

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Petar Vračar

**Učenje in prepoznavanje strukturnih
elementov fasade
iz digitalnih slik**

MAGISTRSKA NALOGA

Mentor: doc. dr. Marko Robnik-Šikonja

Somentor: prof. dr. Igor Kononenko

Ljubljana, 2010

Št.: 128-MAG-RI/2009

Datum: 23. 4. 2009



Petar VRAČAR, univ. dipl. inž. rač. in inf.

Ljubljana

Fakulteta za računalništvo in informatiko Univerze v Ljubljani izdaja naslednjo magistrsko nalogo

Naslov naloge: **Učenje in prepoznavanje strukturnih elementov fasade iz digitalnih slik**

Learning and recognition of façade's structural elements from digital images

Tematika naloge:

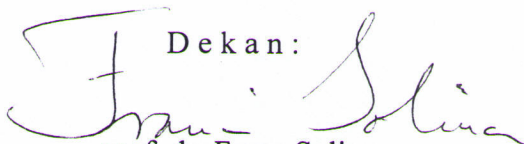
Razvijte metodologijo za razpoznavanje objektov na slikah na podlagi formalizma, ki bo omogočil semantično interpretacijo razpoznanega objekta. Izbrani formalizem zgradite s pomočjo metod strojnega učenja. Praktično rabo metodologije pokažite na problemu prevedbe fotografij fasad zgradb v simbolični opis. Uporabite metode računalniškega vida in strojnega učenja za potrebne transformacije slik, razpoznavanje osnovnih gradnikov in združevanje letih v osnovne strukturne elemente. S pomočjo metod strojnega učenja zgradite formalni opis razpoznane strukture, ki bo omogočila semantično interpretacijo fasade, njeno simetričnost in ponavljajoče se vzorce.

Mentor:


doc. dr. Marko Robnik Šikonja



Dekan:


prof. dr. Franc Solina

Somentor:


prof. dr. Igor Kononenko

Rezultati magistrskega dela so intelektualna lastnina Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavlanje ali izkoriščanje rezultatov magistrskega dela je potrebno pisno soglasje Fakultete za računalništvo in informatiko ter mentorja.

Original izdane teme magistrskega dela s podpisom mentorja in dekana ter
žigom fakultete

IZJAVA O AVTORSTVU

magistrskega dela

Spodaj podpisani/-a Petar Vračar,

z vpisno številko 24015772,

sem avtor/-ica magistrskega dela z naslovom:

Učenje in prepoznavanje strukturnih elementov fasade iz digitalnih slik

S svojim podpisom zagotavljam, da:

- sem magistrsko delo izdelal/-a samostojno pod vodstvom mentorja (naziv, ime in priimek)
doc. dr. Marka Robnik-Šikonje
in somentorstvom (naziv, ime in priimek)
prof. dr. Igorja Kononenka
- so elektronska oblika magistrskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko magistrskega dela
- soglašam z javno objavo elektronske oblike magistrskega dela v zbirki "Dela FRI".

V Ljubljani, dne 26. 1. 2010

Podpis avtorja/-ice:

Zahvala

Na začetku bi se zahvalil mentorju doc. dr. Marku Robnik-Šikonji in somentorju prof. dr. Igorju Kononenku, ki sta me s svojimi nasveti usmerjala na pravo pot. Hvala vama za koristne ideje, trud in potrpljenje, ki sta ga izkazala do mene.

Zahvalil bi se kolegu Eriku Štrumblju, ki probleme drugih tako resno jemlje, kot če bi bili njegovi. Pomagal mi je z mnogimi idejami ter me bodril v trenutkih malodušnosti. Kolegu doc. dr. Zoranu Bosniću se zahvaljujem za razumevanje, ki ga je pokazal v času mojega dela na magistrski nalogi, in je nase prevzel tudi del mojih pedagoških obveznosti. Zahvalil bi se tudi ostalim članom Laboratorija za kognitivno modeliranje, doc. dr. Matjažu Kukarju in Darku Pevcu, za spodbudne besede.

Zahvalil bi se moji mami, ker je vedno verjela vame in mi predstavlja temeljni kamen, oporišče ob življenskih preizkušnjah.

Zahvalil bi se Robertu Korenjaku, ki je žrtvoval veliko prostega časa in ga namenil pogovoru o problematiki s katero sem se ukvarjal, ter na ta način pomagal pri oblikovanju idej, ki so se materializirale v pričujoči magistrski nalogi.

Na koncu bi se še enkrat zahvalil prof. dr. Igorju Kononenku, ker je nekega poletnega dneva prišel s čolnom v trenutku, ko so se mi vse ladje potopile.

Za mog tatu

Kazalo

Povzetek	1
Abstract	2
1 Uvod	3
2 Pregled področja	6
3 Metodologija	10
3.1 Predprocesiranje	11
3.1.1 Perspektivna izravnava	11
3.2 Generiranje hipotez	19
3.2.1 Izračun projekcijskih profilov	22
3.2.2 Glajenje projekcijskih profilov	23
3.2.3 Segmentacija fasade na podlagi projekcijskih profilov	27
3.2.4 Pričakovane dimenzije oken	29
3.3 Vrednotenje hipotez	35
3.3.1 Hevristična ocena na podlagi pozicije v drevesni strukturi	37
3.3.2 Hevristična ocena na podlagi distribucije robov	37
3.3.3 Hevristična ocena na podlagi medsebojne podobnosti	44
3.3.4 Hevristična ocena na podlagi naučene podobe oken	47
3.4 Formiranje končne hipoteze	57
4 Eksperimentalni del	64
4.1 Primeri dobrega delovanja	68
4.2 Problematični primeri	69
5 Zaključek	79
Seznam slik	82

Seznam algoritmov	84
Seznam tabel	85
Literatura	86

Povzetek

V magistrski nalogi je predstavljen postopek za učenje in prepoznavanje oken kot osnovnih strukturnih elementov fasade. Postopek predpostavlja pravokotna okna in je prilagojen segmentaciji fasad, ki se ponašajo z visoko stopnjo poravnosti oken v horizontalni in vertikalni smeri. Omejitve nisi prevelike, saj ima iz praktičnih in estetskih razlogov ogromna večina fasad omenjene lastnosti.

Postopek razpoznavanja oken na fasadi se prične s perspektivno izravnavo vhodne slike, s katero je zagotovljeno, da robovi objektov, ki so v prostoru paralelni, ostajajo paralelni tudi v slikovni ravnini. Obdelava perspektivno izravnane slike se nadaljuje s segmentacijo fasade v hierarhično strukturo in z generiranjem kandidatov za okna. Predstavljene so štiri hevristične funkcije, ki ocenjujejo ustreznost posameznih kandidatov na podlagi položaja v hierarhični strukturi, na podlagi distribucije horizontalnih in vertikalnih robov, na podlagi medsebojne podobnosti in na podlagi naučenega videza okna. Generiranje končne hipoteze se izvede s postopkom, ki na požrešen način izbira najbolj ocenjene kandidate ter favorizira poravnost in homogenost razpoznanih oken na fasadi.

Eksperimentalni rezultati so zadovoljivi vendar bi za praktično rabo bilo potrebno metodo dopolniti. Glavna pomanjkljivost metode je ignoriranje splošnega predznanja o strukturi fasad. Predlagano metodo lahko uporabimo pri izračunu približne rešitve kot izhodiščnega položaja za neko drugo metodo prevedbe fasade v simbolni opis, na primer stohastično gramatiko, ki vključuje predznanje o simetrični strukturi in ponavljajočih se vzorcih fasad.

Ključne besede:

segmentacija fasade, razpoznavanje oken, perspektivna izravnavo, urbano okolje

Abstract

We describe a method for learning and recognizing windows a basic structural elements of façades. We assume rectangular windows and our method is adapted to the segmentation of façades which are horizontally and vertically aligned. These limitations are not very restrictive since vast majority of façades have these properties for practical and eastetic reasons.

The method begins with perspective rectification of the input image, which ensures that the edges of objects, parallel in space, remain parallel in the image plane. The rectified image is then segmented into hierarchical structure of window candidates. Four heuristic functions for estimating the quality of generated window candidates are presented. They are defined on the basis of position in the hierarchical structure, the distribution of horizontal and vertical edges, the similarities between candidates, and the learned window appearances. The generation of the final hypothesis uses a greedy approach, which favors homogeneous and aligned candidates.

Experimental results are satisfactory but the method should be supplemented before the practical use. The main disadvantage of the method is ignorance of general background knowledge of façade structures. The proposed method can be used in the calculation of approximate solutions and as a starting point for methods, like stochastic grammar, which includes background knowledge on the structure of symmetrical and repeated patterns of façades.

Key words:

façade segmentation, window detection, image rectification, urban environment

Poglavje 1

Uvod

Razumevanje vsebine slike je ena izmed osnovnih nalog kognitivnih znanosti in je predmet intenzivnih raziskav na področju računalniškega vida in umetne inteligence. Ima pomembno vlogo pri izdelavi avtonomnega inteligentnega sistema, ki potrebuje funkcije za zaznavanje in interpretacijo informacij iz okolja, v katerem se nahaja, s pomočjo katerih bo izvajal in planiral svoje akcije. Avtonomni inteligentni sistem pridobiva informacije o okolju s pomočjo različnih senzorjev, med drugim tudi za zajem slikovnih podatkov. Obdelava teh podatkov na nivoju slikovnih elementov je zelo kompleksna in nepraktična. Visokonivojsko obdelavo zajetih slik za pridobivanje vizualne informacije lažje izvajamo s prevedbo podatkov iz slikovnih elementov v simbolni zapis. Običajno omenjeno prevedbo izvajamo tako, da na nivoju slikovnih elementov najprej poiščemo osnovne strukturne gradnike, kot so na primer linearni segmenti, koti ali druge lokalne značilke, ki predstavljajo simbole na najnižjem nivoju. Kompleksnejše višjenivojske simbole dobimo z nadaljnjim hierarhičnim združevanjem preprostejših osnovnih simbolov. Simbolni opis slike tako predstavlja urejena hierarhična struktura zaznanih simbolov, ki na najnižjem nivoju vsebuje slikovne elemente.

V zadnjih nekaj letih so se intenzivirale raziskave na področju razpoznavanja in interpretacije urbanega okolja. Urbanistično planiranje, tridimenzionalno modeliranje naselja, simulacije evakuacij prebivalstva v naseljih, simulatorji vožnje, vizualni navigacijski sistemi ter računalniške igre so primeri aplikacij, ki v večji ali manjši meri potrebujejo interpretacijo vizualne informacije urbanega okolja. Bazične funkcije so detekcija, rekonstrukcija in arhitekturna interpretacija stavb kot osnovnih elementov gradnikov urbanega okolja. Pomembno vlogo pri realizaciji omenjenih funkcij imajo postopki za zaznavanje

in interpretacijo fasad kot nosilca vizualne podobe stavbe.

V našem delu smo raziskovali problematiko prevedbe fotografij fasad stavb v simbolični opis. Osredotočili smo se na zaznavanje oken kot osnovnega simbola v opisu fasade. Pri tem nismo izhajali iz nobenih predpostavk o številu ali velikosti oken na sliki. Omejili smo se na razpoznavanje pravokotnih oken, ter na analizo fasad, ki se ponašajo z večjo poravnanoostjo oken. Predpostavljene lastnosti veljajo za veliko večino fasad urbanih središč, tako da niso preveč omejujoče.

Proces razpoznavanja oken je otežen zaradi velike raznolikosti v izgledu. Videz oken spreminjajo delno odprta polkna, delno spuščene žaluzije, zavese in drugi dekorativni objekti, kot so rože na policah in podobno. Na isti fasadi so lahko prisotna okna različnih velikosti. Steklo je reflektivni material in zato se pogosto okna na isti stavbi zelo razlikujejo tudi zaradi odsevov neba ali okolišnjih objektov. Fasade so lahko pisane, vsebujejo različne okraske, ki se prav tako periodično pojavljajo, kar dodatno otežuje zaznavanje oken. Nalogo dodatno otežuje tudi perspektivna transformacija, ki ima za posledico zmanjševanje videza sicer enako velikih objektov z večanjem njihove oddaljenosti od opazovalca.

Kot pri večini metod za razpoznavanje objektov pri detekciji oken izhajamo iz lokalnih struktur na nivoju slikovnih elementov. Večina oken je pravokotne oblike, zato so linearni robovi na slikah fasad dobro izhodišče za generiranje hipoteze o položaju oken.

Pri zaznavanju oken si pomagamo tudi z dejstvom, da so stavbe zasnovane po nadstropjih. Posamezne etaže so razdeljene na prostore. Običajno je organiziranost prostorov po etažah uniformna za celo stavbo. Ker je zaželjena prisotnost naravne svetlobe, običajno vsak ločen prostor vsebuje eno ali več oken. Kot posledica naštetega je prisotna poravnanoost oken v horizontalni smeri (po posameznih etažah) ter v vertikalni smeri (po posameznih prostorih na različnih etažah). Ta lastnost je ključnega pomena za detekcijo kandidatov za okna fasade.

Čeprav se v splošnem razlikujejo po obliki in velikosti, so okna na isti fasadi večinoma podobna med seboj, kar lahko izkoristimo v procesu razpoznavanja. Če se določeni deli fasade velikokrat ponovijo na sliki, se povečuje tudi verjetnost, da tisti deli ustrezajo oknom. Seveda je možno, da s takšnim sklepanjem

napačno razglasimo za okna neke druge strukturne elemente fasade, na primer ornamente ali kake druge okraske.

Nenezadnje je možno uporabiti metode strojnega učenja in ob ustrezni izbrani množici značilk razviti klasifikator za razpoznavanje oken na podlagi naučenega videza.

Naslednje poglavje je namenjeno pregledu področja, v katerem smo predstavili pomembnejše raziskave na temo rekonstrukcije in interpretacije zgradb. Osredotočili smo se na metode za strukturno analizo fasad, ki smo jih uporabili kot izhodišče za naše delo.

V tretjem poglavju je podrobneje predstavljena uporabljena metodologija. Najprej je opisano predprocesiranje vhodnih slik, ki ima za nalogo poenostaviti nadaljnjo obdelavo in zagotoviti večjo zanesljivost uporabljenih postopkov. Predvsem smo se ukvarjali s problematiko odpravljanja posledic perspektivne transformacije, ki je praviloma prisotna na fotografijah stavb. Podrobneje je opisan postopek segmentacije perspektivno izravnane slike fasade, s katero generiramo kandidate za okna. Predstavljene so štiri hevristične ocene, ki nam pomagajo ovrednotiti kvaliteto posameznih kandidatov za okna. Na koncu poglavja je predstavljen postopek formiranja končne hipoteze o poziciji in velikosti oken na fasadi, ki kombinira ocene opisanih hevristik.

Četrto poglavje razkriva rezultate eksperimentalnega dela naloge. Opisana je programska rešitev predstavljene metodologije ter rezultati na dveh testnih množicah. Podanih je nekaj tipičnih tipov fasad in komentirani so doseženi rezultati, ki ilustrirajo prednosti in slabosti uporabljene metodologije.

Na koncu so predstavljene smernice za nadaljnje delo ter ideje za izboljšave doseženih rezultatov.

Poglavje 2

Pregled področja

Tradicionalni pristop k razpoznavanju objektov temelji na njihovih fotometričnih lastnostih, upodobljenih na sliki. V to kategorijo spadajo modeli za razpoznavanje objektov na podlagi lokalnih in globalnih značilik [Lowe, 2004; Turk in Pentland, 1992; Belongie in sod., 2002; Viola in Jones, 2001]. Osnovna pomanjkljivost teh metod je obravnavanje objekta kot celote, kar zahteva relativno veliko učno množico in povzroča večjo občutljivost na šum, spremembe zornega kota, zakrivanje in prekrivanje objektov in drugih vplivov, ki spreminjajo videz objekta na sliki. Večjo robustnost na navedene težave omogočajo metode, ki temeljijo na modeliranju strukture objekta [Felzenszwalb in Huttenlocher, 2005; Fergus in sod., 2003; Fischler in Elschlager, 1973; Weber in sod., 2000]. Objekt je predstavljen kot zbirka sestavnih delov. Posamezni sestavni deli opisujejo lokalne vizualne lastnosti modeliranega objekta, medtem ko prostorske relacije med sestavnimi deli določajo njegovo konfiguracijo. Konfiguracija opazovanega objekta podaja njegovo hierarhično strukturo in omogoča njegovo semantično interpretacijo.

Metode modeliranja strukture objektov na podlagi zajetih podatkov v grobem lahko razdelimo na dve skupini. Prvi pristop ustreza metodologiji od spodaj navzgor ter izhaja iz zajetih podatkov, na podlagi katerih gradi najprej preproste modele, ki jih nato združuje v bolj kompleksne opise.

Drugi pristop, po principu od zgoraj navzdol, izhaja iz razpoložljivega znanja o domeni, predpostavi začetno obliko modela, ki ga potem dopolnjuje in modificira s prilagajanjem zajetim podatkom. Pri tem uporabljeno znanje o domeni usmerja proces iskanja modela ter zagotavlja smiselnost in verodostojnost najdene rešitve. Iz tega sledi, da je način predstavitve prisotnega znanja

o domeni ključnega pomena za uspešnost metod, ki gradijo od zgoraj navzdol.

Za modeliranje odnosov med sestavnimi deli objekta se najbolj pogosto uporabljajo formalizmi iz teorije grafov, v zadnjem času pa se več pozornosti namenja uporabi formalnih gramatik za opis struktur modeliranih objektov [Han in Zhu, 2005; Zhu in sod., 2006; Zhu in Mumford, 2006]. Prednost uporabe formalnih gramatik je v možnosti generiranja velikega števila konfiguracij na podlagi relativno malega števila pravil in osnovnih gradnikov. Primer uspešne uporabe formalnih gramatik so Lyndenmayerjevi sistemi (L-systems) [Prusinkiewicz in Lindenmayer, 1990] za modeliranje razvoja rastlin. Formalne gramatike se uspešno uporabljajo tudi v domeni razpoznavanja in rekonstrukcije objektov v urbanem okolju, saj se celotna naselja, posamezne stavbe ter njihove fasade ponašajo z veliko stopnjo simetričnosti in ponavljajočih se vzorcev.

Že v sedemdesetih letih prejšnjega stoletja sta Stiny in Gips [1971] predstavila idejo uporabe gramatike oblik kot formalnega orodja za opisovanje in interpretacijo objektov v arhitekturnem oblikovanju. Gramatika oblik, definirana po analogiji s formalnimi gramatikami, omogoča proceduralno generiranje n-dimenzionalnih geometrijskih oblik (v praksi ponavadi omejeno na dvo- ali tridimenzionalne prostore). Terminali gramatike predstavljajo primitivne geometrijske oblike, ki jih z uporabo produkcijskih pravil kombiniramo oz. modificiramo in na ta način kreiramo čedalje bolj kompleksne strukture. Dolgo časa so se uporabljale predvsem na področju teoretičnega raziskovanja in analiz, medtem ko je bila aplikativna uporaba otežena zaradi prevelike splošnosti osnovnega formalizma.

V zadnjih letih so se raziskave na področju uporabe gramatik pri arhitekturnem oblikovanju ponovno intenzivirale. Raziskovalci so se osredotočili na poenostavljanje geometrijskih pravil ter razširjanje mehanizma izpeljav.

Wonka [2003] je predstavil sistem za avtomatsko generiranje tridimenzionalnih modelov stavb, ki temelji na uporabi dveh gramatik. Za oblikovanje geometrije objekta je zadolžena t.i. delitvena gramatika ("split grammar"), poimenovana po tipu produkcijskih pravilnih za razbitje objekta v hierarhično strukturo sestavnih delov. Produkcijska pravila delitvene gramatike so parametrizirana ter točno predpisane oblike. Druga gramatika, t.i. kontrolna, pa je zadolžena za postavljanje vrednosti parametrov delitvene gramatike ter na ta način za krmiljenje procesa modeliranja stavbe v skladu z arhitekturnimi

principi. Zahvaljujoč svoji fleksibilnosti je bil opisani sistem sposoben generirati kompleksne in realistične oblike stavb v različnih arhitekturnih slogih. Po drugi strani pa sistem ni bil toliko uspešen pri dodajanju elementov fasade na stavbe zelo kompleksnih oblik, kar je zahtevalo veliko število delitvenih pravil.

Izhajajoč iz opisanih gramatik, je Müller [2006] predstavil kontekstno odvisno gramatiko, ki poleg modeliranja kompleksnih volumetričnih oblik poljubne orientacije omogoča tudi postavitev fasadnih elementov, konsistentnih s modelom stavbe. Na podlagi tega sistema je nastal komercialni programski paket CityEngine [Parish in Müller, 2001] za proceduralno kreiranje prostranih tridimenzionalnih modelov urbanega okolja.

V zadnjem času je opazna tendenca uporabe stohastičnih metod pri procesih rekonstrukcij stavb in posebej njihovih fasad. Dick s sod. [2004] je uporabil Bayesov model za avtomatsko pridobivanje tridimenzionalne predstavitve stavbe na podlagi več slik. Osnovna ideja temelji na gradnji modela stavbe po principu zlaganja kock. Stavba je predstavljena kot množica sestavnih delov, kot so stene, okna, vrata in podobno, ki so parametrizirani geometrijski objekti. Trenutni model stavbe se projicira na vhodne slike ter se na podlagi ugotovljenih razlik sproži stohastični proces modifikacije modela, v smislu dodajanja novih ali odstranjevanja obstoječih sestavnih delov oziroma spreminjanju vrednosti njihovih parametrov, s ciljem približevanja modela vhodnim slikam. Pri izbiri ustrezne modifikacije se upoštevajo apriorne verjetnosti vrednosti posameznih parametrov (širina, višina, barva in drugo) ter tipične karakteristike stavb, ki so deloma naučene iz učne množice, deloma pa določene s strani ekspertov arhitektov. Alegre in Dallaert [2004] sta predstavila stohastično metodo za semantično interpretacijo fasad iz slik. Predznanje o fasadah sta zakodirala v stohastično kontekstno neodvisno gramatiko, ki generira smiselne hipoteze v obliki hierarhičnega razbitja vhodne slike. Metoda z nederministično izbiro produkcijskih pravil gramatike modificira hierarhično strukturo ter na ta način išče najbolj verjeten model fasade, ki se dovolj dobro prilega vhodni sliki.

Ripperda in Brenner [2006] sta na podoben način z uporabo formalne gramatike modelirala hierarhično strukturo fasad, pri čemer sta kot vhod, poleg slik, uporabila tudi globinske podatke. Distribucijo uporabe produkcijskih pravil sta modelirala na podlagi večjih sprememb barve oziroma globine vhodnih podatkov. Drugi sprožilec razbitij je ugotovljena simetričnost strukture na podlagi avtokorelacije posameznih oken.

Skupna lastnost stohastičnih metod rekonstrukcije modela stavb oziroma njihovih fasad je prevelik iskalni prostor z veliko lokalnih minimumov. V takih primerih je običajni pristop pri iskanju globalnih minimumov uporaba simulacij Monte Carlo z Markovskimi verigami (MCMC) [Neal, 1993]. MCMC simulacija predvideva fiksen nabor parametrov, zato se pri vrednotenju modelov z različnim številom parametrov uporablja inačica Reversible Jump MCMC [Green, 1995].

Niso vse raziskave na področju segmentacije in interpretacije fasad stohastične narave. Müller s sod. [2007] je razvil postopek segmentacije fasade na podlagi ugotovljene simetrije ponavljajočih se vzorcev. Za detekcijo ponavljajoče se strukture je uporabljal medsebojno informacijo različnih področij vhodne slike. Samo strategijo delitve fasade je predstavil v obliki optimizacijskega problema iskanja vertikalnih in horizontalnih ločilnic. Kriterijska funkcija optimizacijskega problema nagrajuje postavitve horizontalne ločilnice v tisti del slike, v katerem je veliko horizontalnih robov, ter kaznuje prisotnost vertikalnih robov. Na podoben način je definirana tudi kriterijska funkcija za postavitve vertikalnih ločilnic.

Lee in Nevatia [2004] sta iskala kandidate za okna na fasadi s pomočjo projekcijskih profilov, ki predstavljata akumulacije projekcij horizontalnih in vertikalnih robov vhodne slike. Histogram projekcijskega profila predstavlja močno indikacijo prisotnosti poravnane strukture oken ter ima kot tak odločilno vlogo pri postopku segmentacije fasade. Lee in Nevatia sta segmentacijo fasade izvedla z analizo projekcijskih profilov tako, da sta kandidate za okna definirala v presečiščih pasov, ki ustrezajo vrhovom v histogramih porazdelitve robov. Metoda daje dobre rezultate pri segmentaciji fasad z izrazito poravnano strukturo oken.

Haugeard s sod. [2009] je predstavil postopek, ki z uporabo projekcijskih profilov kombinira globalno in lokalno informacijo pri segmentaciji fasade. Najprej se, na podlagi vertikalnega projekcijskega profila celotne slike, fasada razdeli na etaže. Posamezne etaže se nato neodvisno delijo na okna s pomočjo horizontalnih projekcijskih profilov, ki so izračunani za vsako etažo posebej.

Poglavje 3

Metodologija

V tem poglavju bomo predstavili postopek za učenje in prepoznavanje strukture fasade, pri čemer smo se omejili na razpoznavanje oken kot osnovnih strukturnih elementov fasade. Pri tem smo izhajali iz naslednjih pretpostavk:

- okno je pravokotne oblike,
- okna so medseboj horizontalno in vertikalno poravana na fasadi,
- okna ene fasade so medseboj podobna.

Predpostavke niso preveč omejujoče, saj ima iz praktičnih in estetskih razlogov ogromna večina fasad ter oken omenjene lastnosti.

Postopek prepoznavanja strukture fasade začneja s predprocesiranjem vhodne slike. Osnovni namen te faze je povečanje zanesljivosti in zmanjšanje zahtevnosti uporabljenih postopkov za prepoznavanje strukture fasade. V prvi vrsti je predstavljen postopek za perspektivno izravnavo vhodne slike, s katero zagotovimo, da robovi objektov, ki so v prostoru paralelni, ostajajo paralelni tudi v slikovni ravnini. Omenjena lastnost zelo poenostavi postopke razpoznavanja struktur (v našem primeru so to okna fasade), ki se v prostoru ponašajo z veliko stopnjo paralelnosti in ortogonalnosti.

Po dokončanem predprocesiranju vhodne slike sledi postopek iskanja hipotez o položaju in dimenzijah oken na fasadi. Predstavljen je rekurziven postopek gradnje hierarhične strukture pravokotnih hipotez na podlagi distribucije horizontalnih in vertikalnih robov na sliki. Pravokotne hipoteze ustrezajo izhodiščni predpostavki o obliki oken na fasadi, medtem ko horizontalni in vertikalni

robovi ustrezajo robovom oken na perspektivno izravnani sliki.

Predstavljene so štiri hevristične ocene, s katerimi ovrednotimo posamezne hipoteze o položaju in dimenzijah oken na fasadi. Hevristične ocene izhajajo iz predpostavk o poravnosti in medsebojni podobnosti oken. Predstavljen je tudi postopek učenja kaskadnega klasifikatorja za izvedbo hevristične ocene hipotez na podlagi razpoznavanja videza oken.

Na koncu tega poglavja je predstavljen postopek za formiranje končne hipoteze o poziciji in velikosti oken na opazovani fasadi.

3.1 Predprocesiranje

Fasade se kot del urbanega okolja po navadi odlikujejo z visoko stopnjo paralelnosti in ortogonalnosti elementov gradnikov (okna, okenske police, vrata in podobno). Na fotografijah fasad se zaradi vpliva perspektivne transformacije omenjena urejenost elementov gradnikov ne ohranja. Da bi čim bolj poenostavili analizo strukture fasade, je potrebno najprej odstraniti perspektivno projekcijo na sliki.

3.1.1 Perspektivna izravnava

Perspektiva je pojav navideznega zmanjševanja velikosti opazovanega predmeta z njegovim oddaljevanjem od opazovalca. Z oddaljevanjem v neskončnost se vsi opazovani objekti stekajo v eno točko, imenovano bežišče.

Če opazovalec opazuje opazovano površino frontalno in smer pogleda sovpada s horizontalno ravnino, govorimo o enobežiščni perspektivi. Vse linije stranskih ploskev, vzporedne s horizontalno ravnino, se stekajo v eno samo bežišče na horizontu. Če pa opazovalec opazuje objekt pod določenim kotom vendar še vedno vzporedno s horizontalno ravnino, govorimo o dvobežiščni perspektivi. V tem primeru imata stranski ploskvi vsaka svojo smer ter s tem tudi vsaka svoje bežišče na horizontu. V najbolj splošnem primeru, ko opazovalec opazuje objekt pod kotom tudi glede na horizontalno ravnino, se vzporednice vzdolž navpične osi prostora stekajo v tretje bežišče nad ali pod horizontom in takrat govorimo o trobežiščni perspektivi.

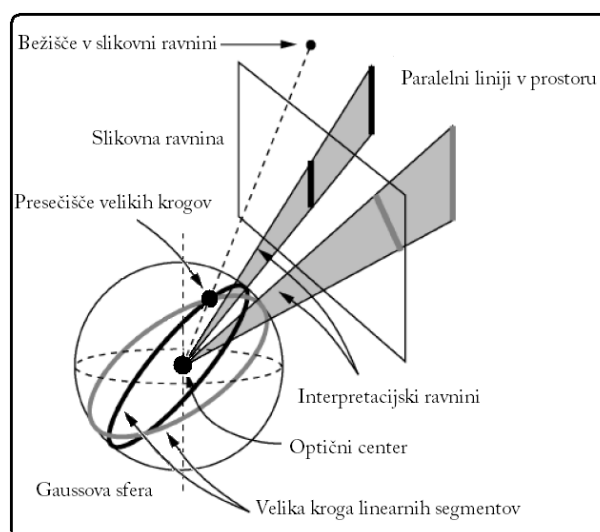
Detekcija bežišč

Prvi korak k avtomatskemu razumevanju tridimenzionalnih scen na podlagi slik, dobljenih s perspektivno projekcijo, je iskanje bežišč. Le-ta nam omogočajo dojetanje relativne orientacije objektov glede na opazovalca. S pomočjo bežišč lahko izvedemo postopek perspektivne izravnave (rektifikacije), s katerim odpravimo perspektivno projekcijo scene ter sliko transformiramo v paralelno projekcijo, ki ohranja paralelnost in ortogonalnost linij iz prostora.

Osnovna ideja iskanja bežišč v slikovni ravnini je naslednja. Najprej poiščemo robove objektov na sliki fasade ter poiščemo premice, na katerih leži največ zaznanih robov. Če bo večina robnih točk ustrezala dejanskim robovom oken oziroma drugim sestavnim delom fasade, ki so v resnici poravnani s horizontalno in vertikalno osjo, bodo namišljene premice skozi te robove paralelne. Ker vemo, da se zaradi perspektivne projekcije vzporednice stekajo v bežišče, pričakujemo da se bodo vse linije skozi robove objektov fasade v slikovni ravnini sekale v isti točki (oziroma zaradi šuma lahko rečemo, da se bodo sekale v neki relativno majhni okolici). Večina postopkov iskanja bežišč temelji na obdelavi linearnih segmentov na sliki. Najprej poiščemo intenzivne spremembe svetlosti slikovnih elementov, do katerih pride na robovih objektov. Omenjeni postopek ponavadi realiziramo z odvajanjem digitalne slike z uporabo gradientnih operatorjev (npr. Sobelov operator [Sobel, 1978]). Tako detektirane robne točke lahko dodatno obdelamo s postopki odpravljanja nenajvečjih vrednosti, tanjšanja robov ali z uporabo Cannyjevega postopka [Canny, 1986]. Ko smo identificirali robne slikovne elemente, jih je potrebno povezati v premice. Standarden postopek pri reševanju nalog tega tipa je Houghova transformacija [Hough, 1962]. Postopki iskanja so običajno zasnovani iz dveh faz. V prvi fazi se najdeni linearni segmenti grupirajo v skupine na podlagi hipotez o stekanju v isto bežišče. V drugi fazi se z vrednotenjem izberejo dominantne skupine linearnih segmentov, ki definirajo dejanska bežišča v slikovni ravnini. Generiranje hipotez o dejanskem položaju bežišč v slikovni ravnini je lahko časovno zelo zahtevno, saj se potencialna bežišča lahko nahajajo kjerkoli v dvodimenzionalnem prostoru \mathbb{R}^2 . Običajni postopek zmanjševanja prostora hipotez je preslikava tega prostora v nek drugi končen parametrični prostor. Le-ta je običajno diskretiziran na končno število celic, ki formirajo akumulacijsko polje.

Barnard [1983] je predlagal parametrični prostor v obliki enotske sfere s središčem v optičnem centru opazovalca, znano tudi kot Gaussova sfera.

Vsak linearni segment na sliki definira premico v slikovni ravnini, ki skupaj z optičnim centrom definira t.i. interpretacijsko ravnino. Presečišče interpretacijske ravnine in Gaussove sfere je na velikem krogu (krožnica na sferi z enakim središčem kot sfera). Vsak velik krog na Gaussovi sferi enolično določa interpretacijsko ravnino ter s tem tudi premico v slikovni ravnini.



Slika 3.1: Uporaba Gaussove sfere kot akumulacijskega polja. Velika kroga, ki ustrezata linearnima segmentoma na sliki, se sekata v eni točki, ki ustreza bežišču.

Dva velika kroga na Gaussovi sferi se sekata v eni točki, ki predstavlja bežišče ustreznih linearnih segmentov v slikovni ravnini. Iz tega izhaja osnovna ideja uporabe Gaussove sfere kot akumulacijskega polja. Gaussovo sfero razbijemo na področja ter vsakemu področju pridružimo akumulacijsko celico. Vsak linearni segment na sliki povečuje vrednosti tistih akumulacijskih celic, skozi katere poteka velik krog, določen kot presečišče interpretacijske ravnine linearne segmenta in Gaussove sfere (glej sliko 3.1). Lokalni maksimumi v akumulacijskem polju ustrezajo hipotezam o položaju bežišč, ki jih podpira večina linearnih segmentov na sliki. Uporaba Gaussove sfere kot akumulacijskega polja omogoča enotno obravnavo bežišč v končnih točkah ter bežišč v neskončnosti.

Glavne pomanjkljivosti klasičnih metod iskanja bežišč na Gaussovi sferi

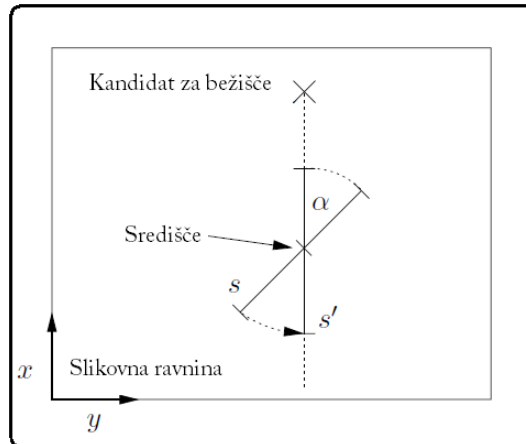
izhajajo iz občutljivosti na parametrizacijo prostora, kar vpliva na natančnost rezultatov in na zahtevnost iskanja lokalnega maksimuma v akumulacijskem polju. Način akumulacije rezultatov je odvisen tudi od relativnega položaja linearnih segmentov glede na njihova bežišča v slikovni ravnini. Posledica tega je, da linearni segmenti, čeprav se stekajo v isto bežišče, ne prispevajo k popolnoma enakemu delu akumulacijskega polja, kar dodatno poveča občutljivost na parametrizacijo. Nenazadnje, klasične metode obravnavajo vse linearne segmente na enak način oz. vsem segmentom pripisujejo enako pomembnost, kar povečuje občutljivost na šum.

V našem delu smo za iskanje bežišč uporabili postopek, ki ga je predlagal Rother [2000]. Akumulacijsko polje je definirano v slikovni ravnini. Akumulacijske celice (kandidati za bežišča) so postavljene v presečiščih premic skozi vse pare linearnih segmentov na sliki. Kot potencialne kandidate obravnavamo tudi presečišča v neskončnosti (med paralelnimi segmenti). Za razliko od kandidatov v končnih točkah, ki so predstavljeni s svojimi koordinatami v slikovni ravnini, kandidate v neskončnosti predstavimo s smernim koeficientom. V tem koraku je možno takoj izločiti kandidate, ki ležijo na samih linearnih segmentih (med robnima točkama), ker vemo, da bežišča v teh točkah niso možna.

Najdene kandidate ovrednotimo po principu glasovanja na podlagi dolžine in števila linearnih segmentov, ki jih podpirajo. Zaradi večje robustnosti na šum vrednotimo tudi stopnjo podpore, ki jo posamezni linearni segment ponuja določenemu kandidatu. Zaradi različnih razlogov (na primer končna ločljivost slikovnih elementov ali numerične napake pri postopku iskanja robnih elementov) se ne bodo vsi linearni segmenti, ki ustrezajo paralelnim objektom v prostoru, v slikovni ravni stekali v isto točko. Da bi mehko vrednotili podporo tudi tistih linearnih segmentov, ki nekoliko zgrešijo kandidata za bežišče, definiramo oceno razdalje $d(vp, s)$ s pomočjo kota α med linearnim segmentom s in premico skozi kandidata vp in središče linearne segmenta s (glej sliko 3.2).

Razdalja $d(vp, s)$ je primerna tudi za obravnavanje kandidatov v neskončnosti ter predstavlja kot med premico s smernim koeficientom kandidata vp ter premico skozi segment s . Na ta način dosežemo enotno obravnavo bežišč v končnih točkah ter tistih v neskončnosti.

Sedaj lahko definiramo proces glasovanja, ki bo ovrednotil posamezne kandidate za bežišča. Za kandidata vp bodo glasovali vsi segmenti s , če je nji-



Slika 3.2: Mehka razdalja podpore kandidatu za bežišče je definirana s pomočjo kota α , ki ga linearni segment s oklepa s premico, ki povezuje njegovo središče s kandidatom za bežišče.

hova medsebojna razdalja $d(vp, s)$ manjša od vnaprej določenega praga t_{vp} (npr. 5%). Prispevek posameznega segmenta s je odvisen od razdalje (bolj usmerjeni segmenti prispevajo več) ter dolžine (daljši segmenti, za katere predpostavljamo da so bolj zanesljivi, bodo prispevali več kot krajši segmenti, ki so lahko posledica šuma). Delež prispevkov kot posledico usmerjenosti in dolžine segmenta uravnava parametra w_1 in w_2 (npr. 0.3 in 0.7). Končno oceno kandidata lahko zapišemo kot vsoto posameznih prispevkov:

$$vote(vp) = \sum_x (w_1(1 - \frac{d(vp, s)}{t_{vp}}) + w_2(\frac{\bar{s}}{max\bar{s}})) \quad (3.1)$$

kjer je vp kandidat za bežišče, s je linearni segment na sliki, \bar{s} je dolžina linearnega segmenta, $max\bar{s}$ je dolžina najdaljšega segmenta na sliki, $d(vp, s)$ je razdalja med kandidatom vp in segmentom s , izražena s kotom, t_{vp} je prag dovoljene razdalje, w_1 in w_2 sta uteži.

Časovna kompleksnost iskanja kandidatov ter izvajanja procesa glasovanja je $O(n^3)$, kjer je n število linearnih segmentov (v kvadratnem času poiščemo vsa presečišča kot potencialne kandidate, ki jih potem v linearnem času ocenimo).

Ko je proces glasovanja zaključen, je potrebno izbrati najboljše kandidate

za bežišča. Poleg dobre ocene morajo izbrani kandidati izpolnjevati še druge kriterije, kot je na primer ortogonalnost:

$$\langle cv_1, cv_2 \rangle = 0, \langle cv_2, cv_3 \rangle = 0, \langle cv_1, cv_3 \rangle = 0 \quad (3.2)$$

kjer je c optični center, v_i bežišča ter $\langle \cdot, \cdot \rangle$ skalarni produkt.

Pomemben je tudi kriterij na podlagi bežiščnice - linije skozi dve bežišči, pri čemer vsaj eno bežišče ni v neskončnosti. Kriterij lahko definiramo kot pogoj, da vsi linearni segmenti, ki so hkrati volili za bežišča v_1 in v_2 , ležijo na ustrezni bežiščnici. V primeru, ko sta obe bežišči v neskončnosti, zahtevamo, da sta množici linearnih segmentov, ki so volili za ta bežišča, medseboj tuji. Kriteriji za izbiro kandidatov za bežišča so podrobno opisani v [Rother, 2000].

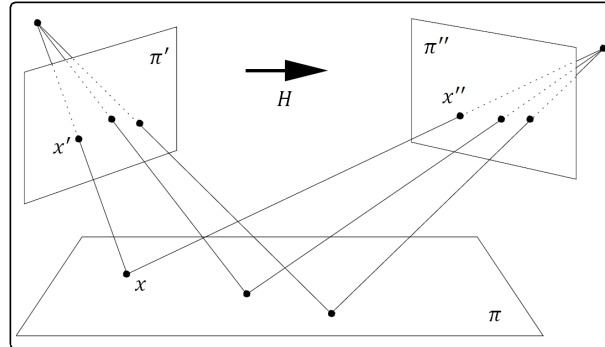
Homografija

S pomočjo detektiranih dominantnih smeri linearnih segmentov na sliki fasade in ob predpostavki, da večina teh linearnih segmentov ustreza robovom oken, ki so ponavadi poravnani s horizontalno in vertikalno osjo prostora, lahko rekonstruiramo prostorsko orientacijo ravnine, ki aproksimira površino fasade. Nadaljnja analiza strukture opazovane fasade se poenostavi, če odpravimo perspektivno projekcijo slike tako, da rekonstruiramo pravokotni pogled na ravnino fasade s poravnanimi osmi slikovne ravnine s horizontalno in vertikalno osjo prostora. Torej potrebujemo transformacijo, ki bo preslikala točke izhodiščne slikovne ravnine v točke slikovne ravnine, ki ustreza pravokotnemu pogledu na fasado.

Transformacijo, ki slika točke ene ravnine v točke druge ravnine, imenujemo 2D homografija. Slika 3.3 prikazuje postopek konstrukcije homografije med ravninama π' in π'' .

Ravnina π predstavlja aproksimacijo površine opazovane fasade. Vektor $\vec{x} = (u, v, z)^T$, zapisan v homogenih koordinatah, predstavlja neko točko ravnine π . Vektorja $\vec{x}_1 = (u_1, v_1, z_1)^T$ in $\vec{x}_2 = (u_2, v_2, z_2)^T$, prav tako zapisana v homogenih koordinatah, pa predstavljata projekciji točke \vec{x} na ravnino π_1 oziroma π_2 .

S pomočjo homografije H lahko preslikamo točko \vec{x}_1 v točko \vec{x}_2 brez poznavanja dejanskih koordinat točke \vec{x} v prostoru.



Slika 3.3: Konstrukcija homografije med ravninama π' in π'' .

Izračun homografije. Transformacijo posamezne točke iz ene ravnine v drugo lahko zapišemo po elementih

$$\begin{bmatrix} u_2 \\ v_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ v_1 \\ z_1 \end{bmatrix} \quad (3.3)$$

oziroma v vektorski obliki

$$a_x^T h = 0 \quad (3.4)$$

$$a_y^T h = 0 \quad (3.5)$$

kjer so

$$h = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33})^T \quad (3.6)$$

$$a_x = (-u_1, -v_1, -1, 0, 0, 0, u_2 u_1, u_2 v_1, u_2)^T \quad (3.7)$$

$$a_y = (0, 0, 0, -u_1, -v_1, -1, v_2 u_1, v_1 v_2, v_2)^T \quad (3.8)$$

Če poznamo N parov korespondenčnih točk (projekcije iste točke iz ravnine π na ravnino π_1 oziroma π_2), lahko sestavimo sistem enačb oblike

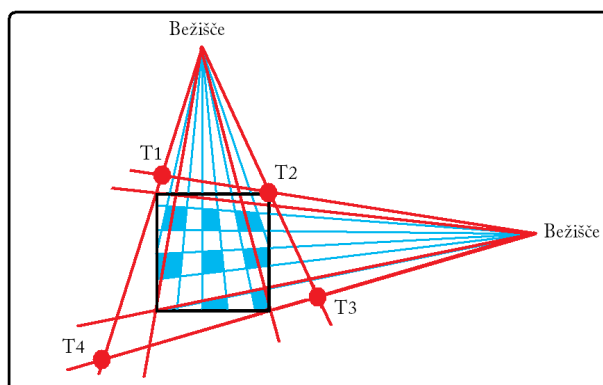
$$Ah = 0 \quad (3.9)$$

$$A = \begin{bmatrix} a_{x1}^T \\ a_{y1}^T \\ a_{x2}^T \\ a_{y2}^T \\ \vdots \\ a_{xN}^T \\ a_{yN}^T \end{bmatrix} \quad (3.10)$$

Iskana transformacija H je v obliki matrike 3×3 elementov. Element h_{33} matrike H določa faktor razmerja med točkami ter ga lahko postavimo na 1. Iz tega sledi, da ima matrika H osem neodvisnih parametrov. Za določanje vrednosti elementov matrike H potrebujemo štiri pare korespondenčnih točk. Pri izbiri korespondenčnih točk moramo dodatno zagotoviti, da nobena kombinacija treh točk ni kolinearna v opazovanih ravninah, saj v tem primeru ne moremo enoznačno določiti vrednosti matrike H . Sistem linearnih enačb lahko rešimo s standardno metodo singularne dekompozicije (SVD) [Golub in Van Loan, 1996].

Izbira korespondenčnih točk. Kot končni rezultat perspektivne rektifikacije slike želimo rekonstrukcijo pravokotnega pogleda na ravnino fasade. Za izračun homografije potrebujemo štiri pare korespondenčnih točk. V slikovni ravnini fotografije fasade bomo definirali take štiri točke, ki naj bi bile projekcije vozlišč pravokotnika, s katerim modeliramo zunanje robove opazovane fasade. Pri iskanju projekcij vozlišč opisanega pravokotnika, ki leži v ravnini fasade, si pomagamo z najdenim parom bežišč (v vertikalni in horizontalni smeri). Potencialne stranice iskanega štirikotnika potekajo skozi linearne segmente slike, ki se stekajo v izbrana bežišča. Štirikotnik konstruiramo tako, da maksimiziramo pokriti del slike. Na večini slik fasad bežišča ležijo zunaj področja slike. V tem primeru lahko dobimo iskana vozlišča s konstrukcijo štirikotnika, ki je orisan okoli celotnega področja slike, s stranicami usmerjenimi proti bežiščem. Slika 3.4 ilustrira tako konstrukcijo iskanih točk. Na ta način rekonstruemo pravokotno projekcijo celotne slike na modelirano ravnino fasade.

Najdena vozlišča štirikotnika določajo področje slikovne ravnine, ki ga pravokotno projiciramo na drugo slikovno ravnino, v kateri nadaljujemo s strukturno analizo fasade. Korespondenčne točke so torej vogali ciljne slike, ki prikazuje pravokotno projekcijo izbranega področja izhodiščne slike.



Slika 3.4: Na slikovni ravnini bomo izbrali take štiri točke $T_1 \dots T_4$, ki določajo štirikotnik s stranicami skozi dominantna bežišča, v katerem je vrisana slika fasade.

Na sliki 3.5 so prikazani primeri perspektivne izravnave slik fasad.

3.2 Generiranje hipotez

Nadaljnjo analizo strukture izvajamo na paralelni projekciji fasade, dobljeni s postopkom perspektivne izravnave, ki smo jo opisali v prejšnjem poglavju. Paralelna projekcija fasade ohranja paralelnost in ortogonalnost robov elementov fasade, kar nam v veliki meri poenostavi iskanje strukturnih elementov in s tem tudi interpretacijo same strukture fasade. Kot končni rezultat te faze rekonstrukcije fasade želimo zgradi seznam potencialnih kandidatov za strukturne gradnike fasade. Iskani kandidati oziroma hipoteze o strukturnih gradnikih določajo regije slike, ki naj bi te gradnike vsebovale. V opravljenem delu smo se ukvarjali predvsem z iskanjem oken na fasadi, zato bo v nadaljnjem besedilu opisan postopek generiranja hipotez o položaju oken. Z ustreznimi spremembami posameznih hevristik je možno postopek prilagoditi iskanju drugih strukturnih elementov fasad, kot so na primer vrata ali balkoni.

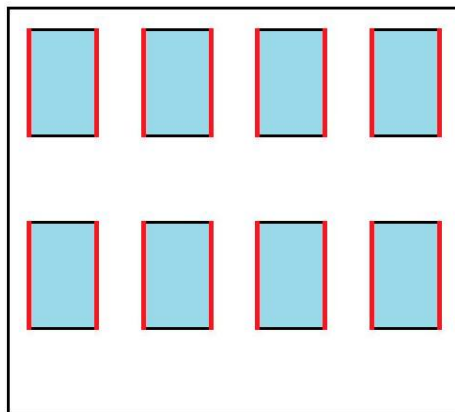
Osnovno izhodišče pri definiciji postopka iskanja hipotez je predpostavka o obliki oken ter njihovi urejeni strukturi. V praksi so daleč najpogostejša pravokotna okna, ki so zaradi praktičnih ter estetskih razlogov med seboj tudi horizontalno in vertikalno poravnana, zato smo se v našem delu omejili na ta model. Če želimo razpoznavati okna drugih oblik, na primer okrogla okna, bi



Slika 3.5: Primeri rektifikacije slik. V levem stolpcu so slike fasad s perspektivno projekcijo. V desnem stolpcu so perspektivno izravnane slike.

bilo potrebno uporabiti prilagojeno hevristiko, recimo Houghov transform za krožnice.

Perspektivno izravnana slika ohranja paralelnost in ortogonalnost elementov fasade. Na podlagi te lastnosti in privzetega modela oblike in urejenosti oken sledi, da bodo robovi oken poravnani s koordinatnimi osmi slikovne ravnine. Ob predpostavki, da večina horizontalnih in vertikalnih robov na sliki fasade pripada okvirjem oken, lahko prisotnost oziroma odsotnost takih robov uporabimo kot osnovno oporišče za generiranje hipotez. Osnovna ideja je ilustrirana na sliki 3.6, ki predstavlja shematski prikaz fasade stavbe, sestavljene iz dveh etaž.



Slika 3.6: Shematski prikaz fasade z okni v dveh etažah. Prisotnost oken v horizontalni smeri lahko detektiramo na podlagi večje koncentracije vertikalnih robov, označenih z rdečo barvo

Zaradi poravnaniosti oken v horizontalni smeri lahko pričakujemo več vertikalnih robov (označenih z rdečo barvo) na delih slike, ki ustrezajo horizontalnim pasovom oken enega nadstropja, ter malo vertikalnih robov v delih slike, ki ustrezajo horizontalnim pasovom med nadstropji oziroma zidu brez oken. Enako velja, če okna fasade opazujemo kot strukture, organizirane po stolpcih. Zaradi poravnaniosti oken v vertikalni smeri opažamo povečano prisotnost horizontalnih robov v delih fasade, ki ustrezajo stolpcem oken, ter odsotnost takih robov na zidu, kjer oken ni.

Lee in Nevatia [2004] sta predlagala postopek iskanja oken na slikah fasad,

ki temelji na ugotovitvah iz prejšnjega odstavka. Definirala sta koncept projekcijskega profila, ki predstavlja histogram porazdelitve robov na sliki. Vertikalni projekcijski profil ustreza porazdelitvi vertikalnih robov v navpični smeri. Izračunamo ga tako, da projiciramo vertikalne robove slike na vertikalno os. Vertikalni robovi poravnanih oken se projicirajo na isti del vertikalne osi. Z akumulacijo teh projekcij dobimo histogram porazdelitve vertikalnih robov. Horizontalni projekcijski profil dobimo na analogen način z akumulacijo projekcij horizontalnih robov na horizontalno os. Tokrat se projekcije robov poravnanih oken projicirajo na isti del horizontalne osi.

3.2.1 Izračun projekcijskih profilov

Izračun vrednosti histogramov lahko ponazorimo z naslednjimi izrazi:

$$Ver(y) = \sum_x \left| \frac{\partial I}{\partial x} \right| \quad (3.11)$$

$$Hor(x) = \sum_y \left| \frac{\partial I}{\partial y} \right| \quad (3.12)$$

kjer sta $\frac{\partial I}{\partial x}$ in $\frac{\partial I}{\partial y}$ parcialna odvoda slike po osi x in y.

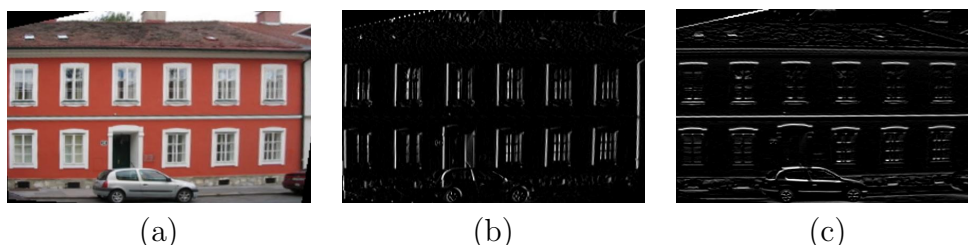
Za izračun parcialnih odvodov slike $\frac{\partial I}{\partial x}$ in $\frac{\partial I}{\partial y}$ lahko uporabimo Sobelov operator [Sobel, 1978], ki predstavlja dvorazsežno konvolucijo slike z ustreznima maskama (slika 3.7).

$$G_x = \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -2 & 0 & 2 \\ \hline -1 & 0 & 1 \\ \hline \end{array} \quad G_y = \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

Slika 3.7: Maski Sobelovega operatorja za izračun diskretnega odvoda slike po osi x in y.

Z diskretno dvorazsežno konvolucijo vhodne slike in gradientne maske G_x izračunamo parcialni odvod slike po osi x oziroma hitrost spremembe svetilnosti slikovnih elementov v horizontalni smeri. Velike vrednosti gradienta spremembe svetilnosti kažejo na prisotnost vertikalnega roba v opazovanih točkah. Na podoben način z uporabo gradientne maske G_y iščemo horizontalne robove. Slika 3.8 prikazuje primer izravnane slike fasade ter robove v horizontalni in

vertikalni smeri, poiskane s parcialnim odvajanjem slike s pomočjo Sobelovega operatorja.



Slika 3.8: Perspektivno izravnana slika fasade ter sliki dolžin gradienta v horizontalni in vertikalni smeri. a) Slika fasade. b) Slika dolžin gradienta dobljena s parcialnim odvajanjem po osi x. c) Slika dolžin gradienta dobljena s parcialnim odvajanjem po osi y.

3.2.2 Glajenje projekcijskih profilov

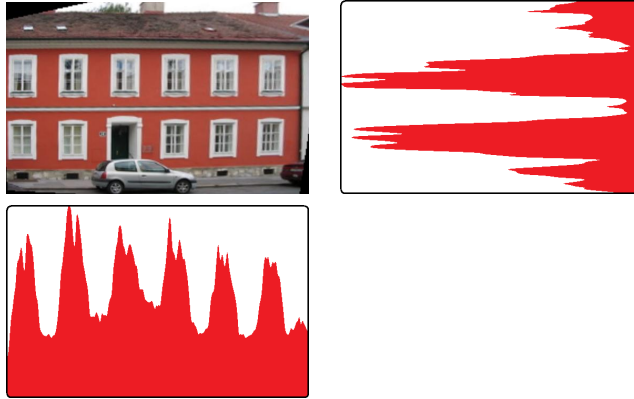
Digitalno odvajanje slike je razmeroma občutljivo na prisotnost šuma. Neustrezna osvetlitev, sence drugih stavb ali okolišnji objekti, ki delno zakrivajo fasado, so pogost vir šuma. V splošnem lahko za šum razglasimo prisotnost kakršnihkoli objektov na sliki, ki niso strukturni elementi fasade, na primer električni in telefonski kabli, odprta polkna, žaluzije, zavese, rože na okenskih policah, oznake hišnih števil, reklamni napisi, grafiti na stenah in podobno. Nenazadnje, šum dodatno prinesemo tudi s postopkom perspektivne izravnave slike zaradi numerične narave postopka.

Vsi nadaljnji koraki analize strukture fasade v določeni meri temeljijo na projekcijskih profilih, zato je potrebno zagotoviti čim večjo robustnost na prisotnost šuma. Eden izmed načinov, kako to naredimo, je glajenje histograma z normirano Gausovo funkcijo $g_\sigma(\cdot) = \frac{1}{2\pi\sigma^2} e^{-\frac{|\cdot|^2}{2\sigma^2}}$. Postopek izvede enodimenzionalno konvolucijo projekcijskega profila in binomskega filtra (slika 3.9), ki predstavlja aproksimacijo Gaussove funkcije za $\sigma = 1$.

$$\frac{1}{16} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$

Slika 3.9: Binomski filter za glajenje projekcijskih profilov

Na sliki 3.10 sta prikazana vertikalni in horizontalni projekcijski profil fasade, izračunana na podlagi izrazov (3.11) in (3.12). Histograma sta zglajena z Gaussovo funkcijo za $\sigma = 1$. Večje vrednosti v histogramih porazdelitve robov ustrezajo pasovom poravnanih oken.



Slika 3.10: Z akumulacijo projekcij robov na vertikalno in horizontalno os dobimo projekcijska profila fasade. Z rdečo barvo so prikazane vrednosti histogramov porazdelitve robov na sliki.

V našem delu smo uporabljali morfološke operacije za glajenje in odtranjevanje šuma iz projekcijskih profilov. Osnovni morfološki operaciji sta razjedanje ali erozija (erosion) ter razširjanje ali dilatacija (dilation). Če histogram porazdelitve robov projekcijskega profila predstavimo kot enodimenzionalni signal f , operator erozije zapišemo z naslednjim izrazom:

$$f \ominus D = \min\{f(z) : z \in D\} \quad (3.13)$$

pri čemer je D enodimenzionalni strukturni element. Operator erozije izvedemo tako, da strukturni element D premikamo vzdolž signala f . Vrednost izhodnega signala $f(x)$ po eroziji je enaka najmanjši vrednosti izmed vseh točk signala z v okolici opazovane točke x , ki je definirana s strukturnim elementom D . Kot tudi samo ime pove, operator erozije povzroča ožanje vrhov histograma. Stopnja ožanja je odvisna od dolžine strukturnega elementa D .

Na podoben način je definiran operator dilatacije:

$$f \oplus D = \max\{f(z) : z \in D\} \quad (3.14)$$

Pri izvajanju dilatacije je izhodna vrednosti signala $f(x)$ enaka največji vrednosti v okolici opazovane točke x , ki je definirana s strukturnim elementom D . Dilatacija povzroča širjenje vrhov histograma v odvisnosti od dolžine strukturnega elementa D .

Zaporedno izvajanje najprej operacije erozije in zatem dilatacije z istim strukturnim elementom imenujemo morfološko odpiranje (opening):

$$f \circ D = (f \ominus D) \oplus D \quad (3.15)$$

Po izvajanju operatorja odpiranja dobimo bolj gladek histogram, iz katerega so ozki vrhovi odstranjeni.

Dualna operacija odpiranju je morfološko zapiranje (closing), ki je definirano kot zaporedno izvajanje najprej dilatacije in zatem erozije z istim strukturnim elementom:

$$f \bullet D = (f \oplus D) \ominus D \quad (3.16)$$

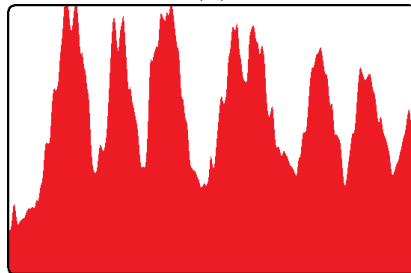
Tudi po izvajanju operatorja zapiranja dobimo bolj gladek histogram, v katerem so ozke doline zapolnjene.

Operator odpiranja odstranjuje pozitivne špice v histogramu, ki so naj pogosteje posledica šuma. Negativne špice, ki so prav tako najverjetneje posledica šuma, odpravlja operator zapiranja. Zaporedna uporaba odpiranja in zapiranja z ustrezno izbranim strukturnim elementom učinkovito odstanjuje enakomerno razporejene špice v histogramu. Toda v praksi se šum pojavlja naključno in so špice v histogramu posejane z različnim medsebojnim zamikom. Za filtriranje takega šuma lahko izmenično izvajamo zaporedje odpiranja in zapiranja z naraščujočim strukturnim elementom (alternating sequential filter, ASF). Slika 3.11 prikazuje histogram projekcijskega profila pred in po obdelavi s filtrom ASF.

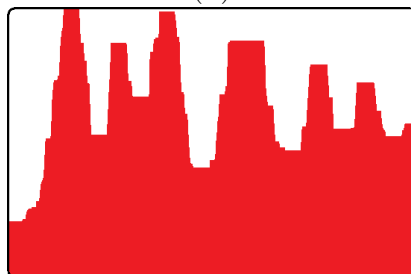
Z uporabo filtra ASF poudarimo izrazite lokalne ekstreme histograma projekcijskega profila. V nadaljevanju se bomo večkrat srečali s problemom razdelitve histograma na področja, ki ustrezajo posameznim lokalnim ekstremom. Z drugimi besedami, za vsako točko projekcijskega profila želimo določiti, ali pripada lokalnemu vrhu ali dolini histograma. Za kriterijsko funkcijo lahko uporabimo morfološki gradient $\delta_D f$ ali Laplaceov operator $\Delta_D f$:



(a)



(b)



(c)

Slika 3.11: Ob uporabi filtra ASF na histogramu projekcijskega profila poudarimo najbolj izrazite ekstreme. a) Slika fasade. b) Histogram horizontalnega projekcijskega profila pred glajenjem s filtrom ASF. b) Histogram horizontalnega projekcijskega profila po glajenju s filtrom ASF.

$$\delta_D f = (f \oplus D) - (f \ominus D) \quad (3.17)$$

$$\Delta_D f = (f \oplus D) - 2 * f + (f \ominus D) \quad (3.18)$$

V našem delu smo izvajali segmentacijo histograma z uporabo Laplaceovega operatorja, ki predstavlja diskretno aproksimacijo drugega odvoda signala f . Področja, za katera velja $\Delta_D f < 0$, ustrezajo vrhovom histograma, medtem ko vrednost $\Delta_D f > 0$ določa dele, ki ustrezajo dolinam histograma.

Postopek izračuna vertikalnega projekcijskega profila vhodne slike je podan za algoritmom (1).

Algoritem 1 Izračun vertikalnega projekcijskega profila

Vhod: Slika dolžin gradienta G_x , dimenzij $X \times Y$,
število iteracij filtra ASF.

Izhod: Histogram vertikalnega projekcijskega profila H_v .

- 1: **za** $y \leftarrow 1, Y$ **izvajaj**
 - 2: $H_v(y) := \sum_{i=1}^X G_x(i, y)$.
 - 3: **konec** {**za**}

 - 4: **za** $n \leftarrow 1$, število iteracij filtra ASF **izvajaj**
 - 5: $H_v := (H_v \circ D) \bullet D$,
 pri čemer je D strukturni element dolžine $2n - 1$.
 - 6: **konec** {**za**}
-

Postopek izračuna horizontalnega projekcijskega profila vhodne slike je podan z algoritmom (2).

3.2.3 Segmentacija fasade na podlagi projekcijskih profilov

Lee in Nevatia sta segmentacijo fasade izvedla z analizo projekcijskih profilov tako, da sta kandidate za okna definirala v presečiščih pasov, ki ustrezajo vrhovom v histogramih porazdelitve robov. Pri našem delu smo se odločili za

Algoritem 2 Izračun horizontalnega projekcijskega profila

Vhod: Slika dolžin gradienta G_y , dimenzij $X \times Y$,
število iteracij filtra ASF.

Izhod: Histogram horizontalnega projekcijskega profila H_h .

- 1: **za** $x \leftarrow 1, X$ **izvajaj**
 - 2: $H_h(x) := \sum_{j=1}^Y G_y(x, j)$.
 - 3: **konec** {**za**}

 - 4: **za** $n \leftarrow 1$, število iteracij filtra ASF **izvajaj**
 - 5: $H_h := (H_h \circ D) \bullet D$,
 pri čemer je D strukturni element dolžine $2n - 1$.
 - 6: **konec** {**za**}
-

pristop, ki je podoben tistemu iz [Haugeard in sod., 2009], pri čemer se nismo omejili na vnaprej določen vrstni red delitev fasade. V želji po izkoriščanju čim več lokalne informacije na sliki smo se odločili za segmentacijo na podlagi požrešne metode, ki rekurzivno deli fasado v globalnih minimumih lokalnih projekcijskih profilov.

Postopek segmentacije fasade začnemo tako, da izračunamo vertikalni in horizontalni projekcijski profil na celotni vhodni sliki. Z upoštevanjem obeh dobljenih histogramov najdemo globalni minimum ter s tem določimo ločilnico, ki poteka po delu slike z najmanjšo količino robov (vertikalnih ali horizontalnih, odvisno od tega, v katerem histogramu se nahaja globalni minimum). Na podlagi drugega odvoda projekcijskega profila določimo področje histograma, ki mu pripada najdena ločilnica ter oblikuje dolino okoli globalnega minimuma. Z upoštevanjem najdenih meja opazovane doline histograma razdelimo vhodno sliko na tri pasove: sredinsko področje, ki ustreza dolini okoli globalnega minimuma, ter področji z vsake strani opazovane doline. Na sliki 3.12 je prikazana delitev fasade na podlagi doline okoli ločilnice, ki ustreza globalnemu minimumu obeh projekcijskih profilov.

Za sredinsko področje predvidevamo, da ne vsebuje strukturnih elementov fasade, in ga zato izločimo iz nadaljnje analize. Večina prisotnih robov vhodne slike se nahaja pred ali za opazovano dolino histograma in predvidevamo, da se v tistih delih slike nahajajo tudi okna. Stranska področja rekurzivno razbijamo z istim postopkom, pri čemer upoštevamo samo tisti del vhodne slike, ki ga pokriva obravnavano področje. Na ta način gradimo hierarhično strukturo opazovane fasade z uporabo lokalne informacije, ki bi se z globalnimi metodami mogoče zgubila. Razbitje na manjše dele ustavimo, ko dimenzije področja padejo pod vnaprej določen minimum oz. ko je histogram robov približno enakomerno porazdeljen in so nadaljnje cepitve nesmiselne.

Z rekurzivnim razbitjem fasade gradimo drevesno strukturo, ki ponazarja strukturo dekompozicijo fasade. Vsako vozlišče tega drevesa pokriva del slike fasade. Vozlišče v korenu ustreza področju, ki pokriva celotno sliko. Vsako notranje vozlišče ima tri sinove, ki ustrezajo trem področjem, na katera smo razbili del fasade, ki pokriva opazovano vozlišče. Slika 3.13 prikazuje opisano drevesno strukturo strukturne dekompozicije fasade.

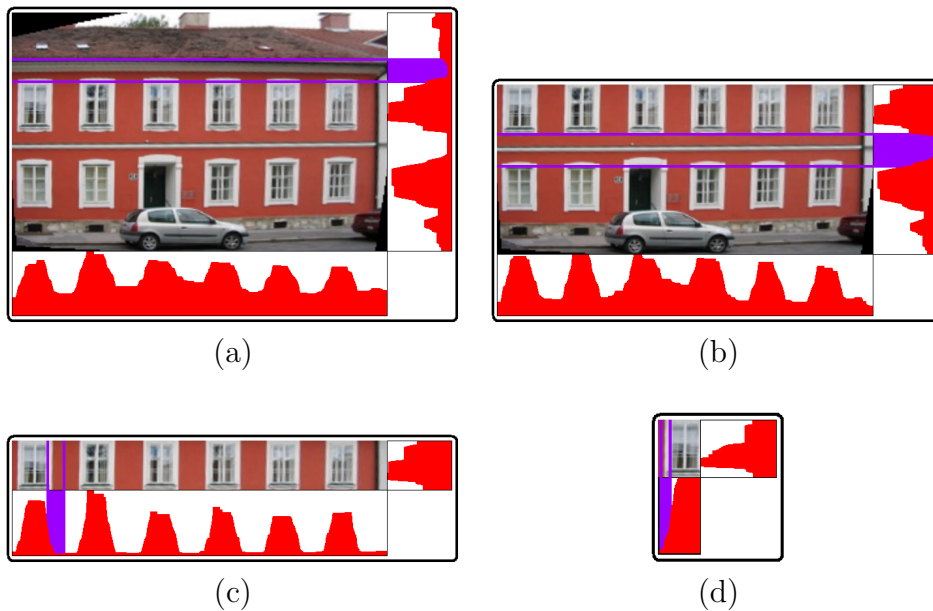
Iz načina konstrukcije sledi, da so vozlišča v notranjosti drevesa čedalje bolj obrobjena, saj z vsakim razbitjem na področja izločamo dele slike brez oziroma z malo robov. Konstrukcija drevesa zagotavlja tudi gradnjo pravokotnih področij, ki so poravnana z osmi fasade, kar sovпада s poenostavljenim opisom okna kot pravokotnega obrobjenega področja, poravnane z vertikalno in horizontalno osjo. Torej vsa vozlišča zgrajene drevesne strukture predstavljajo potencialne kandidate za okna opazovane fasade (glej sliko 3.14).

Postopek hierarhične segmentacije fasade je podan z algoritmom (3), nekaj primerov segmentiranih fasad prikazuje slika (3.15).

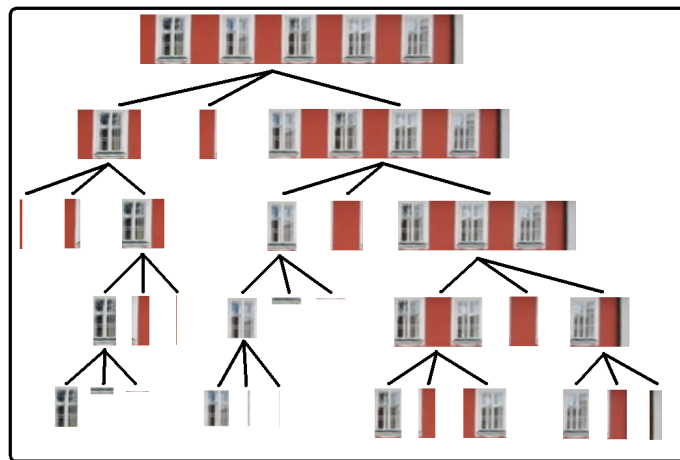
V nadaljevanju bomo področje, ki ga pokriva posamezno vozlišče drevesa, poimenovali hipoteza. Za dokončno segmentacijo fasade je potrebno oceniti posamezne hipoteze ter na podlagi najbolje ocenjenih določiti pozicijo oken.

3.2.4 Pričakovane dimenzije oken

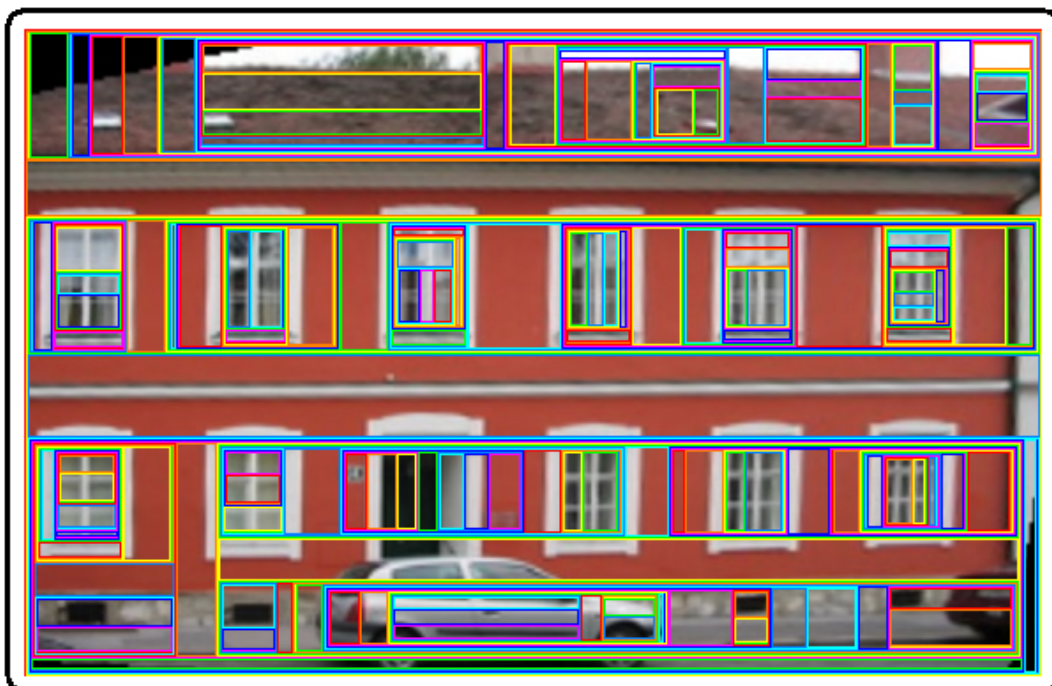
Eden izmed ustavitvenih pogojev pri opisanem postopku segmentacije fasade je določen z minimalno velikostjo področja, ki pripada hipotezi. To vrednost lahko določimo na podlagi segmentacije histogramov obeh projekcijskih pro-



Slika 3.12: Primer generiranja hipotez z rekurzivnim postopkom delitve slike. Na sliki (a) je prikazana segmentacija na podlagi globalnega minimuma, ki se nahaja v vertikalnem projekcijskem profilu. Vijolično področje na histogramu ustreza dolini okoli globalnega minimuma. Vijolični liniji označujeta cepitev slike na tri dele. Središčni pas izločimo iz nadaljnje obravnave, medtem ko rekurzivno nadaljujemo razbitje zgornjega in spodnjega pasu. Segmentacija spodnjega pasu vhodne slike je prikazana pod (b). Projekcijska profila izračunamo samo na obravnavanem delu slike. Globalni minimum se spet nahaja v vertikalnem projekcijskem profilu, zato ponovno razdelimo sliko na horizontalne pasove. Sredinski pas ustreza zidu med etažami in iz tega razloga vsebuje najmanj robov. Nadaljnja segmentacija, v tem primeru zgornjega pasu, je prikazana pod (c). Tokrat se globalni minimum nahaja v horizontalnem profilu, zato sliko razdelimo na vertikalne pasove. Sredinski pas v tem primeru ustreza zidu med okni. Zaradi numerično nestabilnega postopka segmentacije histograma na podlagi drugega odvoda opazimo, da širina sredinskega zidu ni optimalno izbrana. Na sliki (d) je prikazana segmentacija levega pasu iz prejšnje slike. Z rekurzivnim razbitjem slike se čedalje bolj približujemo pravokotnim obrobjenim področjem na sliki, kar ustreza opisu okna.



Slika 3.13: Drevesna struktura hipotez. Vsako notranje vozlišče drevesa ima tri sinove, dobljene z razbitjem slike na pasove v horizontalni ali vertikalni smeri, glede na to, v katerem projekcijskem profilu se nahaja globalni minimum. Sredinski sin korena drevesa ustreza področju slike z najmanj robovi ter predstavlja list drevesa. Področja, ki ga pokrivata druga dva sina, rekurzivno razbijamo naprej, dokler njuna velikost ne pade pod vnaprej določeno mejo (pričakovana velikost oken) oziroma dokler ne dobimo približno enakomerno porazdeljen histogram robov.



Slika 3.14: Segmentacija fasade na hierarhijo pravokotnih področij, poravnanih s koordinatnimi osmi. Področja so obrobljena z različnimi barvami. Manjša področja, vrisana v večje področje, predstavljajo sinove le-tega v drevesni strukturi. Vsako pravokotno področje predstavlja hipotezo o položaju in velikosti okna.

Algoritem 3 Hierarhična segmentacija fasade

Vhod: Perspektivno izravnana slika fasade,
Minimalne dimenzije oken d_x in d_y .

Izhod: Hierarhična segmentacija fasade v obliki drevesne strukture.

- 1: Izračunaj sliki dolžin gradienta G_x in G_y .
 - 2: Kreiraj novo vozlišče V_r , ki bo koren drevesa.
 - 3: Inicializiraj V_r tako, da področje pokriva celotno sliko.
 - 4: Dodaj V_r na sklad.

 - 5: **dokler** sklad ni prazen **izvajaj**
 - 6: $V \leftarrow$ element na vrhu sklada.
 - 7: Odstrani vrh sklada.

 - 8: Izračunaj projekcijska profila H_v in H_h za področje iz V .

 - 9: $gm \leftarrow$ globalni minimum v projekcijskih profilih.
 - 10: Določi meji m_1 in m_2 doline histograma okoli gm tako,
 da je $[m_1, m_2]$ najdaljši interval okoli gm , ter
 $\forall x \in [m_1, m_2] : \Delta_D H(x) > 0$.

 - 11: **če** gm iz H_v **potem**
 - 12: Razbij področje iz V na tri horizontalne pasove p_1, p_2, p_3 .
 - 13: **sicer**
 - 14: Razbij področje iz V na tri vertikalne pasove p_1, p_2, p_3 .
 - 15: **konec {če}**
 pri čemer uporabi m_1 in m_2 za meje med pasovi.

 - 16: Kreiraj vozlišče V_1 kot levega sina vozlišča V .
 - 17: Inicializiraj V_1 tako, da pokriva področje p_1 .
 - 18: **če** $p_1 > d$ **potem** $\triangleright d$ je d_x ali d_y , odvisno od smeri delitve
 - 19: dodaj V_1 na sklad.
 - 20: **konec {če}**

 - 21: Kreiraj vozlišče V_2 kot srednjega sina vozlišča V .
 - 22: Inicializiraj V_2 tako, da pokriva področje p_2 .

 - 23: Kreiraj vozlišče V_3 , kot desnega sina vozlišča V .
 - 24: Inicializiraj V_3 tako, da pokriva področje p_3 .
 - 25: **če** $p_3 > d$ **potem** $\triangleright d$ je d_x ali d_y , odvisno od smeri delitve
 - 26: dodaj V_3 na sklad.
 - 27: **konec {če}**
 - 28: **konec {dokler}**
-



Slika 3.15: Primeri segmentacije fasade.

filov. Ob predpostavki, da vrhovi histogramov ustrezajo poravnanim oknom fasade, lahko sklepamo o dimenzijah teh oken. Širine vrhov histograma horizontalnega projekcijskega profila ustrezajo dimenzijam oken po osi x . Podobno širine vrhov histograma vertikalnega projekcijskega profila ustrezajo dimenzijam oken po osi y . Pričakovane dimenzije oken na sliki lahko definiramo sorazmerno z najožjim in najširšim vrhom posameznih histogramov. Faktor sorazmerja vpliva na stopnjo odpornosti proti šumu. V naši implementaciji smo uporabljali faktor 0.5 za izračun minimalne velikosti okna ter faktor 1.1 za izračun maksimalne velikosti okna. Postopek za izračun pričakovanega intervala širine oken na sliki je podan z algoritmom (4). Na podoben način je realiziran postopek za izračun pričakovanega intervala višine oken s to razliko, da na vходу potrebuje vertikalni projekcijski profil.

3.3 Vrednotenje hipotez

V tem koraku želimo z različnimi hevrističnimi funkcijami oceniti kvaliteto posameznih hipotez, oziroma določiti verjetnost, da hipoteza ustreza oknu na sliki fasade. Pričakovane dimenzije oken, ki smo jih določili med fazo segmentacije vhodne sklike, so podlaga za izključitev prevelikih in premajhnih hipotez. Dodatno lahko izločimo hipoteze z neustreznim razmerjem med višino in širino, pri čemer dovoljene meje določimo empirično ali se jih naučimo na podlagi učne množice slik oken. Preostale hipoteze ocenjujemo s hevristikami na podlagi:

- pozicije v drevesni strukturi,
- distribucije robov,
- medsebojne podobnosti,
- naučene podobe oken.

Končno oceno kvalitete Q hipoteze H definiramo kot linearno kombinacijo ocen posameznih hevrističnih funkcij:

$$Q(H) = \sum_{i=1}^N \alpha_i f_i(H) \quad (3.19)$$

Algoritem 4 Pričakovana minimalna in maksimalna širina oken

Vhod: Horizontalni projekcijski profil H_h širine X ,
 Interval sprejemljivih vrednosti:
 $minPrag$ - minimalna sprejemljiva širina okna,
 $maxPrag$ - maksimalna sprejemljiva širina okna.

Izhod: Pričakovane dimenzije oken:
 $minDist$ - minimalna širina,
 $maxDist$ - maksimalna širina.

```

1:  $bPoz \leftarrow \text{FALSE}$ 
2:  $minDist \leftarrow \infty$ 
3:  $maxDist \leftarrow 0$ 

4: za  $i \leftarrow 1, X$  izvajaj
5:   če  $(i < X) \wedge (\Delta_D H_h(i) > 0)$  potem
6:     če  $bPoz = \text{FALSE}$  potem
7:        $start \leftarrow i$ 
8:        $bPoz \leftarrow \text{TRUE}$ 
9:     konec {če}
10:  sicer
11:    če  $bPoz = \text{TRUE}$  potem
12:       $dist \leftarrow i - start$ 
13:      če  $dist < minDist$  potem
14:         $minDist \leftarrow dist$ 
15:      konec {če}
16:      če  $dist > maxDist$  potem
17:         $maxDist \leftarrow dist$ 
18:      konec {če}
19:       $bPoz \leftarrow \text{FALSE}$ 
20:    konec {če}
21:  konec {če}
22: konec {za}

23: če  $minDist < minPrag$  potem
24:    $minDist \leftarrow minPrag$ 
25: konec {če}
26: če  $maxDist > maxPrag$  potem
27:    $maxDist \leftarrow maxPrag$ 
28: konec {če}

```

pri čemer je N število uporabljenih hevrističnih funkcij f_i , α_i so uteži ocen, ki jih določimo eksperimentalno, oziroma z učenjem na učni množici.

Vsako vozlišče drevesne strukture določa hipotezo o položaju in velikosti okna na fasadi, zato morajo hevristične ocene ovrednotiti vse generirane hipoteze. Za lažjo sledljivost vrednotenja posameznih hipotez bomo do konca tega poglavja uporabljali izris drevesne strukture po nivojih (glej sliko 3.16).

3.3.1 Hevristična ocena na podlagi pozicije v drevesni strukturi

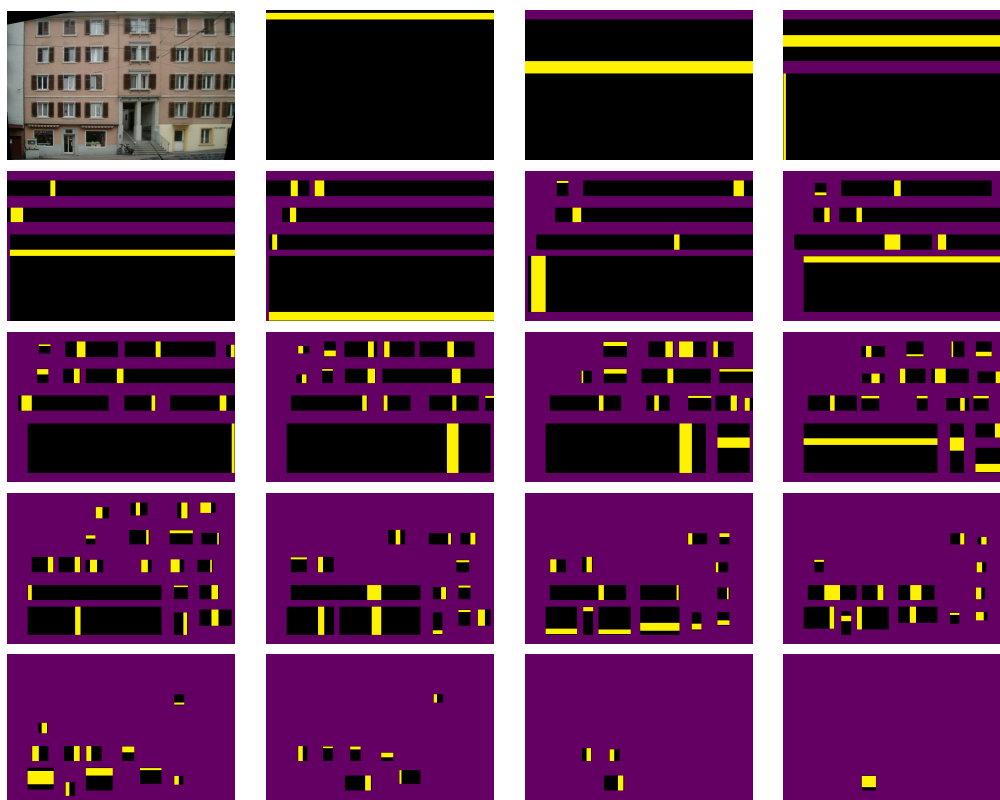
Ta preprosta hevristika izhaja iz opisanega načina konstrukcije hierarhične strukture fasade v obliki drevesa hipotez. Vsako notranje vozlišče drevesa ima tri sinove, pri čemer sredinski sin ustreza področju z minimalno količino robov. Za sredinskega sina predvidevamo, da pokriva del fasade brez oken, in mu lahko pripišemo negativno oceno ter s tem znižamo pričakovanje, da hipoteza ustreza oknu. Z rekurzivnim sprehodom po drevesni strukturi vsem hipotezam, ki ustrezajo sredinskim sinovom, dodelimo negativno oceno. Opisani postopek je podan z algoritmom (5). Na sliki 3.17 je grafični prikaz tako dobljenih ocen hipotez. Rumena območja slike so hipoteze, za katere predvidevamo, da ne ustrezajo oknom. Opisana hevristika je nevtralna do hipotez, ki ustrezajo temnim področjem.

3.3.2 Hevristična ocena na podlagi distribucije robov

Dualna hevristika zgoraj opisani hevristiki temelji na predvidevanju področij, kjer pričakujemo pojavitev okna. Ponovno izhajamo iz projekcijskih profilov, pri čemer nas tokrat zanimajo maksimumi histogramov porazdelitve robov. Postopek, podoben originalnem pristopu iz [Lee in Nevatia, 2004], gradi akumulacijsko polje, v katerem posamezne celice ustrezajo slikovnim elementom poravnane slike fasade. Lokalni maksimumi v histogramih projekcijskih profilov glasujejo za ustrezne pasove akumulacijskega polja. Na presečiščih pasov vertikalnih in horizontalnih projekcijskih profilov se akumulira največ glasov in zato ta področja ustrezajo slikovnim elementom, kjer pričakujemo pojavitev okna. Slika 3.18 ilustrira osnovno idejo opisanega postopka.



Slika 3.16: Izris drevesne strukture po nivojih. Vsako vozlišče drevesne strukture določa hipotezo o položaju in velikosti okna na fasadi, zato morajo heuristične ocene ovrednotiti vse generirane hipoteze. Za lažjo sledljivost vrednotenja posameznih hipotez, bomo uporabljali izris drevesne strukture po nivojih. V zgornjem levem kotu so prikazana področja, ki ustrezajo posameznim hipotezam, sledijo izrisi področij ločeno po nivojih. Vsaka naslednja slika (gledano z leve proti desni) prikazuje sinove vozlišč prejšnje slike.



Slika 3.17: Slika prikazuje ocene hipotez s heuristiko na podlagi pozicije v drevesni strukturi. Z rumeno barvo so prikazane hipoteze, ki jih je heuristika ocenila kot zid. Heuristika je nevtralna do hipotez, ki so predstavljene v črni barvi. Prostor med hipotezami je prikazan v vijolični barvi.

Algoritem 5 Hevristična ocena na podlagi pozicije v drevesni strukturi

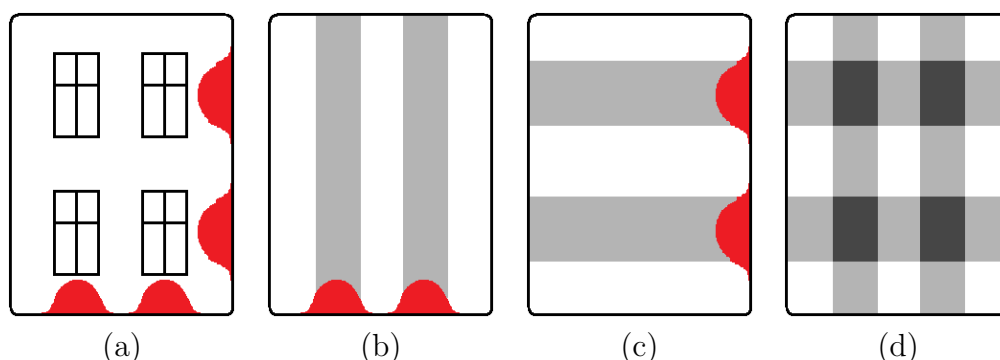
Vhod: Koren drevesne strukture V_r .

Izhod: Ocenjena vozlišča drevesa V_i .

- 1: Postavi oceno za V_r kot NEVTRALNO.
 - 2: Dodaj V_r na sklad.
 - 3: **dokler** sklad ni prazen **izvajaj**
 - 4: $V \leftarrow$ element na vrhu sklada.
 - 5: Odstrani vrh sklada.
 - 6: **če** V ni list drevesa **potem**
 - 7: Postavi oceno levemu sinu V_1 na NEVTRALNO.
 - 8: Dodaj V_1 na sklad.
 - 9: Postavi oceno srednjemu sinu V_2 na ZID.
 - 10: Postavi oceno desnemu sinu V_3 na NEVTRALNO.
 - 11: Dodaj V_3 na sklad.
 - 12: **konec** {**če**}
 - 13: **konec** {**dokler**}
-

Uspešnost metode je odvisna od robustnosti določanja lokalnih maksimumov histogramov projekcijskih profilov. Velika občutljivost pri analizi histogramov bo povzročila preveliko razdrobljenost akumulacijskega polja. Posledica tega je napačna zaznava velikega okna kot skupine manjših oken. Še večji problem nastane, ko zaradi prevelike občutljivosti zaznamo lokalni ekstrem, ki je posledica šuma in ne dejanske prisotnosti oken na fasadi, kar bo v skrajnem primeru pripeljalo do napačnega rezultata. Po drugi strani pa mala občutljivost pri analizi histograma robov lahko privede do izgube detajlov. Posledica tega je napačno zaznavanje skupine majhnih oken kot enega velikega okna. Zaradi male občutljivosti lahko spregledamo celotno vrsto ali stolpec oken, če lokalni ekstrem v histogramu robov napačno okarakteriziramo kot šum.

Z uporabo filtra ASF poudarimo izrazite lokalne ekstreme histograma projekcijskega profila. Preden povečamo vrednosti v akumulacijskem polju, je potrebno določiti, kateri deli histograma ustrezajo lokalnim maksimumom. Za kriterijsko funkcijo uporabimo Laplaceov operator $\Delta_D f$. Področja, za katere velja $\Delta_D f < 0$, ustrezajo vrhovom histograma, medtem ko vrednost $\Delta_D f > 0$ določa dele, ki ustrezajo dolinam histograma. Iz tega sledi, da povečujemo



Slika 3.18: Z akumulacijo lokalnih maksimumov projekcijskih profilov lahko določimo področja slike, v katerih pričakujemo pojavitev oken. a) Shematična slika fasade in ustrezna projekcijska profila, prikazana z rdečo barvo. b) Pasovi slike, ki ustrezajo lokalnim maksimumom horizontalnega projekcijskega profila. c) Pasovi slike, ki ustrezajo lokalnim maksimumom vertikalnega projekcijskega profila. d) Presečišča pasov, kjer se akumulira največ vrednosti, ustrezajo oknom fasade.

vrednosti tistim točkam akumulacijskega polja, ki ustrezajo področjem projekcijskega profila, za katere je vrednost Laplaceovega operatorja negativna. Na sliki 3.19 je prikazan projekcijski profil (z rdečo barvo) ter področja akumulacijskega polja (z belo barvo), katerim se povečuje vrednost na podlagi predznaka Laplaceovega operatorja.

Postopek za izračun pričakovane pozicije oken na fasade je podan z algoritmom (6). Postopek potrebuje kot vhod perspektivno izravnano sliko fasade, izhod predstavlja akumulacijsko polje, v katerem posamezna celica ustreza slikovnemu elementu perspektivno izravnane slike fasade. Večja vrednost akumulacijske celice pomeni večjo verjetnost, da ustrezni slikovni element pripada oknu fasade.

Rezultat izvajanja opisanega postopka je prikazan na sliki 3.20. Če je fasada na vhodni sliki zelo strukturirana in so okna poravnana v vertikalni in horizontalni smeri, dobimo zadovoljiv rezultat. Težave nastopijo, ko je struktura fasade bolj kompleksna in okna niso povsem poravnana. V tem primeru se projekcije oken delno prekrivajo in maksimumi projekcijskih profilov niso nujno poravnani s strukturo oken na fasadi. Do podobnih težav pride tudi zaradi šuma kot posledice zakrivanja dela fasade s strani okolišnjih objektov (drevesa,

Algoritem 6 Pričakovana pozicija oken

Vhod: Sliki dolžin gradienta G_x in G_y , dimenzij $X \times Y$.

Izhod: Akumulacijsko polje A pozicij oken.

- 1: Izračunaj vertikalni projekcijski profil H_v na podlagi G_x .
- 2: Izračunaj horizontalni projekcijski profil H_h na podlagi G_y .
- 3: Inicializiraj akumulacijsko polje A dimenzij $X \times Y$.

- 4: **za** $j \leftarrow 1, Y$ **izvajaj**
- 5: **če** $\Delta_D H_v(j) < 0$ **potem**
- 6: **za** $i \leftarrow 1, X$ **izvajaj**
- 7: $A(i, j) := A(i, j) + 1$.
- 8: **konec** {**za**}
- 9: **konec** {**če**}
- 10: **konec** {**za**}

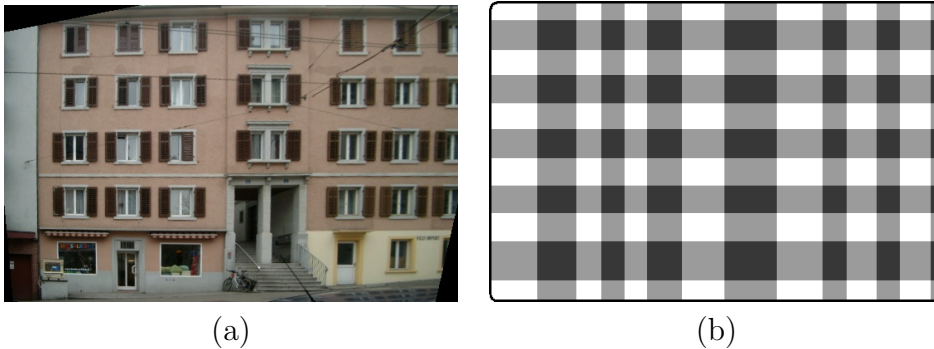
- 11: **za** $i \leftarrow 1, X$ **izvajaj**
- 12: **če** $\Delta_D H_h(i) < 0$ **potem**
- 13: **za** $j \leftarrow 1, Y$ **izvajaj**
- 14: $A(i, j) := A(i, j) + 1$.
- 15: **konec** {**za**}
- 16: **konec** {**če**}
- 17: **konec** {**za**}



Slika 3.19: Predznak Laplaceovega operatorja določa dele akumulacijskega polja, katerim se povečuje vrednost. Beli pasovi ustrezajo delom histograma, za katere velja $\Delta_D f < 0$.

prometni znaki, vozila in podobno). Splošno poravnano oken lahko kvarijo tudi odprta polkna, kot je prikazano na sliki 3.20. Zaradi pokvarjene splošne poravnano oken opazimo, da so pasovi na levi strani akumulacijskega polja različne širine, čeprav je dejanska širina oken v tem delu fasade enaka.

Za odpravljanje posledic šuma bomo uporabili rekurziven postopek posodabljanja akumulacijskega polja. Začnemo na podoben način kot v zgoraj opisanem postopku. Izračunamo vertikalni projekcijski profil celotne slike ter drugi odvod histograma z uporabo Laplaceovega operatorja. Povečamo vrednosti tistim celicam akumulacijskega polja, ki ustrezajo delom histograma, v katerih je vrednost drugega odvoda negativna. Sedaj za vsak posodobljen pas rekurzivno pokličemo funkcijo, ki bo izračunala horizontalen projekcijski profil samo tistega dela slike, ki ustreza opazovanemu pasu. Na podlagi predznaka drugega odvoda horizontalnega profila posodobimo ustrezne dele akumulacijskega polja ter rekurzivno kličemo funkcijo, ki bo tokrat obravnavala vertikalne projekcijske profile posodobljenih delov slike. Rekurzivni klici izmenoma obravnavajo vertikalne in horizontalne projekcijske profile delov slike, ki so v prejšnjem klicu posodobljeni. Z rekurzijo prenehamo, ko je obravnavani del slike manjši od vnaprej določene velikosti najmanjšega pričakovanega okna. Nato isti postopek ponovimo s horizontalnim projekcijskim profilom celotne slike.



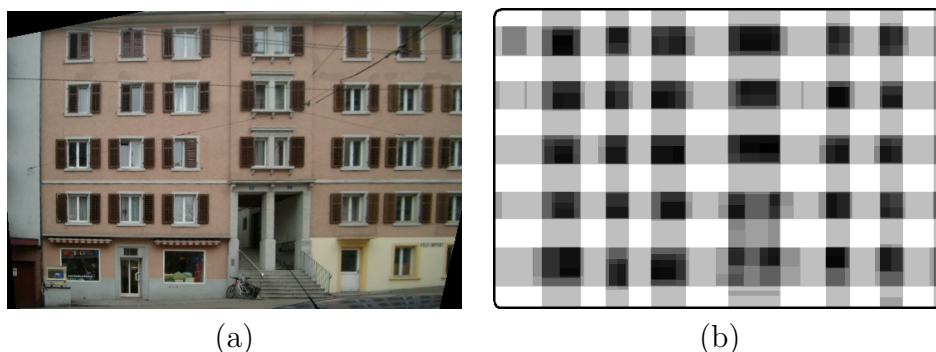
Slika 3.20: Primer akumulacije lokalnih maksimumov projekcijskih profilov pri določanju pričakovane pozicije oken na fasadi. a) Perspektivno izravnana slika fasade. b) Akumulacijsko polje pojavitve oken na sliki. Temnejši odtenki slike ustrezajo večji verjetnosti, da ustrezni slikovni elementi pripadajo oknom. Zaradi odprtih polken, ki kvarijo poravnanoost oken na fasadi, so pasovi na levi strani akumulacijskega polja različne širine, čeprav je dejanska širina oken enaka.

Hevristika posamezno hipotezo oceni kot povprečno vrednost akumulacijskih celic v tistem delu akumulacijskega polja, ki ustreza področju obravnavane hipoteze.

3.3.3 Hevristična ocena na podlagi medsebojne podobnosti

Ideja te hevristične ocene sledi iz ugotovitve, da so posamezna okna na eni fasadi vizualno podobna med seboj. Iz tega izhaja hevristika, ki ocenjuje kvaliteto hipoteze na podlagi stopnje vizualne podobnosti z drugimi hipotezami. Večkrat najdeno ujemanje med hipotezami na različnih koncih slike povečuje verjetnost, da te hipoteze ustrezajo oknom na fasadi. Kot mero podobnosti med slikovnimi elementi, ki jih pokrivata področja opazovanih hipotez, lahko uporabimo normaliziran korelacijski koeficient. Če sliki dimenzij $X \times Y$ slikovnih elementov obravnavamo kot dvorazsežnostna vzorca $f_1(x, y)$ in $f_2(x, y)$, podobnost med njima kvantitativno merimo z naslednjim izrazom:

$$C(f_1, f_2) = \frac{\sum_{x=1}^X \sum_{y=1}^Y (f_1(x, y) - \bar{f}_1)(f_2(x, y) - \bar{f}_2)}{\sqrt{\sum_{x=1}^X \sum_{y=1}^Y (f_1(x, y) - \bar{f}_1)^2 \sum_{x=1}^X \sum_{y=1}^Y (f_2(x, y) - \bar{f}_2)^2}} \quad (3.20)$$



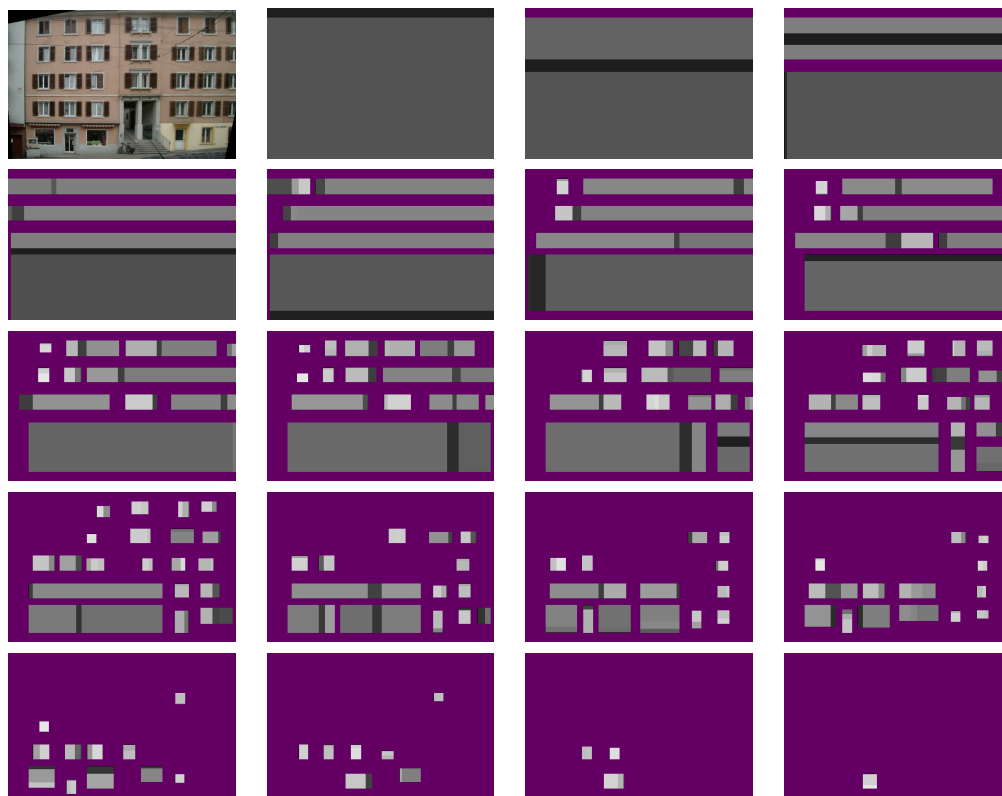
Slika 3.21: Primer akumulacije lokalnih maksimumov projekcijskih profilov pri določanju pričakovane pozicije oken na fasadi. a) Perspektivno izravnana slika fasade. b) Akumulacijsko polje pojavitve oken na sliki dobljeno z rekurzivnim posodabljanjem. Temnejši odtenki slike ustrezajo večji verjetnosti, da ustrezni slikovni elementi pripadajo oknom. Opazimo, da temnejša področja slike bolj ustrezajo dejanskim položajem oken na fasadi.

kjer sta \bar{f}_1 in \bar{f}_2 povprečni vrednosti vzorcev f_1 in f_2 :

$$\bar{f}_i = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y f_i(x, y) \quad (3.21)$$

Ta metrika računa navkrižno korelacijo med slikovnimi elementi ter normalizira rezultat s kvadratnim korenom avtokorelacij posameznih slik. Zaloga vrednosti normiranega korelacijskega koeficienta so realna števila na intervalu $[-1, 1]$. Če je normirani korelacijski koeficient enak 0, ni linearne odvisnosti med slikama. Večja kot je linearna odvisnost med opazovanima slikama, večja je absolutna vrednost normaliziranega korelacijskega koeficienta. Vrednost 1 kaže na popolno korelacijo med vzorcema, vrednost -1 pa na popolno anti-korelacijo med njima. Lepa lastnost te mere je poleg normiranosti neobčutljivost na razlike v svetlosti in kontrastu slik. Na slikah fasad so opisane spremembe relativno pogost pojav, tako da je ta lastnost normiranega korelacijskega koeficienta glavni razlog, zakaj smo se odločili za uporabo te mere podobnosti med slikami.

Ocenjevanje hipotez poteka po naslednjem postopku. Za vsako hipotezo iz drevesa (poimenujmo jo "trenutna" hipoteza) izračunamo stopnjo vizualne



Slika 3.22: Ocene hipotez s hevristiko na podlagi distribucije robov. Z odtenki sive barve je predstavljeno pričakovanje, da hipoteza ustreza oknu na fasadi. Svetlejši odtenki ponazarjajo večjo povprečno vrednost dela akumulacijskega polja, ki ustreza posamezni hipotezi. Z vijolično barvo je predstavljen prostor med hipotezami.

podobnosti s posameznimi hipotezami približno enakih dimenzij (dovoljujemo npr. 5% odstopanje zaradi večje robustnosti do šuma). Tem hipotezam povečamo oceno za izračunano stopnjo ujemanja s "trenutno" hipotezo. Postopek je opisan z algoritmom (7).

Primerjava hipotez z normiranim korelacijskim koeficientom predvideva enako velikost vzorcev. Če sta področja hipotez različnih dimenzij, jih je potrebno izenačiti. To lahko dosežemo z uporabo bilinearne interpolacije.

Algoritem 7 Hevristična ocena na podlagi medsebojne podobnosti

Vhod: Vhodna slika I ,

Drevesna struktura hipotez T ,

Toleranca razlike dimenzij hipotez.

Izhod: Ocenjene hipoteze drevesne strukture T .

```

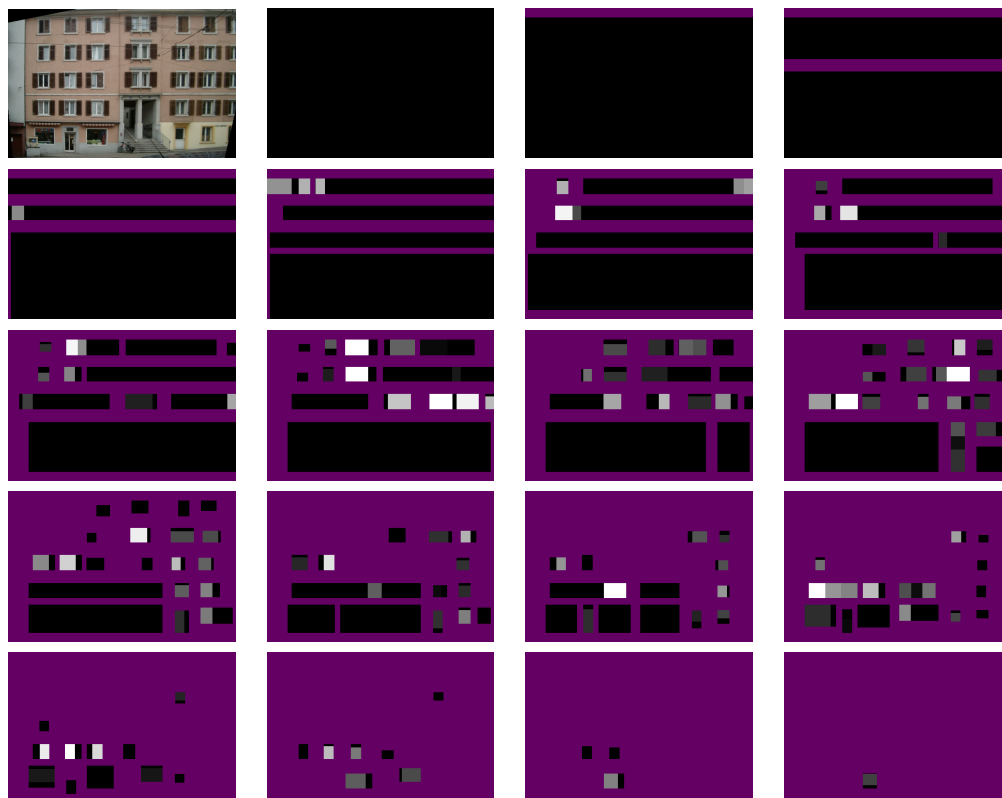
1: za vse hipoteze  $h_1 \in T$  izvajaj
2:    $I_1 \leftarrow$  podslika  $I$ , ki ustreza področju hipoteze  $h_1$ .
3:   za vse hipoteze  $h_2 \in T$  izvajaj
4:     če sta  $h_1$  in  $h_2$  podobnih dimenzij (v mejah tolerance) potem
5:        $I_2 \leftarrow$  podslika  $I$ , ki ustreza področju hipoteze  $h_2$ .
6:       Povečaj oceno hipoteze  $h_2$  za vrednost  $C(I_1, I_2)$ .
7:     konec {če}
8:   konec {za}
9: konec {za}

```

3.3.4 Hevristična ocena na podlagi naučene podobe oken

S pomočjo podanih značilnk, ki opisujejo učno množico pozitivnih in negativnih slik, je možno z metodami strojnega učenja poiskati klasifikacijsko funkcijo za razločevanje med njimi. Če učna množica opisuje podobo nekega objekta, potem je možno naučeni klasifikator uporabiti za detekcijo tega objekta na novih slikah.

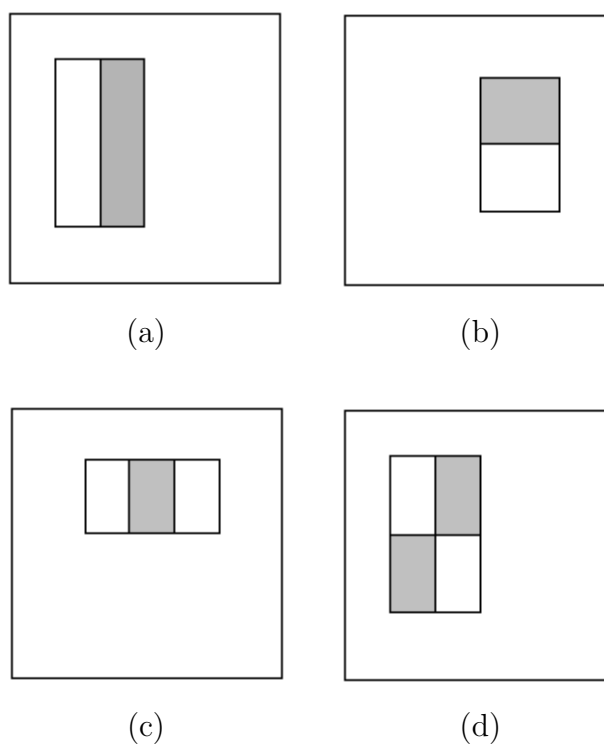
Viola in Jones [2001] sta razvila postopek za detekcijo človeških obrazov na slikah, ki je zaradi svoje splošnosti primeren za detekcijo poljubnih tipov objektov. Mi smo predstavljeno metodologijo uporabili za učenje videza oken.



Slika 3.23: Slika prikazuje ocene hipotez s hevrstiko na podlagi medsebojne podobnosti. Stopnja vizualne podobnosti hipoteze s preostalimi hipotezami podobnih dimenzij je predstavljena z odtenki sive barve. Z drugimi besedami, svetlejša področja se večkrat pojavijo v drugih delih slike fasade. To oceno lahko interpretiramo kot pričakovanje, da hipoteza ustreza oknu, saj med posameznimi okni na fasadi v splošnem velja visoka stopnja podobnosti. Prostor med hipotezami je prikazan v vijolični barvi.

Značilke Viola-Jones

Metoda učenja ne temelji neposredno na slikovnih elementih temveč na velikem številu preprostih značilk, ki spominjajo na Haarove stopničaste funkcije. Značilke predstavljajo razliko vsote intenzitet slikovnih elementov različnih pravokotnih področij slike. Viola in Jones sta uporabila tri vrste značilk, prikazanih na sliki 3.24.



Slika 3.24: Značilke Viola-Jones. Vrednost značilke je določena kot razlika med vsoto sivinskih nivojev temnih in svetlih pravokotnih področij. Primera značilk z dvema pravokotnikoma sta prikazana pod a) in b). Primer značilke s tremi pravokotniki je prikazan pod c). Pod d) je primer značilke s štirimi pravokotniki.

Značilka z dvema pravokotnikoma (two-rectangle feature) predstavlja razliko vsote intenzitet slikovnih elementov v temnem pravokotniku od vsote intenzitet v svetlem pravokotniku. Temen in svetel pravokotnik sta enakih dimenzij ter horizontalno ali vertikalno poravnana. Značilka s tremi pravokotniki (three-rectangle feature) se izračuna kot razlika vsote intenzitet slikovnih

elementov središnjega pravokotnika in vsote intenzitet dveh stranskih pravokotnikov. Značilka s štirimi pravokotniki predstavlja razliko vsote intenzitet slikovnih elementov med diagonalnimi pari pravokotnikov. Značilke se računajo v poljubni velikosti in na poljubnih položajih znotraj slike, zato je njihovo število zelo veliko. Na sliki velikosti 24×24 slikovnih elementov je možno izračunati več kot 180000 značilk, kar je veliko več od števila pikslov slike. Značilke implicitno opisujejo določene lastnosti opazovanega dela slike, kot so robovi, spremembe v teksturi in podobno.

Integralna slika

Avtorja sta definirala nov način predstavitve vhodne slike, tako imenovano integralno sliko, ki omogoča zelo učinkovit način izračuna vrednosti značilk. Slikovni element integralne slike predstavlja vsoto intenzitet vseh slikovnih elementov levo in nad opazovanim slikovnim elementom, vključno s samim sabo:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3.22)$$

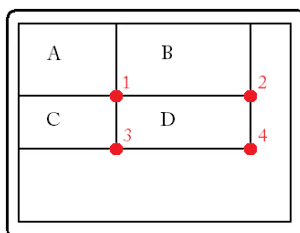
kjer sta $ii(x, y)$ integralna slika in $i(x, y)$ originalna slika. Z uporabo rekurenčnih enačb:

$$s(x, y) = s(x, y - 1) + i(x, y) \quad (3.23)$$

$$ii(x, y) = ii(x - 1, y) + s(x, y) \quad (3.24)$$

kjer je z $s(x, y)$ označena vsota prvih x slikovnih elementov vrstice y ter velja $s(x, -1) = 0$ in $ii(-1, y) = 0$, lahko izračunamo integralno sliko z enim samim prehodom čez originalno sliko.

S štirimi dostopi do podatkov v integralni sliki je možno izračunati vsoto slikovnih elementov kateregakoli pravokotnega področja originalne slike. Slika 3.25 ilustrira ta izračun. Iz tega sledi, da za izračun značilk z dvema pravokotnikoma potrebujemo šest dostopov, značilke s tremi pravokotniki zahtevajo osem dostopov, medtem ko za tiste s štirimi pravokotniki zadošča devet dostopov do podatkov v integralni sliki.



Slika 3.25: Za izračun vsote intenzitet slikovnih elementov pravokotnega področja D potrebujemo štiri dostope do integralne slike. Vrednost integralne slike v točki 1 predstavlja vsoto pikselov področja A. Vrednost v točki 2 je $A + B$, v točki 3 je $A + C$, ter v točki 4 znaša $A + B + C + D$. Vsoto slikovnih elementov področja D lahko izračunamo kot $4 + 2 - (2 + 3)$.

Grajenje klasifikatorja

Na podlagi vrednosti vsake značilke je možno definirati šibki klasifikator $h_j(x)$ oblike:

$$h_j(x) = \begin{cases} 1, & \text{če } p_j f_j(x) < p_j \theta_j \\ 0, & \text{sicer} \end{cases} \quad (3.25)$$

kjer je f_j obravnavana značilka izračunana na podsliki x , p_j naučen predznak ter θ_j naučen prag odločitvene funkcije. Šibki klasifikator f_j bo podsliko x uvrstil v pozitiven razred (vsebuje objekt), ko je vrednost njegove značilke f_j večja (oziroma manjša, odvisno od predznaka p_j) od praga θ_j .

Šibki učni algoritem za vsako značilko f_j optimalno določi pragovni klasifikator h_j , ki najbolje loči pozitivne in negativne učne primere. V praksi en sam šibki klasifikator ne more zadovoljivo opravljati klasificijske funkcije. Po drugi strani pa je bilo eksperimentalno ugotovljeno, da kombinacija zelo malega števila značilk omogoča sestavo uspešnega klasifikatorja. Ker je celotna množica značilk veliko večja od števila slikovnih elementov slike, je izbira ustrezne podmnožice ključnih značilk zamudna operacija. Viola in Jones sta za izbiro ključnih značilk ter učenje in sestavljanje šibkih klasifikatorjev v močni klasifikator uporabila varianto meta-učnega algoritma AdaBoost [Freund in Schapire, 1995].

Boosting je iterativni postopek grajenja močnega klasifikatorja na podlagi učenja množice šibkih klasifikatorjev. Šibki klasifikatorji so tisti, ki imajo

klasifikacijsko točnost vsaj malo večjo od naključne izbire razreda. Močni klasifikator kot urejeno zaporedje šibkih klasifikatorjev ima bistveno večjo klasifikacijsko točnost. Ključni koncept pri boostingu je vrednotenje težavnosti posameznih primerov znotraj učne množice. Vsak učni primer je označen z utežjo, ki ustreza njegovi težavnosti. Na začetku postopka so vsi učni primeri enako obteženi. V vsakem koraku iteracije se najprej na podlagi trenutne distribucije učne množice, ki ustreza težavnosti posameznih primerov, zgradi šibka hipoteza. Le-ta potem napove razred primerom iz učne množice. Vsem napačno klasificiranim primerom se poveča utež ter na ta način usmeri postopek učenja na težavnejše učne primere. Postopek za učenje klasifikatorja je podan z algoritmom (8)

Algoritem v vsaki iteraciji najprej zgradi šibke pragovne klasifikatorje na podlagi posameznih značilk, nato pa izbere tistega z največjo točnostjo klasificiranja učnih primerov. Ocena klasifikacijske točnosti upošteva trenutne uteži oziroma pomembnost posameznih učnih primerov. Po vsaki iteraciji se znižajo uteži vsem učnim primerom, ki jih je izbrani klasifikator pravilno klasificiral. Na ta način se povečuje pomembnost napačno klasificiranih primerov in učni proces se usmerja v obravnavanje le-teh. Vsak naslednji šibki klasifikator bo poskušal pravilno klasificirati tiste primere, ki so v prejšnjih iteracijah imeli napačno napoved. Rezultat postopka je močni klasifikator, ki pri napovedovanju uporablja uteženo glasovanje vseh izbranih šibkih klasifikatorjev $h_t(x)$. Utež napovedi α_t šibkega klasifikatorja $h_t(x)$ je dobljena na podlagi njegove točnosti na uteženi učni množici, iz katere je bil zgrajen.

Časovno najzahtevnejši del zgornjega postopka je izbira najuspešnejšega šibkega klasifikatora glede na trenutne uteži učne množice. Viola in Jones sta uporabila izčrpno iskanje, ki oceni vse možne značilke na vseh primerih učne množice.

Kaskada močnih klasifikatorjev

Viola in Jones sta učinkovitost svoje metode za detekcijo objektov na slikah dodatno izboljšala z uporabo kaskade močnih klasifikatorjev. Osnovna ideja zniževanja računske zahtevnosti postopka je v tem, da čim prej prenehamo z obdelovanjem in zavrtnemo sliko, ki ne vsebuje ciljnega objekta. Močni klasifikator, zgrajen z uporabo malega števila značilk (na primer dveh ali treh) na način, da minimizira število napačno klasificiranih pozitivnih primerov (false

Algoritem 8 AdaBoost postopek za učenje klasifikatorja

Vhod: Učna množica slik $(x_1, y_1), \dots, (x_n, y_n)$,
za pozitivne učne primere velja $y = 1$ ter za negativne $y = 0$.

Izhod: Močni klasifikator $H(x)$.

- 1: Inicializiraj uteži $w_{1,i} \leftarrow \frac{1}{2m}$ za $y_i = 0$ in $w_{1,i} \leftarrow \frac{1}{2l}$ za $y_i = 1$,
kjer sta m in n število negativnih in pozitivnih učnih primerov.
- 2: **za** $t \leftarrow 1, T$ **izvajaj**
- 3: Normaliziraj uteži $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$.
- 4: **za vse** značilke j **izvajaj**
- 5: Zgradi šibki klasifikator h_j .
- 6: Izračunaj napako klasifikatorja h_j , $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
- 7: **konec {za}**
- 8: Izberi klasifikator h_t z najmanjšo napako ϵ_t .
- 9: Posodobi uteži $w_{t+1,i} \leftarrow w_{t,i} \beta_t^{1-e_i}$, pri čemer velja:
 $e_i = 0$, če je x_i pravilno klasificiran, $e_i = 1$ sicer,
 $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- 10: **konec {za}**

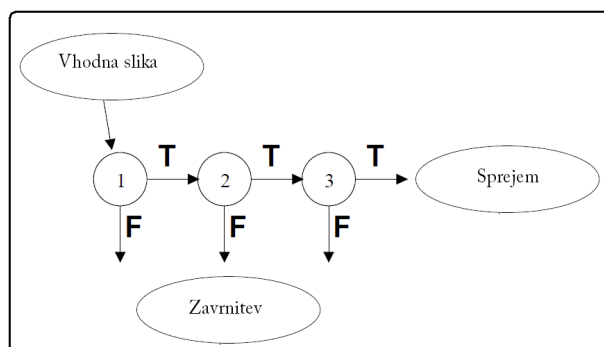
Končna oblika močnega klasifikatorja je:

$$H(x) = \begin{cases} 1, & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0, & \text{sicer} \end{cases},$$

pri čemer je $\alpha_t = \log \frac{1}{\beta_t}$.

negatives), bo računsko nezahteven. Negativen odgovor takega preprostega in učinkovitega klasifikatorja je zelo zanesljiv in je možno nadaljnjo obdelavo slike takoj ustaviti, saj prav gotovo ne vsebuje ciljnega objekta. Po drugi strani pa pozitiven odgovor takega preprostega klasifikatorja ni zanesljiv, saj bo sprejel tudi veliko slik iz negativnega razreda. Zato se v primeru pozitivnega odgovora preprostega klasifikatorja obdelava slik nadaljuje z bolj kompleksnim klasifikatorjem, ki je zgrajen z uporabo večjega števila značilk in je računsko tudi bolj zahteven. Na ta način je zgrajena kaskada zaporedno vezanih, čedalje bolj kompleksnih močnih klasifikatorjev.

Obdelava slike začnja pri najbolj preprostem klasifikatorju. Kompleksnejši klasifikator se aktivira, če njegov predhodnik v kaskadi ne zavrne slike. V primeru, ko tudi zadnji klasifikator sprejme sliko, jo dokončno uvrstimo v pozitivni razred. Negativen odgovor kateregakoli klasifikatorja pomeni takojšnjo uvrstitev slike v negativni razred ter ustavitev nadaljnje obdelave. Večina negativnih slik bo tako zavrnjena v zgodnjih stopnjah kaskade z relativno malo procesiranja. Kompleksnejši klasifikatorji v nadaljevanju kaskade zagotavljajo nizko stopnjo napačno detektiranih negativnih primerov (false positive). Na sliki 3.26 je prikazan postopek klasifikacije s kaskado močnih klasifikatorjev.



Slika 3.26: Kaskada je sestavljena iz zaporedno vezanih močnih klasifikatorjih, ki so urejeni po naraščajoči kompleksnosti. Pozitiven odgovor enega klasifikatorja sproži nadaljnjo obdelavo slike z naslednjim klasifikatorjem v kaskadi. Sliko dokončno uvrstimo v pozitiven razred, če jo sprejme klasifikator na zadnji stopnji kaskade. Negativen odgovor v katerikoli stopnji pomeni takojšnjo zavrnitev slike in njeno uvrstitev v negativni razred.

Stopnja napačno klasificiranih negativnih primerov celotne kaskade je enaka

produktu stopenj napačno klasificiranih negativnih primerov posameznih klasifikatorjev v kaskadi. Na podoben način je določena tudi stopnja pravilno klasificiranih pozitivnih primerov za celotno kaskado. To pomeni, da bo deset stopenjska kaskada pravilno klasificirala 90% pozitivnih primerov ter napačno sprejela enega od milijon negativnih primerov, če imajo posamezni klasifikatorji v kaskadi 99% senzitivnost ($0.9 \approx 0.99^{10}$) ter sprejmejo 30% negativnih primerov ($0.3^{10} \approx 6 \times 10^{-6}$).

Učenje videza okna

Za učenje kaskade klasifikatorjev za detekcijo oken smo uporabili implementacijo, ki je javno dostopna v sklopu knjižnice OpenCV. Učno množico smo izdelali s pomočjo slik iz referenčnih zbirk ZuBuD [Shao in sod., 2003] in TSG-60¹, ki vsebujeta fotografije stavb iz Zuricha (Švica) in Graza (Avstrija). Uporabili smo 1529 pozitivnih učnih primerov, dobljenih z ročnim označevanjem posameznih oken na izbranih fotografijah. Na sliki 3.27 je prikazano nekaj pozitivnih učnih primerov za učenje videza okna. Učna množica vsebuje tudi 4668 negativnih primerov, ki so prav tako ročno označeni na isti zbirki slik. Na negativnih učnih primerih so prikazani deli fasade, ki ne vsebujejo oken, reklamni napisi, drevesa, prometni znaki, parkirana vozila, mimoidoči ter drugi vzorci, ki so ponavadi prisotni na fotografijah fasad. Na sliki 3.28 je prikazano nekaj negativnih učnih primerov.

Pred začetkom učenja so pozitivni učni primeri skalirani na velikost 20×20 slikovnih elementov. Na podlagi tako izbrane učne množice je zgrajena 18 stopenjska kaskada, pri čemer je zahtevana senzitivnost posameznih klasifikatorjev v kaskadi znašala 0.995, maksimalna stopnja napačno klasificiranih negativnih primerov pa 50%.

Ocenjevanje hipotez z uporabo naučene kaskade klasifikatorjev

Detekcija objektov poteka tako, da se po sliki sistematično premika interesno področje različnih velikosti. V vsaki poziciji interesnega področja se aktivira kaskada klasifikatorjev, ki preveri ali se na ustrezni podsliki nahaja ciljni objekt. Na ta način lahko poiščemo pojavitve naučenega objekta v različnih velikostih in na različnih pozicijah znotraj opazovane slike.

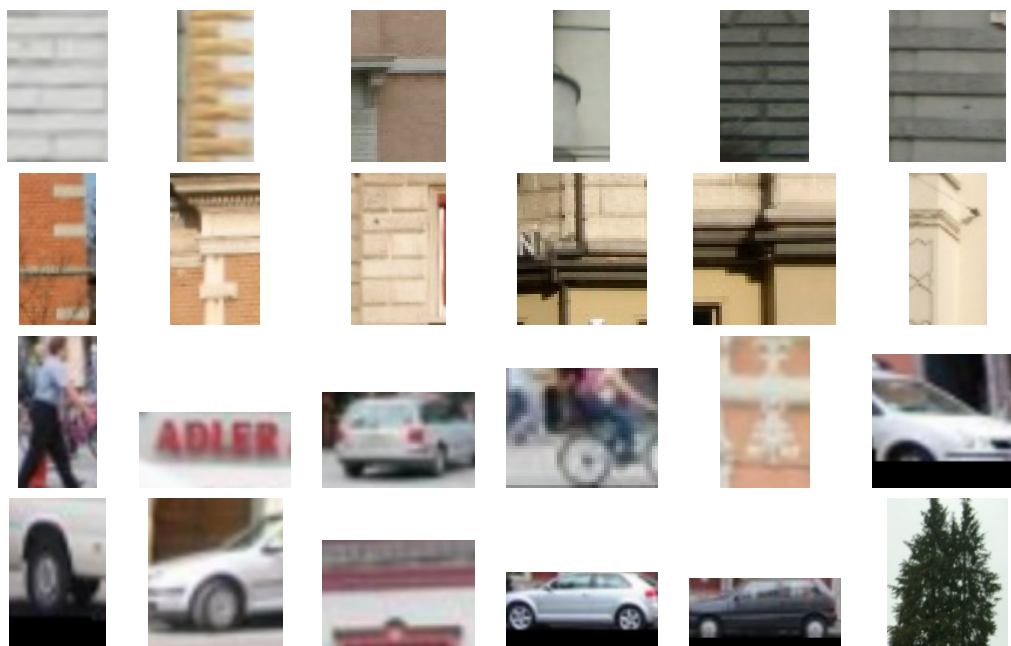
¹dostopno na <http://dib.joanneum.at/cape/TSG-60/>



Slika 3.27: Nekaj pozitivnih učnih primerov oken, ki so ročno označeni na fotografijah fasad.

Zgoraj opisan postopek smo uporabili za iskanje oken na sliki fasade. Začetna velikost interesnega področja je določena kot polovica pričakovane velikosti oken. Po koncu prehoda po celotni sliki se povečajo dimenzije interesnega področja za 10% in postopek iskanja se ponovi. Iskanje se zaključi, ko velikost interesnega področja preraste dimenzije slike.

Hevristiko za ocenjevanje hipotez na podlagi naučenega videza oken smo realizirali s pomočjo akumulacijskega polja velikosti vhodne slike. Za vsako najdeno okno v določeni podsliki posodobimo ustrezno območje akumulacijskega polja. Na sliki 3.29 je prikazano akumulacijsko polje po končanem iskanju oken s pomočjo naučene kaskade klasifikatorjev. Hevristično oceno hipoteze določimo kot povprečno vrednost celic tistega dela akumulacijskega polja, ki ustreza področju opazovane hipoteze.



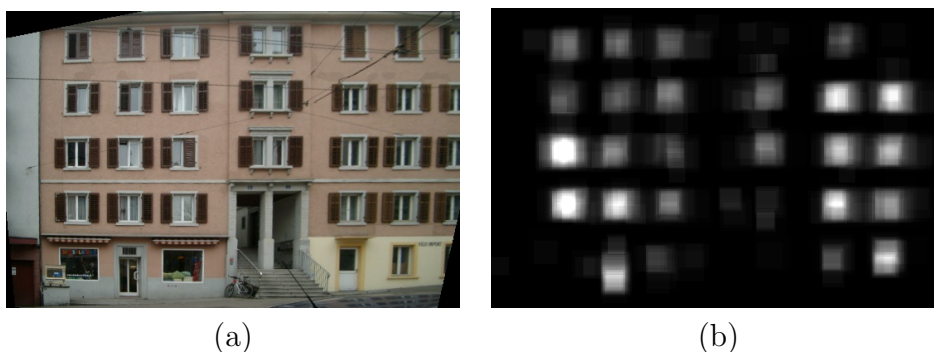
Slika 3.28: Nekaj negativnih učnih primerov, ki so ročno označeni na fotografijah fasad. Prikazujejo vzorce, ki so prisotni na slikah ter ne vsebujejo oken fasad.

3.4 Formiranje končne hipoteze

Končno hipotezo o položaju in velikosti oken na sliki fasade dobimo z izbiro najboljše ocenjenih kandidatov. Ocene posameznih kandidatov dobimo kot kombinacijo ocen uporabljenih hevristik. Možnih je več načinov kombiniranja, mi smo se odločili za linearno kombinacijo. V skladu z izrazom (3.19), končno oceno posameznega kandidata za okno na fasadi definiramo kot:

$$Q(H) = \alpha_1 f_1(H) + \alpha_2 f_2(H) + \alpha_3 f_3(H) + \alpha_4 f_4(H) \quad (3.26)$$

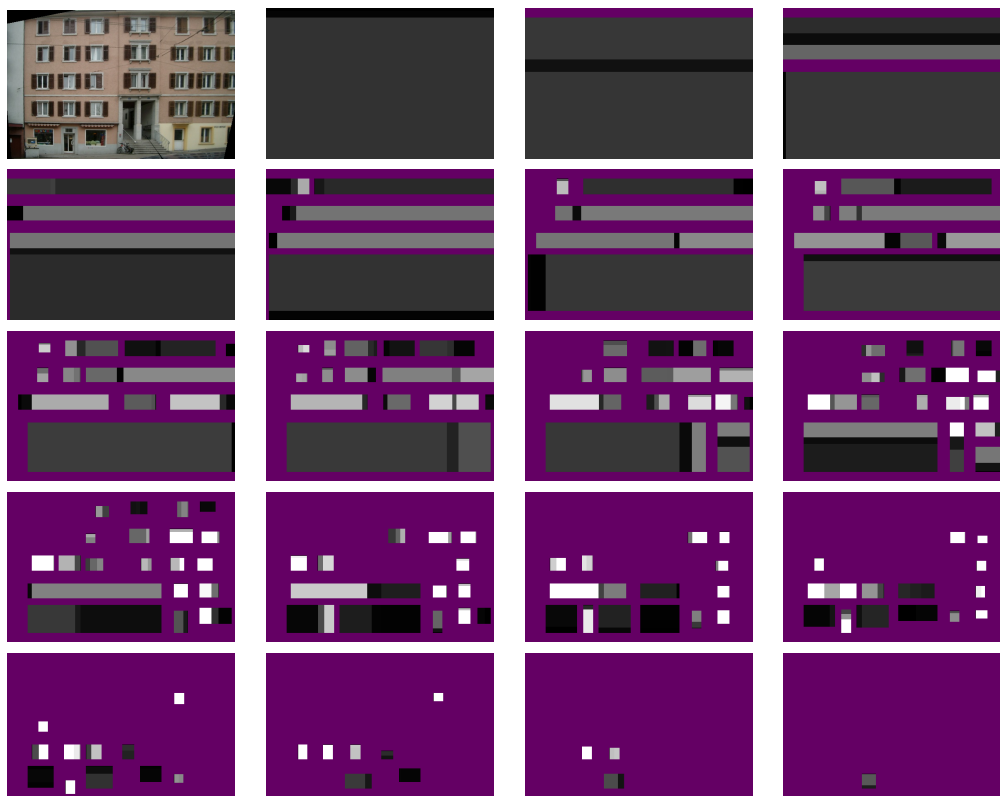
pri čemer H predstavlja kandidata za okno, f_1 je ocena hevristike na podlagi pozicije v drevesni strukturi, z f_2 je označena oceno hevristike na podlagi distribucije robov, f_3 predstavlja oceno hevristike na podlagi medsebojne podobnosti ter f_4 označuje oceno hevristike na podlagi naučene podobe oken. Z utežmi α_i določimo, koliko posamezna hevristika prispeva h končni



Slika 3.29: Z akumulacijo pozicij in velikosti detektiranih oken s pomočjo naučene kaskade klasifikatorjev lahko določimo področja slike, v katerih pričakujemo pojavitev oken. a) Slika fasade b) Akumulacijsko polje po končanem postopku iskanja oken. Svetlejša področja akumulacijskega polja ustrezajo delom slike kjer je kaskada večkrat prepoznala okna.

oceni kandidata. Vrednosti posameznih uteži je možno določiti s strojnimi učenjem. Potrebovali bi učno množico sestavljeno iz pozitivnih in negativnih učnih primerov. Vsak učni primer je predstavljen z štirimi atributi, ki vsebujejo vrednosti ocen posameznih hevristik. Vrednost odvisne spremenljivke y je pri pozitivnih učnih primerih postavljena na 1, za negativne primere pa velja $y = 0$. Z metodo linearne regresije bi lahko določili vrednosti posameznih uteži, ki minimizirajo srednjo kvadratično napako napovedane in dejanske vrednosti odvisne spremenljivke y . Pri tem bi bilo potrebno preprečiti preveliko prileganje učnim primerom, ki bi zagotovilo veliko natančnost na učni množici in praviloma slabše rezultate na novih, še ne videnih primerih.

V našem delu smo vrednosti uteži α_i določili na empiričen način z uporabo validacijske množice slik. Hevristika na podlagi pozicije v drevesni strukturi dobro razpozna dele zidu brez oken v višjih nivojih drevesa (bližje korenu), medtem ko njena natančnost pada z oddaljevanjem od korena drevesa. Druga pomanjkljivost te hevristike je njena binarnost, saj kandidate ocenjuje le z dvema vrednostima. Zaradi teh razlogov smo omenjeni hevristiki določili nizko pomembnost ter postavili $\alpha_1 = 0.1$. Tudi hevristika na podlagi medsebojne podobnosti kandidatov se je izkazala za nezanesljivo. Hevristika dobro ocenjuje kandidate na slikah fasad, ki vsebujejo veliko enakih oken brez odsevov, žaluzij, rolet in podobnih virov šuma. Če je fasada slikana iz zornega kota, ki prinaša veliko odseva na steklu ter s tem posredno znižuje podobnost med posamezn-



Slika 3.30: Slika prikazuje ocene hipotez s heuristiko na podlagi naučenega videza oken. Z odtenki sive barve je predstavljeno število zadetkov naučenega kaskadnega klasifikatorja na delu slike, ki ustreza posamezni hipotezi. Svetlejši odtenki pomenijo večje število zadetkov. Prostor med hipotezami je prikazan v vijolični barvi.

imi okni, so kandidati relativno slabše ocenjeni ter ne izstopajo od delov slike, ki ne ustrezajo oknom. Ocene kandidatov se znižujejo tudi v primeru, ko je medsebojna podobnost med posameznimi okni pokvarjena zaradi različno postavljenih žaluzij, rolet na različnih višinah, rož na okenskih policah, zaves in podobno. Hevristika se slabše obnese na slikah fasad, ki ne vsebujejo veliko oken (na primer družinske hiše), kar ima za posledico dejstvo, da kandidati niso podobni med seboj. Po drugi strani bo hevristika napačno ocenila dele zidu z ornamenti in drugimi dekorativnimi elementi. Čeprav ne gre za okna, se tak vzorec zidu velikokrat ponovi na sliki fasade ter s tem zviša oceno napačnim kandidatom. Iz opisanih razlogov smo tudi tej hevristiki določili nižjo težo in smo zato postavili $\alpha_2 = 0.1$. Hevristika na podlagi distribucije robov deluje zelo zanesljivo v primeru fasad, ki se ponašajo z veliko mero poravnosti oken v obeh dimenzijah. Slabše se obnese pri fasadah z nepravilno strukturo, kar smo poskusili odpraviti z upoštevanjem distribucije lokalnih robov. Ker je velika večina fasad v urbanem okolju zelo poravnana, je v splošnem hevristična ocena zelo zanesljiva in smo ji zato določili utež $\alpha_3 = 0.4$. Za zelo zanesljivo se je izkazala tudi hevristika na podlagi naučene podobe oken. Ob ustrezno izbrani učni množici je možno zgraditi kaskadni klasifikator, ki zna razpoznavati okna določenih arhitekturnih stilov (v skladu s pričakovano domeno uporabe) in na ta način še dodatno povečati zanesljivost ocene. Hevristiki na podlagi naučene podobe oken smo pripisali utež $\alpha_4 = 0.4$. Na sliki 3.31 so prikazane ocene kandidatov na podlagi opisane utežene linearne kombinacije posameznih hevristik.

Zadnji korak pri formiranju hipoteze o položaju in velikosti oken na fasadi je izbira najbolje ocenjenih kandidatov. Uporabili smo preprosto požrešno metodo. V vsaki iteraciji najprej izberemo najbolje ocenjenega kandidata, ter ga začasno razglasimo za okno fasade. Zaradi favoriziranja urejene strukture oken povečamo ocene vsem kandidatom, ki so podobni pravkar izbranemu kandidatu. Celotni postopek ponavljamo, dokler niso ocene vseh kandidatov manjše od vnaprej določenega praga.

Kandidata sta podobna med seboj, ko sta:

- približno enakih velikosti,
- približno vertikalno ali horizontalno poravnana,
- imata podobni oceni vsaj pri eni izmed hevristik na podlagi distribucije robov ter naučenega videza oken.

Ko se iterativni postopek zaključi, dokončno razglasimo za okna tiste izbrane kandidate, ki nimajo izbranega nobenega prednika v drevesni strukturi. Z drugimi besedami, zavržemo tiste kandidate, ki so prekriti z nekim drugim izbranim kandidatom. Postopek formiranja končne hipoteze o oknih je podan z algoritmom (9). Na sliki (3.32) je prikazan primer končne hipoteze o položaju in velikosti oken na sliki fasade.

Prednosti te metode so intuitivnost, sledljivost in hitrost, saj je postopek računsko nezahteven. Pomanjkljivosti so nižja izkoriščenost splošnega znanja o strukturi fasade ter ignoriranje določenih strukturnih omejitev v smislu med-sebojnega položaja in velikosti oken.

Algoritem 9 Formiranje končne hipoteze

Vhod: Drevesna struktura hipotez T ,

$minOcena$ - spodnja meja sprejemljivih ocen hipotez,

f - faktor povečevanja ocen,

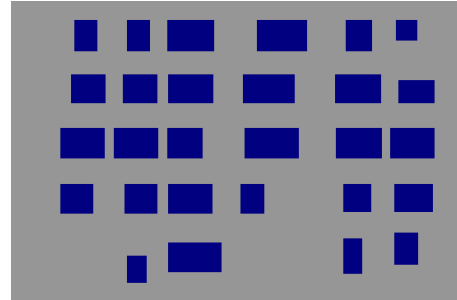
Tolerance podobnosti hipotez

Izhod: Izbrane hipoteze drevesne strukture T .

- 1: **ponavlja**j
 - 2: $h_1 \leftarrow$ najbolje ocenjena še neizbrana hipoteza iz T .
 - 3: **če** ocena $h_1 \geq minOcena$ **potem**
 - 4: Označi hipotezo h_1 kot izbrano.
 - 5: **za vse** še neizbrane hipoteze $h_2 \in T$ **izvajaj**
 - 6: **če** sta h_1 in h_2 podobni (znotraj tolerance) **potem**
 - 7: Povečaj oceno hipoteze h_2 za faktor f .
 - 8: **konec {če}**
 - 9: **konec {za}**
 - 10: **konec {če}**
 - 11: **dokler** ni več sprejemljivih še neizbranih hipotez v T
 - 12: **za vse** izbrane hipoteze $h \in T$ **izvajaj**
 - 13: **če** h ima izbranega prednika v T **potem**
 - 14: Zavrži hipotezo h .
 - 15: **konec {če}**
 - 16: **konec {za}**
-



(a)



(b)

Slika 3.32: Končna hipoteza o položaju in velikosti oken na vhodni sliki fasade. a) Vhodna slika fasade. b) Slika končne hipoteza, na kateri so okna označena z modro barvo.

Poglavje 4

Eksperimentalni del

Postopke, opisane v prejšnjem poglavju, smo implementirali v programskem jeziku C++ z uporabo odprtokodne knjižnice OpenCV [Bradski in Kaehler, 2008]. Uspešnost postopka razpoznavanja oken smo testirali na 60 fotografijah iz zbirke eTRIMS [Korč in Förstner, 2009], ter 96 fotografijah fasad, posnetih v Ljubljani. Vse fotografije vsebujejo ročno označena področja, ki ustrezajo oknom na posnetih fasadah.

Pri predobdelavi slik smo uporabljali tri različne nastavitve parametrov za Houghovo transformacijo (prag akumulacijskega polja, minimalna zahtevana dolžina in maskimalna prekinitev linearnega segmenta), ki je podlaga za izračun kandidatov za bežišča in perspektivno izravnavo. Omenjene parametre smo določili ročno.

Za vse slike smo pri nadaljnji obdelavi uporabljali enake nastavitve, ročno določene na podlagi analize dobljenih hipotez na validacijski množici. Le-to sestavlja 40 slik iz podatkovnih zbirk TSG-60 in ZuBuD. Uporabili smo faktorje 0.5 in 1.1 pri določanju pričakovanih minimalnih in maksimalnih velikosti oken. Izbrali smo 10% povečevanje interesnega področja pri uporabi kaskadnega klasifikatorja za iskanje oken na sliki. Uporabljali smo uteži 0.1, 0.1, 0.4 in 0.4 pri vrednotenju prispevka predstavljenih hevrističnih funkcijah za formiranje končne ocene kandidata. Med postopkom generiranja končne hipoteze smo izločili vse kandidate, ki imajo razmerje dimenzij nad 2.5. Izločili smo tudi vse kandidate, ki so ocenjeni z manj kot 0.01 s strani hevristike na podlagi distribucije robov in hevristike na podlagi naučenega videza. Med grajenjem končne hipoteze smo za faktor 0.1 povečevali ocene vsem še vedno neizbranim kandidatom, ki so podobni nazadnje izbranemu kandidatu. Fak-

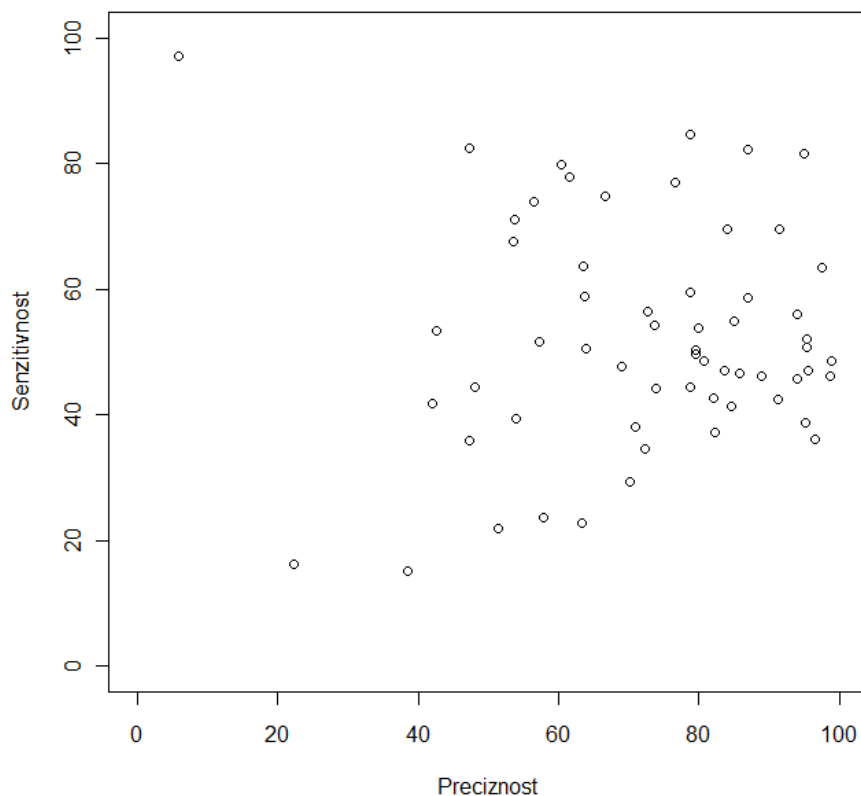
tor 0.1% smo uporabljali tudi pri definiranju pragov tolerance za podobnost dimenzij kandidatov in medsebojne poravnosti. Požrešni postopek grajenja končne hipoteze smo ustavili, ko je ocena vseh preostalih kandidatov padla pod 40% najvišje ocene v drevesu.

V eksperimentih smo merili senzitivnost in preciznost generiranih hipotez o položaju in velikosti oken na vhodnih slikah. Senzitivnost smo definirali kot delež slikovnih elementov, ki dejansko pripadajo oknom in jih generirana hipoteza pokriva. Preciznost hipoteze je delež slikovnih elementov, pokritih s hipotezo, ki dejansko pripadajo oknom na fasadi.

Na slikah 4.1 in 4.2 sta prikazani senzitivnost in preciznost končnih hipotez na uporabljenih testnih množicah slik. Povprečna senzitivnost hipotez na testni množici eTRIMS je 52.34%, povprečna preciznost pa je 72.43%. Na množici Ljubljana je povprečna senzitivnost hipotez 60.92%, preciznost pa znaša 77.84%. Iz rezultatov je razvidno, da je naša metoda konzervativna. Generirane hipoteze pokrijejo manj slikovnih elementov, ki dejansko pripadajo oknom na slikah (manjša senzitivnost), so pa zato bolj precizne. Rezultati kažejo, da je slikovna zbirka Ljubljana nekoliko lažja za detekcijo oken. Ta lastnost je tudi pričakovana, ker omenjena podatkovna zbirka večinoma vsebuje slike stanovanjskih blokov, ki imajo večje število poravnanih oken.

V drugem eksperimentu smo preštevali okna, ki so pokrita s končno hipotezo. Tokrat končne hipoteze nismo obravnavali kot celoto, temveč kot množico ločenih hipotez o posameznih oknih. Okno smo označili kot pokrito, če ena hipoteza pokriva vsaj 50% in nobena druga ne več kot 10% slikovnih elementov, ki pripadajo temu oknu. Definicija delno pokritega okna je analogna prejšnji, s to razliko, da zahtevamo 30% pokritost. Če vsaj dve hipotezi prekrivata, vsaka zase, več kot 10% okna, potem ga označimo kot večkrat pokrito. V vseh ostalih primerih okno označimo kot nepokrito. Na slikah 4.3 in 4.4 so prikazani rezultati na uporabljenih testnih slikovnih zbirkah. Iz rezultatov je razvidno, da je malo oken delno ali večkrat pokritih. To še enkrat potrjuje ugotovitev, da so generirane hipoteze konzervativne. Njihova glavna pomanjkljivost je necelovitost, kar kaže na potrebo po dopolnitvi postopka generiranja končne hipoteze, ki očitno ne izrabi vseh razpoložljivih informacij.

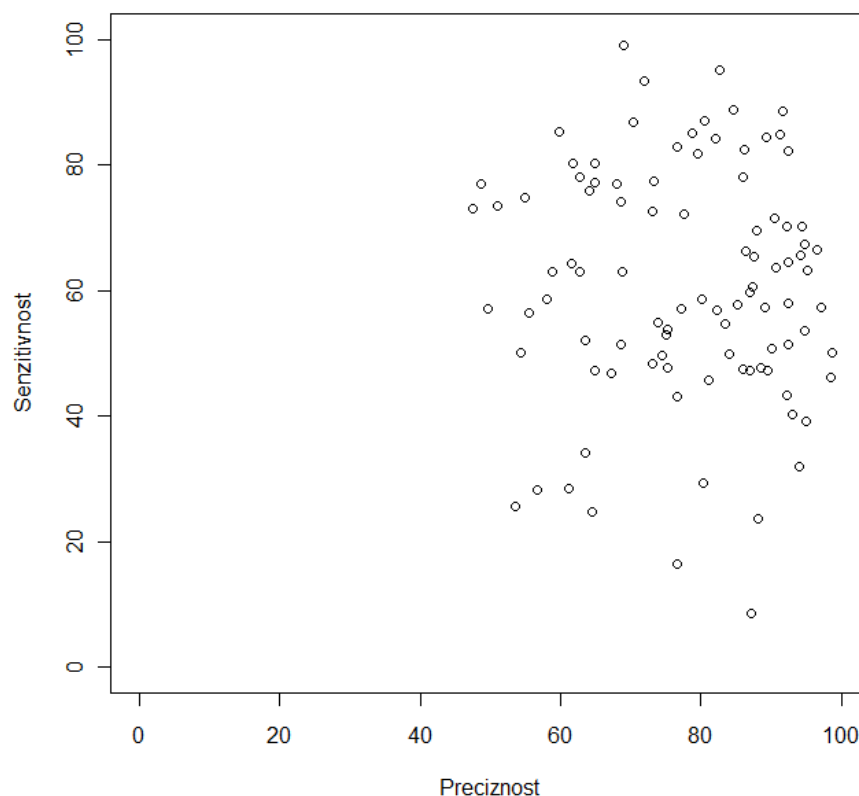
Na podoben način smo ovrednotili preciznost posameznih hipotez, ki nastopajo v končni hipotezi. Za precizne hipoteze smo označili tiste, katerih 50% površine pokriva eno okno ter nobeno drugo okno ne pokriva več kot 10%. Na



Slika 4.1: Eksperimentalni rezultati na podatkovni zbirki eTRIMS.

sliki 4.5 je prikazan rezultat dobljen na dveh testnih množicah. Iz rezultatov se vidi, da je večina generiranih hipotez preciznih.

Tudi izvajalni časi merjeni na 2.4GHz procesorju kažejo na večjo zapletenost slik iz zbirke eTRIMS. Povprečno trajanje posameznih faz pri obdelavi vhodne slike je prikazano v tabeli 4.1. Slike iz zbirke eTRIMS so nekoliko večje (768×512 proti 640×480) ter posledično vsebujejo več linearnih segmentov kar poveča število kandidatov za bežišča in na ta način podaljša tudi postopek perspektivne izravnave. Večje število robov na sliki podaljša tudi segmentacijo, ki bo generirala tudi večjo drevesno strukturo oziroma večje število kandidatov za okna. Več robov na sliki lahko podaljša tudi postopek izračuna



Slika 4.2: Eksperimentalni rezultati na podatkovni zbirki Ljubljana.

pričakovane pozicije oken na fasadi, ki je podlaga za heuristično oceno na podlagi distribucije robov. Iz meritev vidimo, da sta računsko najbolj zahtevni heuristiki za ocenjevanje kandidatov na podlagi medsebojne podobnosti in na podlagi naučenega videza. Prva je kvadratično odvisna od števila kandidatov, ki ju primerjamo. Tudi izračun normaliziranega korelacijskega koeficienta, ki je podlaga za medsebojno primerjanje kandidatov, ima kvadratično časovno zahtevnost oziroma ustreza površini področja, ki ga pokrivata kandidata. Iz opravljenih meritev sledi, da je računsko najbolj zahtevna heuristika na podlagi naučenega videza. To je razumljivo, saj je potrebno klasificirati vse možne podslike, ki jih dobimo s sistematičnim premikanjem interesnega področja različnih velikosti po vhodni sliki, kar prinese kubično časovno kompleksnost.

	eTRIMS	Ljubljana
Perspektivna izravnava	219.87	192.07
Segmentacija slike in generiranje kandidatov za okna	1167.65	844.19
Ocenjevanje kandidatov s hevrstiko na podlagi distribucije robov	337.97	235.81
Ocenjevanje kandidatov s hevrstiko na podlagi medsebojne podobnosti	661.97	752.04
Ocenjevanje kandidatov s hevrstiko na podlagi naučenega videza	1039.23	892.22
Izdelava končne hipoteze	1.83	4.11
Skupni čas	3428.52	2920.44

Tabela 4.1: Povprečni izvajalni časi obdelave posamezne slike izraženi v milisekundah. Ocenjevanje kandidatov s hevrstiko na podlagi pozicije v drevesni strukturi je zanemarljivo kratko in zato ni prikazano.

4.1 Primeri dobrega delovanja

Predlagana metoda se je izkazala za uspešno pri analiziranju fasad z veliko stopnjo horizontalne in vertikalne poravnosti. To je razumljivo, saj postopek generiranja kandidatov in pomembna hevrstična funkcija pri njihovem ocenjevanju temelji na histogramih projekcijskih profilov, ki so bolj izraziti ob močno poravnanih robovih na sliki. V tem smislu je zaželjena čim bolj enolična fasada brez lis, opek, ornamentov in drugih okraskov, ki povzročajo dodatne robove in s tem gladijo projekcijska profila. Dejanska poravnost oken zagotavlja tudi večjo uspešnost postopka formiranja končne hipoteze, ki favorizira kandidate, razvrščene po vrsticah in stolpcih. Na sliki 4.6 so prikazani nekateri primeri dobrih končnih hipotez.

Končna ocena hipoteze je sestavljena iz različnih hevrstik tako, da je možno dobiti pravilno končno hipotezo kljub pisanosti fasade. Na sliki 4.7 je prikazan rezultat obdelave pisane fasade. Rezultat je zadovoljiv zaradi visoke stopnje poravnosti, podobnosti med okni ter ustreznosti naučenemu videzu.

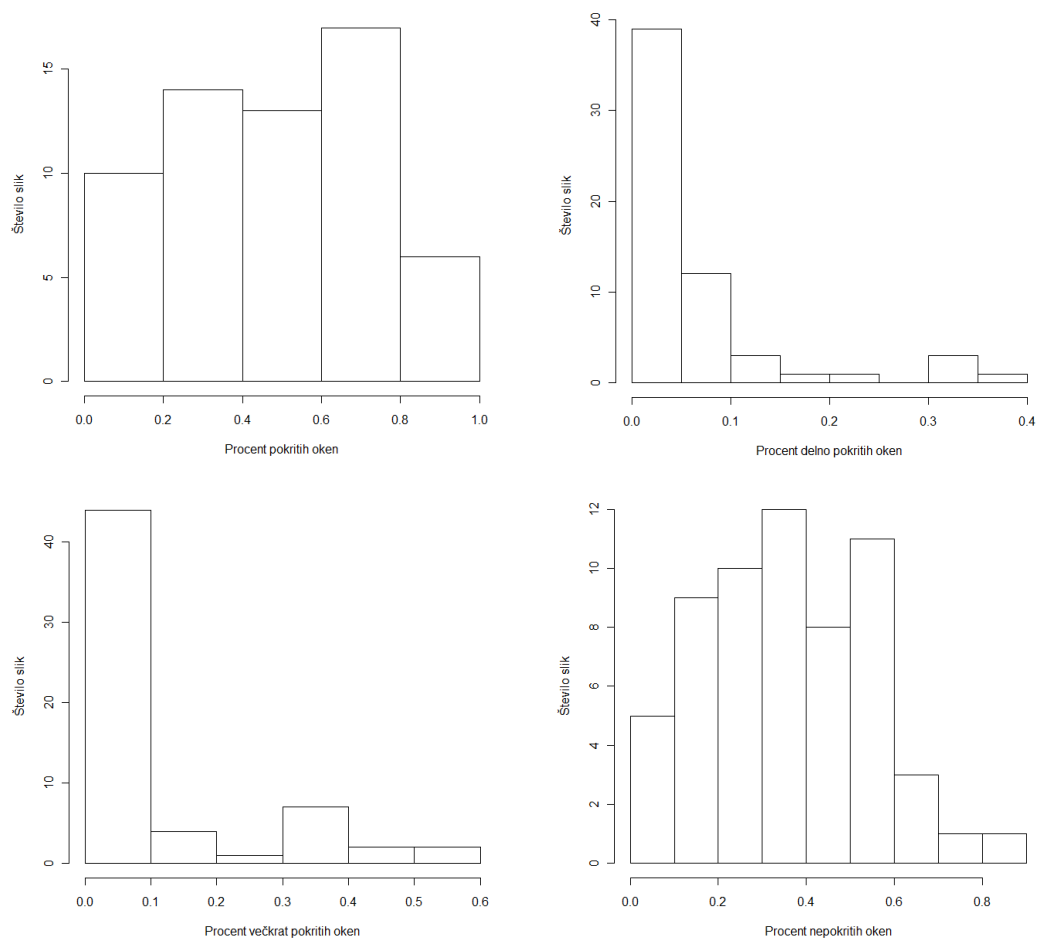
4.2 Problematični primeri

Pomanjkljivost predlagane metode je v omejevanju končne hipoteze samo na generirane kandidate. Osnovni kriterij generiranja kandidatov je segmentacija histograma projekcijskih profilov z uporabo Laplaceovega operatorja, ki je numerično precej nestabilna operacija. Tudi do samih projekcijskih profilov pridemo z akumulacijo robov, dobljenih s Sobelovim operatorjem, ki je tudi občutljiv na šum. V taki situaciji je možno, da bo določen del vhodne slike nepravilno segmentiran, kar ima za posledico generiranje neustreznih kandidatov. Ti kandidati bodo v nadaljnji obdelavi tudi slabo ocenjeni in zato ne bodo sodelovali pri formiranju končne hipoteze. Slika 4.8 prikazuje nekaj takih primerov. Deli slike, pri katerih je prišlo do napačne segmentacije, bodo ostali nepokriti s končno hipotezo, čeprav potencialno vsebujejo okna.

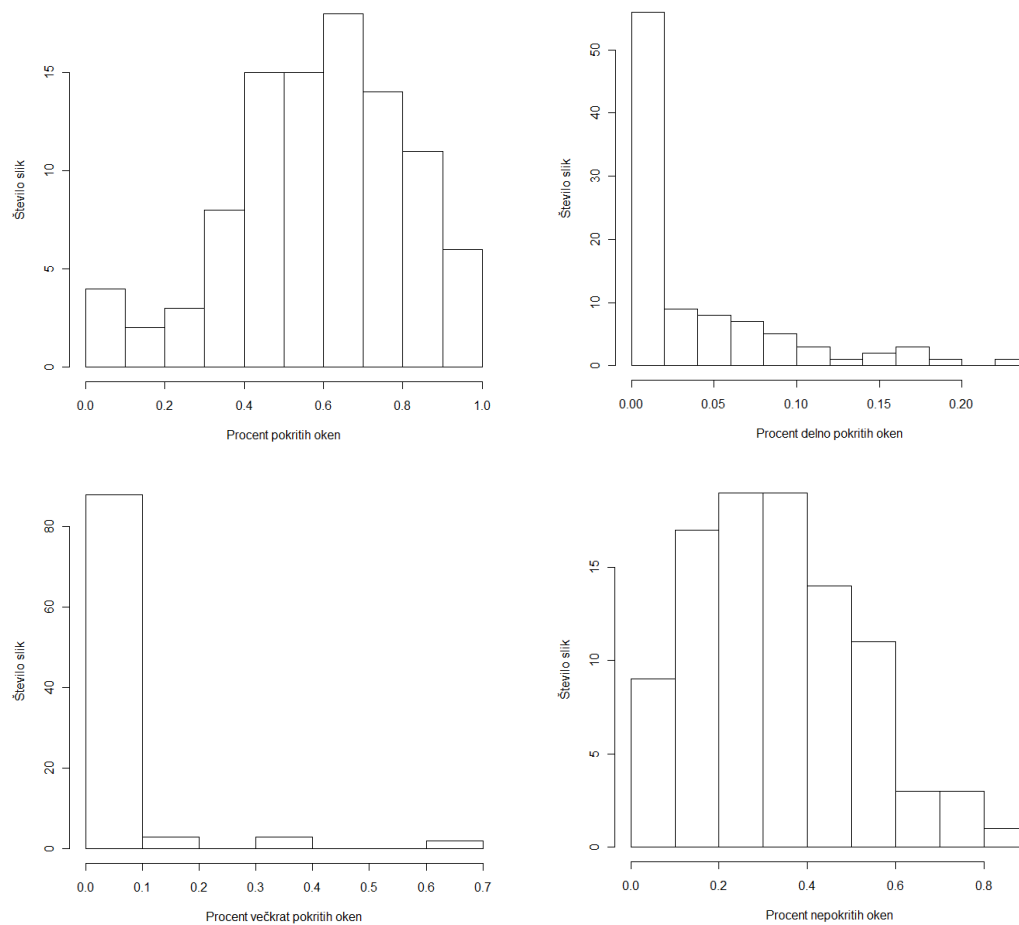
Predlagana metoda bo slabo delovala na fasadah, ki ima izrazito malo oken. V tem primeru je zelo vprašljiva pravilna segmentacija slike, posebej ob prisotnosti izrazitega šuma. Segmentacija se izvaja na podlagi histogramov projekcijskih profilov, ki v opisani situaciji najverjetneje ne bodo dovolj izraziti. Iz istega razloga bo nezanesljiva tudi ocena hevristike na podlagi distribucije robov. Zaradi malega števila oken bo neizrazita tudi hevristična ocena na podlagi medsebojne podobnosti. Slika 4.9 prikazuje slabo končno hipotezo fasade z malo okni.

Neustrezna končna hipoteza bo generirana za fasade z okni, ki se držijo preveč skupaj. V teh primerih bo napačno ocenjena pričakovana velikost oken, kar bo privedlo do nepravilnega združevanja posameznih oken. Za reševanje takih situacij nujno potrebujemo dodatno znanje o strukturi fasade oziroma uporabo globalnih hevrističnih funkcij, ki bi zagotavljale simetričnost in poravnano kreiranje hipoteze. Na sliki 4.10 je prikazana fasada z veliko okni z malo prostora med njimi. Včasih se zgodi, da določene dele slike napačno razglasimo za okna. Primer take fasade je prikazan na sliki 4.11. Tudi napake tega tipa bi lahko zmanjšali z uporabo globalnih hevristik, ki forsirajo simetričnost in poravnano fasade.

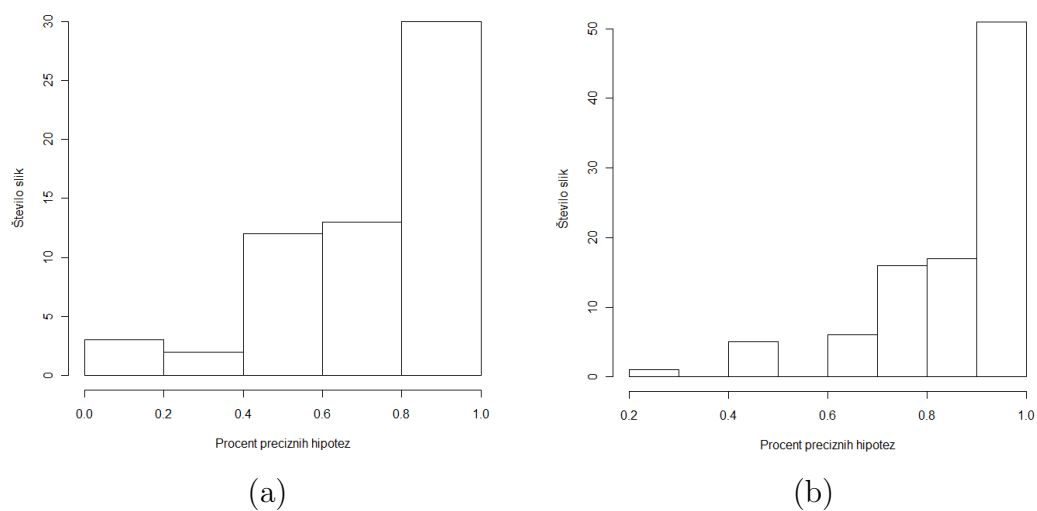
Omejitev predlagane metode je nezmožnost združevanja razdrobljenih končnih hipotez v bolj smiselne rešitve. Na sliki 4.12 je prikazana končna hipoteza, ki bi jo bilo možno izboljšati z uporabo relativno enostavnih pravil združevanja. Primer na sliki 4.13 kaže pomanjkljivosti postopka generiranja končne hipoteze, ki se v določenih primerih prehitro ustavi in ne uporabi vseh kandidatov.



Slika 4.3: Na sliki je prikazana frekvenca deleža pokritih oken, izmerjena na testni množici eTRIMS.



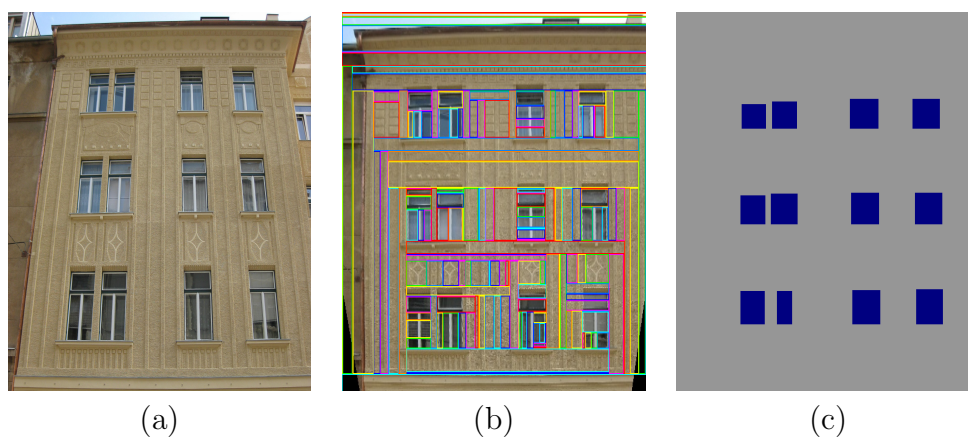
Slika 4.4: Na sliki je prikazana frekvenca deleža pokritih oken, izmerjena na testni množici Ljubljana.



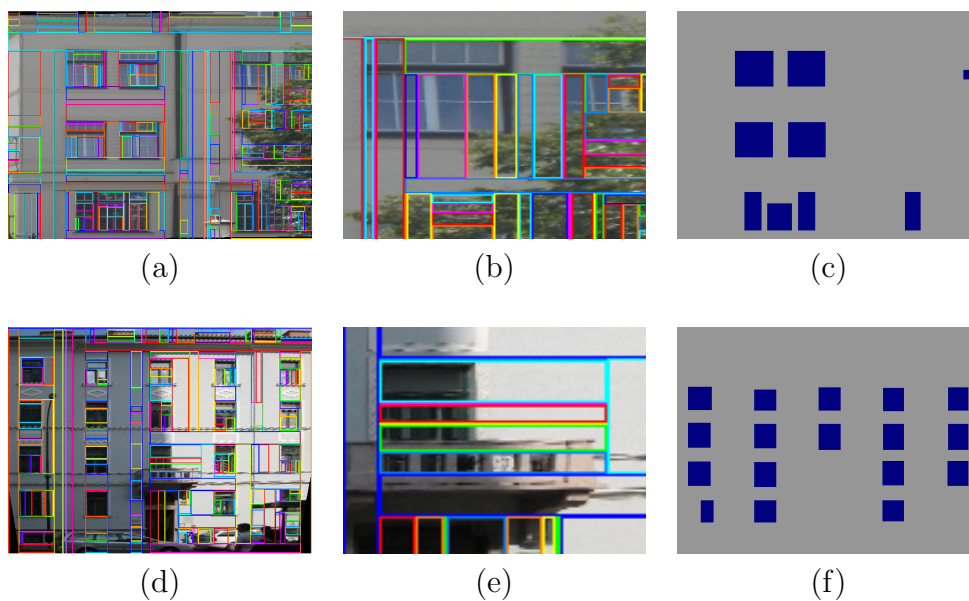
Slika 4.5: Na sliki je prikazana frekvenca deleža preciznih hipotez, a) izmerjena na testni množici eTRIMS, b) izmerjena na testni množici Ljubljana.



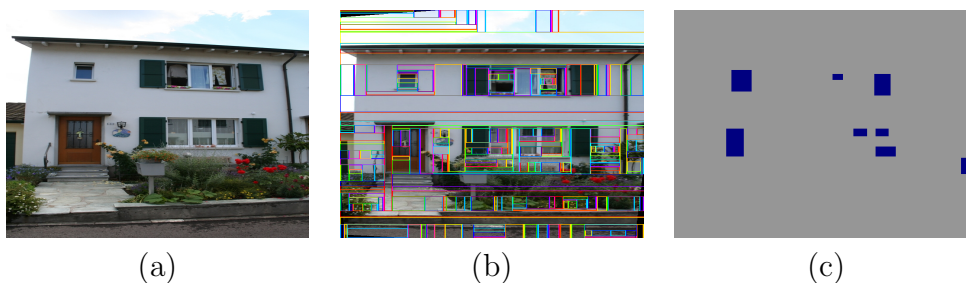
Slika 4.6: Slika prikazuje primere uspešno generiranih hipotez. V levem stolpcu so vhodne fotografije. Srednji stolpec so segmentirane perspektivno izravnane slike. Desni stolpec so generirane končne hipoteze. Opažamo, da se tudi na isti fasadi okna razlikujejo po izgledu (nekatera so odprta, imajo spuščene rolete ali se nahajajo v senci). Kljub temu so dobljene končne hipoteze kvalitetne zaradi izrazite poravnaniosti oken.



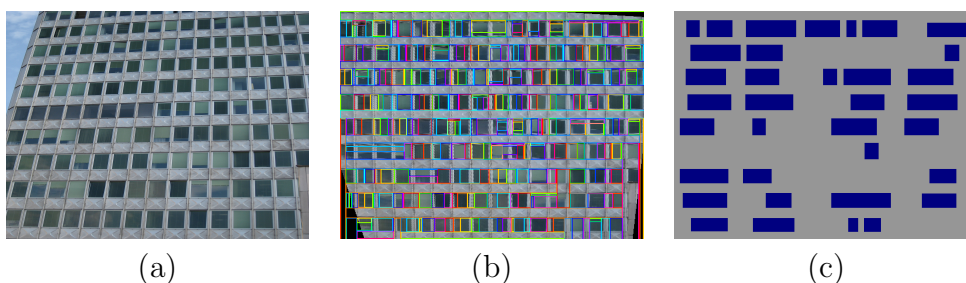
Slika 4.7: Poravnanoost ter podobnost med okni sta zagotovila generiranje uspešne hipoteze, kljub pisani fasadi. a) Vhodna slika fasade, na kateri je možno opaziti okraske, ki se periodično ponavljajo, kar lahko zavede postopek razpoznavanja okna. b) Fragmentacija perspektivno izravnane slike. c) Končna hipoteza je pravilna zaradi velike stopnje poravnanoosti in podobnosti med okni.



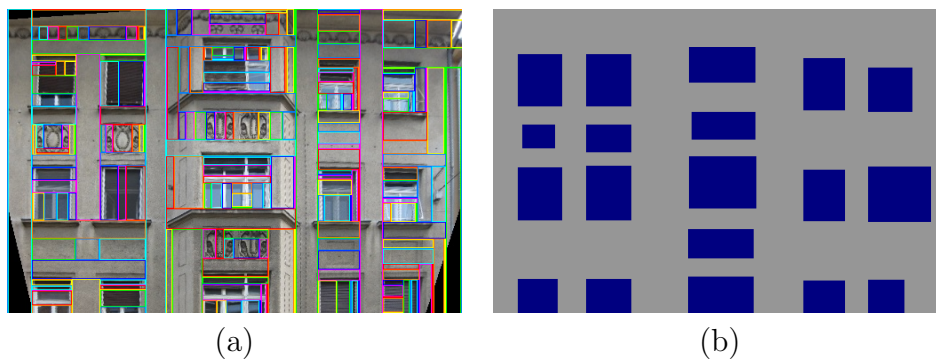
Slika 4.8: Primera posledic omejitve končne hipoteze na generirane kandidate. a) Segmentacija poravnane slike fasade, ki je delno zakrita z drevesom. b) Detajl s prejšnje slike, ki prikazuje napačno segmentacijo okoli okna zaradi šuma, ki ga povzroča drevo. c) Končna hipoteza ne bo pokrila okna v zgornjem desnem kotu slike, čeprav se okno vidi v celoti. d) Segmentacija poravnane slike fasade z balkonom nad vhodnimi vrati. e) Detajl s prejšnje slike, ki prikazuje napačno segmentacijo v bližini balkona. f) Končna hipoteza ne bo pokrila okna nad vrati, ker so vsi kandidati iz tega dela slike neustrezni. Problemom tega tipa bi se izognili, če bi v postopek segmentacije vhodne slike na podlagi projekcijskih profilov vključili še segmentacijo na podlagi dejanskih robov na sliki. Situacijo bi dodatno popravilo tudi ponovno generiranje kandidatov, tokrat na globalni ravni.



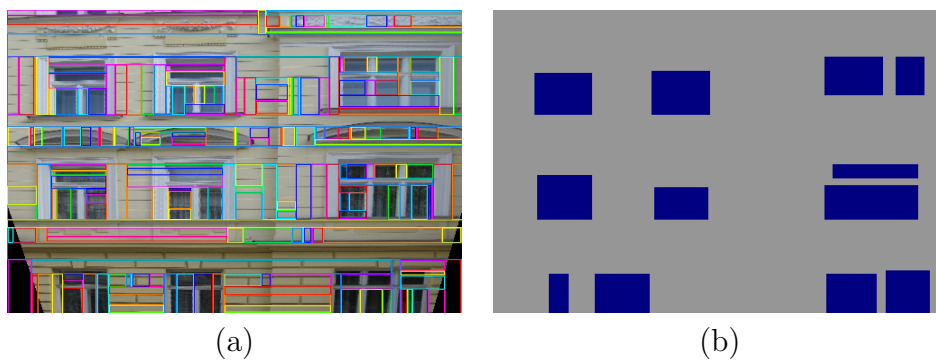
Slika 4.9: Slabo delovanje postopka pri analizi fasade z malo okni. a) Vhodna slika fasade z malo okni in veliko šuma. b) Fragmentacija perspektivno izravnane slike. c) Končna hipoteza je napačna, med drugim tudi zaradi napačno ocenjene pričakovane velikosti oken. Ocene hevrističnih funkcij so neizrazite in nezanesljive, tako da so vsi kandidati slabo ocenjeni.



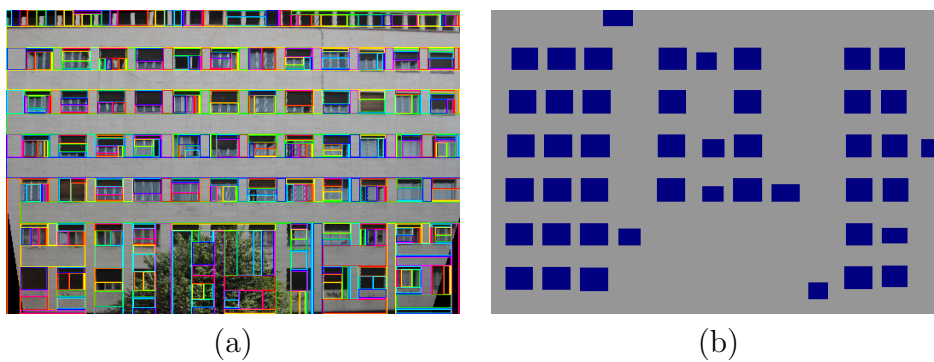
Slika 4.10: Slabo delovanje postopka pri analizi fasade z veliko oken, ki se držijo skupaj. a) Vhodna slika fasade z veliko oken brez izrazitega prostora med njimi. b) Fragmentacija perspektivno izravnane slike. c) Končna hipoteza je napačna, ker se združujejo posamezna okna v neposredni bližini. Za reševanje takih situacij potrebujemo globalne hevristike, ki favorizirajo simetrične in poravnane končne hipoteze.



Slika 4.11: Ponavljajoče se vzorce na fasadi lahko napačno razglasimo za okna. a) Segmentacija poravnane slike. b) Ornamenti so napačno razpoznani kot okna.



Slika 4.12: Z relativno enostavnimi pravili združevanja bi izboljšali razdrobljene hipoteze. a) Segmentacija poravnane slike. b) Razdrobljena končna hipoteza.



Slika 4.13: Pomanjkljivost postopka za generiranje končne hipoteze. a) Segmentacija poravnane slike. b) Postopek dodajanja kandidatov v končno hipotezo se je očitno prehitro nehal. Posledica tega je, da so veliki deli slike ostali ne pokriti.

Poglavje 5

Zaključek

V magistrski nalogi smo predstavili postopek za učenje in prepoznavanje oken kot osnovnih strukturnih elementov fasade. Strukturni zapis fasade omogoča prevedbo vizualne informacije iz nivoja slikovnih elementov na simbolni nivo in predstavlja pomemben korak proti visokonivojskem procesiranju in razumevanju slike.

Prepoznavanje oken se začne s predprocesiranjem vhodne slike, s katerim povečamo zanesljivost in zmanjšamo zahtevnost postopkov v naslednjih fazah obdelave. Ukvarjali smo se z odpravljanjem perspektivne transformacije vhodne slike, s katero zagotovimo paralelnost vseh robov, ki bi bili paralelni ob pravokotnemu pogledu na fasado.

Predstavili smo koncept projekcijskih profilov, ki predstavljajo histograme porazdelitve robov v vertikalni in horizontalni smeri poravnane slike. Ob predpostavki o medsebojni poravnosti oken, ki drži za veliko večino fasad, nam vrhovi histogramov projekcijskih profilov predstavljajo močno indikacijo o prisotnosti oken.

Opisali smo postopek segmentacije fasade v drevesno strukturo, v kateri posamezna vozlišča ustrezajo pravokotnim področjem poravnane slike ter predstavljajo kandidate za okna. Postopek temelji na analizi projekcijskih profilov ter išče področja slike brez robov, ki naj bi ustrezala pasovom zidu brez oken.

Opisali smo štiri hevristične funkcije za ocenjevanje kvalitete posameznih kandidatov in podali postopek za generiranje končne hipoteze o položaju in velikosti oken na sliki fasade.

Rezultati eksperimentalnega dela naloge so zadovoljivi, vendar predlagana metoda še vedno ni ustrezna za praktično rabo.

Ugotovljene lastnosti predstavljene metode so:

- boljše delovanje pri analizi fasad z veliko stopnjo poravnosti oken;
- boljše delovanje v situacijah, ko se okna ne razlikujejo po velikosti;
- boljše delovanje na fasadah z velikim številom oken;
- velika odvisnost od detekcije robov na vhodni sliki;
- omejitve na kandidate, generirane v postopku segmentacije fasade;
- premalo zanesljiv postopek formiranja končne hipoteze;
- metoda ne uporablja splošnega znanja o strukturi fasad.

Veliko je možnosti za izboljšavo predlaganega postopka. V postopek segmentacije fasade bi bilo koristno vključiti tudi razbitje na podlagi dejanskih robov in ne le na podlagi projekcijskih profilov, vsaj na nivoju listov drevesne strukture. Namreč sami robovi so tudi močna indikacija prisotnosti strukturnih elementov fasade. V trenutni implementaciji postopka se pogosto dogaja, da se segmentacija slike prehitro ustavi, ker so projekcijski profili gladki, čeprav so prisotni močni robovi, na podlagi katerih bi bilo možno nadaljevati segmentacijo.

Vsekakor bi bilo koristno preizkusiti druge hevristične funkcije za ocenjevanje kandidatov za okna. Poraja se ideja o uporabi induktivnega logičnega programiranja (ILP), s katerim bi prišli do logične definicije okna kot določene konfiguracije robov. Do boljših rezultatov bi bilo mogoče priti z določanjem uteži prispevkov posameznih hevristik s strojnim učenjem.

Hierarhično razbitje fasade, ki smo ga vpeljali s ciljem izkoričanja lokalne informacije, se je na začetku izkazalo zelo koristno. V končni fazi pa se je pokazalo kot breme, saj neuniformna razrobitev onemogoča uporabo globalnih ocen, kot sta simetričnost in poravnost, z namenom formiranja končne hipoteze. Opisani način drobljenja vhodne slike z upoštevanjem lokalne informacije je uporaben za določanje začetne hipoteze, na podlagi katere bi izvedli

globalno segmentacijo fasade. Ta bi v nadaljevanju omogočila uporabo globalnih hevristik o strukturi fasade.

Največja izboljšava bi bila eksplicitna vključitev dodatnega znanja o splošnih pravilih, ki določajo običajno strukturo fasade. V tem smislu je obetavna uporaba stohastičnih kontekstno neodvisnih gramatik, ki s svojimi produkcijskimi pravili zagotavljajo simetričnost, poravnano ter ponavljajoče se vzorce končne hipoteze kot lastnost arhitekturnih slogov. Stohastične gramatike igrajo vlogo mehanizma za usmerjanje iskanja končne hipoteze v prostoru vseh strukturnih opisov vhodne slike. Prevelika časovna kompleksnost - ob kombinatorični eksploziji vseh možnih strukturnih opisov - predstavlja glavno težavo pri njihovi uporabi. Končna hipoteza, generirana z našim postopkom, lahko predstavlja dobro izhodiščno točko za stohastično gramatiko, ki bi drastično omejila iskalni prostor ter s tem bistveno skrajšala čas procesiranja. Koristno bi bilo uporabiti kriterij minimalne dolžine opisa (MDL), ki bi zagotovil kompromis med točnostjo in zapletenostjo končne hipoteze. Po eni strani želimo rešitev, ki dobro opisuje vhodno sliko ter dobro razlaga prisotne robove ali vsaj histograme projekcijskih profilov. Po drugi strani se želimo izogniti prevelikem prilagajanju vhodne slike, kar bi pripeljalo do preveč kompleksnih opisov strukture fasad, ki so po kriteriju MDL tudi manj verjetni. Zato bi dober kompromis dosegli z iskanjem najmanjšega drevesa izpeljave končne hipoteze (v smislu števila uporabljenih vozlišč), ki še vedno dovolj dobro opisujejo vhodno sliko. Uspešnost opisane metode bi v veliki meri bila odvisna od natančno odmerjenih parametrov pri vrednotenju hipotez in iskanju optimalnega kompromisa, za kar bi verjetno potrebovali strojno učenje na podlagi skrbno izbrane učne množice.

Slike

3.1	Gaussova sfera kot akumulacijsko polje za detekcijo bežišč. . . .	13
3.2	Mera razdalje za mehko stopnjo podpore pri iskanju bežišč. . . .	15
3.3	Konstrukcija homografije.	17
3.4	Izbira korespondenčnih točk homografije	19
3.5	Primeri rektifikacije slik.	20
3.6	Shematski prikaz dela fasade.	21
3.7	Maski Sobelovega operatorja.	22
3.8	Parcialna odvoda slike v horizontalni in vertikalni smeri z uporabo Sobelovega operatorja.	23
3.9	Binarni filter za glajenje vrednosti projekcijskih profilov	23
3.10	Projekcijska profila fasade.	24
3.11	Glajenje projekcijskega profila z ASF filtrom.	26
3.12	Primer generiranja hipotez.	30
3.13	Drevesna struktura hipotez.	31
3.14	Razbitje fasade.	32
3.15	Primeri segmentacije fasade.	34
3.16	Izris hipotez po nivojih drevesne strukture.	38
3.17	Ocene hipotez s hevrstiko na podlagi pozicije v drevesni strukturi.	39
3.18	Akumulacija lokalnih maksimumov projekcijskih profilov.	41
3.19	Laplaceov operator nad projekcijskim profilom.	43
3.20	Akumulacijsko polje pričakovane lokacije oken na fasadi.	44
3.21	Akumulacijsko polje pričakovane lokacije oken na fasadi dobljeno z rekurzivnim posodabljanjem.	45
3.22	Ocene hipotez s hevrstiko na podlagi distribucije robov.	46
3.23	Ocene hipotez s hevrstiko na podlagi medsebojne podobnosti.	48
3.24	Značilke Viola-Jones	49
3.25	Izračun vsote intenzitet pikslov na osnovi integralne slike	51
3.26	Kaskada močnih klasifikatorjev	54
3.27	Pozitivni učni primeri.	56

3.28	Negativni učni primeri.	57
3.29	Akumulacija detektiranih oken s kaskado klasifikatorjev.	58
3.30	Ocene hipotez s hevristiko na podlagi naučenega videza oken.	59
3.31	Končne ocene hipotez na podlagi utežene linearne kombinacije predstavljenih hevristik.	61
3.32	Končna hipoteza o oknih fasade.	63
4.1	Eksperimentalni rezultati na podatkovni zbirki eTRIMS.	66
4.2	Eksperimentalni rezultati na podatkovni zbirki Ljubljana.	67
4.3	Frekvenca deleža pokritih oken - eTRIMS.	70
4.4	Frekvenca deleža pokritih oken - Ljubljana.	71
4.5	Frekvenca deleža preciznih hipotez.	72
4.6	Primeri uspešno generiranih hipotez.	73
4.7	Dobra hipoteza na pisani fasadi.	74
4.8	Slabo delovanje zaradi napačne segmentacije.	75
4.9	Slabo delovanje na fasadi z malo okni.	76
4.10	Slabo delovanje na fasadi z veliko oken.	76
4.11	Ornamenti izbrani kot okna.	77
4.12	Razdrobljena hipoteza.	77
4.13	Pomanjkljivost generiranja končne hipoteze.	78

Algoritmi

1	Izračun vertikalnega projekcijskega profila	27
2	Izračun horizontalnega projekcijskega profila	28
3	Hierarhična segmentacija fasade	33
4	Pričakovana minimalna in maksimalna širina oken	36
5	Hevristična ocena na podlagi pozicije v drevesni strukturi	40
6	Pričakovana pozicija oken	42
7	Hevristična ocena na podlagi medsebojne podobnosti	47
8	AdaBoost postopek za učenje klasifikatorja	53
9	Formiranje končne hipoteze	62

Tabele

4.1	Povprečni izvajalni časi obdelave slike.	68
-----	--	----

Literatura

- Fernando Alegre, Frank Dellaert. A probabilistic approach to the semantic interpretation of building facades. Technical report, Georgia Institute of Technology, 2004.
- S. T. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(4): 435–462, November 1983.
- Serge Belongie, Jitendra Malik, Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.
- Dr. Gary Rost Bradski, Adrian Kaehler. *Learning opencv, 1st edition*. O’Reilly Media, Inc., 2008. ISBN 9780596516130.
- J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986. ISSN 0162-8828.
- A. R. Dick, P. H. S. Torr, R. Cipolla. Modelling and interpretation of architecture from several images. *International Journal of Computer Vision*, 60(2):111–134, November 2004.
- P. F. Felzenszwalb, D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, January 2005.
- R. Fergus, P. Perona, A. Zisserman. Object class recognition by unsupervised scale-invariant learning. V *CVPR*, str. II: 264–271, 2003.
- M. A. Fischler, R. A. Elschlager. The representation and matching of pictorial structures. *IEEE Trans. Computer*, 22(1):67–92, January 1973.
- Yoav Freund, Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. V Paul M. B. Vitányi, urednik,

- EuroCOLT*, volume 904 of *Lecture Notes in Computer Science*, str. 23–37. Springer, 1995. ISBN 3-540-59119-2.
- Gene H. Golub, Charles F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996. ISBN 0801854148.
- P. Green. Reversible jump markov chain monte carlo computation and bayesian model determination, 1995.
- F. Han, S. C. Zhu. Bottom-up/top-down image parsing by attribute graph grammar. V *ICCV*, str. II: 1778–1785, 2005.
- J. E. Haugeard, S. Philipp Foliguet, F. Precioso, J. Lebrun. Extraction of windows in facade using kernel on graph of contours. V *SCIA*, str. 646–656, 2009.
- P. V. C. Hough. Methods and means for recognizing complex patterns. US Patent 3,069,654, 1962.
- F. Korč, W. Förstner. eTRIMS Image Database for interpreting images of man-made scenes. Technical Report TR-IGG-P-2009-01, April 2009. URL <http://www.ipb.uni-bonn.de/projects/etrimdb/>.
- Sung Chun Lee, Ramakant Nevatia. Extraction and integration of window in a 3D building model from ground view image. V *CVPR (2)*, str. 113–120, 2004.
- D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.
- Pascal Müller, Peter Wonka, Simon Haegler, Andreas Ulmer, Luc Van Gool. Procedural modeling of buildings. 25(3):614–623, 2006. ISSN 0730-0301.
- Pascal Müller, Gang Zeng, Peter Wonka, Luc Van Gool. Image-based procedural modeling of facades. 26(3), 2007.
- R. M. Neal. Probabilistic inference using markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, University of Toronto, 1993.
- Yoav I. H. Parish, Pascal Müller. Procedural modeling of cities. V Eugene Fiume, urednik, *Proceedings of ACM SIGGRAPH 2001*, str. 301–308, New York, NY, USA, 2001. ACM Press. ISBN 1-58113-374-X.

- Przemyslaw Prusinkiewicz, Aristid Lindenmayer. *The Algorithmic Beauty of Plants*. Springer-Verlag, 1990.
- Nora Ripperda, Claus Brenner. Reconstruction of façade structures using a formal grammar and rjmc. V Katrin Franke, Klaus-Robert Müller, Bertram Nickolay, Ralf Schäfer, uredniki, *DAGM-Symposium*, volume 4174 of *Lecture Notes in Computer Science*, str. 750–759. Springer, 2006. ISBN 3-540-44412-2.
- C. Rother. A new approach for vanishing point detection in architectural environments. V *BMVC*, str. xx–yy, 2000.
- Hao Shao, Tomáš Svoboda, Luc van Gool. ZuBuD – Zurich Buildings Database for Image Based Recognition. Technical Report Technical Report No. 260, Computer Vision Lab, Swiss Federal Institute of Technology, Switzerland, Zurich, Switzerland, April 2003.
- J. Sobel. Neighborhood coding of binary images for fast contour following and general binary array processing. *Computer Graphics Image Processing*, 8(1): 127–135, 1978.
- George Stiny, James Gips. Shape grammars and the generative specification of painting and sculpture. V *IFIP Congress (2)*, str. 1460–1465, 1971.
- M. A. Turk, A. P. Pentland. Face recognition using eigenfaces. V *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, str. 586–590, Hawaii, June 1992.
- P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. *Proc. CVPR*, 1:511–518, 2001.
- M. Weber, M. Welling, P. Perona. Towards automatic discovery of object categories. V *CVPR*, str. II: 101–108, 2000.
- Peter Wonka, Michael Wimmer, François Sillion, William Ribarsky. Instant architecture. *ACM Trans. Graph.*, 22(3):669–677, 2003. ISSN 0730-0301.
- Long Zhu, Yuanhao Chen, Alan L. Yuille. Unsupervised learning of a probabilistic grammar for object detection and parsing. V Bernhard Schölkopf, John C. Platt, Thomas Hoffman, uredniki, *NIPS*, str. 1617–1624. MIT Press, 2006. ISBN 0-262-19568-2.
- Song Chun Zhu, David Mumford. A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision*, 2(4):259–362, 2006.