. Veliki broj naučnih i stručnih timova iz raznih organizacija bavi se ovom problematikom – pored ostalog to je i tema evropskog projekta COST 292 u kojem učestvuje 18 zemalja Evrope, među njima i Srbija i Crna Gora. U ovom radu su, ukratko, izloženi osnovni postupci pretraživanja zasnovani na sadržaju scene (content-based retrieval) uz intervenciju korisnika (user's relevance feedback). Kroz nekoliko primera je ilustrovana efikasnost takvog postupka pretraživanja.

Ključne reči: Značajna obeležja slika, vektor obeležja, indeksiranje, semantičko označavanje, pretraživanje baze slika, subjektivna percepcija, povratna sprega korisnika, neuralne mreže radijalne osnove.

1. Uvod

Eksplodirajući razvoj multimedijalne tehnologije i mogućnost razmene audio-vizuelnih sadržaja preko Interneta, izazvali su značajno povećanje količine multimedijalnih podataka – kako broja dokumenata, tako i veličine pojedinačnih fajlova. To je, dalje, postavilo niz različitih i veoma teških zahteva timovima istraživača širom sveta. Pojavila se potreba za efikasnim tehnikama kompresije, kao i za efikasnim postupcima označavanja, pretraživanja i pronalaženja željenih multimedijalnih dokumenata.

Tehnike kompresije se, generalno, mogu podeliti u dve grupe: na kompresiju bez gubitaka (*lossless*) i kompresiju sa gubicima (*loosy*). Metode kompresije bez gubitaka omogućavaju perfektnu rekonstrukciju slike: nakon kompresije pa, zatim, dekompresije, dobijena slika je identična originalu. Takav vid kompresije je u nekim slučajevima jedino dozvoljen (na primer, u medicini, za slike namenjene dijagnostici, gde se ne smeju izgubiti nikakvi detalji, niti se dozvoljava formiranje artefakata). Za multimedijalan saobraćaj se koriste metode kompresije sa gubicima, kao što su JPEG, MPEG-x, H.264x,

s obzirom da dopuštaju znatno veći stepen kompresije (do oko 100:1, i više), bez subjektivno vidljive degradacije slika.

Konvencionalno pretraživanje baze podataka zasnovano je na tekstualnom upitu na osnovu ključnih reči. Međutim, kod pretraživanja multimedijalnog materijala takav pristup ne daje uvek zadovoljavajuće rešenje. Naime, subjektivni doživljaj istog pojma se razlikuje, zavisno od korisnika. Na primer, pod istim pojmom "kuća" neko misli na kuću u planini, neko na kuću na moru ili u selu, zatim, neko misli na kuću sive boje i od čvrstog materijala, neko na brvnaru, itd. Nadalje, slike u bazi podataka, često, nisu ni označene tekstualnim opisom ili korisnik ne poznaje konkretan rečnik koji je pridružen datoj bazi slika, pa ne koristi ključne reči koje prepoznaje taj rečnik. Stoga poseban značaj ima pretraživanje slika na osnovu njenog sadržaja (CBIR = *Content-Based Image Retrieval*). U CBIR pretraživanju svaka slika je opisana nekim karakterističnim objektivnim obeležjima nižeg nivoa (*low-level descriptors*), kao što su boja, tekstura, oblik i/ili pozicija objekata. Nakon formiranja vektora obeležja, slike iz baze podataka se indeksiraju – formira se jednoznačna veza vektora obeležja i baze slika. Pretraživanje se, zatim, vrši na osnovu poređenja vektora obeležja slike upita (*query*) i vektora obeležja iz baze slika. Traže se oni vektori iz baze koji su najbliži vektoru slike upita, što predstavlja objektivnu meru sličnosti. Svakako, za ovakvo pretraživanje vektori obeležja se moraju formirati na isti način za upitnu sliku i slike iz baze. Dodatni problem u CBIR pretraživanju je slučaj pretraživanja komprimovanih slika jer, zavisno od vrste i stepena kompresije, neka od obeležja niskog nivoa mogu biti znatno izmenjena u odnosu na nekomprimovanu sliku.

Pretraživanje slika samo na osnovu njenog sadržaja nije uvek efikasno. Naime, treba odabrati relevantna obeležja koja najbolje opisuju sadržaj slike. Nadalje, nisu sva obeležja podjednako značajna – potrebno je uvesti odgovarajuće težinske faktore. Najzad, i dalje je, u konkretnom slučaju, presudan subjektivan osećaj korisnika pri vrednovanju slika što prouzrokuje tzv. semantički jaz (*semantic gap*) – nastaje razlika između objektivne metrike i subjektivnog osećaja korisnika. Ocena korisnika (RF = *Relevance Feedback*) se koristi za korigovanje vektora obeležja, najčešće uz asistenciju pogodno odabrane mreže veštačke inteligencije. Ovim postupkom se ublažava problem semantičkog jaza i omogućava se efikasnije pretraživanje, sa subjektivnog stanovišta. Međutim, i dalje ostaje dodatni problem veličine vektora obeležja. Za efikasno označavanje i pretraživanje dobro je imati što veći broj obeležja. To, sa druge strane, otežava pretraživanje jer je baza obeležja velika i povećava se vreme pretraživanja – javlja se problem poznat kao 'kopanje po podacima' (*data mining*). Pristupa se redukciji vektora obeležja primenom nekih od poznatih postupaka kao što su: analiza glavnih komponenta (PCA = *Principal Component Analysis*), uvođenje neuralnih mreža radijalne osnove (RBF = *Radial Basis Functions*), korišćenje SVM (*Support Vector Machine*) za obučavanje RBF mreža, vrši se klasterizacija i hijerarhijsko pretraživanje, itd.

2. Princip postupka pretraživanja multimedijalnog materijala

Multimedijalni materijal sadrži video (pokretne slike), audio i alfanumeričke podatke. Pretraživanje alfanumeričkih podataka je najjednostavnije jer je zasnovano na semantici. Doduše, potrebno je definisati i standardizovati rečnik pojmova i obezbediti brz i efikasan sistem poređenja i vrednovanja rezultata. Pretraživanje video i audio

materijala zahteva prethodnu obradu signala u cilju definisanja i formiranja vektora karakterističnih obeležja, koji objektivno opisuje signal.

Za audio se ranije koristilo semantičko pretraživanje na osnovu naziva odgovarajućeg muzičkog dela. U radu [1] koristi se prepoznavanje govora – sistem SpeechBot je namenjen indeksiranju i pretraživanju govornog audio materijala preko mreže. Verzija sistema koji je opisan u tom radu (koji je pisan 2000. godine) je sposobna da indeksira do 800 sati audio materijala nedeljno, na računarima klase Pentium III sa taktom od 600 MHz. Naravno da je savremena tehnologija znatno unapredila mogućnosti indeksiranja i pretraživanja baze podataka.

Za pretraživanje audio materijala, u opštem slučaju, koriste se vektori obeležja, slično kao za video materijal. Zvuk je veoma kompleksan i neregularan signal, karakterisan specifičnom bojom tona (*timbre*) i, svakako, subjektivnim doživljajem slušaoca. Klasična Furijeova analiza je moćan alat za opisivanje sličnosti između signala koji su u vidu dugačkih sekvenci i pretežno stacionarni, sastavljeni od harmonika (prostoperiodičnih komponenata različitih frekvencija). Uostalom, ta analiza je primenjena u opisivanju svojstava audio signala u standardu MPEG-7. Međutim, kod standardne Furijeove analize se javlja klasičan problem rezolucije po vremenu i frekvenciji. Na primer, za analizu signala pomoću širokog prozora komponente iz domena niskih frekvencija (NF) se dobro analiziraju ali se više frekvencije (VF) signala slabo razlikuju. Obrnuto, korišćenjem kratkih prozora ostvaruje se dobra rezolucija signala iz VF domena dok su NF signali maskirani i ne razlikuju se međusobno. Talasna transformacija (*wavelet transform*, WT), budući da se bavi vremensko-frekvencijskom (ili, prostorno-frekvencijskom, u slučaju slika) analizom, je veoma efikasna u obradi nestacionarnih signala kako u VF tako i u NF domenu, čak i pri veoma kratkom trajanju signala. Saglasno tome, WT se prepoznaje kao dobar alat za opisivanje realnih audio signala. U analizi i opisivanju audio sadržaja se, uobičajeno, koriste koeficijenti iz nekoliko nivoa WT (6 do 15) [2], [3].

U daljem tekstu ćemo se, pretežno, baviti opisom pretraživanja slika, s obzirom da se slična tehnologija koristi i u pretraživanju video signala, a sa malim korekcijama u pogledu opisivanja signala, isti princip se koristi i u pretraživanju audio materijala.

Svi CBIR sistemi pretraživanja zasnovani su na formiranju karakterističnih obeležja slike kao što su boja, tekstura, oblik, pozicija, itd. Takav pristup se, ponekad, referiše i kao *image-centric* sistem. Istraživanja su vršena u tri osnovna pravca. U početku su sistemi bili zasnovani na pažljivo odabranim globalnim karakteristikama slika i fiksnoj metrici za opisivanje sličnosti. Ti sistemi su bili efikasni pri pretraživanju onih slika kod kojih je ceo sadržaj bio od značaja, npr. slike tipa 'scena' ili grupnih slika. Međutim, u slučajevima kada je samo mali deo scene bio od značaja, npr. neki određeni objekat, ovi sistemi su bili neefikasni jer je ostatak scene, inače neinteresantan za pretraživanje, izazivao maskirajući efekat. Druga klasa pretraživanja je zasnovana na objektima (*object-based*). Ovaj metod je zahtevao pronalaženje i segmentaciju objekata i ekstrakciju obeležja samo unutar selektovanih objekata. Sami objekti se mogu opisati na osnovu regiona ili kontura. Mada je opisivanje na osnovu regiona robusnije, opisivanje pomoću kontura je preciznije [4]. Ovakvi sistemi su efikasni u pretraživanju homogenih baza slika, dakle, ako su slike sličnog sadržaja. Za dati upit će se uvek dobiti isti skup odabranih slika čak i ako različiti korisnici imaju različite zahteve.

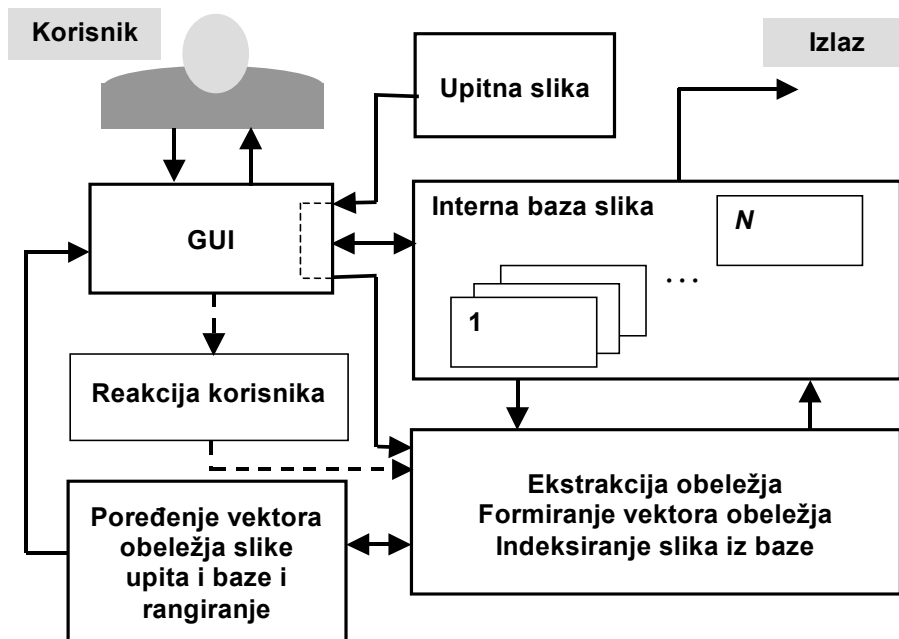
Karakteristike sistema za pretraživanje sa fiksnom metrikom sličnosti nisu zadovoljavajuće, prvenstveno zbog semantičkog jaza: subjektivna percepcija korisnika se

može znatno razlikovati od objektivnih elementarnih svojstava slike. To se može prevazići uvođenjem povratne sprege korisnika (*user's relevance feedback*), što je osnova treće klase metoda pretraživanja. Slike se najpre rangiraju na osnovu objektivne mere sličnosti a zatim korisnik vrednuje ponuđen skup slika označavajući one koje su najbližije upitu čime se, u povratnoj sprezi, koriguje uticaj pojedinih komponenata vektora obeležja slike. Pretraživanje se, zatim, ponavlja u interaktivnom vidu.

Umesto 'computer-centered' sistema, kao što je klasičan CBIR, prelazi se na 'human-centered' sistem odlučivanja, mada se pokazuje kao najboljom integracija ta dva pristupa, što se može obezbediti odgovarajućim 'human-computer' interfejsom, koji omogućava da se pretraživanje i odlučivanje obavi u skladu sa zahtevima i trenutnim željama korisnika. Primetimo da dva različita korisnika mogu, pod istim upitom, podrazumevati različite stvari. Štaviše, budući da je vizuelni nadražaj (kao i drugi nadražaji koje primamo preko naših čula) veoma individualan, može se desiti da i ista osoba ima različite utiske o istoj sceni, zavisno od raspoloženja i trenutnih uslova. Nove metode pretraživanja aktivno uključuju sudelovanje korisnika, tako što se, na osnovu povratne sprege, podešava strategija pretraživanja i odlučivanja. Kako danas gotovo sve metode koriste RF, izložićemo, ukratko, princip takvog pretraživanja. Primetimo, ujedno, da takvi sistemi sadrže, kao prvi korak, 'klasičan' objektivni način CBIR pretraživanja. Arhitektura sistema za pretraživanje data je na slici 1. Komunikacija korisnika sa sistemom vrši se preko odgovarajućeg grafičkog interfejsa (GUI).

Slike iz interne baze se, najpre, unificirano obrađuju tako što se izdvoje relevantna obeležja, formira se vektor obeležja za svaku sliku, a slike se indeksiraju. Korisnik bira upitnu sliku (koja može biti i iz date baze), za koju se formira vektor obeležja na isti način kao za bazu slika. Zatim se vrši poređenje vektora obeležja upitne slike i slika iz baze, određuju se rastojanja vektora obeležja u odgovarajućoj metrici (to je, najčešće, Euklidsko rastojanje, mada može biti i neka druga metrika: Mahalanobis rastojanje, Bhattacharya rastojanje, itd.). Slike iz baze se rangiraju po opadajućim vrednostima rastojanja od upita i prvih R najbolje rangiranih, tj. objektivno najbližijih, slika (obično se bira $R=10$ do 30) se prikazuje korisniku. U ovom, tzv., objektivnom koraku pretraživanja komponente vektora obeležja imaju podjednake uticaje na određivanje rastojanja – raspodela verovatnoća težina uz komponente vektora je uniformna. Za sistem bez korisničke povratne sprege proces pretraživanja je ovime završen. Od ponuđenih R objektivno najbližijih slika korisnik bira onu koja subjektivno najbolje odgovara upitu.

Za sistem koji koristi reakciju korisnika (RF), proces se nastavlja time što od ponuđenih R najbližih slika korisnik bira $C \leq R$ subjektivno najbližijih njegovom zahtevu i pogodno ih označava, na primer binarno, sa 1, kao relevantne, a ostalih ($R-C$) nerelevantnih sa 0. Moguća je i finija podela, kao što se koristi u sistemu MUVIS [5], gde postoji gradacija od 0 (nema sličnosti), do 5 (ista ili skoro identična slika). U svakom slučaju, tako rangirane slike se koriste za korekciju vektora obeležja, saglasno subjektivnom osećaju korisnika. Vektori obeležja označenih slika će uticati na raspodelu verovatnoća težina uz komponente vektora, što se vrši posredstvom pogodno odabrane neuralne mreže (najčešće RBF tipa), prilagođavajući ih subjektivnom zahtevu korisnika. Zatim se vrši ponovno merenje rastojanja i rangiranje slika, a postupak se ponavlja, ako je to neophodno [6], [7]. Uobičajeno je dovoljna jedna intervencija korisnika.



Slika 1. Arhitektura sistema za pretraživanje baze slika.

Opisani postupak se odnosi na pretraživanje slika, a ne video materijala. Međutim, video materijal se može raščlaniti na sledeće nivoe: pojedinačna slika (*frame*), kadar (*shot*) – niz slika pri jednom kretanju kamere, scena – niz kadrova koji su u semantički bliskoj vezi (opisuju sličnu situaciju), i isečak (*clip*) – kolekcija scena. Uobičajeno je video materijal označen semantički (podsetimo se filmskog rečnika “kadar 128, treći put”), pa se u pretraživanju video materijala može koristiti kombinovano odlučivanje na osnovu semantike i sadržaja slike. Kadrovi se mogu i automatski selektovati iz video materijala jer se, najčešće, pri promeni kadra naglo menja statistika slike. Selekcija i pretraživanje se, zatim, svode na slučaj pretraživanja pojedinačnih slika, s obzirom da kadar započinje jednom slikom koja se zove i ‘ključna slika’ (*key frame*).

3. Neki rezultati pretraživanja baze slika sa intervencijom korisnika

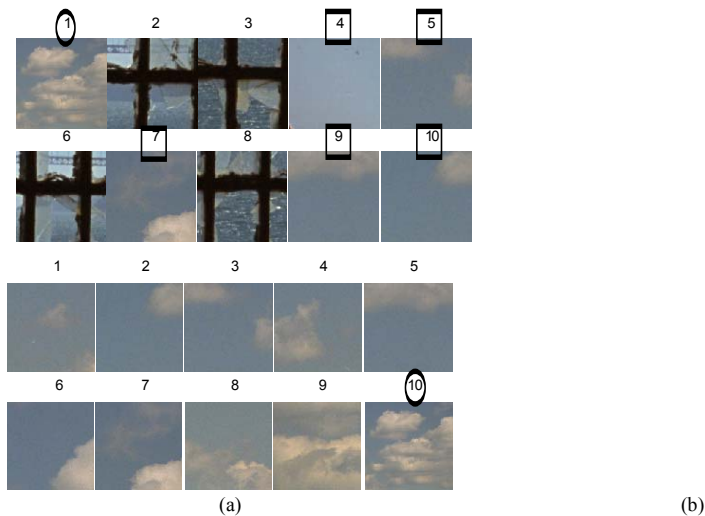
Efikasnost i brzina pretraživanja multimedijalnog materijala, pretežno slika, su veliki problem i predmet su intenzivnog proučavanja širom sveta. Pored ostalog, od septembra 2004 godine započet je evropski projekat COST* 292, *Semantic Multimodal Analysis of Digital Media* [8], kojem je, do sada, pristupilo 18 zemalja Evrope, među njima i naša zemlja. Cilj projekta je da se obezbedi automatsko označavanje audiovizuelnog sadržaja koristeći koncepte i nivoe apstrakcije koji su svojstveni ljudskom

* COST = *European Co-Operation in the field of Scientific and Technical research*

rasuđivanju, i da se u modul za pretraživanje ugradi baza znanja na osnovu iskustva i kroz interaktivnu asistenciju korisnika.

U radu [9] je opisan sistem za pretraživanje slika sa intervencijom korisnika sa arhitekturom koja odgovara slici 1. Sistem koristi obeležja boje i tekture i neuralnu mrežu radijalne osnove na koju deluju vektori obeležja slika koje je korisnik označio kao najslbličnije upitu, nakon prvog, objektivnog, rangiranja. Za testiranje je korišćena referentna MIT baza slika [10]. Originalne slike su bile dimenzije 512 x 512 piksela, a od njih je dobijena proširena baza tako što je svaka slika podeljena na 16 sub-slika dimenzije 128 x 128 piksela. Tako dobijena baza od oko 3.500 slika je korišćena za pretraživanje. U narednom tekstu biće izloženi neki karakteristični primeri. U svim izloženim primerima upitna slika je označena kružićem iznad odgovarajućeg broja, dok su slike koje je, nakon prvog prolaza, korisnik rangirao kao relevantne, označene kvadratićem. Radi uštede prostora u ovom radu, upitna slika se nije posebno prikazivala već je data u sklopu grupe od $R=10$ slika koje su najbliže upitnoj.

Na slici 2 je prikazan prvi primer. Slika 2(a) prikazuje prvih $R=10$ slika koje su objektivno najbliže upitnoj. Slika upita se, kao što se i očekivalo, našla na prvom mestu. Kao subjektivno najslbličnije upitu korisnik je odabrao slike rangirane kao 4, 5, 7, 9, i 10. Vektori obeležja ovih slika su, zatim, izvršili korekciju težina komponenata vektora obeležja baze slika, nakon čega je postupak pretraživanja ponovljen. Rezultat novog pretraživanja je prikazan na slici 2(b). Primitimo da je upitna slika sada izgubila prvu poziciju, jer se težište klastera relevantnih slika pomerilo, ali kompletan set prvih 10 najbližih slika pripada klasi upitne slike.

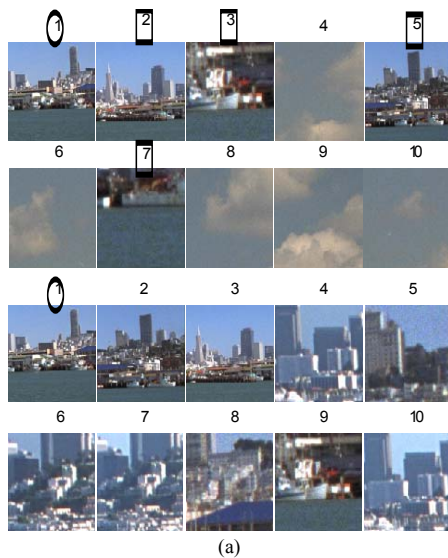


Slika 2. Ilustracija prvog primera pretraživanja: (a) Objektivno rangiranje, (b) rezultat nakon intervencije korisnika. Kružićem je označena upitna slika.

Na slikama 3 i 4 ilustrovana su još dva primera koja su preuzeta iz [9]. Upitna slika, nakon intervencije korisnika može izgubiti prvu poziciju, slika 3(b) ali su i u ovim primerima prvih $R=10$ slika bile iz klase slika upita, što se vidi na slikama 3(b) i 4(b).



Slika 3. Ilustracija drugog primera pretraživanja: (a) Objektivno rangiranje, (b) rezultat nakon intervencije korisnika. Kružićem je označena upitna slika.



Slika 4. Ilustracija trećeg primera pretraživanja: (a) Objektivno rangiranje, (b) rezultat nakon intervencije korisnika. Kružićem je označena upitna slika.

4. Zaključak

U radu je izložena problematika pretraživanja multimedijalnog materijala sa posebnim osvrtom na pretraživanje slika, kao najzahtevnijih dokumenata. Ogromna količina podataka, kako u pogledu velikog broja fajlova tako i njihove pojedinačne

veličine, predstavlja veliki tehnički problem poznat kao *data mining*. Tačnost pretraživanja zahteva formiranje velikog broja pojedinačnih obeležja, ali to povećava vreme pretraživanja. Stoga se posebna pažnja posvećuje izboru relevantnih obeležja i redukciji dimenzije vektora obeležja. Ujedno, u pretraživanje se unosi subjektivno vrednovanje kvaliteta pretraživanja na osnovu reakcije korisnika. Na osnovu objektivnog sadržaja dokumenta koji se analizira, postižu se veoma dobri rezultati pretraživanja, a očekuje se da će kombinacija semantičkog i objektivnog označavanja, uz intervenciju korisnika, zajedno sa klasterizacijom obeležja i hijerarhijskim pretraživanjem, dovesti do efikasnih modula za indeksiranje, rangiranje i pretraživanje multimedijalnog materijala.

Literatura

- [1] J.-M. Van Thong, P. J. Moreno, B. Logan, B. Fidler, K. Maffey, M. Moores, "SpeechBot: An experimental speech-based search engine for multimedia content on the web", *IEEE Trans. on Multimedia*, Vol. 4, No. 1, pp. 88-96, March 2002.
- [2] Stephan Rein, Martin Reisslein, "Identifying the classical music composition of an unknown performance with wavelet dispersion vector and neural nets" (Extended version), December 2004.
- [3] V. Đorđević, N. Reljin, I. Reljin, "Identifying and retrieving of audio sequences by using wavelet descriptors and neural network with user's assistance", *Special Session COST 292, IEEE Conference EUROCON 2005*, Belgrade, SCG, November 21-24, 2005.
- [4] G. Aggarwal, T. V. Ashwin, S. Ghosal, "An image retrieval system with automatic query modification", *IEEE Trans. on Multimedia*, Vol. 4, No. 2, pp. 201-214, June 2002.
- [5] S. Kiranyaz, K. Caglar, E. Guldogan, M. Gabbouj, "MUVIS: A content-based multimedia indexing and retrieval framework", in *Proc. 3rd Int. Workshop on Content-Based Multimedia Indexing*, CBMI 2003, Rennes, France, pp. 405-412, 22-24 Sept. 2003.
- [6] P. Muneesawang, and L. Guan, "An interactive approach for CBIR using a network of radial basis functions", *IEEE Trans. Multimedia*, Vol. 6, No. 5, pp. 703-716, Oct. 2004.
- [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*, EFPL, Montreux, Switzerland, April 2005.
- [8] COST Action 292, *Semantic Multimodal Analysis of Digital Media*, www.cost292.org
- [9] S. Čabarkapa, N. Kojić, V. Radosavljević, G. Zajić, B. Reljin, "Adaptive content-based image retrieval with relevance feedback", *Special Session COST 292, IEEE Conference EUROCON 2005*, Belgrade, SCG, November 21-24, 2005
- [10] <http://vismod.media.mit.edu/pub/VisTex/>

Abstract: *Rapid growth of electronics, communications and computers, culminating by the explosion of Internet, has created the enormous number and size of multimedia databases. Annotation and classification of such material, which is known as indexing, and efficient search and retrieval of audio-visual data has become a great problem*

either for professional organizations or for users. A large number of scientific and professional teams from different organizations investigated this problem – among others, this is a theme of the European project COST 292 participated by 18 European countries, including Serbia and Montenegro. In this paper basic methods for image retrieval, based on the image content with user's relevance feedback, are briefly described. Several examples illustrate the efficiency of such retrieving procedure.

Keywords: *Relevant image features, feature vector, indexing, semantic description, image retrieval, subjective perception, user's relevance feedback, radial basis function neural networks.*

MULTIMEDIA DATABASE INDEXING AND RETRIEVING

Branimir Reljin