

EKSTRAKCIJA I PRAĆENJE KLJUČNIH TAČAKA DALEKE POZADINE U DINAMIČKOJ MONOKULARNOJ VIZIJI

Branko Karan

Institut tehničkih nauka SANU, Beograd; e-mail: branko.karan@itn.sanu.ac.rs

SAŽETAK

Praćenje tačaka udaljene pozadine je od višestrukog značaja u vizuelnoj navigaciji, posebno u zadacima koji pretpostavljaju brzu dinamiku kretanja kamere. U radu je opisan algoritam kojim se ostvaruje efikasna identifikacija i praćenje ovih tačaka korišćenjem monokularne vizije. Algoritam se zasniva na modelu udaljene pozadine koji se sastoji iz skupa ključnih tačaka i asociiranih stepena članstva. Efektivnost predloženog pristupa demonstrirana je na video sekvencama dobijenim na otvorenom prostoru. Takođe su razmotrena neka ograničenja i dalji rad u cilju operacionalizacije metoda.

1. UVOD

U zadacima vizuelne navigacije, daleku pozadinu čini skup statičkih objekata koji se nalaze na velikoj udaljenosti od kamere. Tokom kretanja kamere, slika ovih objekata ostaje praktično nepromenjena – menja se samo njihova pozicija i orijentacija na slici kao posledica rotacije kamere.

U idealnom slučaju u kome bi se objekti udaljene pozadine nalazili na beskonačnom rastojanju od kamere, slika udaljene pozadine ne bi zavisila od translatornog pomeranja kamere. Osim toga, relativno je lako zadržati istaknute tačke udaljene pozadine u vidnom polju kamere, što ove tačke čini pogodnim za vizuelno određivanje i održavanje zadate orijentacije kamere.

Osim što pruža dobru osnovu za merenje orijentacije, značaj udaljene pozadine leži i u tome što su detekcija i praćenje objekata na velikoj udaljenosti jednostavniji u odnosu na bliske objekte. Ovo dalje pruža osnov za efikasniju ekstrakciju pozadinske slike, posle čije eliminacije na slici ostaju bliski objekti koji su obično od značaja u planiranju kretanja i izbegavanju prepreka kod mobilnih robota.

Ekstrakcija i eliminacija pozadinske slike je dobro razrađena oblast, obzirom da je od značaja u nizu aplikacija kompjuterske vizije, na primer u zadacima nadzora [3]. Međutim, ove aplikacije obično se ograničavaju na slučaj stacionarne ili eventualno kvazistacionarne kamere. Nasuprot tome, u ovom radu se tretira opštiji slučaj u kome se pretpostavlja brza dinamika kretanja kamere, što je uobičajeno za niz mobilnih robota.

Postupak ekstrakcije pozadinske slike razmatran u ovom radu bazira se na korišćenju monokularne vizije. Moguće su i alternativne tehnike, na primer korišćenje dodatnog žiroskopa za određivanje orijentacije kamere i zatim korišćenje nađene promene orijentacije za poravnavanje susednih slika u sekvenci u cilju nalaženja segmenata s najboljim poklapanjem. Mogućnost ekstrakcije pozadine isključivo korišćenjem kamere, bez drugih

senzora, zanimljiv je zbog toga što ne zahteva dodatno investiranje u druge senzore.

Treba primetiti da u ovakvim zadacima stereo kamere ne predstavljaju ozbiljnu prednost u odnosu na monokularne kamere, u svakom slučaju ne u navigaciji na otvorenom prostoru, obzirom da je paralaksa koja se može detektovati pomoću stereo kamere normalno znatno manja u odnosu na promene koje se dobijaju s moguće udaljenih položaja tokom kretanja kamere.

Udaljene tačke mogu da se detektuju i korišćenjem *a priori* znanja o robotskom okruženju (na primer, paralelne linije puta seku se u tački u beskonačnosti). Međutim, ovakav pristup primenljiv je samo u specifičnim situacijama, nasuprot pristupu zasnovanom na kretanju koji je opšti po prirodi.

Detekcija i praćenje pozadinske scene od primarnog je značaja kod kretanja na otvorenom prostoru. Osim toga, predloženi algoritam je posebno zamišljen za rad u uslovima u kojima je kretanje kamere podvrgnuto jakoj dinamici, pri čemu je dinamika rotacije znatno veća u odnosu na translaciju. Ovo je tipična situacija kod kamere montiranih na humanoidne i druge hodajuće robote ili platforme namenjene za kretanje po teško pristupačnom terenu.

Udaljena pozadina se modelira u ovom radu kao skup ključnih tačaka koje se odlikuju po tome što se unapred definišane i fiksna okolina ključne tačke na slici razlikuje na neki način i ističe u odnosu na okoline ostalih tačaka slike. Ključne tačke detektuju se u dva koraka: u prvom koraku se za svaku tačku slike po određenom algoritmu izračunava odziv uzimajući u obzir karakteristike (intenzitet) same tačke i njoj okolnih tačaka u fiksnoj, unapred definisanoj okolini date tačke, pri čemu se okolina selektuje korišćenjem kvadratne ili cirkularne maske. U drugom koraku izdvajaju se lokalni maksimumi, obično tzv. tehnikom potiskivanja nemaksimuma. Konačno, za svaku od detektovanih ključnih tačaka izračunava se deskriptor koji se dalje koristi za uparivanje ključnih tačaka nađenih u različitim slikama (odnosno uzastopnim kadrovima dobijenim s kamere). Tehnike detekcije ključnih tačaka, izbor pogodnih deskriptora i algoritmi uparivanja predmet su intenzivnih istraživanja zadnjih desetak godina i predložen je niz efikasnih detektora s različitim performansama i oblastima primene. Među bitne atribute detektora spadaju afina invarijantnost i invarijantnost po skali [6], koje se često postižu po cenu sniženja računске efikasnosti. Jedna važna karakteristika tačaka daleke scene je upravo da ovi i neki drugi teško ostvarivi kriterijumi nisu od značaja, što pruža mogućnost korišćenja brzih detektora kao što je FAST [8] i računski efikasnih deskriptora kao što su ORB [9] i FREAK [1].

Udaljena pozadina je zapravo aproksimacija idealnih objekata na beskonačnom odstojanju. Sledeći ovu činjenicu, u ovom radu koristi se stepen članstva ključne tačke kao pokazatelj u kom obimu prividno kretanje razmatrane tačke slike može da se tretira kao projekcija stacionarne tačke u beskonačnosti. Dodeljivanje stepena članstva praćenim tačkama je ujedno centralna ideja predloženog algoritma. Jednom estimirani, stepeni članstva mogu da se koriste za poboljšanje estimacije orijentacije robota. Za uzvrat, estimirana orijentacija se u sledećem ciklusu rada koristi za ažuriranje stepena članstva.

U okviru naredne sekcije detaljnije je opisan princip rada algoritma ekstrakcije i praćenja tačaka. U trećoj sekciji opisan je eksperimentalni sistem koji je realizovan u cilju provere predloženog postupka i čiji je rad testiran na više video sekvenci dobijenih na otvorenom prostoru. U zaključku su razmotreni uočeni nedostaci i predložen je dalji rad na poboljšanju metoda.

2. EKSTRAKCIJA I PRAĆENJE TAČAKA

U kontekstu ovog rada, udaljena pozadina je naprosto skup tačaka u beskonačnosti. Tačka u beskonačnosti pogodno može da se opiše korišćenjem jediničnog vektora $\mathbf{e} = [e_x; e_y; e_z]^T$ zraka koji polazi iz centra kamere i prolazi kroz datu tačku (primetimo da komponente vektora \mathbf{e} jedinstveno mogu da se odrede na osnovu normalizovanih koordinata otiska tačke u ravni slike kad god je tačka u vidnom polju kamere). Kad je tačka vidljiva iz dva položaja kamere koji su povezani matricom rotacije \mathbf{R} onda jedinični vektori položaja \mathbf{e} , \mathbf{e}' koji odgovaraju redom prvom i drugom položaju kamere zadovoljavaju:

$$\mathbf{e}' = \mathbf{R} \cdot \mathbf{e} \quad (1)$$

Kombinovanjem ove relacije za tri tačke u opštem međusobnom položaju dobijamo:

$$\begin{bmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 & \mathbf{e}'_3 \end{bmatrix} = \mathbf{R} \cdot \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \quad (2)$$

odnosno:

$$\mathbf{E}'_{3 \times 3} = \mathbf{R} \cdot \mathbf{E}_{3 \times 3} \quad (3)$$

gde su $\mathbf{E}_{3 \times 3} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3]$ i $\mathbf{E}'_{3 \times 3} = [\mathbf{e}'_1 \ \mathbf{e}'_2 \ \mathbf{e}'_3]$. Oдавde matrica relativne rotacije može da se izračuna kao:

$$\mathbf{R} = \mathbf{E}'_{3 \times 3} \cdot \mathbf{E}_{3 \times 3}^{-1} \quad (4)$$

Relacija (4) takođe može da se iskoristi kao test da li tri praćene tačke u generalnom položaju pripadaju udaljenoj pozadini, pod uslovom da je kamera između dva položaja posmatranja izvršila i neki translatorni pomeraj: za proveru je dovoljno da se utvrdi da li je matrica dobijena relacijom (4) ortogonalna. Osim toga, ako je na raspolaganju estimacija matrice rotacije, onda već i relacija (1) može da se iskoristi za proveru da li izolovana tačka pripada udaljenoj pozadini.

Tačnost određivanja orijentacije može da se poveća ako se umesto tri tačke iskoristi veći broj tačaka. Ako su na raspolaganju podaci o n tačaka u beskonačnosti, relacija (2) može da se proširi u:

$$\begin{bmatrix} \mathbf{e}'_1 & \dots & \mathbf{e}'_n \end{bmatrix} = \mathbf{R} \cdot \begin{bmatrix} \mathbf{e}_1 & \dots & \mathbf{e}_n \end{bmatrix} \quad (5)$$

Smenom $\mathbf{E} = [\mathbf{e}_1 \ \dots \ \mathbf{e}_n]$ i $\mathbf{E}' = [\mathbf{e}'_1 \ \dots \ \mathbf{e}'_n]$, dobija se matricna jednačina:

$$\mathbf{E}' = \mathbf{R} \cdot \mathbf{E} \quad (6)$$

koja, uz uslov da je rang matrica \mathbf{E} i \mathbf{E}' jednak 3, može da se reši metodom najmanjih kvadrata: množenjem s \mathbf{E}^T dobija se:

$$\mathbf{E}' \cdot \mathbf{E}^T = \mathbf{R} \cdot \mathbf{E} \cdot \mathbf{E}^T \quad (7)$$

odnosno, u razvijenom obliku:

$$\sum_i \mathbf{e}'_i \cdot \mathbf{e}_i^T = \mathbf{R} \cdot \sum_i \mathbf{e}_i \cdot \mathbf{e}_i^T \quad (8)$$

odakle se dolazi do rešenja:

$$\mathbf{R} = \left(\sum_i \mathbf{e}'_i \cdot \mathbf{e}_i^T \right) \cdot \left(\sum_i \mathbf{e}_i \cdot \mathbf{e}_i^T \right)^{-1} \quad (9)$$

Relacije (8)–(9) strogo važe samo ako se primene na tačke u beskonačnosti. Međutim, one će ostati u važnosti i ako se sabiranje (8) vrši po *svim* praćenim tačkama uz dodavanje skalarnih koeficijenata $w_i \in \{0,1\}$ koji pokazuju da li odgovarajuća tačka pripada udaljenoj pozadini:

$$\sum_i w_i \cdot \mathbf{e}'_i \cdot \mathbf{e}_i^T = \mathbf{R} \cdot \sum_i w_i \cdot \mathbf{e}_i \cdot \mathbf{e}_i^T \quad (10)$$

$$\mathbf{R} = \left(\sum_i w_i \cdot \mathbf{e}'_i \cdot \mathbf{e}_i^T \right) \cdot \left(\sum_i w_i \cdot \mathbf{e}_i \cdot \mathbf{e}_i^T \right)^{-1} \quad (11)$$

I više od toga, relacija (11) ostaje validno rešenje najmanjih kvadrata za *bilu koje* vrednosti $w_i \in (0,1]$ dodeljene tačkama u beskonačnosti, pod uslovom da su ostalim tačkama (izvan beskonačnosti) dodeljeni težinski koeficijenti nula [Relacija (10), obzirom da je dobijena sumiranjem relacija oblika (1), trivijalno važi za proizvoljne vrednosti w_i .]

S druge strane, tačke u beskonačnosti su samo korisna idealizacija stvarnih 3D tačaka. Kao posledica, ni (9) ni (11) u stvarnosti ne dovode do egzaktno matrice rotacije. Umesto toga, rezultujuća matrica je deformisana (tj. kolone dobijene matrice u opštem slučaju ne moraju da budu jedinični međusobno ortogonalni vektori) i potrebno je da se iz nje ekstrahuje rotacioni deo. U računarskoj viziji, standardne procedure za ekstrakciju rotacionog dela su QR dekompozicija ili SVD dekompozicija [4,10]. U kontekstu ovog rada, SVD je preferentno rešenje s obzirom da daje bolju osnovu za estimaciju deformacija superponiranih na matricu rotacije. Dekompozicija ima oblik:

$$\mathbf{R} = \mathbf{L} \cdot \mathbf{D} \cdot \mathbf{U}^T \quad (12)$$

gde su \mathbf{L} i \mathbf{U} matrice rotacije a \mathbf{D} je dijagonalna. Oдавde, za matricu rotacije može da se usvoji estimacija:

$$\hat{\mathbf{R}} = \mathbf{L} \cdot \mathbf{U}^T \quad (13)$$

pri čemu dijagonalni elementi \mathbf{D} služe kao mera deformacija istezanja/sabijanja duž odgovarajućih osa. Za dobru estimaciju, svi dijagonalni elementi moraju da budu bliski jedinici.

Adekvatnim izborom težinskih koeficijenata poboljšava se estimacija orijentacije. Tačkama koje bolje aproksimiraju tačke u beskonačnosti treba da se pridruže veće vrednosti koeficijenata čime se pojačava njihov doprinos rezultujućoj estimaciji rotacije. Pritom, pojam „bolje aproksimacije“ prvenstveno znači zadovoljavajući rezultat testova (1) i (4) tokom dužeg vremena.

Predloženi sistem praćenja ostvaruje se tako što se u svakom radnom ciklusu ažurira model udaljene pozadine i estimacija tekuće orijentacije kamere u skladu s prividnim kretanjem ključnih tačaka identifikovanih u sekvenci slika dobijenih kamerom. U najjednostavnijem obliku, model udaljene pozadine je naprosto skup trojki $(w_i, \mathbf{e}_i, \mathbf{d}_i)$ pridruženih praćenim ključnim tačkama, gde

w_i označava stepen pripadnosti, \mathbf{e}_i reprezentuje zrak duž koga je ključna tačka detektovana po prvi put, a \mathbf{d}_i je deskriptor koji reprezentuje vizuelne karakteristike fragmenta slike u okolini ključne tačke. Inicijalno, u modelu mogu da se unesu sve ključne tačke identifikovane u polaznom kadru uz postavljanje koeficijentata članstva na neku neutralnu vrednosti, recimo 0.5.

Pošto se na osnovu analize novog kadra s kamere ustanovi korespondencija između tačaka modela i novodektovanih ključnih tačaka, estimacija rotacije kamere dobija se iz (11)–(13). Ako je procenat korektnih korespondencija dovoljno visok, estimacija će biti bliska stvarnoj rotaciji tako da se može iskoristiti relacija (1) za izračunavanje uglova između očekivanih $\hat{\mathbf{R}} \cdot \mathbf{e}$ i detektovanih orijentacija \mathbf{e}' uparenih ključnih tačaka:

$$\varphi_i = \angle(\mathbf{e}', \hat{\mathbf{R}} \cdot \mathbf{e}) \quad (14)$$

Distribucija ugaonih devijacija φ_i može pritom da se iskoristi za uklanjanje izbačaja (eng. outliers, rus. выбросы) iz očekivane distribucije, tj. odbacivanje svih uparivanja za koje je napr. dobijena ugaona devijacija znatno veća (u praksi 2 do 3 puta) od izračunate vrednosti standardne devijacije.

Poslednji važan korak radnog ciklusa sastoji se u ažuriranju koeficijentata pripadnosti. Važno je da se ovaj korak vrši *samo* onda kad je distribucija ugaonih odstupanja (14) dovoljno široka, tj. kad je standardna devijacija σ_φ veća od neke unapred propisane granice φ_w :

$$\sigma_\varphi > \varphi_w \quad (15)$$

Drugim rečima, ažuriranje stepena članstva treba da se vrši tek onda kad je kretanje kamere dovoljno bogato da može da se napravi razlika između bliskih i udaljenih ključnih tačaka. Kad je ovaj preduslov ispunjen, potrebno je dodatno ustanoviti kriterijum za razgraničenje između bliskih i udaljenih tačaka. Granica je po prirodi postepena i zavisi od niza faktora kao što su priroda robotskog zadatka i brzina kretanja robota.

U jednom jednostavnom obliku, stepen članstva može da se implementira kao kontinualna opadajuća funkcija ugaonog odstupanja. U eksperimentalnoj implementaciji opisanoj u narednoj sekciji rada korišćena je funkcija:

$$w_i = \frac{1}{1 + (\varphi_i / \varphi_0)^2} \quad (16)$$

gde φ_0 predstavlja prag odstupanja φ_i za koje se dobija $w_i = 0,5$. Inicijalni testovi pokazali su da rezultati praćenja ne zavise mnogo od oblika funkcije stepena članstva. Nasuprot tome, pokazala se velika osetljivost na vrednosti pragova φ_w i φ_0 .

3. EKSPERIMENTALNA IMPLEMENTACIJA

Implementacija je ostvarena u okruženju OpenCV [2] korišćenjem radnog okvira 2D Features [7]. Ekstrakcija ključnih tačaka realizovana je pomoću detektora FAST [8], danas široko prihvaćenog kao najsavremenijeg, pri čemu je deskriptor SIFT [5] iskorišćen za uparivanje ključnih tačaka. Iako daleko od najbržeg, SIFT je izabran

jer je za njega demonstrirano da nadmašuje druge deskriptore u pogledu obima potiskivanja izbačaja [6].

Algoritam praćenja primenjen je na nekoliko video sekvenci dobijenih na otvorenom prostoru pomoću jeftine kamere ugrađene u mobilni telefon. Kamera generiše komprimovani MPEG-4 CIF video niske rezolucije s 352×288 piksela po slici pri brzini od 15 kadrova u sekundi. Osim šuma uvedenog kompresijom, izlazne slike ove kamere dodatno su zašumljene usled kašnjenja CMOS senzora.

Sl. 1 prikazuje nekoliko kadrova iz jedne tipične sekvence. Prikazane su slike dobijene u vremenskim intervalima od jedne sekunde i preko njih su ucrtani kružići koji markiraju detektovane udaljene ključne tačke s stepenima članstva jednakim ili većim 0.5; dodatno, u cilju ilustracije dobijenih estimacija rotacije, u centru svake slike ucrtane su koordinatne ose sistema koji odgovara inicijalnoj orijentaciji kamere.

Izborom odgovarajućih parametara detektora FAST broj praćenih tačaka održavan je između 200 i 500. Po inicijalnom uparivanju tačaka korišćenjem deskriptora SIFT, parovi s ugaonim odstupanjima φ_i većim od $2\sigma_\varphi$ označeni su kao izbačaji. Rezultati prikazani na sl. 1 dobijeni su za vrednost praga φ_w jednakom 10% vidnog polja kamere, dok je prag φ_0 postavljen na $\varphi_w/2$.

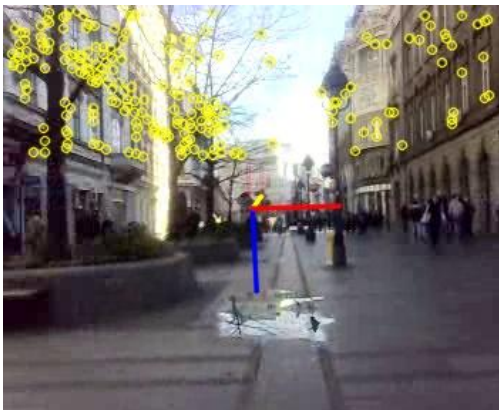
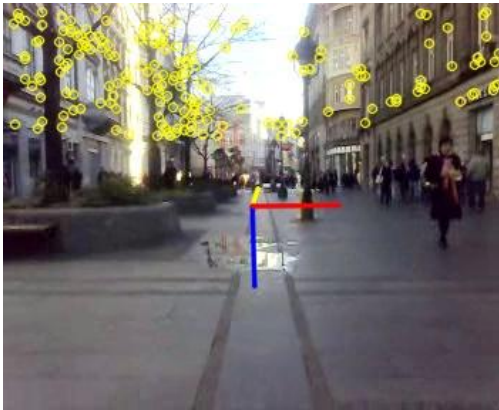
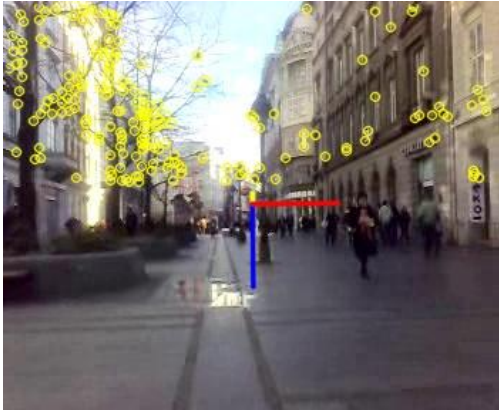
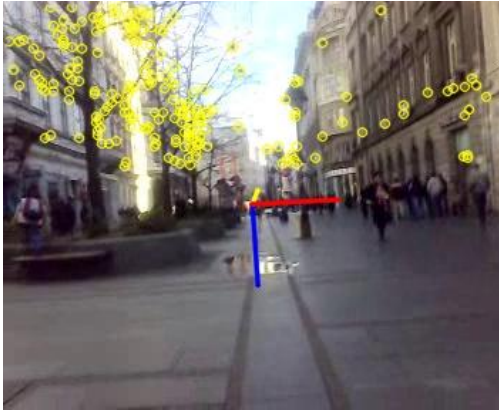
Dobijeni rezultati izgledaju zadovoljavajući uprkos značajnom šumu kamere. Međutim, detaljnijim uvidom otkriva se nestabilnost praćenih tačaka koja je izvan domena predloženog mehanizma praćenja. Nestabilnost je rezultat nemogućnosti čistog deskriptorskog uparivanja da prepozna korektnne parove u prisustvu šuma. Kao posledica, lanac ka ranije detektovanim ključnim tačkama se gubi i ova nestabilnost dovodi do klizanja u estimaciji orijentacije koja je katkad vidljiva u nekim video sekvencama.

4. ZAKLJUČAK

Inicijalni testovi pokazuju da je s predloženim algoritmom ostvarivo efikasno praćenje udaljenih tačaka uz dobru estimaciju orijentacije što dalje ukazuje da bi se njegovom ugradnjom u sisteme vizuelne navigacije moglo postići poboljšanje njihovih performansi. Međutim, za punu operacionalizaciju pristupa potreban je dodatni rad. Prvo, potrebno je ostvariti robusnije uparivanje tačaka. Jedan način da se ovo postigne je da se estimacija orijentacije kamere čvršće integriše u algoritam uparivanja, umesto da služi kao prost spoljašnji kriterijum za statističko odbacivanje izbačaja. Ovo bi takođe moglo da rezultuje u manjoj zavisnosti od složenih deskriptora ključnih tačaka što bi vodilo računski efikasnijoj implementaciji. Drugi smer bio bi u razvoju složenijih kriterijuma utvrđivanja pripadnosti udaljenoj pozadini što bi omogućilo tačniju estimaciju i bolje pokrivanje pozadinskih regiona.

ZAHVALNICA

Ovaj rad potpomognut je od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije kroz projekte III-44008 i TR-35003.



Slika 1: Tačke udaljenje pozadine i estimirana orijentacija u kadrovima primera video sekvence

LITERATURA

- [1] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition*, Providence, USA, June 2012, pp. 510-517.
- [2] G.R. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*.: O'Reilly Media, 2008.
- [3] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino, "Background subtraction for automated multisensor surveillance: A comprehensive review," *EURASIP J. Advances in Signal Processing*, vol. 2010, pp. 43:1-43:24, February 2010.
- [4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed.: Cambridge University Press, 2004.
- [5] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91-110, November 2004.
- [6] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, October 2005.
- [7] OpenCV Dev Team. (2012, Juli) OpenCV v2.4.2 Documentation, 2D Features Framework. [Online]. <http://docs.opencv.org/modules/features2d/doc/features2d.html>
- [8] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. 2006 European Conf. Computer Vision*, Graz, May 2006.
- [9] E. Rublee, V. Rabaud, K. Konolige, and G.R. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. 2011 IEEE Int. Conf. Computer Vision*, Barcelona, Spain, November 2011, pp. 2564-2571.
- [10] R. Szeliski, *Computer Vision: Algorithms and Applications*.: Springer, 2011.

EXTRACTION AND TRACKING OF FAR BACKGROUND KEYPOINTS IN DYNAMIC MONOCULAR VISION

ABSTRACT

Tracking of points lying in far background is of multiple importance in visual navigation, especially in tasks assuming fast camera motion dynamics. This paper describes an algorithm for efficient identification and tracking of such points using single monocular camera. The algorithm is based on a model of far background consisting of a set of tracked keypoints and associated membership degrees. The effectiveness of the proposed approach is demonstrated on open air video sequences. Some restrictions and further work necessary to operationalize the method are also considered.