

Uporaba lokalnih značilnic v aplikacijah spoznavnega vida za urbana okolja*

Dušan Omerčević, Roland Perko in Aleš Leonardis
Univerza v Ljubljani, Fakulteta za računalništvo in informatiko
Tržaška 25, 1001 Ljubljana, Slovenija
{dusan.omercevic, roland.perko}@fri.uni-lj.si

Local features in cognitive vision applications for urban environments

In this paper we present a performance evaluation of MSER and Hessian-Affine local feature detectors in a typical use case of cognitive vision applications in urban environments. By using a wide baseline stereo matching approach we try to find camera motion between a user image and images stored in a database. Running this application on test images twice while only changing the underlying local feature type has shown that the MSER local feature detector outperforms the Hessian-Affine detector. Additionally, we have shown that local features can perform well in cognitive vision applications for urban environments.

1 Uvod

Zlivanje digitalnih tehnologij in razvoj mobilnih telekomunikacij v zadnjih 15 letih je omogočilo uporabo metod s področja spoznavnega vida na nove načine in v novih okoljih. V okviru EU-projekta MOBVIS [1] proučujemo možnosti, ki jih spoznavni vid ponuja mobilnemu uporabniku v urbanih okoljih. Na podlagi vizualne informacije, globalnega položajnega sistema, inercialskih in drugih senzorjev ter sklepanja želimo ugotoviti, kje se uporabnik nahaja in kaj ga obkroža. Na podlagi te informacije je možno uporabniku ponuditi kontekstno oziroma lokacijsko pogojene storitve.

Med cilji naših raziskav je tudi razvoj metod, ki bodo omogočale določitev geografskega položaja uporabnika na osnovi slike njegove okolice, zajete na primer s telefonom opremljenim z digitalnim fotoaparatom. Eden izmed možnih načinov določanja položaja uporabnika temelji na iskanju ujemanja med sliko uporabnikove okolice in

slikami posnetimi z znanih položajev, ki so shranjene v podatkovni zbirki. To lahko storimo tako, da primerjamo celotne slike med seboj na podlagi nekih globalnih značilnic, ali da primerjamo med seboj samo manjša področja slik, to je lokalne značilnice. Pristopi na podlagi globalnih značilnic omogočajo gradnjo odzivnih in enostavnih informacijskih sistemov, vendar so le-ti občutljivi na spremembe točke gledišča in spremembe v okolju. Pristopi na podlagi lokalnih značilnic so manj občutljivi na takšne spremembe, vendar so računsko bolj zahtevni, saj moramo množico lokalnih značilnic zaznanih v vhodni sliki primerjati s še večjo množico lokalnih značilnic shranjenih v neki podatkovni zbirki.

Lowe [7] je opredelil lokalne značilnice kot značilnice, ki so dovolj razlikovalne za iskanje ujemanja se značilnic v veliki zbirki značilnic, a obenem dovolj lokalne, da so neobčutljive na gnečo (angl. clutter) in zakrivanje. Njegova opredelitev ustreza prevladujočem načinu uporabe lokalnih značilnic za iskanje ujemanja med dvema slikama ali za prepoznavo slik, kot sta ga predstavila Schmid in Mohr [15]. Njun pristop je sestavljen iz (1) zaznave množice lokalnih značilnic v vhodni sliki, (2) opisovanja vsake lokalne značilnice z nekim opisnikom, (3) iskanja na podlagi opisnika v zbirki značilnic in (4) verifikacije tako pridobljenih hipotez.

V grobem lahko lokalne značilnice razdelimo glede na to, ali temeljijo na zanimivih točkah, skupkih daljic ali odlikovanih področjih. Detektorji za prvo vrsto lokalnih značilnic (npr. [15, 16, 8, 2, 10]) najprej poiščejo v sliki neko zanimivo točko in nato poskušajo najti značilno okolico te točke. Če je zanimivih točk malo ali je njihova okolica neizrazita (npr. pročelje modernističnih zgradb), najdemo le malo lokalnih značilnic temelječih na teh točkah. V takih primerih pridejo do izraza lokalne značilnice temelječe na skupkih daljic (npr. [14, 6, 3]), pri čemer daljice povečini predstavljajo robove ploskev. Če so se detektorji lokalnih značilnic prejšnjih dveh vrst povečini opirali na kote in robove, poskušajo detektorji za tretjo vrsto lokalnih značilnic (npr. [17,

* To delo so deloma podprli MVZT (Raziskovalni program Računalniški vid P2-0214) ter EU projekta CoSy (FP6-004250-IP) in MOBVIS (FP6-511051-2).

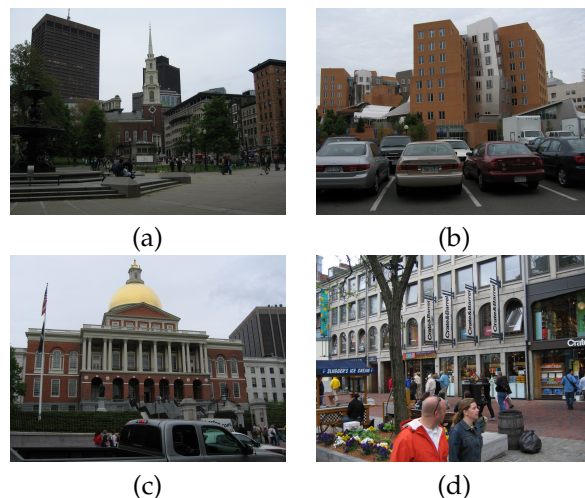
9]) na sliki poiskati področja, ki jih odlikuje neka značilna lastnost. Primer odlikovanega področja je ekstremalno področje, ki ga sestavljajo povezani slikovni elementi, ki so vsi svetlejši oz. temnejši od slikovnih elementov, ki mejijo na to področje.

V tem članku želimo oceniti, katera vrsta lokalnih značilnic je najbolj primerna za gradnjo aplikacij spoznavnega vida namenjenih uporabi v urbanih okoljih. V razdelku 2 opišemo uporabljeni postopek ocenjevanja lokalnih značilnic. V razdelku 3 predstavimo izide ocenjevanja in v razdelku 4 podamo zaključek.

2 Ocena zmožnosti lokalnih značilnic

Na podlagi ocen predstavljenih v treh različnih člankih [11, 4, 12] smo izbrali dve vrsti lokalnih značilnic za podrobnejšo obravnavo. Avtorji se strinjajo, da je za predmete z mnogo ravnimi površinami, pogostimi v urbanih okoljih, najbolj primeren detektor MSER [9], sledi mu Hessian-Affine detektor [10]. Ta detektorja in lokalne značilnice, ki jih zaznavata, smo izbrali za podrobnejšo obravnavo. Za ocenjevanje njunih zmožnosti smo uporabili aplikacijo iskanja ujemanja med sliko uporabnikove okolice in slikami posnetimi z znanih položajev, ki so shranjene v podatkovni zbirki. Uspešnost iskanja ujemanja smo uporabili kot mero zmožnosti lokalnih značilnic. Ocenjujemo, da ta mera dobro odraža željene lastnosti lokalnih značilnic, ki so ponovljivost zaznave, razlikovalnost/informativnost, natančnost določitve položaja lokalne značilnice, ne preveliko število zaznanih lokalnih značilnic, natančnost prekrivanja soodzivnih lokalnih značilnic in čim manjša računska zahtevnost postopka zaznave [11]. Primerjavo med izbranimi lokalnima značilnicama smo naredili tako, da smo dvakrat izvedli enako aplikacijo iskanja ujemanja na izbranih testnih slikah, pri čemer je bila edina razlika med izvedbama v izbrani lokalni značilnici. Za testne slike smo izbrali nabor slik iz ICCV 2005 Computer Vision Contest (<http://research.microsoft.com/iccv2005/contest/>). Izbrali smo najbolj zahteven nabor slik (št. 3), ki ga sestavlja 60 učnih slik in 37 testnih slik. Slike prihajajo iz štirih različnih delov Bostona, ZDA (slika 1). Aplikacijo in testne slike smo izbrali tako, da dobro predstavljajo značilno uporabo lokalnih značilnic v aplikacijah spoznavnega vida za urbana okolja.

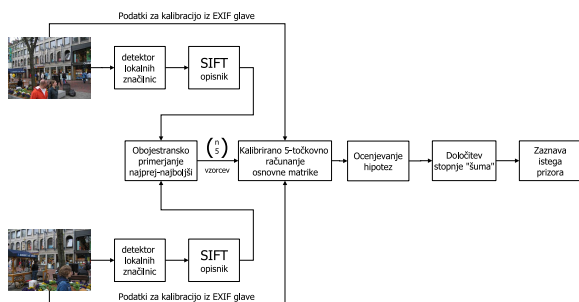
Aplikacija iskanja ujemanja primerja uporabnikovo sliko s slikami shranjenimi v podatkovni zbirki. Iskanje ujemanja pove, ali uporabnikova slika prikazuje isti prizor kot slika iz podatkovne zbirke, in v primeru ujemanja, tudi gibanje kamere med slikama. Postopek iskanja ujemanja med



Slika 1: slike iz slikovne zbirke ICCV 2005 Computer Vision Contest. (a) park, (b) parkirišče, (c) zgradba parlamenta in (d) ulica namenjena pešcem.

slikama (slika 2) se prične z zaznavo lokalnih značilnic. To nam da seznam eliptičnih področij slike, ki predstavljajo neke zanimive lokalne značilnice. V naslednjem koraku vsako izmed teh eliptičnih regij opišemo z opisnikom SIFT [8], ki je 128 dimenzionalni vektor. Dosedaj sta bili slike obdelovani neodvisno druga od druge. V naslednjem koraku izvršimo obojestransko primerjanje najprej najboljši. V tem koraku se za vsak vektor SIFT, ki opisuje neko lokalno značilnico v prvi sliki, poišče najbližji vektor SIFT v drugi sliki. Za iskanje najbližjega vektorja uporabljamo evklidsko razdaljo in K-D drevo [5]. Med najdenimi pari vektorjev SIFT ohranimo samo pare, ki so si vzajemno najbližje. Izhod tega koraka je urejen seznam potencialnih soodzivnosti med slikama. Seznam je urejen naraščajoče glede na evklidsko razdaljo med paroma vektorjev SIFT. Naše izkušnje kažejo, da verjetnost prave soodzivnosti narašča z manjšanjem evklidske razdalje med paroma vektorjev SIFT. Soodzivnost pomeni dve ujemajoči se lokalni značilnici v prvi oziroma drugi sliki. Dve področji v sliki sta pravi soodzivnosti, če sta projekciji iste strukture v 3D-svetu.

V naslednjem koraku z uporabo omejitev epipolarne geometrije izberemo prave soodzivnosti s seznamom potencialnih soodzivnosti pridobljenega v prejšnjem koraku. Nad vsako kombinacijo petih izmed prvih n elementov z urejenega seznama potencialnih soodzivnosti uporabimo 5-točkovni algoritem [13] za ustvarjanje hipoteze, to je osnovne matrice E , ki določa epipolarno geometrijo med dvema slikama. Ker 5-točkovni algoritem zahteva kalibracijo kamere, uporabljamo informacijo o goriščni razdalji pridobljeni iz EXIF glave slike za normalizacijo koordinat slikovnih elementov. Poleg tega



Slika 2: Postopek iskanja ujemanja

predpostavimo, da so slike zajete s CCD kamero, da je glavna točka v središču slike in da je razmerje stranic slikovnega elementa enako 1. V naši aplikaciji smo določili $n = 15$, kar pomeni 3.003 hipoteze. Med vsemi ustvarjenimi hipotezami izberemo tisto z največjo podporo, to je z največjim številom soodzivnosti, ki ustrezajo izračunani osnovni matriki. Ustreznost z izračunano osnovno matriko pomeni, da je razdalja med središčem eliptičnega področja in ustrezno epipolarno črto manjša od nekega praga. V naši aplikaciji smo ta prag določili na približno 2,5 slikovnega elementa za sliko velikosti 1600×1200 slikovnih elementov.

Hipoteza z največ podporo ni nujno pravilna hipoteza. Tudi hipoteza ustvarjena na podlagi nepravilnih soodzivnosti lahko dobi nekaj podpore. Zato izmed soodzivnosti, ki niso v skladu z zmagovalno hipotezo, povlečemo določeno število vzorcev petih elementov in izračunamo podporo, ki jo ima tako ustvarjena hipoteza. Na tej podlagi in v odvisnosti od nivoja specifičnosti, ki ga želimo, izračunamo stopnjo "šuma", to je, kolikšno podporo lahko največ dobi nepravilna hipoteza. Ta prag uporabimo za oceno, ali je podpora, ki jo je dobila zmagovalna hipoteza, zadostna. Če je podpora zadostna privzamemo, da sliki predstavljata isti prizor. Nivo specifičnosti pri tem določa razmerje med številom pravilno določenih parov slik, ki v resnici ne prikazujeta istega prizora, in številom nepravilno določenih parov slik, ki v resnici prikazujeta isti prizor, čeprav uporabljeni postopek trdi drugače.

3 Potek ocenjevanja in izidi

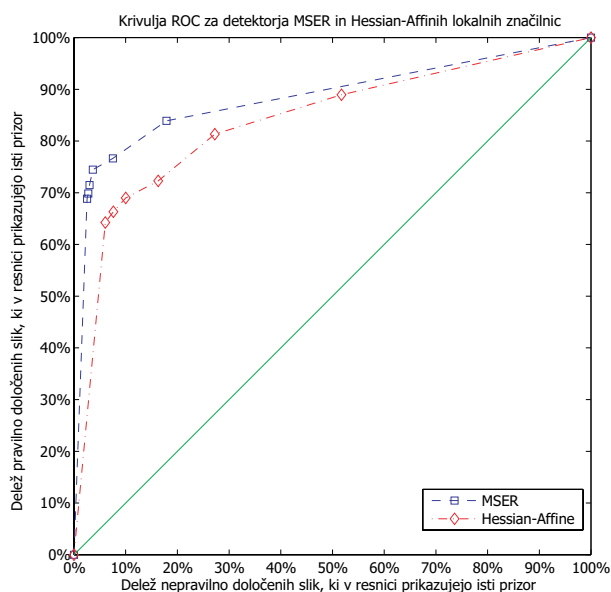
Uporabili smo lastno izvedbo detektorja MSER, za katero smo preverili, da je po zmožnostih primerljiva z izvorno izvedbo [9]. Nastavitve detektorja MSER smo pustili enake privzetim nastavitvam izvirne izvedbe. Za Hessian-Affine detek-

Tabela 1: Najmanjše, največje in povprečno število skupaj s standardnim odklonom števila zaznanih lokalnih značilnic za obe vrsti lokalnih značilnic na množici 97 slik iz uporabljenega nabora slik.

Vrsta značilnic	Min	Max	Mean	StdDev
MSER	272	3416	1570	770
Hessian-Affine	168	4313	1725	1065

tor smo uporabili izvedbo, ki je javno dostopna na <http://www.robots.ox.ac.uk/~vgg/research/affine/>. Ker število zaznanih lokalnih značilnic močno vpliva na odzivnost aplikacije iskanja ujemanja, smo nastavili prag "kotnosti" Hessian-Affinega detektorja na 1000 in s tem uravnotežili število zaznanih lokalnih značilnic med detektorjem MSER in Hessian-Affinim detektorjem. Število zaznanih lokalnih značilnic za obe vrsti lokalnih značilnic na izbranem naboru slik je predstavljeno v tabeli 1.

Aplikacijo iskanja ujemanja smo zagnali dvakrat. Prvič smo uporabili detektor MSER lokalnih značilnic in drugič detektor Hessian-Affinih lokalnih značilnic. 60 slik smo uporabili kot referenčne slike in preostalih 37 slik za testiranje. Ročno smo določili pravilne rezultate, to je, za vsako testno sliko smo izbrali referenčne slike, ki vsaj v nekem delu prikazujejo isti prizor kot testna slika. Za vsako testno sliko smo iskali ujemanje z vsako izmed 60 referenčnih slik. S primerjavo ročno določenih pravilnih rezultatov in dobljenih rezultatov za različne nivoje specifičnosti smo ocenili zmožnosti aplikacije tako, da smo prešteli število pravilno in nepravilno določenih referenčnih slik, ki v resnici prikazujejo isti prizor (true/false positives), in število pravilno in nepravilno določenih referenčnih slik, ki v resnici ne prikazujejo istega prizora (true/false negatives). Pri tem smo najprej uporabili detektor MSER lokalnih značilnic, nakar smo ponovili postopek še z uporabo detektorja Hessian-Affinih lokalnih značilnic. Pripadajoča krivulja ROC je prikazana na sliki 3. Razlika v zmožnostih aplikacije ob uporabi detektorja MSER ali Hessian-Affinega detektorja lokalnih značilnic je opazna za vse nivoje specifičnosti. Na osnovi teh izidov posredno sklepamo, da so ponovljivost zaznave, razlikovalnost/informativnost, natančnost določitve položaja in natančnost prekrivanja lokalnih značilnic MSER in pripadajočega detektorja mnogo boljši od Hessian-Affinih lokalnih značilnic in pripadajočega detektorja. Kot smo že omenili, smo nastavili Hessian-Affine detektor tako, da oba izbrana detektorja zaznata približno enako število lokalnih značilnic, pri čemer je standardni odklon števila zaznanih lokalnih značilnic manjši pri detektorju MSER (tabela 1). Glede računske zahtevnosti postopka za-



Slika 3: Krivulja ROC za detektorja MSER in Hessian-Affinih lokalnih značilnic

znave se sklicujemo na [11], kjer poročajo, da detektor MSER izvrši zaznavo mnogo hitreje od Hessian-Affinega detektorja.

4 Zaključek

Na podlagi pregleda literature in ocen zmožnosti lokalnih značilnic predstavljenih v treh različnih člankih [11, 4, 12] smo izbrali detektorja MSER in Hessian-Affinih lokalnih značilnic za podrobnejšo obravnavo. Aplikacijo uporabljeno za ocenjevanje in testne slike smo izbrali tako, da dobro predstavljajo značilno uporabo lokalnih značilnic v aplikacijah spoznavnega vida za urbana okolja. V primeru izbrane aplikacije iskanja ujemanja med dvema slikama in na izbranem naboru slik je detektor MSER lokalnih značilnic dal boljše rezultate od detektorja Hessian-Affinih lokalnih značilnic. Ocenjevanje zmožnosti lokalnih značilnic je nadalje pokazalo, da so primerne za aplikacije spoznavnega vida namenjene uporabi v urbanih okoljih.

Literatura

- [1] <http://www.mobvis.org>.
- [2] A. Baumberg. Reliable feature matching across widely separated views. V *CVPR*, 1. del, str. 774–781, 2000.
- [3] H. Bay, V. Ferrari, in L. Van Gool. Wide-baseline stereo matching with line segments. V *CVPR*, 1. del, str. 329–336, 2005.
- [4] F. Fraundorfer in H. Bischof. A novel performance evaluation method of local detectors on non-planar scenes. V *Workshop Proc. Empirical Evaluation Methods in Computer Vision (CVPR)*, str. 33, 2005.
- [5] J. H. Friedman, J. L. Bentley, in R. A. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw.*, letnik 3, št. 3, str. 209–226, 1977.
- [6] Y. Li in L. G. Shapiro. Consistent line clusters for building recognition in CBIR. V *ICPR*, 3. del, str. 952–956, 2002.
- [7] D. G. Lowe. Local feature view clustering for 3D object recognition. V *CVPR*, 1. del, str. 682–688, 2001.
- [8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, letnik 60, št. 2, str. 91–110, 2004.
- [9] J. Matas, O. Chum, M. Urban, in T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, letnik 22, št. 10, str. 761–767, 2004.
- [10] K. Mikolajczyk in C. Schmid. An affine invariant interest point detector. V *ECCV*, 1. del, str. 128–142, 2002.
- [11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, in L. Van Gool. A comparison of affine region detectors. *IJCV*, letnik 65, št. 1-2, str. 43–72, 2005.
- [12] P. Moreels in P. Perona. Evaluation of features detectors and descriptors based on 3D objects. V *ICCV*, 1. del, str. 800–807, 2005.
- [13] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE PAMI*, letnik 26, št. 6, str. 756–777, 2004.
- [14] P. Pritchett in A. Zisserman. Wide baseline stereo matching. V *ICCV*, str. 754–760, 1998.
- [15] C. Schmid in R. Mohr. Local grayvalue invariants for image retrieval. *IEEE PAMI*, letnik 19, št. 5, str. 530–534, 1997.
- [16] T. Tuytelaars in L. Van Gool. Content-based image retrieval based on local affinity invariant regions. V *Proc. VISUAL '99: Information and Information Systems*, str. 493–500, 1999.
- [17] T. Tuytelaars in L. Van Gool. Wide baseline stereo matching based on local, affinity invariant regions. V *BMVC*, str. 412–425, 2000.