

Prijedlog teme doktorske disertacije

13. studenog 2013.

Univerzitet u Sarajevu

Elektrotehnički fakultet Sarajevo

Vijeće doktorskog studija

Kandidat: mr Vedran Ljubović, dipl. ing. el.

Saradnik na definisanju teme: V. prof. dr Haris Šupić, dipl. ing. el.

Predmet: Metode pretraživanja informacija

Sadržaj

1 Radni naslov teme doktorske disertacije	2
2 Obrazloženje teme	2
2.1 Motivacija	2
2.2 Opšti pristup i klasifikacija CBIR sistema	3
2.3 Princip rada CBIR sistema	5
2.4 Algoritmi u JPEG kompresovanom domenu	6
3 Stanje oblasti	11
3.1 Histogram boja i druge značajke na osnovu boje	11
3.2 “Naivne” metode u kompresovanoj domeni	12
3.3 Značajke zasnovane na teksturi	14
3.4 Detekcija rubova	17
3.5 Oblik	18
3.6 Kombinovane značajke	19
3.7 Ostali algoritmi i metode u kompresovanom domenu	19
4 Ciljevi i plan istraživanja	20
5 Metodologija istraživanja i potrebni istraživački resursi	20
5.1 Skupovi podataka	20
5.2 Metrike poređenja metoda pretraživanja	21
5.3 Eksperimentalni softver	23
5.4 Potrebni istraživački resursi	23
6 Očekivani izvorni naučni doprinos	23

1 Radni naslov teme doktorske disertacije

“Unaprjeđenje performansi pretraživanja slika na osnovu sadržaja u JPEG kompresovanom domenu”

“Improving performance of Content based image retrieval in JPEG compressed domain”

2 Obrazloženje teme

2.1 Motivacija

Od davnina ljudi se bave arhiviranjem i klasifikacijom dokumenata znajući da znanje koje nije lako dostupno ima umanjenu vrijednost. Računarska tehnologija znatno je olakšala ovaj posao i omogućila digitalno arhiviranje neprevazidenih količina podataka. Međutim, većina računarskih aplikacija ovog tipa bavi se obradom tekstualnih i numeričkih podataka. Proteklih decenija pojavljuje se problem narastajućih količina multimedijalnih podataka u digitalnom obliku, a prije svega zvučnih zapisa, slika i videa.

Pretraživanje slika na osnovu sadržaja (eng. Content-Based Image Retrieval, CBIR) je naučna oblast koja se bavi proučavanjem metoda i algoritama namijenjenih rješavanju problema organizacije arhive digitalnih slika na osnovu njihovog vizuelnog sadržaja. Problem CBIR možemo formulisati na sljedeći način: na koji način automatski obraditi slike i omogućiti njihovu klasifikaciju i lakše pronalaženje?

Tradicionalan pristup ovom problemu zasniva se na pridruživanju određenih metapodataka slikama, te na analizi teksta koji se nalazi u blizini slike kako bi se prepoznao npr. naslov slike. Ovakav pristup koristi većina savremenih Internet pretraživača (Google Image Search, Yahoo! Image Search...), no on ima svojih ograničenja. Metapodaci koji se slikama pridružuju automatski pokazuju se kao neadekvatni za potrebe pretraživanja slika, a ručni unos metapodataka je naporan i u pravilu izostaje.

CBIR metode pokušavaju nadopuniti dostupne informacije o slici podacima nastalim analizom sadržaja slike, odnosno njenih piksela. Jedan jednostavan slučaj upotrebe CBIR metoda može biti sistem koji pronalazi identične slike (plagijate) ili slike sa neznatnim izmjenama. Naprednije primjene su moguće u oblasti biometrije (prepoznavanje otiska prsta), medicinskim aplikacijama (analiza rentgenskih i drugih snimaka i pronalazak sličnosti), nadzor (prepoznavanje zračnog snimka teritorije na osnovu rasporeda cesta) itd. Krajnji cilj odnosno “sveti Gral” multimedijalnog pretraživanja [1] bio bi sistem koji je u stanju odgovoriti na proizvoljne upite korisnika, npr. softver koji na osnovu slike određene osobe pronalazi sve fotografije na kojima se nalazi ta osoba. Ovaj zadatak je još uvijek daleko od ostvarenja.

2.2 Opšti pristup i klasifikacija CBIR sistema

Veliko pregledno istraživanje [2] navodi dva “procjepa” koje CBIR sistemi moraju savladati da bi bili efikasni u ostvarenju svojih ciljeva. *Senzorni procjep* (eng. sensory gap) nastaje zbog razlika između načina na koji ljudi percipiraju vidljivi svijet i računarskog zapisa slike. Npr. slike u računar su obično predstavljene u RGB modelu boje koji je razvijen radi efikasnog prikaza na ekranu i ima brojna razilaženja sa ljudskim doživljajem boje. *Semantički procjep* (eng. semantic gap) je pak nedostatak poklapanja između sadržine slike koja se može automatski odrediti algoritamskim putem i njenog shvatanja od strane korisnika. Npr. računar je u stanju vrlo jednostavno odrediti poluprečnik i obim kruga, no teško će uočiti razliku između kuće sa krovom od slame i plasta sijena.

Smeulders i dr. [2] su 2000. godine procijenili da će napredak tehnologija računarske vizuelne percepcije ubrzo u potpunosti zatvoriti senzorni procjep, no semantički procjep ostaje kao neriješen problem. Stoga, imajući u vidu veliki uticaj citiranog rada u naučnom svijetu,¹ nije iznenađujuće da je veliki broj radova nastalih nakon njega posvećeno rješavanju problema semantičkog procjepa.

No nakon više od 30 godina istraživanja u ovoj oblasti pokazuje se da rješavanje opšteg problema CBIR nije jednostavno, pa većina autora preporučuju sužavanje na neki specifični problemski domen [3]. Čovjek koristi bazu svojih životnih iskustava kako bi prepoznao sadržaj slike, pa je najperspektivniji pristup u prevladavanju semantičkog procjepa zasnovan na metodama mašinskog učenja i velikih baza znanja (Internet). U posljednje vrijeme napredak u ovoj oblasti postignut je korištenjem “dubokih” neuralnih mreža [4, 5].²

Od posebnog značaja su istraživanja ponašanja korisnika i njihovih očekivanja od CBIR sistema. Pokazuje se da u nekim situacijama korisnici smatraju čisto vizuelnu sličnost zadovoljavajućim kriterijem pretrage, dok u drugim očekuju i određeni nivo semantičke sličnosti. Datta i dr. [6] pokušavaju objasniti ovu pojavu koristeći klasifikaciju CBIR sistema prema njihovom opsegu, korisničkoj namjeri i načinu zadavanja upita.

Opseg podataka (scope) varira od lične kolekcije do web pretraživača. Datta i dr. navode da se sistemi namijenjeni za rad sa ličnom kolekcijom trebaju fokusirati na personalizaciju, korisnički interfejs i sl., dok sistemi za web trebaju biti optimizirani za veliku količinu saobraćaja. No u radu [7] pokazali smo da se čak i kod ličnih kolekcija javlja problem vremena izvršenja, budući da je u proteklom periodu količina multimedijalnih podataka rasla znatno brže nego procesne mogućnosti ličnih računara.

Po pitanju *namjere korisnika* (intent), Datta i dr. razvrstavaju korisnike u tri kategorije: pregledač (browser) nema konkretan cilj, surfer ima osrednje precizno definisan cilj, dok istraživač (searcher) ima jasno definisan cilj. Većina popularnih pretraživača web stranica poput Google, Yahoo! ili Bing pokušavaju da osmisle korisnički interfejs na način da on zadovolji sve kategorije korisnika. No u CBIR sistemima, možda će se za različite kategorije korisnika morati os-

¹Citiran 4.822 puta prema podacima sa sajta Google Scholar (6. 8. 2013).

²Rezultati prezentovani u citiranom radu uspješno su iskorišteni u Google Photos aplikaciji.

misлити potpuno različiti algoritmi pretrage, sa većim ili manjim udjelom semantike itd. Pored uobičajenog načina rangiranja rezultata pretrage po relevantnosti, CBIR sistem može prezentovati rezultate u formi hijerarhijskih struktura, gomila (clusters), u vidu vremenske linije (historijata) i slično.

Modalitet vršenja upita je uobičajeno pitanje u istraživanjima iz oblasti CBIR. Prije svega, korisnik može postaviti upit u formi teksta, bilo ključnih riječi, bilo upitne rečenice ili kompletnog naslova. U ovom slučaju sistem se oslanja na kvalitet priloženih metapodataka i sposobnost da se prepozna sadržaj (semantika) slike. Druga vrsta upita je *pretraživanje na osnovu primjera* (query by example, QbE). Korisnik pretražuje bazu tako što zadaje sliku koja predstavlja upit, a zatim se u bazi pronalazi slika ili slike koje su najslabije upitu.

Osnovni problem je definicija pojma sličnosti. Mišljenja smo da su korisnici koji koriste ovaj modalitet pretrage primarno zainteresirani za vizuelnu sličnost, dok se kod pretrage na osnovu semantike korisnici intuitivno opredjeljuju za tekstualno zadavanje upita. Npr. ako korisnik želi da pronade sve fotografije na kojima se nalazi Bill Clinton (kao u radu [1]), logično je da će koristiti tekstualni modalitet pretrage i kucati ime “Bill Clinton” a ne vršiti upit koristeći neku postojeću sliku sa Billom Clintonom.

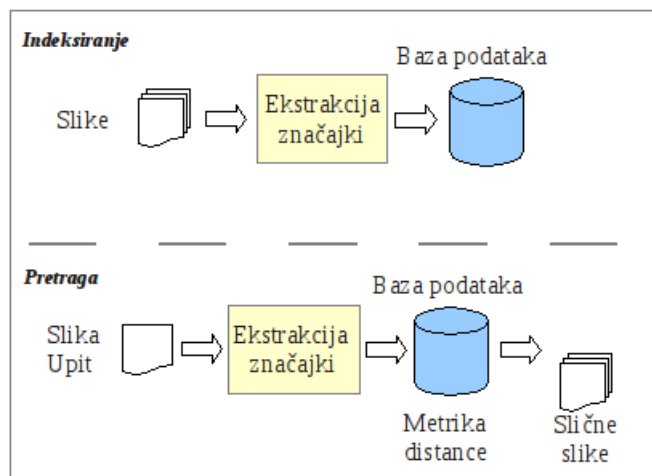
Još jedan interesantan modalitet pretrage je *pretraga na osnovu skice* (sketch). Ovaj način rada je postao posebno interesantan razvojem mobilnih uređaja sa ekranima osjetljivim na dodir. Korisnik na ekranu može grubo skicirati sliku koja ga interesuje, nakon čega se u bazi pronalaze slike koje odgovaraju skici. Kod ovakvih sistema poseban naglasak se treba staviti na prostorni raspored vizuelnih elemenata, a u manjoj mjeri na njihovu boju ili semantičko značenje. U literaturi su zastupljene i razne vrste kombinovanih načina zadavanja upita (primjer i tekst ili skica i tekst).³

Posebna tema u literaturi su interaktivni sistemi koji koriste *povratnu informaciju* (feedback) korisnika kako bi usavršavali i rafinirali rezultate. Nakon što se korisniku prezentuje prvih nekoliko rezultata, sistem dobija povratnu informaciju od korisnika na osnovu koje se ponovo procjenjuje relevancija svakog rezultata i time vjerovatno dobijaju rezultati bliži onome što je korisnik tražio. U literaturi su opisani i sistemi koji uče iz baze slika, pokušavajući usavršiti prepoznavanje oblika na osnovu priložene anotacije, ili automatski predložiti novu anotaciju za postojeće slike.

U posljednje vrijeme posebnu popularnost u naučnom svijetu dobili su sistemi zasnovani na konceptu *vreće vizualnih riječi* (bag-of-visual-words, BoVW). Osnovna ideja je pokušati primijeniti metode pretrage teksta na slike tako što će se slika posmatrati kao dokument sastavljen od “riječi”. U ovom slučaju “riječi” su elementi slike izolovani algoritmima za segmentaciju i prepoznavanje oblika. Ove “riječi” se zatim zapisuju u formi deskriptora kao što je SIFT, te se konačno na osnovu tih deskriptora formira kodna knjiga (codebook) [9].

U radu [6] istaknuto je i pitanje stepena specijalizacije sistema. Metode CBIR našle su svoju primjenu u medicini, sigurnosti, robotici itd. te su razvi-

³Ovdje nećemo detaljnije citirati takve radove, a zainteresirane referiramo na pregledne radove kao što su [2, 8, 6].



Slika 1: Blok šema CBIR QbE sistema

jeni brojni specijalizirani sistemi sa uskim fokusom na određenu primjenu. S druge strane, nastavljeno je istraživanje opšteg problema organizacije slika na Internetu kao i opštih kolekcija fotografija, umjetničkih kolekcija itd. Drugi primjer specijaliziranih CBIR sistema su sistemi za automatsku klasifikaciju slika, kao što je razlikovanje fotografija od vještački generisanih slika (grafikoni, tekst, snimci ekrana), prepoznavanje osjetljivog sadržaja, prepoznavanje lica i sl. [10]

Fokus ove doktorske teze je pretraživanje kolekcije slika na osnovu primjera, i to u korisničkom slučaju pretraživanja lične kolekcije fotografija, mada nećemo zanemariti ni slučaj pretraživanja weba. Iako su mnoge od tvrdnji i metoda istraženih u ovom radu vjerovatno primjenjive i na druge modalitete CBIR, nećemo tome pridavati pažnju zbog ograničenog opsega teze. Poseban naglasak biće stavljen na pitanje zauzeća resursa,⁴ dijelom i zato što je to glavni izazov kod pretrage kolekcije videa.

2.3 Princip rada CBIR sistema

Na slici 1 data je blok šema CBIR sistema koji vrši pretraživanje na osnovu primjera, što je i fokus predložene teme doktorskog rada. Možemo prepoznati

⁴Naglašavamo da se termin *performanse* (eng. performance) u literaturi na temu pretraživanja informacija (IR) uglavnom koristi u značenju “sposobnost metode pretraživanja da vrati relevantne rezultate”. Zbog toga, da bismo izbjegli zabunu, kada budemo govorili o algoritamskoj složenosti ili tipičnom vremenu izvršenja algoritma ekstrakcije značajki, računanja distance itd. koristićemo termin *zauzeće resursa* imajući u vidu da je procesor najinteresantniji računarski resurs sa aspekta CBIRa, te da većina sadašnjih računara posjeduje više fizičkih jezgri kao i druge oblike hardverske podrške za paralelno izvršavanje (hiperniti, eng. hyperthreading).

dvije osnovne funkcije koje obavlja CBIR sistem, a to su *indeksiranje* i *pretraga*. Indeksiranje je početna operacija pripreme sistema za pretragu. Pored same originalne slike, potrebno je u bazu podataka pohraniti i značajke koje su nekim algoritmom ekstraktovane iz slike. U pravilu ove značajke su u formi niza numeričkih veličina, pa govorimo o *vektoru značajki* (feature vector).

Kod pretrage, nakon što korisnik priloži sliku koja je primjer odnosno upit, iz nje se također ekstraktuju značajke a zatim se takve značajke pretražuju u bazi podataka. U nekim aplikacijama interesuje nas samo identična slika tako da je u pitanju jednostavno binarno poređenje vektora značajki. No mnogo je češći slučaj kada korisnik traži niz slika poredanih po sličnosti sa upitom kako bi sam procijenio/la da li je to ono što traži. Zato se uvodi poseban algoritam koji utvrđuje stepen sličnosti između dva vektora značajki, obično izražen kao realan broj u intervalu $[0,1]$. Ovaj algoritam nazivamo *metrika distance*. Konačno, vrlo važno pitanje je i način predstavljanja ovih informacija korisniku, odnosno korisnički interfejs, te eventualno mogućnost korisnika da dostavi povratnu informaciju (relevance feedback) pomoću koje će rafinirati rezultate.

Literatura oblasti CBIR fokusira se na algoritme za rješavanje dva ključna problema, a to su ekstrakcija značajki iz slike (feature extraction) i određivanje najbliže slike odnosno metrika distance (distance metric). Značajke (features) slike predstavljaju sažete podatke o slici čijim poređenjem se može utvrditi u kojoj mjeri su dvije slike *slične*. Problem koji pokušavamo riješiti može se formulisati na sljedeći način: Koji aspekti slike su značajni za pretraživanje tako da pomoću njih možemo sažeto predstaviti sliku, a moguće ih je matematskim putem izdvojiti iz uobičajene rasterske reprezentacije slike (mreža piksela)?

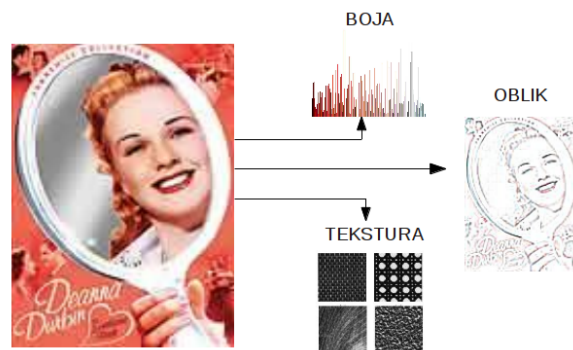
Jedan od prvih cjelovitih sistema za pretraživanje slika opisanih u literaturi je IBMov QBIC [11]. U ovom sistemu kao osnovne značajke za pretragu slika uvedeni su *boja, oblik i tekstura* (slika 2), a to je pristup koji usvaja i većina kasnijih radova. Dostupni su i radovi sa drugim metodama formiranja značajki, koje ćemo također ukratko opisati u ovoj tezi. Ova doktorska teza je strukturirana na način da se za svaki od navedenih tipova značajki daje pregled aktuelnih naučnih dostignuća te se opisuju obavljena istraživanja u toj oblasti.

Budući da je primjenljivost i validnost CBIR metoda već odavno dokazana, naglasak predložene teme doktorskog rada biće na njihovoj efikasnoj praktičnoj implementaciji. Savremene tehnologije postavljaju nove izazove pred sisteme za obradu multimedijalnih podataka. Današnji korisnici multimedije rukuju sa neprevaziđenim količinama podataka. Većina današnjih digitalnih fotografija uslikane su mobilnim telefonima sa 6-8 MP (miliona piksela) ili digitalnim kamerama sa više desetina MP,⁵ a smještene su na medijima od nekoliko terabajta.

2.4 Algoritmi u JPEG kompresovanom domenu

Zauzeće resursa je važan i često zanemaren problem u pretraživanju slika [7]. Problem resursa možemo podijeliti na tri podproblema: zauzeće memorijskog

⁵Na stranici <http://www.flickr.com/cameras/> (preuzetoj 5. 8. 2013.) mogu se naći statistike najčešće korištenih kamera i mobitela na ovom popularnom sajtu za razmjenu fotografija – flickr. Također se mogu dobiti i tehnički podaci o ovim kamerama kao što je rezolucija.



Slika 2: Tri osnovna tipa značajki niskog nivoa u pretraživanju slika su značajke zasnovane na boji (npr. histogram boje), obliku i teksturi.

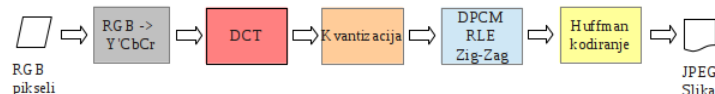
prostora (veličina vektora značajki), vrijeme pronalaska najbliže slike i vrijeme indeksiranja. Prilikom diskutovanja različitih pristupa generisanju vektora značajki (poglavlja 3 i 4) naglašavali smo i veličinu vektora kao jedan od faktora. Metode pronalaska najbliže slike opisali smo u prošlom poglavlju, a sada ćemo prodiskutovati vrijeme indeksiranja.

Problem indeksiranja slika je “sramotno paralelan” (eng. embarrassingly parallel) problem jer većinu vremena predstavlja izračunavanje vektora značajki za datu sliku iz skupa koje je neovisno od ostatka skupa slika, pa možemo jednostavno podijeliti skup slika na različite procesne jedinice. No u slučaju indeksiranja vrlo velikog broja slika, cijena izračunavanja vektora može ipak biti značajna. U radu [7] pokazali smo da čak i kod nesofisticiranog metoda smještaja slika (jeftini hard diskovi) i u slučaju indeksiranja lične kolekcije, te uz primjenu paralelizacije (4-jezgreni procesor), vrijeme procesiranja odnosno opterećenje procesora kod uobičajenih metoda za pretragu slika znatno dominira nad vremenom čitanja slika sa diska.

Već odavno⁶ je u literaturi primijećeno da je većina slika danas pohranjena koristeći JPEG format kompresije, što otvara mogućnosti optimizacije. Naime, JPEG format koristi diskretnu kosinusnu transformaciju (DCT) koja sliku transformiše iz domena piksela (tačaka) u frekventni domen, što omogućuje da se zanemare visokofrekventni podaci koji su malo ili nikako uočljivi ljudskom posmatraču. Ova osobina DCT može biti korisna za efikasnu reprezentaciju slike u vidu vektora značajki, pa je u mnogim radovima na temu CBIR korišten DCT neovisno od kompresije.

Prilikom dekodiranja slike u JPEG formatu, npr. kako bismo odredili njen histogram boja, najprije se vrši inverzni DCT (IDCT) kako bi se dobila slika u piksel domenu. Metode pretrage u kompresovanom domenu preskaču ovu IDCT fazu i izračunavaju statistiku podataka u frekventnom domenu. IDCT

⁶Najstariji rad koje smo pronašli koji opisuje metode za pretragu u kompresovanom domenu je [12] iz 1995. godine.



Slika 3: Proces kodiranja JPEG slike

je vrlo procesorski skupa operacija, pa se efikasnom implementacijom metode u kompresovanom domenu vrijeme indeksiranja može drastično smanjiti [7]. Na manjinu slika koje nisu u JPEG formatu može se primijeniti DCT koji je podjednako složen kao IDCT.

JPEG format podataka opisan je standardom [13]. U pripremnom istraživanju za ovu tezu implementiran je potpuno funkcionalan JPEG koder i dekode⁷, te su ovdje prezentovana iskustva i zaključci zasnovani na njemu. No ovaj dekode⁷ nije imao dovoljno kratko vrijeme izvršenja, pa smo za kasnije eksperimente koristili modifikovanu verziju libjpeg. libjpeg je standardna biblioteka za (de)kodiranje slika u JPEG formatu koju je razvila organizacija Independent JPEG Group (IJG),⁸ a naša modifikacija će biti objašnjena ispod. Pored ove biblioteke dostupna je i biblioteka libjpeg-turbo⁹ koja koristi SIMD instrukcije kako bi dodatno skratilo vrijeme dekodiranja na računarima koji podržavaju SIMD.

U nastavku ćemo ukratko dati opis metoda JPEG kompresije i dekompresije sa aspekta koji je bitan za pretraživanje slika.

Slika u memoriji računara je predstavljena kao niz 32-bitnih vrijednosti koje odgovaraju RGBA vrijednostima (po 8 bita za svaki od RGB kanala, te alfa kanal koji zanemarujemo). Da bismo dobili JPEG datoteku potrebno je da izvršimo pet koraka ilustrovanih na 3:

1. *Konverzija modela boje*. JPEG standard specificira da su slike u Y'CbCr modelu radi veze sa digitalnom televizijom.
2. *Diskretna kosinusna transformacija (DCT)*. U JPEG formatu koristi se dvodimenzionalna varijanta DCT-II koja je opisana formulom

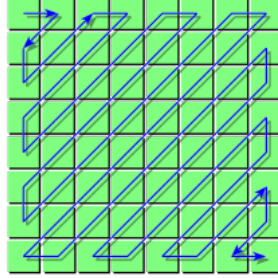
$$F(u, v) = \frac{2C(u)C(v)}{\sqrt{MN}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \cos \frac{(2i-1)u\pi}{2M} \cos \frac{(2j-1)v\pi}{2N} f(i, j). \quad (1)$$

Za primjenu formule (1) najprije je potrebno da sliku izdijelimo u blokove veličine 8x8. U slučaju da dimenzije slike nisu djeljive sa 8, ostatak piksela se dopunjuje crnom bojom a dekode⁷ će takve piksele zanemariti. Zatim se na svaki blok primjenjuje (1) pri čemu su M i N dimenzije bloka ($M =$

⁷Zahvaljujemo se studentu Amiru Duranu na pomoći.

⁸<http://www.ijg.org/> i <http://www.jpeg.org/jpeg/>.

⁹<http://libjpeg-turbo.virtualgl.org/>



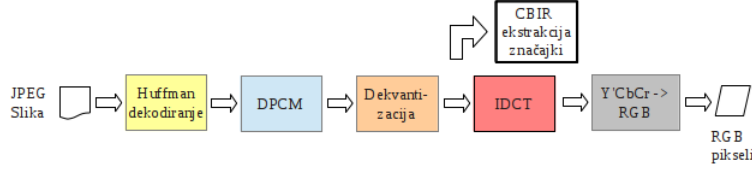
Slika 4: Cik-cak kodiranje u JPEG standardu

$N = 8$), $f(i, j)$ su vrijednosti piksela, a $C(x)$ su normalizacijski koeficijenti dati standardom kao

$$C(x) = \begin{cases} \sqrt{2}/2, & x = 0 \\ 1, & x > 0 \end{cases}.$$

DCT se primjenjuje na svaki od kanala boje odvojeno. Prije izračunavanja (1) može se primijeniti *chroma subsampling*, odnosno umanjeње rezolucije kanala C_B i C_R . JPEG standard podržava više chroma subsampling šema ali najčešće se koristi šema 4:1:1 kod koje se rezolucija C_B i C_R kanala umanjuje dva puta horizontalno i vertikalno, što znači da svakom DCT bloku luma (Y) komponente odgovara po četiri DCT bloka chroma komponenti.

3. *Kvantizacija* osigurava da se vrijednosti visokofrekventnih DCT koeficijenata smanje kako bi se postiglo da većina elemenata u DCT bloku ima vrijednost nula, što olakšava kompresiju. Svaki DCT blok dijeli se kvantizacijskom matricom i zaokružuje na cjelobrojne vrijednosti. Vrijednosti u kvantizacijskoj matrici su određene tako da visokofrekventnim koeficijentima odgovaraju veće vrijednosti. JPEG standard definiše niz kvantizacijskih matrica što omogućuje autoru JPEG slike da odabere “koeficijent kvalitete”. Veća kvaliteta odgovara matrici sa manjim vrijednostima, odnosno manjem efektu kvantizacije nakon zaokruživanja vrijednosti.
4. *Diferencijalna PCM, cik-cak kodiranje i run-length kodiranje* su tri koraka koja se obično implementiraju istovremeno. Na prvi koeficijent svakog DCT bloka (koji se obično naziva DC koeficijent) primjenjuje se diferencijalno kodiranje, odnosno od njega se oduzima vrijednost prethodnog DC koeficijenta. Matrica se linearizuje po cik-cak sistemu ilustrovanom na 4. Pošto su većina koeficijenata pri kraju ovakvog niza nula, uvodi se simbol “kraj bloka”. Kada nađe na ovaj simbol dekodirer će pretpostaviti da su ostali koeficijenti nula čime je većina blokova znatno skraćeno, što smo označili kao run-length kodiranje.



Slika 5: Proces dekodiranja JPEG slike na kojem je naznačena tačka u kojoj se ekstraktuju značajke za potrebe pretraživanja

5. Na kraju, JPEG standard specificira varijantu Huffmanovog algoritma kompresije. Standardom su definisane četiri kodne tablice za DC odnosno AC koeficijente luma i chroma kanala.

Zaglavlje JPEG datoteke čuva mnoge korisne metapodatke kao što su dimenzije slike, korišteni stepen kvalitete (kvantizacijska matrica), chroma subsampling itd. Pored ovoga, JPEG format podržava i nekoliko podformata koji su rjeđe korišteni uključujući kompresiju bez gubitaka (u potpunosti se izostavlja DCT i kvantizacija), korištenje drugih modela boja itd. no podrška za ove podformate u aplikacijskom softveru je slaba. Exif standard definiše specifikaciju dodatnih metapodataka koji se mogu uključiti u zaglavlje kao što su: model i tip kamere, rezolucija slike, datum i vrijeme slikanja, parametri kamere, geolokacijski podaci itd.

Da bismo ovako kodiranu JPEG datoteku mogli koristiti npr. za prikaz na ekranu ili izračunavanje histograma boja, moramo izvršiti dekodiranje JPEG slike, inverzan proces koji je ilustrovan na 5:

1. Najprije se dekodira slika po spomenutom Huffmanovom algoritmu.
2. Na krajeve blokova se dodaju nule, blokovi se po cik-cak sistemu pretvaraju iz niza u matricu, te se ponavlja DPCM kako bi se poništio efekat.
3. Svaki blok se množi za istom kvantizacijskom matricom koja je korištena prilikom kodiranja.
4. Primjenjuju se formule za Inverzni DCT (IDCT).

$$\tilde{f}(i, j) = \sum_{u=0}^M \sum_{v=0}^N \frac{C(u)C(v)}{\sqrt{MN}} \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} F(u, v). \quad (2)$$

5. Slika se iz Y'CbCr modela konvertuje u RGB model.

Algoritmi u kompresovanom domenu operišu nad podacima u DCT domenu, što znači da se proces dekodiranja opisan iznad prekida nakon trećeg koraka. U praktičnoj implementaciji JPEG dekodera, formula (2) je procesorski najzahtjevniji dio. Naivna implementacija uključuje računanje 128 kosinusa za svaku tačku na slici što se unaprjeđuje LUTama.

Ovdje je vrlo važno napomenuti da je potrebno izvršiti dekvantizaciju (treći korak) osim ako je u radu izričito navedeno suprotno (npr. [14] koristi neke koeficijente u kvantiziranom a neke u nekvantiziranom obliku). Ukoliko se dekvantizacija preskoči, algoritam može prividno dati bolje rezultate na nekim skupovima (npr. Wang1000).

3 Stanje oblasti

U prethodnom poglavlju dat je opšti pregled oblasti CBIR, definicije bitnih pojmova sa posebnim naglaskom za pretraživanje slika u JPEG kompresovanom domenu. Na odgovarajućim mjestima citirana je i relevantna literatura (pregledni i drugi interesantni radovi). Iscrpan pregled oblasti CBIR zahtijevao bi znatno više prostora i biće dat u finalnoj varijanti doktorske teze, a u nastavku će biti dat pregled uže oblasti pretraživanja slika na osnovu sadržaja u JPEG kompresovanom domenu.

Pokušali smo dati klasifikaciju metoda u kompresovanom domenu prema tipu značajke, što se može vidjeti u nastavku teksta. Tu spadaju i metode čiji autori nisu naglasili njihovu primjenu u JPEG kompresovanom domenu, ali je u njima korišten DCT te smo ih uspješno primijenili u kompresovanoj domeni. Posebno smo istakli probleme u implementaciji različitih metoda kojima se poništava dobitak performansi.

Prilikom pregleda radova na datu temu, treba obratiti pažnju da neki radovi opisuju metode u kompresovanom domenu u širem smislu, odnosno to ne mora biti DCT kompresovani domen koji se koristi u JPEG kompresiji nego recimo domen Fourierove transformacije, jedne od wavelet transformacija ili neki drugi oblik kompresije.¹⁰ S druge strane, radovi u DCT domenu ne moraju uvijek imati smisla pri radu sa JPEG datotekama, npr. zato što se koriste blokovi koji nisu dimenzija 8x8 ili zato što su procesorski tako zahtjevni da se gubi prednost u performansama.

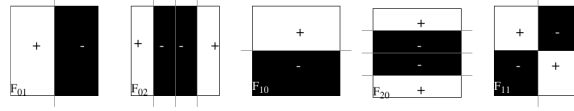
Još jedna stvar na koju treba obratiti pažnju su metode koje simultano vrše indeksiranje i kompresiju, kao što je CVPIC [16]. Ove metode ne operišu u DCT domenu nego u piksel domenu, a paralelno se vrši kompresija što može biti od interesa za multimedijalne baze podataka, no takve metode nisu u fokusu ove doktorske teze.

3.1 Histogram boja i druge značajke na osnovu boje

Iz formule (1) možemo zaključiti da koeficijent na koordinatama (0,0) odnosno DC koeficijent predstavlja srednju vrijednost čitavog bloka. Ovo nam omogućuje da formiramo histogram boja u $Y'CbCr$ modelu na slici umanjenoj osam puta po horizontalnoj i vertikalnoj dimenziji.

Opisani histogram boja korišten je između ostalih u radovima [17, 18]. Neki radovi koriste "histogram prvog DCT koeficijenta" ne prepoznajući da je u pitanju običan histogram boja (npr. [19, 20]).

¹⁰Za pregled koji između ostalog navodi i mnoge takve radove vidjeti [15].



Slika 6: Fizička značenja pet karakterističnih AC koeficijenata (slika preuzeta iz [14]). Npr. koeficijent F_{01} (u našoj notaciji $C(0, 1)$) ima pozitivnu vrijednost ako dominiraju vrijednosti u lijevoj polovini bloka 8x8, a negativnu ako dominiraju vrijednosti u desnoj polovini.

Jiang i dr. [21] pokazali su da se sljedećim jednostavnim formulama mogu izračunati srednje vrijednosti podblokova veličine 4x4, pa se tako može dobiti histogram boja na slici umanjenoj četiri puta:

$$\begin{aligned}
M_{11} &= \frac{2C(0,0) + 2C(1,0) + 2C(0,1) + 2C(1,1)}{16} \\
M_{12} &= \frac{2C(0,0) + 2C(1,0) - 2C(0,1) - 2C(1,1)}{16} \\
M_{21} &= \frac{2C(0,0) - 2C(1,0) + 2C(0,1) - 2C(1,1)}{16} \\
M_{22} &= \frac{2C(0,0) - 2C(1,0) - 2C(0,1) + 2C(1,1)}{16}
\end{aligned}$$

gdje je $C(x, y)$ koeficijent DCT bloka na koordinatama (x, y) , a M_{11} , M_{12} , M_{21} , M_{22} su srednje vrijednosti za podblokove veličine 4x4. Vidimo da su ove formule i dalje znatno jednostavnije od (2). Pokušaji da dalje poboljšamo rezoluciju doveli bi do povratka na formule (2) čime gubimo prednosti rada u DCT domenu.

U radu [22] korišten je dijeljeni histogram boja na slici umanjenoj četiri puta po formulama datim iznad. Kreiran je histogram sa 256 korpi po kanalu, što ukupno daje vektor od 768 značajki, te je korištena L_2 norma kao metrika distance. Pokazali smo da prelaskom na kombinovani (linkovani) histogram sa 256 koeficijenata i Matsushita distancu možemo dobiti drastično poboljšanje performansi, odnosno performanse bliske sličnom histogramu u piksel domeni.

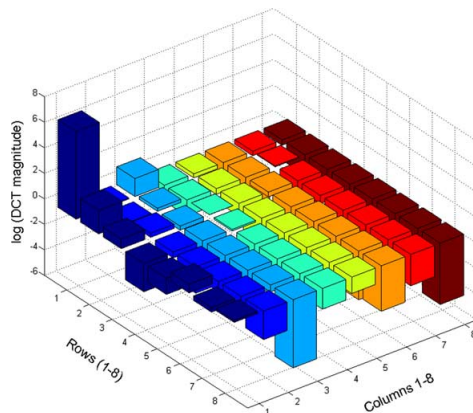
Jiang i dr. [23] opisuju način za ekstrakciju MPEG-7 DCD (Dominant Color Descriptor) iz kompresovane domene te dokazuju da ovakav metod daje približne rezultate kao u piksel domeni. Islam i Ali [24] daju metod za ekstrakciju značajke izuzetno slične MPEG-7 CLD (Color Layout Deskriptor).

3.2 “Naivne” metode u kompresovanoj domeni

Jedan broj radova zasnovano je na manipulaciji koeficijentima u kompresovanom domenu bez razumijevanja njihovog matematskog značenja.

Hatzigiorgaki i Skodras [20] mjerili su performanse pretrage sa histogramima DCT koeficijenata sa 512, 1024 i 2048 korpi.¹¹ U ovom eksperimentu korištena

¹¹Iz (1) i činjenice da su $f(i, j) \in [0, 255]$ cijeli brojevi, možemo zaključiti da su $F(u, v) \in$



Slika 7: Distribucija vrijednosti DCT koeficijenata u tipičnom 8x8 bloku (slika preuzeta iz [25]).

je samo luminansa, odnosno pretraga je vršena na crno-bijelim slikama. Pored ranije opisanih histograma DC koeficijenata, provjerili su i histograme kod kojih su ubrojani i prva tri odnosno prvih osam AC koeficijenata. Očekivano, rezultat je bio da nema prevelike razlike između ovih tipova histograma, štaviše uočen je blagi pad performansi nakon uključenja AC koeficijenata jer je njihova semantika zanemarena.¹² Također uočen je neznatan pad performansi prelaskom na histogram sa 512 korpi što je saglasno sa [19].

Lay i Guan [17] prave drugačiju grešku: na isti histogram ubrojavaju koeficijente za različite kanale boje, pa dolaze do vrlo loših performansi pretrage uz zaključak suprotan od [20], da je kombinovanje određenog broja AC koeficijenata korisno za pretragu.

U više radova [26, 27, 28] opisana je metoda formiranja histograma ne iz vrijednosti DC koeficijenata nego iz njihovih razlika, preskačući faze diferencijalnog dekodiranja i dekvantizacije. Pored toga, Chang i dr. [27] primjenjuju sličnu metodu i na prvih devet AC koeficijenata svakog DCT bloka. Algoritmi predstavljeni u ovim radovima nisu bili dovoljno testirani u literaturi. Komparativna studija [29] pronalazi da je metod iz rada [26] dao slabije rezultate od običnog histograma boja, a ukupan broj procesorskih operacija potrebnih za ekstrakciju značajke je sličan kao kod drugih metoda u kompresovanom domenu.

Climmer i Bhatia [30] koriste DCT koeficijente kao mehanizam za efikasno filtriranje skupa slika. Za svaku sliku u bazi formira se quad-stablo DCT blokova, pri čemu interni čvorovi stabla čuvaju srednju vrijednost DC koeficijenata svoja četiri djeteta te vrijednost najvećeg AC koeficijenta. Dakle, u korijenu stabla

[−1024, 1023]. Osim toga, u JPEG datotekama $F(u, v)$ su zaokruženi na cijele brojeve u fazi kvantizacije. Dakle, u histogramu sa 2048 korpi praktično nema nikakve dodatne kvantizacije osim one koju uvodi sam JPEG format.

¹²Mišljenja smo da korištenje prvih nekoliko AC koeficijenata nema previše smisla obzirom da se koristi cik-cak kodiranje (4). Pretpostavljamo da pod “prvih osam koeficijenata” autori misle na koeficijente sa koordinatama (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1) i (2,2).

nalazi se srednja vrijednost DC koeficijenta i najveći AC koeficijent za čitavu sliku. Prilikom pretrage primjenjuje se sljedeća metrika na korijen stabla:

$$d(q, n) = \frac{|c_q - c_n|}{\max(d_q, d_n)}$$

pri čemu c_q, c_n označava srednju vrijednost DC koeficijenta, a d_q, d_n dominantni AC koeficijent. Svi rezultati za koje je distanca veća od neke granice se odbacuju, a zatim se poređenje prenosi na drugi nivo quad-stabla gdje se računa distanca za odgovarajuće čvorove i uzima njihova srednja vrijednost. Proces se rekurzivno nastavlja prema dnu stabla.

Kao što znamo, DC koeficijent sadrži prosječan intenzitet slike, dok semantika najvećeg AC koeficijenta nije najjasnija. Ova metoda bi mogla pogrešno klasificirati dvije vrlo slične slike sa različitim prosječnim intenzitetom (npr. ista scena pod različitim osvjetljenjem). Osim toga, autori nisu diskutovali implementaciju na slikama u boji, odnosno posmatra se samo Y komponenta. U literaturi nismo pronašli komparaciju ove metode sa drugim metodama.

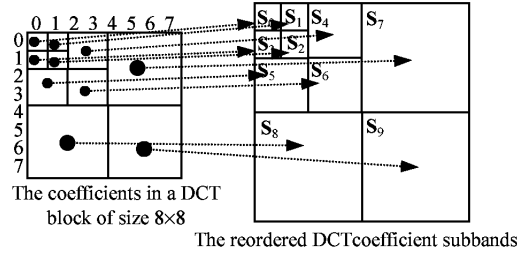
Zhong i Defée [25] predlažu metodu zasnovanu na histogramu čitavih DCT blokova. Kompletan blok se posmatra kao jedinstvena značajka i formira se histogram pojavljivanja te značajke na slici. Naravno ovakav histogram bi bio nepraktičan pošto je broj različitih kombinacija DCT blokova ogroman, pa se vrši sljedeće procesiranje: najprije se formiraju blokovi veličine 4x4 koristeći metodu koju su predložili Jiang i dr. [21]. DC koeficijent se eliminiše i posmatraju se samo AC koeficijenti. Svi koeficijenti se dodatno kvantiziraju na način da se dijele jedinstvenim faktorom kvantizacije (npr. $QP = 22$).

Ovakvim procesiranjem većina blokova će imati sve vrijednosti koeficijenata nule. No bez obzira na to, histogram na nivou kompletne baze slika će i dalje biti vrlo velik, a većina njegovih korpi će biti prazne. Stoga autori predlažu da se u bazi pohranjuje samo prvih nekoliko korpi histograma (npr. 40). Kao metrika distance predlaže se L_1 norma.

Ova značajka je u radovima autora dala dosta dobre rezultate, no bitno je naglasiti da su se oni fokusirali isključivo na problem prepoznavanja lica i to na skupovima crno-bijelih fotografija. Pored toga, predloženi algoritam zahtijeva treniranje na skupu slika koje znatno mijenja njegove performanse. Između ostalog, treniranje se koristi da bi se normalizovalo osvjetljenje slike, odredio koeficijent kvantizacije, broj, pa čak i konkretan odabir korpi histograma (npr. svaka druga korpa). Ovako intenzivnim treniranjem algoritam se prilagođava skupu do te mjere da je upitna generalnost njegove primjene.

3.3 Značajke zasnovane na teksturi

Procesiranje slike korištenjem DCT transformacije (npr. posmatranjem samo prvog koeficijenta) nema skoro nikakav efekat na značajke zasnovane na boji, kao što smo vidjeli. No kod značajki na bazi teksture ili rubova bitne informacije sadržane su u visokofrekventnom opsegu koji DCT transformacija eliminiše [31]. Stoga jedan od bitnijih izazova u kompresovanom domenu je pronalazak adekvatnih teksturnih značajki.



Slika 8: Preraspodjela DCT koeficijenata opisana u radu [32] (slika preuzeta iz istog rada).

Savremene značajke zasnovane na teksturi uglavnom koriste neku vrstu wavelet transformacije kako bi se izolovala frekvencija i orijentacija izolovanih segmenata slike. No već odavno je prepoznato da DCT posjeduje određene sličnosti sa wavelet transformacijama [12, 32]. Analiza teksture u transformacijskom domenu proučava se još od ranih 70tih godina, s tim što su radovi iz tog perioda uglavnom koristili diskretnu Fourierovu transformaciju (DFT) [33].

Najjednostavnija značajka sastojala bi se od varijanse (srednje vrijednosti i standardne devijacije) nekoliko početnih AC koeficijenata (recimo 8) [34, 35]. U citiranim radovima navodi se da ovakva značajka karakteriše teksturu slike bez detaljnije diskusije. Posebno treba istaći jednostavnost kalkulacije ovakve značajke, no kao i iznad, mišljenja smo da koeficijente treba pametnije odabrati.

Na tom tragu, u jednom od prvih radova na temu pretraživanja slika u kompresovanom domenu [12], Chang predlaže teksturnu značajku zasnovanu na statističkoj analizi DCT koeficijenata, npr. koristeći momente prvog reda nad DCT koeficijentima. Nakon što se odredi tekstura jednog 8x8 DCT bloka, Smith i Chang predlažu korištenje quad-tree strukture podataka za grubu segmentaciju slike prema sličnosti tekstura [36]. Ovo quad-stablo ujedno omogućuje pretragu slika po sličnim teksturnim segmentima. Za smanjenje dimenzionalnosti koristi se analiza Fisherovih diskriminanti koja zahtijeva treniranje nad nekim fiksnim skupom slika. Reeves i dr. [37] koriste isti pristup s tim što umjesto Fisherovih diskriminanti smanjuju dimenzionalnost tako što koriste samo prvih osam AC koeficijenata.

Mišljenja smo da citirani radovi ne pružaju dovoljno informacija za implementaciju i provjeru validnosti, ali da nude niz ideja za teksturne značajke od kojih su neke isprobane u kasnijim radovima. Islam i Ali [24] navode vrlo slabe rezultate pretrage za ovaj algoritam, ali zbog brojnih metodoloških pogrešaka u ovom radu taj rezultat se ne može smatrati ozbiljnim.

Huang i Chang [32] pokazuju da preraspodjela DCT koeficijenata po određenoj šemi daje strukturu vrlo sličnu multirezolucijskom piramidalnom waveletu. Vektor značajki sastoji se od srednjih vrijednosti i standardnih devijacija tako dobijenih devet koeficijenata, što daje ukupno 18 značajki. Distanca se određuje kao suma apsolutnih razlika tako dobijenih značajki, normalizovanih po standardnoj devijaciji same značajke. U radu je pokazano da ovakva značajka daje bolje

rezultate od 16-Daubechies waveleta kao primjera ortogonalne wavelet transformacije.

Sličnu globalnu značajku na bazi tekstone ponudili su Ngo i dr. [18]. U pitanju je varijansa prvih devet AC koeficijenata po cik-cak redoslijedu. Zbog velike varijacije dobijenih devet vrijednosti, izračunava se matrica kovarijanse, a kao mjera sličnosti koristi se Mahalanobis distanca.

Feng i Jiang [38] pokazali su da se iz DCT koeficijenata mogu vrlo jednostavno odrediti srednja vrijednost i varijansa blokova veličine 8×8 , 4×4 i 16×16 . Tako dobijene formule su iskorištene za kreiranje sistema za pretragu slika. Značajka korištena u ovom sistemu je kombinovani histogram srednjih vrijednosti i standardnih devijacija pri čemu je srednja vrijednost kvantizirana na četiri moguće vrijednosti, a standardna devijacija na sedam vrijednosti, čime se dobija histogram od 28 elemenata. Posebna metrika distance pod nazivom “težinska transformacija distance” je predložena sa ciljem boljih performansi pretrage. U komparativnoj studiji metoda u kompresovanom domenu [29], metoda Feng i Jiang dala je vrlo loše rezultate pretrage.

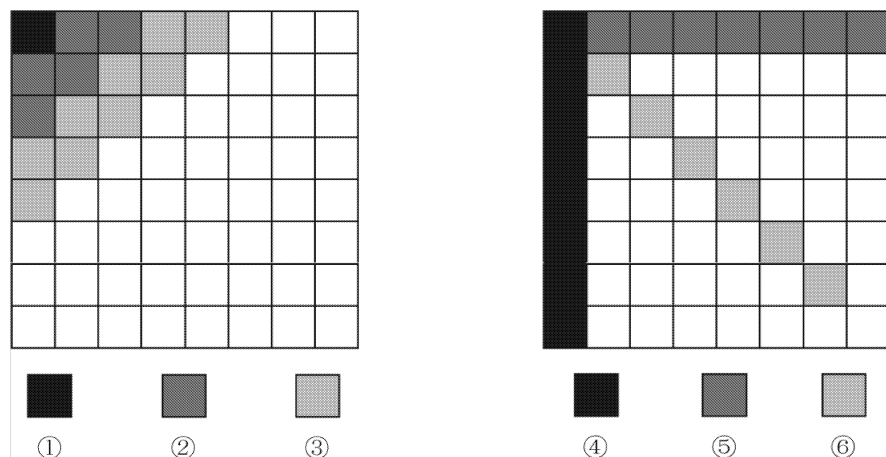
Zhong i Defée su u svojim radovima [25, 39] predložili više značajki koje pokušavaju opisati direkcionalnost posmatrajući odnos DC koeficijenta u bloku 4×4 ¹³ sa DC koeficijentom osam okolnih blokova. U radu [25] značajku čine oznake nekoliko (recimo četiri) susjedna bloka sa najvećom razlikom, a na osnovu te značajke se formira histogram. U radu [39] formira se binarna značajka pri čemu vrijednosti osam bita određuje to da li je razlika sa odgovarajućim susjednim blokom veća od granične (threshold). Autori evaluiraju svoje metode isključivo na bazama sastavljenim od crno-bijelih fotografija lica.

Bae i Jung [40] predstavili su potpuno drugačiji pristup karakterizaciji tekstone u DCT domenu. Sumiranjem prvih 5, zatim sljedećih 9, te sljedećih 13 AC koeficijenata po cik-cak redoslijedu može se dobiti frekventna karakteristika odnosno neka vrsta mjere dimenzija ponavljajućeg uzorka. Zatim autori putem slike definišu pet kategorija AC koeficijenata pomoću kojih se može prepoznati da li blok 8×8 ima dominantno vodoravnu, uspravnu ili dijagonalnu karakteristiku. Uz DC koeficijent ovih osam suma čine vektor značajki bloka 8×8 , no u radu nije specificirano kako se iz vektora blokova izračunava vektor za čitavu sliku.

Lu i dr. [22] su modificirali ovu značajku i dodatno pojasnili njeno izračunavanje. Broj značajki smanjen je sa devet na šest, mada je osnovni princip njihovog odabira vrlo sličan (9). Srednje vrijednosti i standardne devijacije ovih šest suma za sve blokove na slici predstavljaju 12 značajki slike. Te značajke se izračunavaju odvojeno za svaki od tri kanala boje, da bi se dobilo ukupno 48 značajki. Nakon indeksiranja svih slika u bazi izračunava se varijansa svake značajke, te se kao metrika distance koristi se L_2 norma varijansi.

U citiranom radu [22] postoji nekoliko pogrešaka koje otežavaju implementaciju. Računanje suma prema 9 preuzetom iz tog rada davalo je neupotrebljiv rezultat za sumu broj 4 jer je DC koeficijent znatno veći od AC koeficijenata pa je do-

¹³Ovakvi blokovi su dobiveni iz 8×8 blokova koristeći metodu koju su predložili Jiang i dr. [21]



Slika 9: Teksturna značajka opisana u radu [22] (slika preuzeta iz istog rada). Da bi se dobile značajke, najprije je potrebno izračunati sumu šest grupa koeficijenata označenih na slici. Srednje vrijednosti i standardne devijacije ovih šest suma predstavljaju 12 značajki. Eksperimentalno smo utvrdili da na slici postoji greška: u grupi (4) ne treba uključiti DC koeficijent u gornjem lijevom uglu.

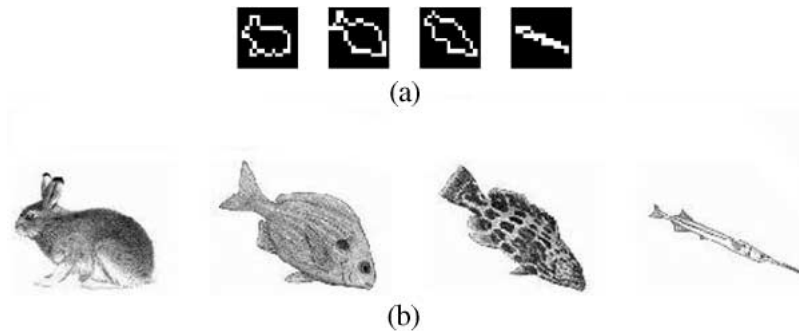
minirao u toj sumi. Nakon što smo izostavili DC koeficijent iz sume, uspjeli smo reproducirati rezultate iz rada. Dalje, većina radova na temu teksture, uključujući i citirani [40] koriste crno-bijele slike (luma kanal) za određivanje teksture, dok Lu i dr. računaju teksture za sva tri kanala boje.

Komparativna studija metoda u kompresovanom domenu [29] pokazuje da metod Lu i dr. daje bolje rezultate od četiri druge metode u kompresovanom domenu ([21, 17, 26, 41]), ali i dalje slabije od običnog histograma boje. Testovi su sprovedeni na MPEG-7 skupu slika.

Primjer strukturnog modela teksture u DCT domenu opisan je u radu [42]. Proces indeksiranja započinje ekstrakcijom DCT 8x8 blokova iz čitavog skupa Brodatz tekstura. Svi blokovi su pohranjeni u bazu, eliminisani su duplikati, a zatim je izvršen clustering u 64-dimenzionalnom prostoru kako bi se smanjio broj tačaka. Prilikom formiranja vektora značajki za datu sliku, svi njeni DCT blokovi se pretražuju u tako formiranom indeksu i zamjenjuju najbližim rezultatom. Problem ovakvog pristupa je njegova ovisnost o rezoluciji slika, budući da su blokovi 8x8 relativno mali i ne mogu uspješno kodirati teksturu, pa bi vrijedilo razmisliti o povećavanju veličine bloka na 16x16 ili više koristeći pristup opisan u radu [21].

3.4 Detekcija rubova

Rani pokušaji detekcije rubova u kompresovanom domenu opisani su u radu [43]. U ovom radu Abdel-Malek i Hershey pokazuju da se posmatranjem DCT koefi-



Slika 10: Konture (a) ekstraktovane iz slika (b) koristeći algoritam za prepoznavanje oblika iz rada [18].

cijenata može uočiti da li posmatrani blok sadrži primarno vodoravne, uspravne ili dijagonalne osobine.

Znatno detaljnije razmatranje uz primjenu na pretragu slika dali su Shen i Sethi [14]. Jednostavnim poređenjem prva dva AC koeficijenta može se utvrditi da li je rub horizontalan, vertikalni ili dijagonalan. Formulama koje uključuju dodatne koeficijente može se utvrditi nagib, ofset i jačina ruba. U radu [14] pokazano je da ovakva metoda može dati rezultate uporedive sa Sobelovim operatorom detekcije ruba.

Na osnovu rada [14], Eom i Choe [41] nude metodu za ekstrakciju MPEG-7 Edge Histogram Descriptor (EHD) značajke u kompresovanom domenu. Prema njihovim rezultatima, njihova EHD značajka daje u prosjeku neznatno bolje rezultate od standardne EHD značajke. No u komparativnoj studiji [29] EHD značajka je dala oko 10% slabije rezultate od EHD.

3.5 Oblik

Nakon zaključka da se DCT koeficijenti mogu reorganizirati tako da predstavljaju Mandala koeficijente, Ngo i dr. [18] ponudili su sljedeću jednostavnu formulu za gradijent slike:

$$\nabla f = \sqrt{C(0, 1)^2 + C(1, 0)^2}$$

gdje su $C(x, y)$ DCT koeficijenti kao i ranije. Na ovaj način se može izvršiti detekcija rubova slike koristeći samo prva dva AC koeficijenta u luma kanalu. Autori zatim koriste ovaj gradijent da bi odredili konturu objekta (10)¹⁴ za koju izračunavaju momente $\mu_{p,q}$, koji su zatim normalizovani kao $\eta_{p,q}$ te je izračunato sedam invarijanti ϕ_1 - ϕ_7 na translaciju, rotaciju i skaliranje, a sve to prema formulama datim u radu. Zbog velike varijacije koeficijenata ϕ_i izračunava se matrica kovarijanse. Za poređenje sličnosti korištena je Mahalanobis distanca.

¹⁴Nije dat algoritam za određivanje konture, no ovaj problem je poznat u literaturi.

3.6 Kombinovane značajke

Ngo i dr. [18] formiraju kombinovanu značajku sastavljenu od histograma boja, oblika i tekstone. Rad [22] opisuje dvije odvojene značajke: histogram boja na slici umanjenoj četiri puta po horizontalnoj i vertikalnoj dimenziji i značajku zasnovanu na teksturi. Ove dvije značajke su kombinovane koristeći težinsku sumu distanci. Koeficijenti za težinsku sumu dati u radu [22] su apsurdni (efekat teksturne značajke je zanemaren) i u našim eksperimentima nisu dali adekvatne rezultate.

3.7 Ostali algoritmi i metode u kompresovanom domenu

U literaturi je opisan niz tehnika u kompresovanom domenu koje izlaze izvan opsega ove disertacije. Posebna pažnja posvećena je problemu pretraživanja videa, a konkretni predloženi algoritmi tiču se prepoznavanja rezova kamere (camera cuts), odnosno segmentacije videa kako bi se izbjeglo nepotrebno indeksiranje svakog zasebnog okvira filma.

Jedan broj radova bavi se pretraživanjem slika koristeći metapodatke sadržane u JPEG datotekama. Npr. rad [44] opisuje metodu filtriranja slika zasnovanu na analizi kodnih tablica za Huffman kodiranje.

Shneier i Abdel-Mottaleb [45] kao značajke za pretragu odnosno “ključeve” (keys) koriste srednje vrijednosti DCT koeficijenata izračunate nad “prozorom” (segmentom slike). Slika je izdijeljena na $2K$ prozora, pri čemu je veličina prozora ovisna o veličini slike i određuje se kao minimalni umnožak broja 8 za koji se dobije neko fiksno, unaprijed određeno K . U svakom prozoru izračunavaju se srednje vrijednosti svakog od 64 DCT koeficijenata. Prozori se zatim uparaju metodom slučajnog izbora i poredi se svaka od srednjih vrijednosti. Ako je vrijednost u prvom prozoru veća od vrijednosti u drugom uzima se vrijednost “1”, a u suprotnom vrijednost “0”. Na ovaj način dobija se vektor sastavljen $64 \times K$ binarnih vrijednosti. Ovakvi vektori se upoređuju metodom Hamming distance.¹⁵

Ova metoda pretrage je interesantna jer u kompresovanom domenu uključuje određeni efekat prostornog rasporeda boje i tekstone. U komparativnoj studiji [29] ova metoda je dala vrlo slabe rezultate.

Armstrong i Jiang [46] daju pristup kojim se bilo koja metoda u piksel domenu može primijeniti na parcijalno dekodirane JPEG slike. Unutar svakog 8×8 bloka DCT koeficijenti se sumiraju prema datoj šemi, čime se dobijaju ukupno 4 koeficijenta (DC i tri sumirana AC koeficijenta). Na ovakve koeficijente se primjenjuju IDCT formule (2). Opisani pristup je procesorski znatno manje zahtjevan od primjene IDCT (2) na čitav izvorni blok. Na primjeru jednostavne teksturne značajke u piksel domeni, autori pokazuju da nema gubitaka u performansama pretrage.

¹⁵Pošto nismo uspjeli doći do originalnog rada, detaljan opis dat je prema [35].

4 Ciljevi i plan istraživanja

Osnovni cilj ove doktorske disertacije, koji proizlazi iz detaljne analize stanja u oblasti istraživanja i lične motivacije, je unaprjeđenje performansi postojećih vektora značajki ekstraktovanih iz slika u JPEG kompresovanom domenu za pretraživanje slika na osnovu sadržaja.

Kako bi se ostvario navedeni cilj, istraživanje će biti koncipirano prema sljedećem planu:

1. Izrada detaljnog pregleda stanja u oblasti CBIR sa posebnim naglaskom na algoritme koji operišu u JPEG kompresovanom domenu. Pri tome će biti dat kritički osvrt na prednosti i nedostatke opisanih metoda imajući u vidu pregledne radove i radove koji daju eksperimentalnu komparaciju metoda.
2. Razviti eksperimentalnu metodologiju za poređenje metoda pretraživanja slika. Za ovo je potrebno evaluirati dostupne otvorene skupove podataka, metrike poređenja, te open-source implementacije.
3. Napraviti sistematičnu komparaciju metoda opisanih u literaturi. Kroz ovo potrebno je implementirati određene izabrane metode za koje se smatra da bi mogle dati dobre rezultate. Osnovni cilj ove faze je dati odgovor na pitanje: Da li je posvećivanje posebne pažnje algoritmima u JPEG kompresovanom domenu opravdano sa naučne strane? Pri tome treba se osvrnuti kako na postizanje skraćenog vremena indeksiranja kao i na mogućnost da se takvim algoritmima ostvare rezultati usporedivi sa state-of-the-art metodama koje ne operišu u JPEG kompresovanom domenu.
4. Izabrati nekoliko metoda u kompresovanom domenu i pokušati napraviti unaprjeđenja na njihovim performansama, vodeći se zaključcima iz prethodnih faza istraživanja.

5 Metodologija istraživanja i potrebni istraživački resursi

Uobičajena metoda međusobne komparacije CBIR algoritama su eksperimenti provedeni na standardnim skupovima podataka, pri čemu se poređenjem ostvarenih rezultata pretrage sa nekim idealnim rezultatom formira ocjena algoritma. Za objektivnu ocjenu koristi se određeni broj standardnih metrika za komparaciju algoritama pretrage.

5.1 Skupovi podataka

Bitan problem u istraživanju CBIR metoda su standardizovani skupovi podataka (eng. datasets). Jedan skup podataka za CBIR sastoji se od kolekcije slika, te tzv. bazne istine (eng. ground truth) odnosno skupa pretraga sa definisanim očekivanim rezultatima koji su relevantni za postavljeni upit. Ova bazna

istina može biti definisana i preko ključnih riječi oblika “suma”, “nebo”, “put”, “ljudi” i slično, pa se prilikom pretrage kod koje je upit neka od slika iz kolekcije, za relevantne rezultate mogu smatrati one slike koje su označene istim ključnim riječima kao slika iz upita.

U prošlosti često se dešavalo da se u naučnim radovima prezentuju rezultati koji su potkrijepljeni eksperimentima nad skupovima podataka koji nisu publikovani (“lična kolekcija”, “fotografije sa putovanja” i sl.) ili su zaštićeni autorskim pravima (npr. Corel clipart kolekcija slika). Također se dešavalo da skupovi podataka koji su bili dostupni u nekom periodu vremena postanu nedostupni jer licenca dozvoljava preuzimanje sa originalnog sajta, ali ne i redistribuciju. Posljedica je da nezavisni istraživači nisu u mogućnosti provjeriti i reproducirati rezultate eksperimenata jer testni podaci nisu legalno dostupni [47]. Također postojali su slučajevi manipulacije nad skupom podataka ili baznom istinom na način da se biraju oni podaci koji pogoduju pozitivnom nalazu eksperimenta.

Zbog ovoga je bitno da se u istraživanjima u oblasti CBIR koriste javno dostupni skupovi podataka pod otvorenim licencama (open-source, public domain), te da bazna istina bude nepristrasna i provjerena, a po mogućnosti da dolazi od samih korisnika [48, 49]. Također je bitno da se testiranje ne zadrži na jednom ili dva skupa podataka, te da se uključi i testiranje na opštem skupu (bilo kojoj slici) [3, 50].

S ovim ciljem pokrenut je Benchathlon projekat¹⁶ koji je trebao stvoriti pogodno okruženje za razmjenu resursa unutar CBIR zajednice. Nažalost, ovaj projekat nije zaživio zbog nespremnosti autora da svoje radove podvrgnu kritičkoj evaluaciji.

U pripremnim istraživanjima za ovu doktorsku tezu korišteni su sljedeći javno dostupni referentni skupovi podataka:

- Wang SIMPLIcity <http://wang.ist.psu.edu/docs/related>
- University of Washington (UW) dataset
<http://www.cs.washington.edu/research/imagedatabase/groundtruth/>
- UCID dataset <http://homepages.lboro.ac.uk/~cogs/datasets/ucid/ucid.html>
- MIRflickr <http://press.liacs.nl/mirflickr>

Pored nabrojanih, evaluirani su i drugi skupovi podataka kao što su: MPEG-7 Common Color Dataset, ZuBuD, Nimur, Brodatz itd. U finalnoj doktorskoj tezi će biti dati rezultati ove evaluacije ili će se kandidat opredijeliti da uvrsti i spomenute skupove u eksperimente.

5.2 Metrike poređenja metoda pretraživanja

Pretraživanje slika na osnovu sadržaja (CBIR) je samo jedna od podoblasti pretraživanja informacija (eng. *information retrieval*, *IR*). Mnoga istraživanja

¹⁶Adresa: <http://www.benchathlon.net/>. Autori projekta su iste osobe koje su napisale radove [49, 47].

u oblasti pretraživanja informacija bave se načinom mjerenja i upoređivanja performansi različitih metoda pretraživanja. Prvobitni radovi iz oblasti pretraživanja slika opisivali su svoje rezultate pomoću primjera: ilustracije upita i rezultata (slika) koji su dobijeni. Očigledno na ovaj način autor može izabrati onaj upit koji daje vizuelno efektivne rezultate.

Vremenom kao najčešće standardne metrike performansi u pretraživanju informacija izdvojili su se preciznost i dohvat [51]. Navedene metrike su razvijene u vrijeme kada su sistemi za pretraživanje dobavljali ograničen broj dokumenata zbog performansi. Skoro svi CBIR sistemi zasnovani su na konceptu distance između dvije slike $d(I_1, I_2) \in [0, 1]$. Određuje se distanca između slike upita i svake pojedinačne slike u skupu, a zatim se skup slika sortira po distanci, tako da CBIR sistemi uvijek dobavljaju sve dokumente pa je dohvat uvijek 1 a preciznost $p = \frac{N_R}{N}$. Mogu se posmatrati neke specifične vrijednosti preciznosti i dohvata za dati broj dokumenata ili dati procenat dokumenata kao i *preciznost-dohvat* $p(r)$ (preciznost za dati dohvat) te *dohvat-preciznost* $r(p)$ (dohvat za datu preciznost). Ovakve jednostavne metrike ocjenjuju samo jedan pojedinačni aspekt metoda za pretraživanje i ne daju kompletnu sliku.

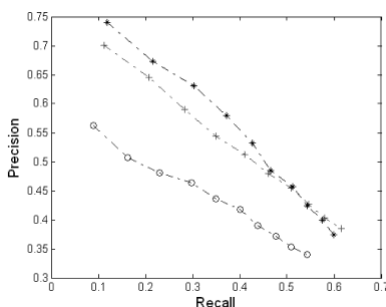
U ovakvim slučajevima možemo koristiti grafički prikaz kao što je *grafikon preciznost-dohvat* (precision-recall graph, PR graph). Da bismo nacrtali PR graf potrebno je da sistem podesimo tako da dobavi sve dokumente i rangira ih po relevantnosti. Zatim određujemo vrijednost preciznosti za onaj broj dokumenata kod kojeg je dohvat imao određene predefinisane vrijednosti (u pravilu: 0,1, 0,2, 0,3 itd. do 1,0 što odgovara potpunom dohvat u svih relevantnih dokumenata u bazi).¹⁷

Dva algoritma možemo jednostavno uporediti tako što spojimo njihove PR grafove kao na slici 11. Primjere ovakve upotrebe možemo naći u radovima [22, 52].

No u slučajevima kada želimo uporediti veći broj CBIR metoda pri tome koristeći niz upita, poželjno je da imamo jedinstven numerički pokazatelj pomoću kojeg možemo tabelarno porediti metode.

U sklopu istraživanja za ovu doktorsku tezu, evaluirano je 13 različitih metrika za poređenje algoritama pretraživanja spomenutih u literaturi iz oblasti pretraživanja [51, 49, 11, 53, 54, 55, 56, 57]. Nakon većeg broja različitih eksperimenata, opredijelili smo se za metrike: MAP [49] i ANMRR [54]. Kako bi se formirala kompletnija slika performansi, biće predložene dvije potpuno nove metrike za poređenje algoritama koje uzimaju u obzir ne-binarnu mjeru relevancije dokumenata.

¹⁷Pri tome je važno napomenuti da je vrijednost preciznosti interpolirana, odnosno za vrijednost preciznosti u određenoj tački grafa uzima se maksimalna preciznost za bilo koju vrijednost dohvata veću ili jednaku vrijednosti u datoj tački. Na taj način postiže se da PR graf monotonno opada te da je definisana njegova vrijednost i za dohvat 0,0 što inače ne bi imalo smisla. Vidjeti [51].



Slika 11: Primjer poređenja nekoliko metoda pomoću PR grafa. Ilustracija preuzeta iz rada [22].

5.3 Eksperimentalni softver

U sklopu istraživanja za ovu doktorsku tezu, evaluirano je šest različitih open-source softverskih alata za CBIR, budući da se analizom closed-source rješenja ne mogu formirati nikakvi naučno relevantni zaključci niti se može dopuniti njihova implementacija kako bi se uporedila nova predložena metoda. Zaključeno je da od evaluiranih alata samo jedan pruža mogućnost implementacije algoritama u kompresovanom domenu (što je cilj definisane teme istraživanja).

Stoga je započet rad na razvoju softvera ETFImageSearch – aplikacije za pretraživanje slika koja će služiti kao testna platforma za usporedbu različitih metoda pretraživanja, reproduciranje rezultata pokazanih u korištenim naučnim radovima. Spomenuti softver dostupan je na adresi:

<http://people.etf.unsa.ba/~vljubovic/etfis/>.

5.4 Potrebni istraživački resursi

Obzirom na dostupnost open-source softverskih rješenja i otvorenih skupova podataka, eksperimenti se mogu provoditi na bilo kojem PC računaru. No vrijeme izvršenja ovakvih eksperimenata je veliko. Budući da su algoritmi CBIR izuzetno pogodni za paralelizaciju, vrijeme potrebno za eksperimente bi se moglo znatno skratiti korištenjem paralelnih računarskih platformi kao što je cluster. Također se eksperimenti mogu izvršavati na laboratorijskim računarima ETFa u periodu kada se ne izvodi nastava.

6 Očekivani izvorni naučni doprinos

Imajući u vidu do sada sprovedena istraživanja, smatram da mogu predložiti novu kombinovanu značajku za pretraživanje u JPEG kompresovanom domenu, odnosno algoritam za ekstrakciju takve značajke i pretragu slika koji može dati bolje rezultate od sličnih značajki koje su do sada predložene u literaturi.

Tokom proteklih godina u naučnom svijetu predloženo je niz različitih algoritama pretrage u kompresovanom domenu. Uspio sam doći do nekoliko desetina naučnih radova na ovu temu (datih u literaturi) te sam pokušao dati njihovu sistematizaciju i evaluaciju (spomenuti rad nije objavljen, a biće poslan na neku od referentnih konferencija planiranih za 2014. godinu). Slična takva evaluacija data je i u radu [29].

Prilikom evaluacije ovih metoda, neke se mogu odmah otpisati zato što prezentovani radovi ne sadrže dovoljno informacija za reproduciranje njihovih rezultata, sadrže neke fundamentalne greške ili se može konstatovati da zaključci nisu iskoristivi. Od preostalih radova, implementirao sam desetak algoritama i reproducirao rezultate prezentovane u radovima. Preostaje još nekoliko radova koje bi trebalo probati implementirati.

Među evaluiranim algoritmima izdvaja se algoritam prezentovan u radu [22]. Ovaj rad je dostupan u “prepress” varijanti koja sadrži nekoliko očiglednih ili manje očiglednih grešaka koje otežavaju implementaciju. Kandidat je uspio uočiti i popraviti te greške čime je dobio rezultate koji su već usporedivi sa state-of-the-art algoritmima u piksel domenu kao što je onaj prezentovan u radu [56]. Pored toga, već su provedeni neki eksperimenti koji daju naslutiti da se rezultati mogu dalje unaprijediti.

Osim toga, mišljenja sam da u literaturi koju sam do sada uspio samo pročitati a ne i eksperimentalno verifikirati postoji nekoliko vrlo obećavajućih pristupa koje treba kritički obraditi.

Pored ovoga, u literaturi koja se tiče pretraživanja slika u prostornoj (piksel) domeni prezentovani su neki zaključci koji bi se mogli po mom mišljenju prenijeti u kompresovanu domenu. Tu prije svega mislim na neizrazito (fuzzy) procesiranje boje, te primjenu genetičkog algoritma za utvrđivanje optimalne kvantizacije. U pitanju su potencijalni pravci istraživanja koji mogu biti uključeni u finalnu formu teze ako eksperimentalno daju zadovoljavajuće rezultate.

Kao sekundarne doprinose ove doktorske teze ističem rad na razvoju metodologije za kritičku evaluaciju algoritama pretrage (što uključuje i softversku testnu platformu), te opsežnu evaluaciju različitih pristupa pretraživanju i njihovu eksperimentalnu komparaciju. Dalje, razvijene su efikasne implementacije različitih metrika distance koristeći look-up tabele (LUT).

7 Korištena literatura

Literatura

- [1] A. Hanjalic, R. Lienhart, W.-Y. Ma, and J. R. Smith, “The holy grail of multimedia information retrieval: So close or yet so far away?” *Proceedings of the IEEE*, vol. 96, no. 4, pp. 541–547, 2008.
- [2] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,” *Pattern*

- Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1349–1380, 2000.
- [3] T. Pavlidis, “Limitations of content-based image retrieval,” in *Invited Plenary Talk at the 19th Internat. Conf. on Pattern Recognition, Tampa, Florida, December, 2008*, pp. 8–11.
 - [4] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1106–1114.
 - [5] C. Rosenberg, “Improving photo search: A step across the semantic gap,” <http://googleresearch.blogspot.com/2013/06/improving-photo-search-step-across.html>, Jun. 2013.
 - [6] R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” *ACM Computing Surveys (CSUR)*, vol. 40, no. 2, pp. 5:1–5:60, 2008.
 - [7] V. Ljubovic and H. Supic, “Issue of resource usage in content-based image retrieval algorithms,” in *Telecommunications (BIHTEL), 2012 IX International Symposium on*. IEEE, 2012, pp. 1–5.
 - [8] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, “Content-based multimedia information retrieval: State of the art and challenges,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM-CAP)*, vol. 2, no. 1, pp. 1–19, 2006.
 - [9] S. A. Chatzichristofis, C. Iakovidou, Y. Boutalis, and O. Marques, “Co. Vi. Wo.: Color visual words based on non-predefined size codebooks,” 2012.
 - [10] J. Z. Wang, J. Li, and G. Wiederhold, “SIMPLIcity: Semantics-sensitive integrated matching for picture libraries,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 9, pp. 947–963, 2001.
 - [11] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, “Efficient and effective querying by image content,” *Journal of Intelligent Information Systems*, vol. 3, no. 3-4, pp. 231–262, 1994.
 - [12] S.-F. Chang, “Compressed-domain techniques for image/video indexing and manipulation,” in *Image Processing, 1995. Proceedings., International Conference on*, vol. 1. IEEE, 1995, pp. 314–317.
 - [13] *ITU-R BT.601 (1993), ITU CCITT T.81 Digital Compression and Coding of Continuous-Tone Still Images – Requirements and Guidelines*, <http://www.astm.org/DATABASE.CART/WITHDRAWN/E1294.htm>, ITU International, 1211 Geneva 20, Switzerland Std.

- [14] B. Shen and I. K. Sethi, "Direct feature extraction from compressed images," in *Electronic Imaging: Science & Technology*. International Society for Optics and Photonics, 1996, pp. 404–414.
- [15] M. K. Mandal, F. Idris, and S. Panchanathan, "A critical evaluation of image and video indexing techniques in the compressed domain," *Image and Vision Computing*, vol. 17, no. 7, pp. 513–529, 1999.
- [16] G. Schaefer, "CVPIC colour/shape histograms for compressed domain image retrieval," in *Pattern Recognition*. Springer, 2004, pp. 424–431.
- [17] J. A. Lay and L. Guan, "Image retrieval based on energy histograms of the low frequency DCT coefficients," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 6. IEEE, 1999, pp. 3009–3012.
- [18] C.-W. Ngo, T.-C. Pong, and R. T. Chin, "Exploiting image indexing techniques in DCT domain," *Pattern Recognition*, vol. 34, no. 9, pp. 1841–1851, 2001.
- [19] B. Furht and P. Saksobhavit, "Fast content-based multimedia retrieval technique using compressed data," in *Photonics East (ISAM, VVDC, IEMB)*. International Society for Optics and Photonics, 1998, pp. 561–571.
- [20] M. Hatzigiorgaki and A. N. Skodras, "Compressed domain image retrieval: A comparative study of similarity metrics," in *Proc. SPIE 5150. Visual Communications and Image Processing 2003*. International Society for Optics and Photonics, 2003, pp. 439–448.
- [21] J. Jiang, A. Armstrong, and G.-C. Feng, "Direct content access and extraction from JPEG compressed images," *Pattern Recognition*, vol. 35, no. 11, pp. 2511–2519, 2002.
- [22] Z.-M. Lu, S. Li, and H. Burkhardt, "A content-based image retrieval scheme in JPEG compressed domain," *International Journal of Innovative Computing, Information and Control*, vol. 2, no. 4, pp. 831–839, 2006.
- [23] J. Jiang, Y. Weng, and P. Li, "Dominant colour extraction in DCT domain," *Image and Vision computing*, vol. 24, no. 12, pp. 1269–1277, 2006.
- [24] M. Islam, M. Ali *et al.*, "A miniature-based image retrieval system," *arXiv preprint arXiv:1008.3346*, 2010.
- [25] D. Zhong and I. Defée, "DCT histogram optimization for image database retrieval," *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2272–2281, 2005.
- [26] G. Schaefer, "Jpeg image retrieval by simple operators," 2001.

- [27] C.-C. Chang, J.-C. Chuang, and Y.-S. Hu, "Retrieving digital images from a jpeg compressed image database," *Image and Vision Computing*, vol. 22, no. 6, pp. 471–484, 2004.
- [28] G. Schaefer and D. Edmundson, "Dc stream based jpeg compressed domain image retrieval," in *Active Media Technology*. Springer, 2012, pp. 318–327.
- [29] D. Edmundson and G. Schaefer, "An overview and evaluation of JPEG compressed domain retrieval techniques," in *ELMAR, 2012 Proceedings*. IEEE, 2012, pp. 75–78.
- [30] S. Climer and S. K. Bhatia, "Image database indexing using jpeg coefficients," *Pattern Recognition*, vol. 35, no. 11, pp. 2479–2488, 2002.
- [31] E. Guldogan, M. Gabbouj, and O. Guldogan, "DCT-based downscaling effects on color and texture-based image retrieval," in *EWIMT 2004 IEE European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, Proceedings of*, 2004.
- [32] Y.-L. Huang and R.-F. Chang, "Texture features for DCT-coded image retrieval and classification," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 6. IEEE, 1999, pp. 3013–3016.
- [33] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.
- [34] J. R. Smith and S.-F. Chang, "Transform features for texture classification and discrimination in large image databases," in *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, vol. 3. IEEE, 1994, pp. 407–411.
- [35] S. Panchanathan, "Compressed or progressive image search," in *Image Databases: Search and retrieval of digital imagery*, V. Castelli and L. D. Bergman, Eds. John Wiley & Sons, Inc., 2002, pp. 465–495.
- [36] J. Smith and S.-F. Chang, "Quad-tree segmentation for texture-based image query," in *Proceedings of the second ACM international conference on Multimedia*. ACM, 1994, pp. 279–286.
- [37] R. Reeves, K. Kubik, and W. M. Osberger, "Texture characterization of compressed aerial images using DCT coefficients," in *Electronic Imaging'97*. International Society for Optics and Photonics, 1997, pp. 398–407.
- [38] G. Feng and J. Jiang, "JPEG compressed image retrieval via statistical features," *Pattern Recognition*, vol. 36, no. 4, pp. 977–985, 2003.

- [39] D. Zhong and I. Defée, “Study of image retrieval based on feature vectors in compressed domain,” in *Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic*. IEEE, 2006, pp. 202–205.
- [40] H.-J. Bae and S.-H. Jung, “Image retrieval using texture based on DCT,” in *Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on*. IEEE, 1997, pp. 1065–1068.
- [41] M. Eom and Y. Choe, “Fast extraction of edge histogram in dct domain based on mpeg7,” in *Proceedings of world academy of science, engineering and technology*, vol. 9, 2005, pp. 209–212.
- [42] A. McIntyre and M. Heywood, “Exploring content-based image indexing techniques in the compressed domain,” in *Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on*, vol. 2. IEEE, 2002, pp. 957–962.
- [43] A. A. Abdel-Malek and J. E. Hershey, “Feature cueing in the discrete cosine transform domain,” *Journal of Electronic Imaging*, vol. 3, no. 1, pp. 71–80, 1994.
- [44] D. Edmundson and G. Schaefer, “Efficient and effective online image retrieval,” in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2312–2317.
- [45] M. Shneier and M. Abdel-Mottaleb, “Exploiting the JPEG compression scheme for image retrieval,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 8, pp. 849–853, 1996.
- [46] A. Armstrong and J. Jiang, “An efficient image indexing algorithm in JPEG compressed domain,” in *Consumer Electronics, 2001. ICCE. International Conference on*. IEEE, 2001, pp. 350–351.
- [47] H. Müller, S. Marchand-Maillet, and T. Pun, “The truth about Corel-evaluation in image retrieval,” in *Image and Video Retrieval*. Springer, 2002, pp. 38–49.
- [48] M. J. Huiskes and M. S. Lew, “The MIR flickr retrieval evaluation,” in *Multimedia information retrieval (MIR ’08), Proceedings of the 1st ACM international conference on*. ACM, 2008, pp. 39–43.
- [49] H. Müller, W. Müller, D. M. Squire, S. Marchand-Maillet, and T. Pun, “Performance evaluation in content-based image retrieval: Overview and proposals,” *Pattern Recognition Letters*, vol. 22, no. 5, pp. 593–601, 2001.
- [50] J. P. Ioannidis, “Why most published research findings are false,” *PLoS medicine*, vol. 2, no. 8, p. e124, 2005.

- [51] E. M. Voorhees and D. Harman, “Common evaluation measures,” in *The Twentieth Text Retrieval Conference (TREC 2011), Proceedings*, 2012, pp. 500–255.
- [52] J. R. Smith and S.-F. Chang, “Tools and techniques for color image retrieval,” in *Proc. SPIE 2670. Storage and Retrieval for Still Image and Video Databases IV*. International Society for Optics and Photonics, 1996, pp. 426–437.
- [53] M. J. Swain and D. H. Ballard, “Color indexing,” *International journal of computer vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [54] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, “Color and texture descriptors,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, no. 6, pp. 703–715, 2001.
- [55] G. Schaefer and M. Stich, “UCID: An uncompressed color image database,” in *Proc. SPIE 5307. Storage and Retrieval Methods and Applications for Multimedia 2004*. International Society for Optics and Photonics, 2003, pp. 472–480.
- [56] S. A. Chatzichristofis, K. Zagoris, Y. S. Boutalis, and N. Papamarkos, “Accurate image retrieval based on compact composite descriptors and relevance feedback information,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 24, no. 02, pp. 207–244, 2010.
- [57] T. Deselaers, D. Keysers, and H. Ney, “Features for image retrieval: An experimental comparison,” *Information Retrieval*, vol. 11, no. 2, pp. 77–107, 2008.
- [58] Y. Rubner, C. Tomasi, and L. J. Guibas, “The Earth mover’s distance as a metric for image retrieval,” *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [59] L. A. Granka, T. Joachims, and G. Gay, “Eye-tracking analysis of user behavior in WWW search,” in *Research and development in information retrieval (SIGIR ’04), Proceedings of the 27th annual international ACM SIGIR conference on*. ACM, 2004, pp. 478–479.
- [60] B. Pan, H. Hembrooke, T. Joachims, L. Lorigo, G. Gay, and L. Granka, “In Google we trust: Users’ decisions on rank, position, and relevance,” *Journal of Computer-Mediated Communication*, vol. 12, no. 3, pp. 801–823, 2007.
- [61] *ITU-R BT.709-5 (2005), Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange*, <http://www.itu.int/rec/R-REC-BT.709/en>, ITU International, 1211 Geneva 20, Switzerland Std.

- [62] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, *sRGB (1996), A Standard Default Color Space for the Internet*, <http://www.w3.org/Graphics/Color/sRGB>, Microsoft and Hewlett-Packard Joint Report Std.
- [63] Z. Zhang, W. Li, and B. Li, "An improving technique of color histogram in segmentation-based image retrieval," in *Information Assurance and Security (IAS '09). Fifth International Conference on*, vol. 2. IEEE, 2009, pp. 381–384.
- [64] M. A. Stricker and M. Orengo, "Similarity of color images," in *Proc. SPIE 2420. Storage and Retrieval for Image and Video Databases III*. International Society for Optics and Photonics, 1995, pp. 381–392.
- [65] K.-M. Wong, L.-M. Po, and K.-W. Cheung, "A compact and efficient color descriptor for image retrieval," in *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007, pp. 611–614.
- [66] K. Konstantinidis, A. Gasteratos, and I. Andreadis, "Image retrieval based on fuzzy color histogram processing," *Optics Communications*, vol. 248, no. 4, pp. 375–386, 2005.
- [67] R. Sudhir and L. D. S. S. Baboo, "An efficient CBIR technique with YUV color space and texture features," *Computer Engineering and Intelligent Systems*, vol. 2, no. 6, pp. 85–95, 2011.
- [68] T. Tsai, Y.-P. Huang, and T.-W. Chiang, "Fast image retrieval using low frequency DCT coefficients," in *Proceedings of the 10th conference on artificial intelligence and applications, Kaohsiung, Taiwan*, 2005.
- [69] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Image Processing. Proceedings of the 2002 International Conference on*, vol. 2. IEEE, 2002, pp. II–589.
- [70] A. R. Smith, "Color gamut transform pairs," in *SIGGRAPH '78 Proceedings of the 5th annual conference on Computer graphics and interactive techniques*, vol. 12, no. 3. ACM, 1978, pp. 12–19.
- [71] V. Ljubovic and H. Supic, "A compact feature for image retrieval in HSL model," 2013, accepted for Information, Communication and Automation Technologies (ICAT), 2013 XX International Symposium on.
- [72] T. Wang, S.-M. Hu, and J.-G. Sun, "Image retrieval based on color-spatial feature," *Journal of Software*, vol. 13, no. 10, pp. 2031–2036, 2002.
- [73] A. Hanbury and J. Serra, "A 3D-polar coordinate colour representation suitable for image analysis," *submitted to Computer Vision and Image Understanding*, 2002.

- [74] G. H. Joblove and D. Greenberg, "Color spaces for computer graphics," in *SIGGRAPH '78. Proceedings of the 5th annual conference on Computer graphics and interactive techniques*, vol. 12, no. 3. ACM, 1978, pp. 20–25.
- [75] T. Gevers, "Color-based retrieval," in *Principles of Visual Information Retrieval*, M. S. Lew, Ed. Springer, 2001, pp. 11–51.
- [76] R. Brunelli and O. Mich, "Histograms analysis for image retrieval," *Pattern Recognition*, vol. 34, no. 8, pp. 1625–1637, 2001.
- [77] V. Ljubovic and H. Supic, "Comparative study of color histograms as global feature for image retrieval," in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2013 36th International Convention on*. IEEE, 2013, pp. 1059–1063.
- [78] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" in *Database Theory-ICDT'99. Proceedings of the 7th International Conference on*. Springer, 1999, pp. 217–235.
- [79] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233–1244, 1996.
- [80] M. Lux and S. A. Chatzichristofis, "LIRE: Lucene Image REtrieval: An extensible Java CBIR library," in *Multimedia (MM' 08). Proceedings of the 16th ACM international conference on*. ACM, 2008, pp. 1085–1088.
- [81] H. Eidenberger, "How good are the visual MPEG-7 features?" in *Proc. SPIE 5150. Visual Communications and Image Processing 2003*. International Society for Optics and Photonics, 2003, pp. 476–488.
- [82] J. Puzicha, T. Hofmann, and J. M. Buhmann, "Non-parametric similarity measures for unsupervised texture segmentation and image retrieval," in *Computer Vision and Pattern Recognition (CVPR). Proceedings of the 1997 IEEE Computer Society Conference on*. IEEE, 1997, pp. 267–272.
- [83] H. J. Kim and J. S. Lee, "HMMD color space and method for quantizing color using HMMD space and color spreading," Patent US 6 633 407 B1, 2003.
- [84] H. Liu, D. Song, S. Rüger, R. Hu, and V. Uren, "Comparing dissimilarity measures for content-based image retrieval," in *Information Retrieval Technology*. Springer, 2008, pp. 44–50.
- [85] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1, no. 4, pp. 300–307, 2007.
- [86] B. S. Manjunath and W.-Y. Ma, "Texture features for browsing and retrieval of image data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 8, pp. 837–842, 1996.

- [87] D. Androutsos, K. Plataniotis, and A. N. Venetsanopoulos, "Distance measures for color image retrieval," in *Image Processing (ICIP 98). Proceedings of the 1998 International Conference on*, vol. 2. IEEE, 1998, pp. 770–774.
- [88] L. D. Griffin, "Scale-imprecision space," *Image and Vision Computing*, vol. 15, no. 5, pp. 369–398, 1997.
- [89] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar, "Spatial information in multiresolution histograms," in *Computer Vision and Pattern Recognition (CVPR 2001). Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. 702–709.
- [90] S. Sural, "Histogram generation from the HSV color space using saturation projection," in *Multimedia Systems and Content-based Image Retrieval*, S. Deb, Ed. Idea Group Publishing, Hershey, PA, USA, 2004, pp. 101–113.
- [91] H. Kekre and K. Sonawane, "Standard deviation of mean and variance of rows and columns of images for CBIR," *WASET International Journal of Computer, Information and System Science and Engineering (IJCISSE)*, vol. 3, no. 1, pp. 8–11, 2009.
- [92] D. P. Huijsmans and N. Sebe, "Extended performance graphs for cluster retrieval," in *Computer Vision and Pattern Recognition (CVPR 2001). Proceedings of the IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. 26–31.
- [93] D. Keysers, T. Deselaers, C. Gollan, and H. Ney, "Deformation models for image recognition," *Pattern Analysis and Machine Intelligence. IEEE Transactions on*, vol. 29, no. 8, pp. 1422–1435, 2007.
- [94] M. Radovanović, A. Nanopoulos, and M. Ivanović, "On the existence of obstinate results in vector space models," in *Research and development in information retrieval (SIGIR '10). Proceedings of the 33rd international ACM SIGIR conference on*. ACM, 2010, pp. 186–193.
- [95] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.
- [96] G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest-neighbor methods in learning and vision: Theory and practice*. MIT Press, Cambridge, MA, USA, 2005.
- [97] P. Indyk, "Nearest neighbours in high-dimensional spaces," in *Handbook of Discrete and Computational Geometry (2nd ed)*, J. E. Goodman and J. O'Rourke, Eds. CRC press, 2004.

- [98] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Wu, “An optimal algorithm for approximate nearest neighbor searching,” in *Discrete algorithms (SODA ’94). Proceedings of the fifth annual ACM-SIAM symposium on*. ACM Society for Industrial and Applied Mathematics, 1994, pp. 573–582.
- [99] M. Slaney and M. Casey, “Locality-sensitive hashing for finding nearest neighbors,” *Signal Processing Magazine*, vol. 25, no. 2, pp. 128–131, 2008.
- [100] P. Indyk and R. Motwani, “Approximate nearest neighbors: Towards removing the curse of dimensionality,” in *Theory of computing (STOC ’98). Proceedings of the thirtieth annual ACM symposium on*. ACM, 1998, pp. 604–613.
- [101] S. A. Chatzichristofis and Y. S. Boutalis, “FCTH: Fuzzy color and texture histogram – A low level feature for accurate image retrieval,” in *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS’08. Ninth International Workshop on*. IEEE, 2008, pp. 191–196.
- [102] “Annotated groundtruth database. Department of Computer Science and Engineering, University of Washington. Retrieved on February 13th, 2013 from:,” <http://googleresearch.blogspot.com/2013/06/improving-photo-search-step-across.html>, 2013.
- [103] N. Sebe and M. S. Lew, “Color-based retrieval,” *Pattern Recognition Letters*, vol. 22, no. 2, pp. 223–230, 2001.
- [104] D. Edmundson and G. Schaefer, “JIRL – A C++ library for JPEG compressed domain image retrieval,” in *Multimedia (ISM), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 210–213.
- [105] M. O. Lam, T. Disney, D. S. Raicu, J. Furst, and D. S. Channin, “BRISC – An open source pulmonary nodule image retrieval framework,” *Journal of digital imaging*, vol. 20, no. 1, pp. 63–71, 2007.
- [106] C. E. Jacobs, A. Finkelstein, and D. H. Salesin, “Fast multiresolution image querying,” in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 1995, pp. 277–286.
- [107] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 8, no. 6, pp. 460–473, 1978.
- [108] B. Shen and I. K. Sethi, “Inner-block operations on compressed images,” in *Proceedings of the third ACM international conference on Multimedia*. ACM, 1995, pp. 489–498.
- [109] W. B. Seales, C. J. Yuan, W. Hu, and M. D. Cutts, “Content analysis of compressed video,” *University of Kentucky Computer Science Department Technical Report*, no. 2, p. 28, 1996.

- [110] W.-Y. Ma and B. Manjunath, "A comparison of wavelet transform features for texture image annotation," in *Image Processing, 1995. Proceedings., International Conference on*, vol. 2. IEEE, 1995, pp. 256–259.
- [111] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern recognition*, vol. 25, no. 2, pp. 173–188, 1992.
- [112] H. Nezamabadi-Pour and S. Saryazdi, "Object-based image indexing and retrieval in DCT domain using clustering techniques," in *Proceedings of World Academy of Science Engineering and Technology*, 2005.
- [113] C.-S. Li, J. R. Smith, V. Castelli, and L. Bergman, "Comparing texture feature sets for retrieving core images in petroleum applications," in *Proc. SPIE*, vol. 3656, 1999, pp. 2–11.
- [114] B. S. Manjunath and W.-Y. Ma, "Texture features for image retrieval," in *Image Databases: Search and retrieval of digital imagery*, V. Castelli and L. D. Bergman, Eds. John Wiley & Sons, Inc., 2002, pp. 312–344.
- [115] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.
- [116] Q. Tian, N. Sebe, M. S. Lew, E. Loupilas, and T. S. Huang, "Image retrieval using wavelet-based salient points," *Journal of Electronic Imaging*, vol. 10, no. 4, pp. 835–849, 2001.
- [117] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, "Blob-world: A system for region-based image indexing and retrieval," in *Visual Information and Information Systems*. Springer, 1999, pp. 509–517.
- [118] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.
- [119] L. Ledwich and S. Williams, "Reduced SIFT features for image retrieval and indoor localisation," in *Australian conference on robotics and automation*, vol. 322. Citeseer, 2004.
- [120] B. Thomee, E. M. Bakker, and M. S. Lew, "TOP-SURF: a visual words toolkit," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1473–1476.
- [121] K. Zagoris, S. A. Chatzichristofis, and A. Arampatzis, "Bag-of-visual-words vs global image descriptors on two-stage multimodal retrieval," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 1251–1252.

- [122] M. Kogler and M. Lux, “Bag of visual words revisited: an exploratory study on robust image retrieval exploiting fuzzy codebooks,” in *Proceedings of the Tenth International Workshop on Multimedia Data Mining*. ACM, 2010, p. 3.
- [123] B. B. Kimia, “Shape representation for image retrieval,” in *Image Databases: Search and retrieval of digital imagery*, V. Castelli and L. D. Bergman, Eds. John Wiley & Sons, Inc., 2002, pp. 345–372.
- [124] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee, “Shape measures for content based image retrieval: a comparison,” *Information Processing & Management*, vol. 33, no. 3, pp. 319–337, 1997.
- [125] M. Bober, “MPEG-7 visual shape descriptors,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, no. 6, pp. 716–719, 2001.
- [126] D. Zhang and G. Lu, “Shape-based image retrieval using generic Fourier descriptor,” *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825–848, 2002.
- [127] —, “A comparative study of curvature scale space and fourier descriptors for shape-based image retrieval,” *Journal of Visual Communication and Image Representation*, vol. 14, no. 1, pp. 39–57, 2003.
- [128] T. Chang and C.-C. Kuo, “Texture analysis and classification with tree-structured wavelet transform,” *Image Processing, IEEE Transactions on*, vol. 2, no. 4, pp. 429–441, 1993.
- [129] A. Guttman, “R-trees: A dynamic index structure for spatial searching,” in *Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, ser. SIGMOD ’84. New York, NY, USA: ACM, 1984, pp. 47–57. [Online]. Available: <http://doi.acm.org/10.1145/602259.602266>
- [130] S. Hwang, K. Kwon, S. K. Cha, and B. S. Lee, “Performance evaluation of main-memory R-tree variants,” in *Advances in Spatial and Temporal Databases*. Springer, 2003, pp. 10–27.
- [131] S. Berchtold, D. A. Keim, and H.-P. Kriegel, “The X-tree: An index structure for high-dimensional data,” *Readings in multimedia computing and networking*, p. 451, 2001.
- [132] D. Zhong and I. Defée, “Performance of similarity measures based on histograms of local image feature vectors,” *Pattern Recognition Letters*, vol. 28, no. 15, pp. 2003–2010, 2007.
- [133] S. Dubuisson, “The computation of the bhattacharyya distance between histograms without histograms,” in *Image Processing Theory Tools and Applications (IPTA), 2010 2nd International Conference on*. IEEE, 2010, pp. 373–378.

- [134] Y. Meyer and D. H. Salinger, *Wavelets and operators*. Cambridge university press, 1995, vol. 1.
- [135] B. Vidakovic and P. Mueller, “Wavelets for kids,” *Instituto de Estadística, Universidad de Duke*, 1994.
- [136] A. Haar, “Zur theorie der orthogonalen funktionensysteme,” *Mathematische Annalen*, vol. 69, no. 3, pp. 331–371, 1910.
- [137] I. Daubechies *et al.*, *Ten lectures on wavelets*. SIAM, 1992, vol. 61.
- [138] S. Mallat, *A wavelet tour of signal processing*. Access Online via Elsevier, 1999.