

Pretraživanje slika po sadržaju

Seminarski rad u okviru kursa
Metodologija stručnog i naučnog rada
Matematički fakultet

Božidar Mitrović, Luka Vujčić,
Marija Erić, Dušan Petrović

Novembar 2021

Sažetak

Multimedijalna analiza sadržaja se primenjuje u različitim problemima kompjuterske vizije, a digitalne slike čine veliki deo multimedijalnog sadržaja. U poslednjim godinama, kompleksnost multimedijalnih sadržaja (posebno slika) je eksponencijalno porasla. Više od milion slika dnevno dolazi do platformi poput Fejsbuka, Tvitera i Instagrama[6]. Efikasan pronalazak određene slike iz ovako brojne arhive je netrivialan zadatak u oblasti kompjuterske vizije.

Tradicionalni pretraživači koje svakodnevno koristimo baziraju svoju pretragu na pretraživanju natpisa koji prati sliku. Međutim, ogroman napredak je napravljen na polju pretrage slika po sadržaju (*PSPS*), klasifikaciji slika i njihovoj analizi. Glavna odlika *PSPS* modela i klasifikacije slika je u tome što su slike visoke rezolucije mapirane u numeričke vektore. Istraživanja su pokazala da postoji velika razlika između ovih vektora i ljudskog smisla semantike u slici, te im je cilj da se ova razlika što je moguće više smanji.

Ovim seminarskim radom pravimo pregled trenutnih tehnika u *PSPS* istraživanju i predstavljanju slika. Proćićemo različite aspekte istraživanja i nadamo se novom proboju na polju ovih tehnika i poboljšanju performansi trenutnih.

Sadržaj

1	Uvod	3
2	Osnovna ideja pretrage slika po sadržaju	4
3	Odlike niskog nivoaa	4
3.1	Segmentacija slike	5
3.2	Odlike boja	5
3.3	Odlike teksture	5
3.4	Odlike oblika	6
3.5	Prostorne odlike	6
3.6	Mere sličnosti	6
4	Odlike visokog nivoaa	6
4.1	Objektna ontologija	7
4.2	Nadgledano učenje	7
4.3	Nenadgledano učenje	7
5	Interakcija korsnika	8
5.1	Formiranje upita	8
5.2	Korisnička povratna informacija	8
6	Evaluacija performansi	9
7	Zaključak	10

1 Uvod

Napretkom tehnologije, korišćenje mobilnih telefona, digitalnih kamera interneta postaje dosta jendostavnije. Zbog toga raste i količina multimedijalnih sadržaja kojeg oni proizvode. Samim tim, tehnike koje omogućavaju efikasno pretraživanje i upoređivanje slika postaju zahtevan problem. Internet pretraživači koje svakodnevno koristimo se služe opisom koji stoji uz slike. Dakle, oni rade poređenje reči i određivanje sličnosti između njih. Ovaj pristup možemo odmah da otpišemo jer je gotovo nemoguće manuelno opisati slike koje imamo a čak i tada naši pretraživači znaju da pogreše. Treba nam nešto drugačije.

Jedan pristup pri ovakoj analizi je da se najpre primeni nekakav automatski sistem za anotaciju slika koji ume da opiše sadržaj slike, te da se na osnovu njega vrši opis slika. Ovde bi glavni model zavisio od sposobnosti sistema za anotaciju da precizno odredi ivice[4], boju, teksturu i druge elemente slike što uopšte nije lak zadatak nekad čak i za ljude. A pošto naš glavni model dosta zavisi od ovih anotacija, tolerancije za grešku su izuzetno male.

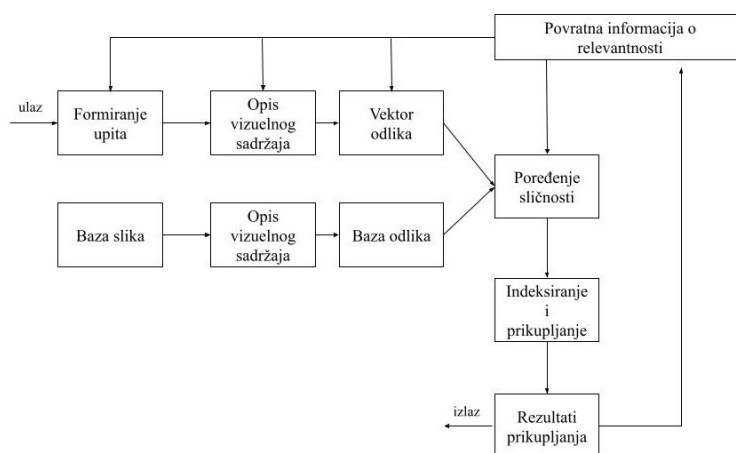
Pretraživanje slika po sadržaju (PSPS)[2] je tehnika pomoću koja prevazilazi gorepomenute probleme, jer se radi vizuelna analiza slike koja se pretražuje. Dakle, glavni uslov za rad ovakvih modela je da postoji slika za koju se radi pretraga, a pomoću nje se određuje sličnost sa slikama koje se nalaze u arhivi. Boja, oblici, tekstura i ostali elementi niskog nivoa slike za pretragu se koriste tek na kraju za sortiranje izlaza. PSPS modeli su našli primenu u medicini, otkrivanju kriminala, analizi video snimaka, vojsci i mnogim drugim delatnostima.

Jedna od glavnih odlika za uspešnost modela za preuzimanje slika je minimalna ljudska interakcija. Vizuelne karakteristike koje će biti korišćene za sortiranje slike (boja, tekstura, oblici) zavise od potreba korisnika koji koristi sistem. Međutim, da bi karakteristike učinili robusnijim i jedinstvenijim, obično su komputacijske cene nezanemarljive. Takođe, izborom pogrešnih osobina na slici možemo dodatno usporiti ili otežati proces učenja, što takođe nije dobro. Dakle, neki od glavnih zadataka istraživanja na ovu temu podrazumevaju[1]:

1. kako se performanse modela za pretraživanje slika po sadržaju mogu poboljšati korišćenjem osobina niskog nivoa slike
2. kako se semantička razlika između osobina niskog nivoa i semantičkih osobina visokog nivoa slike može smanjiti
3. kako tehnike mašinskog učenja mogu poboljšati performanse (PSPS) modela

2 Osnovna ideja pretrage slika po sadržaju

Pretraga slika po sadržaju koristi vizuelni sadržaj slike, kao što su boja, oblik, tekstura, prostorne odlike, za predstavljanje i indeksiranje slike. U klasičnom sistemu, vizuelni sadržaj slika je opisan više-dimenzionim vektorom odlika, od kojih se dalje formira baza odlika. Za prikupljanje slika, korisnik snadbeva sistem sa primerom slike ili skiciranih figura, koji se prevodi u vektor odlika. Zatim se izračunava sličnost između vektora odlika upita primera i slika iz baze. Prikupljanje se vrši uz pomoć indeksne sheme koja omogućava efikasan način pretrage slika u bazi. Noviji sistemi u obzir uzimaju i korisnikov povratnu informaciju o relevantnosti, koja omogućava generisanje semantički i perceptivno značajnih rezultata. Na slici 3 je prikazana shema opisanog sistema.



Slika 1:

Shematski prikaz sistema za pretragu slika po sadržaju ¹.

Sadržaj slike može biti vizuelni i semantički. Vizuelni sadržaj može biti opšti ili vezan za određeni domen. Opšti vizuelni sadržaj (odlike niskog nivoa) obuhvata boju, teksturu itd., dok sadržaj vezan za domen (npr. ljudsko lice) zavisi od primene i zahteva poznavanje domena. Semantički sadržaj se dobija iz tekstualnih anotacija ili kompleksnih procedura za ključevanja na osnovu vizuelnog sadržaja [3].

3 Odlike niskog nivoa

Izdvajanje odlika niskog nivoa predstavlja osnovu PSPS sistema. Izdvajanje se može izvršiti sa cele slike ili sa nekog regiona slike. Predstavljanje slika na nivou regiona je sličnije ljudskom sistemu za opažanje i samim

¹Shema je preuzeta iz [3].

tim više PSPS sistema koristi taj pristup. Da bi se izvršila pretraga slika zasnovanim na regionima (PSZR), prvi korak je segmentacija slike a potom se izdvajaju odlike niskog nivoa kao što su boja, tekstura, oblik ili prostorne odlike.

3.1 Segmentacija slike

Za segmentaciju slika postoji više algoritama, svaki ima svoje prednosti i mane. Neki od njih su:

1. JSEG segmentacija: sastoji iz 2 koraka - klasifikacija boja na slici i zamena piksela slike sa određenom klasom boje. Ovaj algoritam je dobar za izvlačenje sličnih tekstura i boja sa slike.
2. KMCC segmentacija - bazirano na algoritmu za klasterovanje k-sredina. Ovaj algoritam je bolji ukoliko imamo objekte na slici koji nisu homogeni a želimo da pripadaju istom regionu.



Slika 2:
Rezultati JSEG segmentacije ².

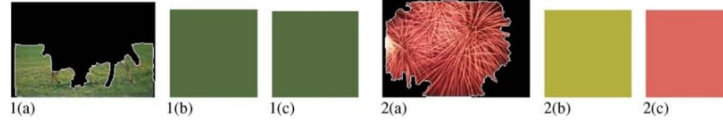
3.2 Odlike boja

Boja je jedna od najčešćih odlika koja se koristi za pretraživanje slika. Ona je definisana na jednom od više mogućih prostora boja. Neki od češće korišćenih prostora boja su RGB, LAB, LUV, HSV, YCrCb i HMMD. Neke odlike boja koje se koriste u PSZR su matrica kovarijanse boja, histogram boje, momenti boje, prosečna boja i dominantna boja. Odabir odlike boje uveliko zavisi od rezultata segmentacije. Na primer ukoliko segmentacijom dobijemo nehomogene regione, prosečna boja nam neće puno pomoći u ovom slučaju.

3.3 Odlike teksture

Tekstura nam daje bitne informacije za klasifikovanje slika. Neke češće korišćene odlike u PSPS sistemima su: spektralne odlike dobijene korišćenjem Gaborovim filtriranjem, statističke odlike kojima pripadaju 6 Tamurinih odlika i globalne odlike koje je predložio Liu et al².

²Slike su preuzete iz [3].



Slika 3:
Prosečna i dominantna boja: (a) početni region; (b) prosečna boja; (c) dominantana boja ².

3.4 Odlike oblika

Oblik nije toliko korišćen u PSZR sistemima kao boja i tekstura, obično je koristan u nekim domensko specifičnim slikama kao što su objekti koje je napravio čovek. Neke osnovne odlike su odnos visine i širine, zaobljenost, Furijeovi deskriptori i uzastopna granica.

3.5 Prostorne odlike

Prostorne odlike su korisne u klasifikaciji regiona. Na primer more i nebo imaju sličnu boju i teksturu, ali prostorno nebo se obično nalazi na vrhu slike dok je more obično na dnu slike. Lokacija se obično jednostavno definiše kao gore ili dole. Postoje i relativni prostorni odnosi kao što su levo/desno i iznad/ispod koji se definišu za neka dva objekta.

3.6 Mere sličnosti

U PSZR sistemi sličnost se meri na dva nivoa. Prvi je nivo regiona koji meri sličnost 2 regiona na osnovu njihovih odlika niskog nivoa. Drugi je nivo slike koji meri sličnost 2 slike sa različitim brojem regiona. Većina istraživača koriste rastojanje Minkovskog. Ako su dva regiona predstavljena sa dva p dimenziona vektora (x_1, x_2, \dots, x_p) , (y_1, y_2, \dots, y_p) , redom. Onda se rastojanje Minkovskog definiše kao

$$d(X, Y) = \left(\sum_{i=1}^p |x_i - y_i|^r \right)^{1/r}$$

U slučaju da je $r = 2$ dobijamo Euklidsko rastojanje, a ako je $r = 1$ dobijamo Menhetn rastojanje.

4 Odlike visokog nivoa

U prethodnom poglavlju videli smo šta predstavljaju odlike niskog nivoa i neke osnovne načine za pronalaženje slika pomoću algoritama koje su bazirane na njima. Ovaj tip odlika daje vizuelni opis slike.

Odlike visokog nivoa slike predstavljaju daleko složeniji koncept, one predstavljaju njen semantički opis. Ljudi gledanjem slike primenjuju slične koncepte. Najpre određuju vizuelni aspekt slike (boje, oblici, uočavanje

regiona,...) a zatim daljim procesiranjem dobijaju njen semantički opis. Naša želja je da razvijemo sistem koji radi na navedenom principu.

U narednom delu govorićemo o tehnikama za određivanje semantičkog opisa.

4.1 Objektna ontologija

Objektna ontologija pruža opis slike na srednjem nivou. Na ovaj način ona predstavlja sponu između opisa na niskom nivou i visokom nivou. Ona daje opis slike koja se može čuti u svakodnevnom jeziku. Ona nam omogućava da formiramo jednostavan rečnik.

Naime možemo odrediti određene ključne reči (atribute) poput: boja, veličina, pozicija, tekstura, oblik itd. Svaki ovaj atribut je kategoričan i uzima vrednosti iz određenog skupa. Šta znači opisati sliku ili region slike u terminima ovih atributa? To znači da ćemo na kraju za sliku dobiti opis poput: "region tamno zelene boje, unifromno obojen, pozicionirana dole u sredini....".

4.2 Nadgledano učenje

Postoje različiti algoritmi nadgledanog učenja koje možemo koristiti za učenje odlika visokog nivoa.

SVM (support vector machine) se koristi za klasifikaciju teksta, prepoznavanje objekata i mnoge druge stvari. Razvijen je kao binarni klasifikator tj. klasifikator koji objekte raspoređuje u dve klase. Da bismo koristili ovaj klasifikator moramo objekte mapirati u vektore. Klasifikacija se svodi na postavljanje optimalne hiperravni.

Drugi metod je upotreba Bajesovih klasifikatora. Bajesovi klasifikatori u osnovi koriste Bajesovu teoremu o uslovnoj verovatnoći.

Pored navedenih klasifikatora, postoje i mnogobrojni drugi kao što su stabla odlučivanja, pravila odlučivanja, neuronske mreže itd. Svi ovi klasifikatori imaju svoje prednosti i mane i uopštem slučaju nije moguće odrediti koji je najbolje koristiti nego je potrebno za svaki izvršiti evaluaciju i uporediti rezultate.

Glavni problemi sa nadgledanim učenjem je potreba za podacima koji imaju označene klase kojima pripadaju. Deo ovih podataka se koristi za treniranje, a deo za evaluaciju kvaliteta modela. Takođe, potrebna je velika količina podataka za nadgledano učenje, potovo za neuronske mreže.

4.3 Nenadgledano učenje

Kod nenadgledanog učenja nije potrebno imati unapred poznate klase niti označene podatke klasama. Klasterovanje je vid nenadgledanog učenja. Pod klasterovanjem slika podrazumevamo grupisanje slika u klaster. U klasteru je maksimizovana sličnost između slika, a slike iz različitih klastera su minimalno slične. Rezultujuće klasterne možemo shvatiti i kao oznake klase.

Jedan od tradicionalnih algoritama za klasterovanje je algoritam K sredina. Za potrebe ovog algoritma neophodno je odrediti unapred broj klastera, to je moguće odrediti empirijski. Navedeni algoritam ne određuje

dobro klastere kada je distribucija podataka kompleksna (npr varijabilna gustina podataka).

Daleko moćniji pristup klasterovanju je verovatnosna klasifikacija zasnovana na Bajesovoj teoriji. Ovakvi algoritmi se zasnivaju na određivanju maksimalne aposteriorne verovatnoće.

5 Interakcija korisnika

U PSPS sistemima, interakcija sa korisnicima je od krucijalnog značaja, jer fleksibilnost formulacije i modifikacije upita se može postići uključivanjem korisnika u proces prikupljanja. Korisnički interfejs u ovim sistemima se najčešće sastoji od formulacije upit i prezentacije rezultat.

5.1 Formiranje upita

Specifikacija tipa slika koje korisnik želi da prikupi iz baze se može vršiti na više načina. Najčešće korišćeni upiti su:

1. *Pretraživanje po kategorijama (category browsing)* - koristi se za pretraživanje baze podataka na osnovu kategorija slike. Za ovu svrhu se vrši klasifikacija slika u bazi u različite kategorije na osnovu njihovih semantičkih odlika ili na osnovu vizuelnog sadržaja.
2. *Pretraživanje na osnovu koncepta (query by concept)* - koristi se za prikupljanje slika na osnovu konceptualnog opisa povezanog sa svakom slikom u bazi podataka.
3. *Pretraživanje na osnovu skice (query by sketch)* - omogućava korisniku da nacrti skicu slike koristeći alat za grafički prikaz koji ne mora biti u okviru sistema za prikupljanje.
4. *Pretraživanje na osnovu primera (query by example)* - omogućava korisniku da formuliše uput na osnovu primera slike. Sistem prevodi primer slike u unutrašnju reprezentaciju odlika. Zatim se pretražuju slike u bazi sa sličnim odlikama.

Takođe u nekim sistemima postoji mogućnost pretraživanja uz pomoć grupe slika. Sistem tada pronalazi slike koje najviše odgovaraju odlikama zadate grupe. Mnogi sistemi omogućavaju upite sa pozitivnim i negativnim primerima [3].

5.2 Korisnička povratna informacija

Ljudska percepcija sličnosti između slika je subjektivna i zavisi od semantike. Iako PSPS metode daju obećavajuće rezultate za prikupljenje slika, rezultati zasnovani na sličnostima vizuelnih odlika nisu uvek percipirani i semantički značajni. Dodatno, svaki tip vizualne odlike obuhvata samo jedan aspekt svojstva slike, i korisnik često ima poteškoće da specifikuje kako se različiti aspekti kombinuju. U cilju rešavanja ovog problema, uvedena je interaktivna povratna informacija relevantnosti - tehnika iz tradicionalnih sistema za prikupljanje tekstualnih informacija. Primena povratne informacije omogućava uspostavljanje veze između koncepta visokog nivoa i odlika niskog nivoa.

Povratna informacija je nadgledana tehnika učenja koja se koristi za poboljšanje efikasnosti informacionih sistema. Glavna ideja je da se za unapređenje sistema koriste pozitivni i negativni primeri korisnika. Za dati upit, sistem prvobitno prikuplja listu slika na osnovu predefinisanih mera sličnosti. Zatim, korisnik označava prikupljene slike kao relevantne upitu (pozitivni primeri) ili nerelevantne (negativni primeri). Sistem će zatim uzeti u obzir i korisničke povratne informacije. Glavni problem je kako inkorporirati pozitivne i negativne primere u prečišćavanje upita i/ili njihovo uključivanje u mere sličnosti.

6 Evaluacija performansi

Za evaluaciju performansi sistema za prikupljanje koriste se dve mere - odziv i preciznost [5]. Skup slika iz baze koji su relevantni za dati upit q , označavamo sa $R(q)$, a prikupljene rezultate upita q sa $Q(q)$.

Preciznost prikupljenih rezultata definišemo kao razlomak prikupljenih slika koje su zaista relevantne za upit.

$$precision = \frac{|Q(q) \cap R(q)|}{|Q(q)|}$$

Odziv prikupljenih rezultata definišemo kao razlomak relevantnih slika koje je sistem prikupio na osnovu upita q .

$$recall = \frac{|Q(q) \cap R(q)|}{|R(q)|}$$

Potrebno je napraviti balans između ove dve mere, jer poboljšanjem jedne žrtvujemo drugu. U klasičnim sistemima, odziv se povećava sa povećanjem brojem prikupljenih slika, dok se preciznost u tom slučaju smanjuje. Dodatno, određivanje relevantnog skupa slika $R(q)$ je manje stabilno, zbog različitih interpretacija slika. Takođe, ukoliko je broj relevantnih slika veći nego broj prikupljenih slika, odziv je beznačajan. Dakle, preciznost i odziv predstavljaju grub opis performansi.

Iz tog razloga je predložena nova mera performansi - average normalized modified retrieval rank (ANMRR), koja kombinuje preciznost i odziv.

Obeležimo ukupan broj slika relevantnih za upit q sa $N(q)$, a maksimalan broj relevantnih slika za svih Q upita $\max(N(q1), N(q2), \dots, N(Q))$ sa M . Onda za svaki dati upit q , svakoj relevantnoj slici k se dodeljuje rang vrednosti $rank(k)$ koja je jednaka rang K ukoliko se nalazi u prvih K (gde je $K + \min[4N(q), 2M]$) rezultata upita q . Inače joj se dodeljuje rang $K + 1$.

Srednja vrednost ranga (*averagerank* - AVR) za upit q se računa:

$$AVR(q) = \sum_{k=1}^{N(q)} \frac{rank(k)}{N(q)}$$

Modifikovani rang (*The modified retrieval rank* - MRR) za upit q :

$$MRR(q) = AVR(q) - 0.5 - 0.5 * N(q)$$

Vrednost $MMR(q)$ ima vrednost 0 kada se sve relevantni slike nalaze u prvih K prikupljenih rezultata.

Dalje normalizujemo MMR , i dobijamo vrednosti između 0 i 1:

$$NMRR(q) = \frac{MMR(q)}{K + 0.5 + 0.5 * N(q)}$$

I na kraju izvršimo usrednjavanje po svim upitima q :

$$ANMRR = \frac{1}{Q} \sum_{q=1}^Q NMRR(q)$$

Detaljnije o ovoj meri evaluacije se može naći u [3].

7 Zaključak

Literatura

- [1] Umer Sajid Jameel Ahmed Nouman Ali Naeem Iqbal Ratyal Bushra Zafar Saadat Hanif Dar Muhammad Sajid Afshan Latif, Aqsa Rasheed and Tehmina Khalil. Content-based image retrieval and feature extraction: A comprehensive review. 2019.
- [2] Yilong Yang Guoyong Duana, Jing Yanga. Content-based image retrieval research. *2011 International Conference on Physics Science and Technology (ICPST 2011)*, pages 471–477.
- [3] Fuhui Long, Hongjiang Zhang, and David Dagan Feng. Fundamentals of content-based image retrieval. In *Multimedia information retrieval and management*, pages 1–26. Springer, 2003.
- [4] M. Mittal S. Susan, P. Agrawal and S. Bansal. New shape descriptor in the context of edge continuity. 4:101–109, 2019.
- [5] Arnold WM Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1349–1380, 2000.
- [6] Dustin Stout. Social media statistics 2021: Top networks by the numbers. <https://dustinstout.com/social-media-statistics/>, 2021.