SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Globalne značajke za raspoznavanje scene

Bojan Popović
Voditelj: *Siniša Šegvić*

Zagreb, lipanj 2011.

SADRŽAJ

1. Uvod

Tradicionalan pristup detekcije objekata u sceni zasnivao se na promatranju zasebnih dijelova scene te ispitivanje postojanju traženog objekta u njima. Međutim, vrijeme je pokazalo da je takav pristup u velikoj mjeri redundantan jer gotovo uvijek pretražujemo regije slike gdje se objekt gotovo sigurno neće nalaziti. Primjer za to je analiza slike otvorene ceste na kojoj tražimo prometne znakove. Pretraživanje svih regija slike obuhvaća i analizu neba ili same ceste gdje u pravilu nećemo naći prometne znakove. Također, ako znak u sceni ali jako daleko, moguće je da ga klasifikator neće pronaći. Treniranje klasifikatora nad netransparentnim, sitnim, ili mutnim primjerima dovodi do prenaučenosti.

Napredak nad tradicionalnim pristupom ostvariv je korištenjem globalnih značajki scene. One pružaju informaciju o najvjerojatnijim lokacijama gdje bi se traženi objekt mogao nalaziti. Također, pružaju mogućnost razlikovanja kategorija scene, primjerice otvorene ceste ili ulice, što nam daje nove informacije koje možemo koristiti kod detekcije objekata. Kontekst u sceni omogućava stvaranje poveznica između lokacija pojedinih objekata u sceni, njihove veličine, te relacije u odnosu na druge objekte.

Prepoznavanje scene i njenih značajki omogućava *zaključivanje* o netransparentnim ili nedovoljno istaknutim dijelovima scene korištenjem njihove okoline.

Ovaj rad je uvod u metode raspoznavanja scene, njenih značajki, te korištenja konteksta pri detekciji objekata u sceni. Drugo poglavlje bavi se metodama koje se zasnivaju na korištenju konteksta. Treće poglavlje je uvod u kategorizaciju i klasifikaciju scene, opisujući dva idejno različita pristupa tom problemu. Četvrto poglavlje prikazuje moć kombinacije lokalnih i globalnih značajki scene pri detekciji objekata.

2. Kontekst pri raspoznavanju objekata

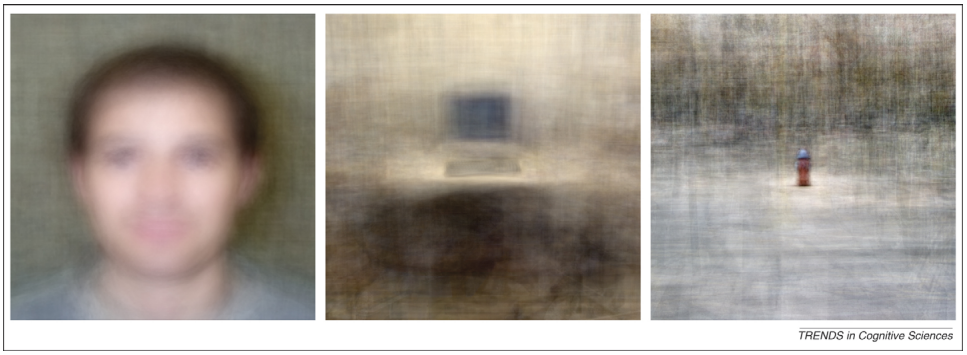
U stvarnom svijetu, objekti se nikad ne javljaju izolirani, već se uklapaju u okruženje s drugim objektima. Na taj način stvaraju izvor međudjelovanja sa okolinom - kontekst. Prirodan način predstavljanja konteksta zasniva se na stvaranju poveznica između objekta i njegovih susjeda. Statistička obrada scene predstavlja bujan izvor informacija za stvaranje kontekstualnih veza, što ljudskim očima omogućava učinkovit način primjećivanja važnih aspekata neke prirodne scene. Bolje razumijevanje načina na koji ljudi grade reprezentaciju scene, te mehanizama analize konteksta, dovest će do nove generacije sustava računalnog vida.

Strukturu mnogih scena (ili grupa objekata), obilježavaju čvrsta pravila slična onima koja se odnose na građu jedinke (tj. zasebnog objekta). Primjerice, ako uzmemo prosjek stotina slika koje predstavljaju ljudski portret, javlja se pravilnost u intenzitetu boja na slici, te dobivamo grub prikaz dijelova lica koje dijele svi subjekti koji stvore skup uzoraka. Ovakvo uzimanje prosjeka uzoraka objekata (gledanih kao jedinki) često dovodi do pojave pravilnosti koje nisu vezane uz sam objekt. Na primjer, u prosječnoj fotografiji tipkovnice često nalazimo radni stol i monitor u pozadini, unatoč tome što fotografije nisu napravljene sa svrhom prikazivanja i tih objekata. Slično pravilo vrijedi i za sliku hidranta. Iako je pozadina takve fotografije rijetko slična, prosjek uzoraka pokazuje obris tla pod hidrantom. Ova su obilježja vidljiva na slici 2.1

Prisutnost određenog objekta uvodi mogućnost postojanja drugih objekata u njegovoj blizini, definirajući kontekst u kojem se nalazi. Utjecaj konteksta postaje još veći ako su objekti koje je potrebno prepoznati maleni ili djelomično sakriveni. U tom slučaju lokalne značajke za raspoznavanje nisu dovoljne.

Istraživanja su pokazala višeslojni utjecaj konteksta:

- *semantički* - stol i stolica imaju veću vjerojatnost da budu na istoj slici, za razliku od slona i čajnika



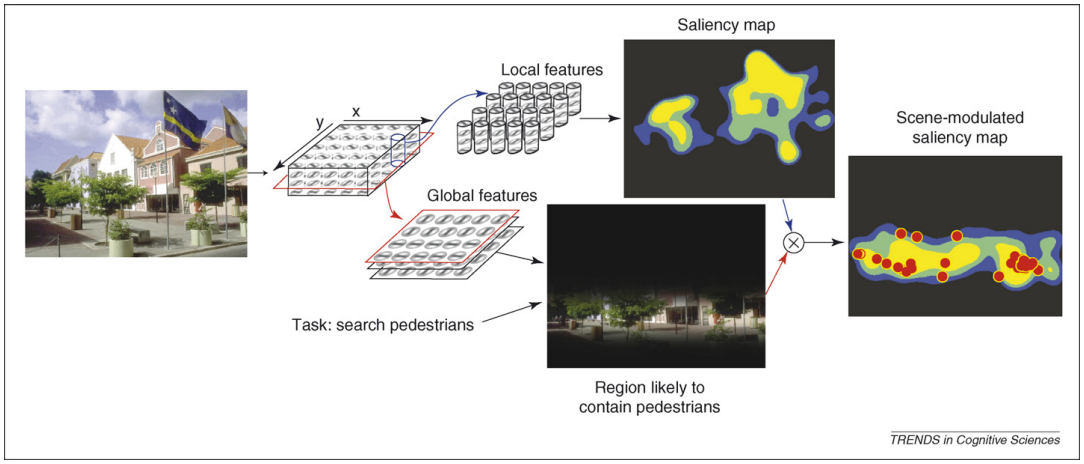
Slika 2.1: *Prosjek* stotine uzoraka slika ljudskog lica, tipkovnice i hidranta

– *prostorna konfiguracija* - tipkovnicu očekujemo pronaći negdje ispod računalnog monitora

– *orijentacija* - monitor obično nije okrenut prema zidu, automobil je orijentiran paralelno s pravcem ceste

U računalnom vidu, najčešći pristup *lokalizacije* objekata u slikama jest *pomični prozor*. Postupak se temelji na particioniranju slike te pomicanju prozora varijabilne veličine po svim participama. U lokalnoj domeni prozora cilj klasifikacije je traženi objekt ili pozadina, ako se objekt ne nalazi u lokalnoj domeni. Ovaj je pristup bio uspješan pri detekciji objekata kao što su ljudsko lice, automobili ili prometni znakovi. Međutim, kontekstualne informacije mogu se koristiti uz bok opisanog lokaliziranog pristupa kako bi se podigla razina uspješnosti detekcije, brzina izvođenja, te tolerancija na kvalitetu ulaznog skupa slika. Jedan od glavnih problema s kojima se susreće automatizirano raspoznavanje objekata pri korištenju konteksta, jest nedostatak jednostavnih prikaza kontekstualne informacije te učinkovitih algoritama za ekstrakciju novih informacija iz kontekstualnih veza.

Idealni će promatrač, kada promatra scenu te traži ciljani objekt, svoju pažnju usmjeriti prema mjestima koja imaju najveću posteriornu vjerojatnost sadržavanja traženog objekta. *Pažnju* promatrača mogu voditi neki globalni atributi scene. Primjerice, kada kada promatra scenu parka tražeći klupe, pažnja će mu biti usmjerena na dio slike bliži tlu. Također, promatrač pokušava uočiti objekte koji su kontekstualno vezani uz traženi objekt. U navedenoj sceni parka, kante za smeće se obično nalaze blizu klupa, pa promatrač obraća pažnju i na njih. Objekti koji su *bliski* (eng. salient) traženom često se transformiraju u poseban prikaz, *karta bliskosti*(eng. saliency map). Na slici 2.2 prikazana je karta bliskosti, koja predstavlja lokalne značajke, te u kombinaciji sa globalnim značajkama (npr. obris tla) daje regiju slike gdje će se traženi objekt najvjerojatnije nalaziti.



Slika 2.2: Prikaz detekcije pješaka korištenjem lokalnih (kontekstualnih) značajki te globalnih atributa scene

2.1. Modeliranje konteksta postupkom temeljenom na Bayesovoj formuli

Kontekstualne informacije moguće je promatrati u obliku koji podsjeća na Bayesovu formulu. Za svaku lokaciju na slici (x), veličinu pomičnog prozora kojim promatramo dio scene (σ), dodjeljuje se vjerojatnost postojanja traženog objekta: $p(o, x, \sigma, \alpha|\vec{v}_l, \vec{v}_c)$. Vjerojatnost uvjetuju lokalni i globalni uzorci. Primjenom Bayesovog teorema izvodi se formula 2.1:

$$p(O|\vec{v}_l, \vec{v}_c) = \frac{1}{p(\vec{v}_l|\vec{v}_c)} p(\vec{v}_l|O, \vec{v}_c) p(O|\vec{v}_c) \tag{2.1}$$

gdje je $O = o, x, \sigma, \alpha$. $\vec{\alpha}$ je vektor parametara koji opisuje prikaz traženog objekta (npr. kut gledanja na objekt).

Normalizacijski vektor $\frac{1}{p(\vec{v}_l|\vec{v}_c)}$ nije ovisan o traženom objektu. On predstavlja mjeru male vjerojatnosti nalaženja vektora lokalnih značajki $\vec{v}_l$ u odnosu na okolni kontekst $\vec{v}_c$. Ova vjerojatnosna definicija *bliskosti* prirodno odgovara detekciji i klasifikaciji objekata [Oliva i Torralba (2002)].

Drugi faktor, $p(\vec{v}_l|O, \vec{v}_c)$, daje mjeru vjerojatnosti pojave lokalnih značajki $\vec{v}_l$ kada je objekt O prisutan u promatranom kontekstu. Ovo povećava važnost regija slike sa značajkama kontekstualno vezanim za traženi objekt, a smanjuje važnost onih koje se ne uklapaju u kontekst traženog objekta.

Treći faktor, $p(O|\vec{v}_c)$ predstavlja *kontekstualne apriorne vrijednosti* (eng. context-based priors) klase objekta, lokacije, veličine, i prikaza traženog objekta, odnosno kontekstualnu modulaciju bliskosti sa traženim objektom [Torralba (2004)]. On dolazi do izražaja kada lokalne značajke daju nesuglasne rezultate, iako ne ovisi o lokalnim

značajkama.

2.2. Pristup usporedbe esencija scena

Ovaj pristup definira vjerojatnosti detekcije traženog objekta kao dijela scene koristeći velike baze podataka označenih slika koje sadrže slične scene [B. Russell (2005)]. Pozadina, umjesto da se uzima kao skup negativnih značajki, služi kao vodilja u postupku detekcije. Pristup se oslanja na pretpostavku da je baza označenih slika dovoljno velika kako bi se s visokom vjerojatnošću mogle pronaći slike čija je *esencija* bliska onoj slike koju istražujemo. Esencija se ovdje definira u samom izgledu scene, njenih čimbenika, te prostornog uređenja objekata unutar scene. Budući da su slike u bazi podataka djelomično obilježene, tj. nisu obilježeni baš *svi* objekti u scenama, dostupne oznake objekata mogu se iskoristiti kao informacije o sceni koju se upravo analizira. Zamišljeni proces provedbe ovog pristupa prikazan je na slici 2.3. Ako uzmemo u obzir gorenavedene činjenice, problem detekcije objekta postaje problem *preklapanja* scena, u smislu da pokušava se dovesti dvije scene u vezu uzimajući u obzir udaljenost, veličinu slika, kut gledanja, itd. U ovom pristupu javlja se nekoliko važnih problema:

1. Može li se pronaći dovoljno velika baza označenih slika kako postojao dovoljan broj konfiguracija sličnih scena?

2. Može li se u bazi slika pronaći skup bliskih scena koje se dobro *preklapaju* sa traženom scenom?

3. Kako prenijeti informacije o označenim objektima iz baze u scenu koju istražujemo?

2.2.1. Scene kandidati

Baza označenih slika *LabelMe* [Russel et al. (2005)] odgovara zahtijevima, budući da sadrži velik broj označenih slika. Istraživanja su pokazala prednosti korištenja neparametarskih metoda računalnog vida i računalne grafike kako bi se iz velikog skupa označenih slika indeksiranjem došlo do slika sa prostornom konfiguracijom objekata sličnom onoj u sceni koju istražujemo [Hays i Efros (2007)].

Jezgra ovog sustava jest prijenos oznaka sa slika najsličnijih sceni koju istražujemo. Pretpostavka je da između označenih objekata postoji koherentnost u vidu sadržaja scena. Prema tome, slike se grupiraju u *scene kandidate* od kojih je potrebno odabrati

Slika 2.3: Prikaz usporedbe tražene scene (lijevo) sa bazom LabelMe, dohvaćanje sličnih scena (sredina), te označavanje tražene scene uz pomoć oznaka iz baze(desno)

onu koja najviše sliči traženoj sceni. Odabir se vrši preko relativno jednostavnog generativnog modela. Odabir scena vidljivo je na slici 2.4.

Slika 2.4: Prikazane su dvije scene te njihove scene kandidati iz baze LabelMe. Kandidati su sortirani kriterijem $L1 - udaljenosti$ po značajki esencije scene, te *bliskosti* njihovih konfiguracija traženoj sceni.

2.2.2. Značajka esencije

Značajku esencije scene (eng. gist of a scene) definiramo kao niskodimenzionalni vektor regije slike. Kako bismo konstruirali taj vektor, na regiju slike se primjenjuje niz Gaborovih filtera koji sadrže 4 veličine i 8 orijentacija. Gaborov filtar (slika 2.5) je linearni filtar koji se koristi za detekciju rubova i ekstrakciju značajki tekstura sa slike, koristeći Gaussovu jezgrenu funkciju moduliranu dvodimenzionalnom sinusoidom [Petkov (2008)]. Regija slike particionira se na 16 disjunktnih podregija. Energija izlaza pojedinih filtara ujednačava kroz svaku podregiju. Naposlijetku dobivamo vektor bliskosti od 512 ($4 \times 8 \times 16$) dimenzija. Značajka esencije čuva prostornu strukturu informacije koju nosi slika. Sličan rezultat postiže se uporabom SIFT deskriptora [Liu et al. (2009)].

Slika 2.5: Primjer izgleda jezgrenih funkcija Gaborovih filtara različitih valnih duljina.

Značajka esencije pokazala se vrlo dobrim mjerilom za usporedbu scena. Kandidate je dovoljno sortirati kriterijem $L1 - udaljenosti$. Evaluacija je pokazala da ovaj pristup daje mnogo bolje rezultate od stroja s potpornim vektorima (SVM), odnosno pristupa gdje se pokušava naučiti klasifikator kako bi raspoznavao slične scene [B. Russell (2005)].

2.2.3. Vjerojatnosni model

Naposlijetku, korišten je probabilistički model koji uključuje lokalne značajke scene, prisutnost određenih objekata u sceni, te informaciju prostorne vjerojatnosti dane oznakama iz skupa najsličnijih scena.

Želimo modelirati vezu između kategorija objekata o, njihove lokacije u prostoru x, te njihovog izgleda g. Za skup od N slika, pod pretpostavkom da svaka slika ima mogućnost sadržavanja M_i objekata koji pripadaju u L kategorija, proizlazi model:

$$p(o, x, g|\theta, \phi, \eta) = \prod_{i=1}^{N} \prod_{j=1}^{M} \sum_{h_{i,j}=0}^{1} p(o_{i,j}|h_{i,j}, \theta)\ p(x_{i,j}|o_{i,j}, h_{i,j}, \phi)\ p(g_{i,j}|o_{i,j}, h_{i,j}, \eta) \tag{2.2}$$

Model se sastoji od tri faktora:

- vjerojatnost pojavljivanja pojedinih kategorija objekata u slici
- vjerojatnost prostornih lokacija pojavljivanja objekta kategorije l u slici
- vjerojatnost *izgleda* objekta kategorije l.

$h_{i,j}$ predstavlja binarnu vrijednost pojave objekta kategorije $o_{i,j}$ na lokaciji $x_{i,j}$. θ, ϕ i η su parametri uvjetne vjerojatnosti te označavaju vjerojatnost grupiranja objekata u određenu grupu, prostornu distribuciju objekata na slici, te parametar funkcije dobivene eksperimentalno, korištenjem SVM algoritma nad skupom za učenje, tako da odgovara pozitivnim i negativnim primjerima.

Dobivena funkcija:

$$(1 + exp(-\eta_{m,l}\,[1\ g_{i,j}]^T))^{-1} \tag{2.3}$$

Ovaj model primjenjuje se pri grupiranju slika u bazi podataka prema njihovim označenim objektima kako bi se dobile scene kandidati za usporedbu sa analiziranom scenom.

3. Metode kategorizacije scena

U ovom poglavlju promatramo dva pristupa koji ne analiziraju kontekst u sceni te relacije među pojedinim objektima, već pokušavaju ocijeniti u koju bi kategoriju pripadala promatrana scena. Pristupi se razlikuju u osnovnoj ideji; **pristup *čestih oblika*** bavi se kategorizacijom scene prema njenim fizičkim karakteristikama, poglavito teksturi. Pokušava se dobiti brza prijenosna funkcija koja će za danu scenu dobiti njenu esenciju te je tako kategorizirati jednu od ponuenih kategorija. **Pristup *prostorne ovojnice*** definira kriterije po kojima se scena kategorizira. Ovaj pristup temelji se na primjeni spektralne analize i metoda strojnog učenja.

3.1. Prepoznavanje čestih *oblika* scene

Scene iz prirode, odnosno njihove *oblike* (eng. shape, albedo), definira količina svjetlosti koja ih obasjava, odnosno intenzitet svjetla koji se odbija od njihove površine. Ti oblici mogu biti vrlo kompleksni, što iziskuje slike visoke rezolucije kako bi ih se pravilo prepoznalo. Ovaj pristup predlaže prikaz oblika u niskoj rezoluciji; česti oblik scene (eng. shape recipe), koji se oslanja na samu sliku te informaciju o kompleksnoj konfiguraciji scene koju slika sadrži. Česti oblici scene su primjeri koji sadrže regresijske koeficijente korištene za predviđanje oblika scene iz same slike.

Iz slike želimo aproksimirati razne atribute scene poput njenog oblika, osvijetljenosti, materijala od kojih se sastoji, te je li scena u pokretu ili je statična. Te bi se značajke mogle čuvati u tablici vrijednosti ili u matematičkom obliku kao serija ekspanzija nad osnovnim modelom skupa površinskih deformacija [Sclaroff i Pentland (1991)]. Detaljni prikaz scene zahtijevao bi tablicu vrijednosti visoke rezolucije ili visoku razinu serije ekspanzija. Kako bi aproksimiranje i čuvanje tih vrijednosti bilo jednostavnije i brže, potrebno je smisliti jednostavniji način prikaza tih atributa. Ako pretpostavimo da je originalna slika uvijek dostupna, scenu koja se nalazi u slici možemo opisati *u odnosu na* sliku. Reprezentaciju scene, odnosno željene atribute, moguće je aproksimirati iz originalne slike korištenjem konačnog skupa pravila, *prav-*

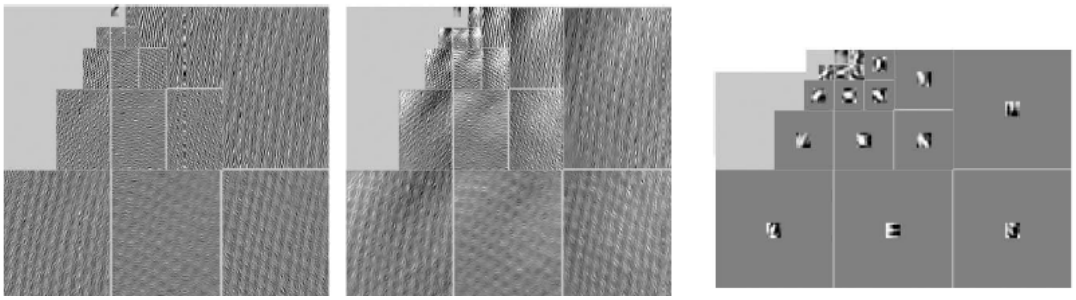
ila scene (eng. scene recipe).

Prikaz *oblika scene* oslanja se na definiranju funkcije koja će prikazati relaciju između slike i scene koju predstavlja. Ta relacija nije generalizirana za sve slike, već specifična s obzirom na osvijetljenje i materijale (teksture) na slici.

Kako bi se došlo do oblika scene u dvodimenzionalnoj slici, ona se dijeli na pojaseve korištenjem rezolucijske piramide, transformacije slike u više veličina i orijentacija. Pojedini pojas oblika je u relaciji sa pojasom originalne slike preko funkcije

$$Z_k = f_k(I_k) \tag{3.1}$$

gdje je f_k lokalna funkcija a Z_k i I_k predstavljaju k-te pojase rezolucijskih piramida [Simoncelli i Freeman (1995)] slike i oblika. Primjer rezolucijske piramide (eng. steerable pyramid) nalazi se na slici 3.1.

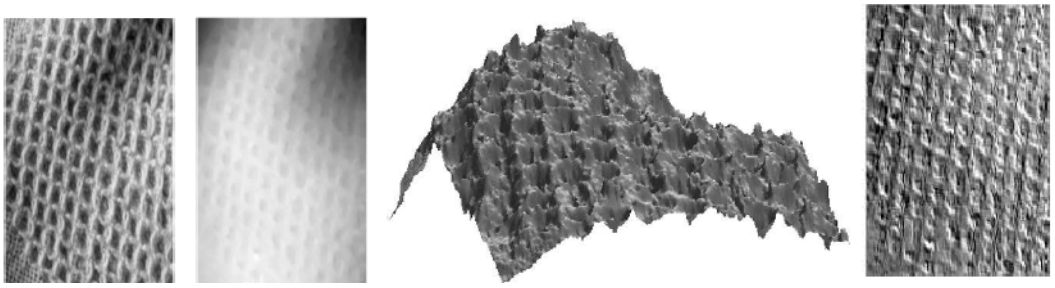


Slika 3.1: Pomična piramide slike (lijevo), pomična piramida oblika (sredina), česti oblici za pojedine pojaseve (desno).

Najjednostavniji način dobivanja relacije između intenziteta slike i njenog oblika je preko linearnog filtra: $Z_k \approx r_k \star I_k$ gdje operator $\star$ označava konvoluciju, a r_k konvolucijsku jezgrenu funkciju specifičnu za svaku veličinu i orijentaciju u piramidi [Torralba i Freeman (2004)]. r_k transformira *pojas slike* I_k u odgovarajući *pojas oblika* Z_k. *Recept* r_k za svaki se pojas dobiva minimizacijom izraza $\sum_x |Z_k - I_k \star r_k|^2$. Rezultat ovih operacija vidljiv je na slici 3.2.

3.2. Detekcija korištenjem *prostorne ovojnice*

Kako bi se definiralo što je scena, nasuprot objekta ili teksture, prvo se uzima u obzir apsolutna udaljenost između promatrača i promatrane scene. Ako slika predstavlja objekt, tada je udaljenost do promatrača od 1 do 2 metra. Scena je definirana udaljenošću od 5 ili više metara do promatrača.



Slika 3.2: Originalna slika i njen oblik scene (lijevo), oblik scene u 3D-u (sredina), oblik scene aproksimiran iz sačuvanih podataka (desno)

Prostorna ovojnica (eng. spatial envelope) nekog okruženja sačinjena je od kompleksnog skupa pravila i granica, kao što su zidovi, regije, tlo, te veće površine koje definiraju oblik prostora. Tako većina slika autoputa prikazuje veliku površinu koja se gubi u horizontu, dok slike šuma izgledaju kao zatvorena okruženja strukturirana vertikalno (dvreće) te horizoontalno (tlo). Prostorna ovojnica je veza između osnovnih crta površina koje sačinjavanju sliku te tekstura koje čine objekte na njoj.

3.2.1. Kategorije scena

Za dobivanje značajki kojima će se scene razlikovati, 17 ispitanika imalo je zadatak razdvojiti slike 81 scene u grupe. Izvršeno je grupiranje po nekom globalnom atributu koji scene međusobno dijele. Kriteriji koji se odnose na pojedine objekte nisu uzeti u obzir. Slijedi 8 kriterija koje su ispitanici najviše korisiti kod kategorizacije scena:

- prirodnost - razlikuje prirodne scene od urbanih
- otvorenost - zatvorena ili otvorena okolina, postoji li vidljiva granica horizonta
- perspektiva - uglavnom korištena kod urbanih scena
- veličina - jesu li pojedini objekti u sceni mali ili veliki
- dijagonalna ravnina - ima li rastućih i padajućih ravnina na slici (npr. kod planina)
- dubina - nivo ekspanzije prostora u sceni, daleke i bliske scene u odnosu na promatrača
- simetrija - postoji li simetričan odnos među dijelovima scene
- kontrast

Zanimljivo je da ispitanici nisu mnogo koristili kriterije simetrije i kontrasta, što dovodi do zaključka da ih ljudi ne uzimaju u obzir kada brzo trebaju odlučiti kakva je

scena, tj. doznati njenu suštinu [Oliva i Torralba (2001)].

Pomoću tih rezultata, dobiven je skup od 5 osnovnih značajki za definiranje prostorne ovojnice:

- Razina prirodnosti
- Razina otvorenosti
- Razina grubosti - odnosi se na veličinu glavnih dijelova scene. U korelaciji je sa kompleksnošću scene
- Razina ekspanzije - ljudske tvorevine većinom se sastoje od vertikalnih i horizontalnih linija. Konvergencija paralelnih linija dovodi do perpecije dubine u prostoru.
- Razina rigidnosti - devijacija tla u odnosu na horizont. Rigidna okolina sadrži mnogo kontura u slici te skriva horizont. Većina ljudskih tvorevina sagrađena je ravnici, pa su rigidne scene većinom prirodne.

3.2.2. Korišteni računalni modeli

Jezgru sustava čine dobiveni diskriminacijski spektri koji se koriste za klasifikaciju. To su DST koji je zapravo statistička informacija i $WDST$ iz kojeg je vidljiv prostorni raspored komponenti scene. $WDST$ je inačica DST-a pokrenuta na regiji slike. Pristup se oslanja na pet računalnih metoda za kategorizaciju scene.

1. *Opisne razine modela scene* - Scena se opisuje na tri razine: **niska razina** (lokalne značajke, objekti u sceni, njihova tekstura i boja), **osnovna razina** (podjela u osnovne kategorije - šuma, planina, ulica - prema obliku), te **visoka razina** (najviši nivo apstrakcije, razlikovanje otvorenih i zatvorenih okolina)

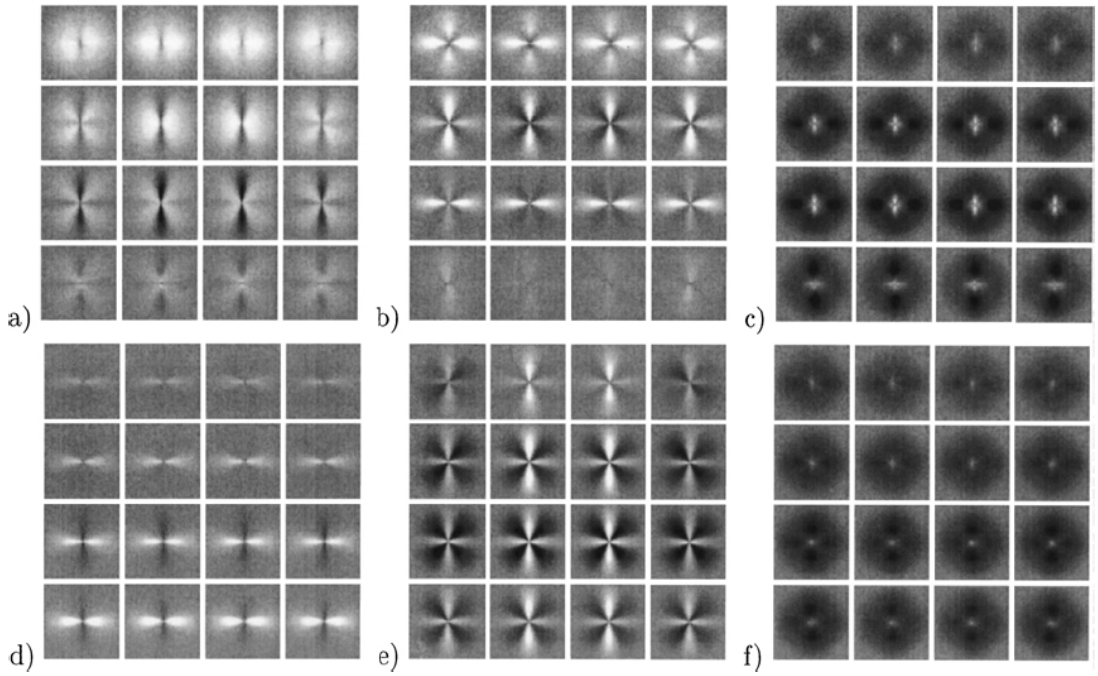
2. *Atributi slike scene* - preko diskretne Fourierove transformacije, uzevši u obzir distribuciju intenziteta slike, prostornih komponenata te Hanningovog prozora dobiva se kompleksna funkcija $I(f_x, f_y)$ koja se sastoji od amplitude i faze. Fazna funkcija $\phi(f_x, f_y)$ predstavlja informacije o lokalnim atributima slike, a amplituda $A(f_x, f_y)$ o globalnim atributima.

3. *Spektar kategorija scena* - Za svaku je kategoriju scena na malom skupu primjera provedena spektralna analiza. Dobivena je aproksimirana funkcija kojom je moguće dobiti spektar svake od kategorija:

$$E[A(f,\theta)^2|S] \simeq \Gamma_s(\theta)/f^{-\alpha_s(\theta)} \tag{3.2}$$

koja predstavlja očekivanje energetskog spektra za kategoriju S.

4. *Računanje atributa prostorne ovojnice scena prirode* - U scenama prirode dominiraju značajke otvorenosti, grubosti i rigidnosti. Te su značajke naučene metodom linearne regresije pomoću podskupa veličine 500 slika iz baze od ukupno 4000 slika. Za pojedinu su značajku dobiveni tipični spektri kao što je vidljivo na slici 3.3.



Slika 3.3: Tipični diskriminantni spektri za prirodne scene: a) otvorenost, b) grubost, c) rigidnost. Za scene ljudskog porijekla: d) otvorenost, e)grubost, f) rigidnost

5. *Računanje atributa prostorne ovojnice scena stvorenih umjetnim putem* - Slučajnim je odabirom uzeto 500 primjera iz skupa od 3500 slika. Kao i za scene prirode, organizacija se temelji na značajkama za raspoznavanje kategorije scene.

Diskriminantni spektri koriste se za klasifikaciju kategorija scena. Uspješnost klasifikacije korištenjem ovog pristupa je oko 90%. Osnovni problem kod korištenja ovog pristupa jest potreba za velikom bazom slika kako bi se dobili diskriminacijski spektri.

4. Metoda korištenja kombinacije globalnih i lokalnih značajki scene pri detekciji objekata

Tradicionalni pristupi detekcije objekata uzimaju u obzir samo regije slike, tj. lokalne informacije. Često se koristi pomični prozor ili promatra regija oko detektirane točke interesa. Međutim, lokalna informacija može biti netransparentna, pogotovo ako je objekt koji želimo detektirati malen (npr. prometni znak na većoj udaljenosti). Ta se netransparentnost može smanjiti korištenjem globalnih značajki scene koje zovemo *esencijom* scene kao dodatnim izvorom informacija.

Kako bi se objekt koji je potrebno detektirati smjestio u kontekst, izvode se dva zadatka:

- Detekcija prisutnosti objekta - je li na slici prisutna jedna ili više instanci klase objekta. Definira se kao $P(O = 1|f(I))$. $O = 1$ znači da se objekt nalazi na slici, a $f(I)$ su lokalne i globalne značajke promatrane slike.
- Lokalizacija objekta - nalaženje lokacije i veličine objekta na slici. Formalna definicija je $P(X = i|f(I))$, gdje je $i \in 1, \ldots, N$ diskretizacija skupa mogućih lokacija i veličina ($\sum_i P(X = 1|\cdot) = 1$).

4.1. Lokalne značajke

Uobičajeni pristup je klasifikacija regije slike kao pozadinu ili objekt, ovisno o tome nalazi se traženi objekt u promatranoj regiji slike. Dvije su odluke ključne: koju vrstu lokalnih značajki promatrati, te koji klasifikator iskoristiti na dobivenom vektoru značajki.

Prvo se provodi konvolucija slike skupom filtara. Sveukupno je korišteno 13 filtara: delta funkcije, derivacije Gaussove funkcije, Laplaceove transformacije Gausso-

Slika 4.1: Objekti u mutnoj slici mogu se klasificirati pogrešno bez kontekstualne informacije. Zaokruženi dijelovi slika imaju slične vrijednosti piksela (osim rotacije), a predstavljaju potpuno različite semantičke informacije ovisno o kontekstu u kojem se nalaze.

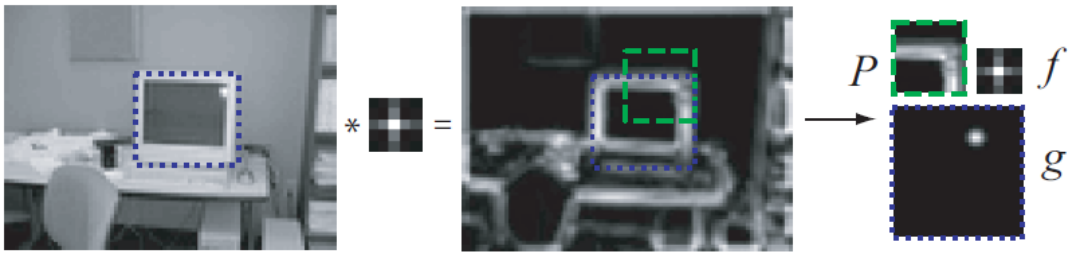
vih funkcija, detektori rubova, te detektori stupaca.

Nakon toga se uzimaju fragmenti slike iz izlaza filtara, nasumično birajući veličinu i lokaciju fragmenata. Ovaj se proces ponavlja za više filtara i fragmenata, te dobivamo rječnik značajki ($N \approx 150$). Prema tome svaki i-ti dio rječnika sadrži filtar f_i, fragment P_i, te Gaussovu masku g_i. Iz ovih informacija gradimo vektor značajki za svaki piksel na slici korištenjem formule

$$v_i = [(I \star f_i) \bigotimes P_i] \star g_i \tag{4.1}$$

gdje $\star$ označava konvoluciju, a $\bigotimes$ normaliziranu kros-korelaciju. $v_i(x)$ je i-ta komponenta vektora značajki na lokaciji x (Slika 4.2). Intuitivno, formula se može objasniti ovako: normalizirana kros-korelacija detektira mjesta gdje se javlja fragment P_i, te se provodi *glasanje* o mjestu središta objekta korištenjem maske g_i (tj. *Houghovom transformacijom*).

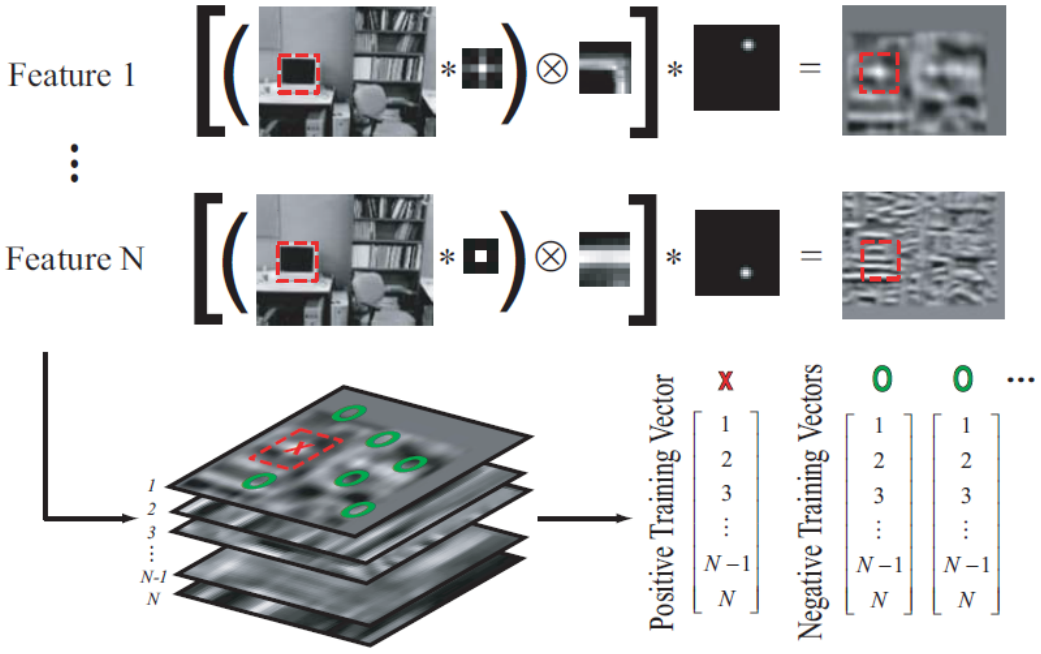
Kao klasifikator koristi se *gentleBoost* algoritam [Friedman et al. (2000)]. Skup za učenje klasifikatora gradi se tako da se stvori vektor značajki za pojedinu sliku, te se filtriraju dijelovi slike - jednom na lokaciji traženog objekta (za dobivanje pozitivnog



Slika 4.2: Izgradnja stavke u rječniku značajki

primjera) te na 20 nasumično odabranih lokacija u slici koje ne sadrže objekt (za dobivanje negativnih primjera). Ti podaci se šalju u klasifikator. Provodi se 50 rundi boostanja, te za skup primjera za učenje veličiine 700 slika potrebno je oko 3 sata kako bi se naučilo klasifikatore. Većina vremena otpada na računanje vektora značajki. Proces izgradnje skupa pozitivnih i negativnih primjera prikazan je na slici 4.3.

Kada je klasifikator naučen na temelju skupa za učenje, njegova primjena na novom primjeru traje oko 3 sekunde za sliku veličine 240×320.



Slika 4.3: Izgradnja skupova pozitivnih i negativnih primjera za učenje.

4.2. Globalne značajke

Esencija scene, odnosno njene globalne značajke dobivaju se postupkom u tri osnovna koraka:

1. Računanje pomične piramide korištenjem 4 orijentacije i 2 veličine, podjela na matricu veličine 4×4

2. Računanje prosječne energije svakog elementa matrice. Dobiva se $4 \times 2 \times 4 \times 4 = 128$ značajki.

3. Redukcija dimenzionalnosti korištenjem metode PCA [Smith (2002)] na 80 dimenzija.

Esencija slike sadrži grubu teksturu i prostornu konfiguraciju scene. Iz dobivene esencije moguće je predvidjeti lokaciju objekata te njihovu veličinu bez korištenja metode lokalne detekcije (Slika 4.4). Ovo omogućava usmjeravanje lokalnog detektora samo na područja u slici gdje je visoka vjerojatnost za nalaženje traženog objekta, što značajno ubrzava proces detekcije objekata [Murphy et al. (2006)].

Slika 4.4: Predviđanje lokacija objekata u sceni. Parovi slika: cijela scena (lijevo), regija gdje će najvjerojatnij biti traženi objekt (desno)

5. Zaključak

Pokazalo se kako su promatrani pristupi efektivni pri prepoznavanju biti scene, njene prostorne konfiguracije, te pomoći pri korištenju lokalnih značajki scene za detekciju objekata.

Kontekstualni pristup dobar je za određivanje relacije među objektima te zaključivanje o vjerojatnim lokacijama traženih objekata ako su pronađeni neki njima bliski objekti. Taj se pristup zasniva na statističkim modelima što uvjetuje brzinu i stabilnost kod izvođenja.

Pristup usporedbe esencija scene zasniva se na podešavanju atributa novog primjera kako bi se što bolje preklopio s primjerima iz baze podataka koja sadrži označene slike. Međutim, ovaj princip iziskuje postojanje velike baze primjera.

Obrađene su i dvije metode kategorizacije scene bez uporabe konteksta i lokalnih informacija o pojedinim objektima. Prepoznavanje oblika scene klasificira scenu po njenoj geometrijskoj strukturi i prostornoj konfiguraciji te je uspoređuje s gotovim niskorezolucijskim klasifikatorima kako bi se scena kategorizirala. Ovaj pristup je vrlo brz i efikasan. Pristup korištenjem prostorne ovojnice sporiji je način kategorizacije iako pruža drugačiji pogled na problematiku.

Metode detekcije globalnih značajki koriste algoritme računalnog vida kako bi se na temelju globalnih značajki ubrzalo korištenje lokalnih značajki pri detekciji objekta.

6. Literatura

C. Liu R. Fergus W. Freeman B. Russell, A. Torralba. Object recognition by scene alignment. *International Journal of Computer Vision*, 2005.

J. Friedman, T. Hastie, i R. Tibshirani. Additive logic regression: a statistical view of boosting. *Annals of Statistics*, 2000.

J. Hays i A. Efros. Scene completion using millions of photographs. *SIGGRAPH*, 2007.

C. Liu, J. Yuen, A. Torralba, J. Sivic, i W. Freeman. Sift flow: Dense correspondence across different scenes. 2009.

K. Murphy, A. Torralba, D. Eaton, i W. Freeman. Object detection and localization using local and global features. *International Journal of Computer Vision*, 2006.

A. Oliva i A. Torralba. Modeling the shape of a scene: The holistic representation of the spatial envelope. *International Journal of Computer Vision*, 2001.

A. Oliva i A. Torralba. The role of context in object recognition. *Trends in Cognitive Science*, 2002.

N. Petkov. Gabor filter for image processing and computer vision, 2008. `http://matlabserver.cs.rug.nl/edgedetectionweb/web/edgedetection_params.html`.

B. Russel, A. Torralba, i W. Freeman. Labelme - the open annotation tool, 2005. `http://labelme.csail.mit.edu`.

S. Sclaroff i A. Pentland. Generalized implicit functions for computer graphics. *Computer graphics*, 1991.

E. Simoncelli i W. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. *2.nd IEEE International Conference on Image Processing*, 1995.

L. Smith. A tutorial on principal component analysis, 2002.

A. Torralba. Contextual influences on saliency. *International Journal of Computer Vision*, 2004.

A. Torralba i W. Freeman. Shape recipes: Scene representations that refer to the image. *International Journal of Computer Vision*, 2004.

7. Sažetak

Ovaj rad predstavlja uvod u metode za raspoznavanje i kategoriziranje scene te njenih globalnih značajki. Analizira se uporaba konteksta pri detekciji objekata, te se uvode metode zaključivanja o lokaciji i veličini objekta iz globalnih atributa scene ako je traženi objekt netransparentan, prekriven ili jako malen pa ga obični lokalni klasifikatori ne mogu detektirati bez rizika prenaučenosti. Promatrane su probabilističke metode, kao i one koje se zasnivaju na algoritmima strojnog učenja te računalnog vida.