

DETEKCIJA, LOKALIZACIJA IN IDENTIFIKACIJA OSEB Z VEČ KAMERAMI TER MAPAMI ZNAČILNIC

Rok Mandeljc, Stanislav Kovačič, Matej Kristan, Janez Persš

Laboratorij za strojni vid
Fakulteta za elektrotehniko, Univerza v Ljubljani
E-pošta: rok.mandeljc@fe.uni-lj.si
URL: <http://vision.fe.uni-lj.si>

POVZETEK: V članku je predstavljen sistem za detekcijo, lokalizacijo in identifikacijo oseb v posameznih trenutkih, brez časovnega filtriranja, ki je prisotno v večini tovrstnih sistemov. Glavni cilj predstavljenega pristopa je odpravljanje katastrofalnih napak, ki onemogočajo popolnoma samodejno obdelavo realistično dolgih video posnetkov. Sistem temelji na zlivanju (fuziji) več šibkih značilnic, zapisanih v obliki map značilnic, zlivanje pa je izvedeno s pomočjo enega ali več naučenih razvrščevalnikov.

1. UVOD

Detekcija, določitev položaja in identitet oseb v prostoru predstavljajo znaten izziv in so zato predmet obsežnih raziskav, med drugim tudi na področju računalniškega vida. Potreba po tovrstnih podatkih se pojavlja na različnih področjih, najizraziteje pri analizah v športu ter v športni medicini, pa tudi v različnih varnostnih in video-nadzornih aplikacijah, na primer pri analizah obnašanja ter rekonstrukcijah dogodkov.

Kljub temu, da ima raziskovalno področje detekcije in sledenja s pomočjo ene ali večih video kamer precej bogato raziskovalno tradicijo in ponuja znaten nabor obstoječih metod, pa zaenkrat še ne ponuja rešitve, ki bi omogočala avtonomno (nenadzorovano) obdelavo realistično dolgih video posnetkov, npr. celotne rokometne ali košarkarske tekme.

Obstoječe metode lahko v grobem razdelimo v dve skupini. Sekvenčne metode (npr. [4,3]) uporabljajo Kalmanov filter ali splošnejši filter delcev, zato lahko pride do nepopravljivih napak v primeru slabe detekcije oseb oziroma v primeru, ko algoritem nepravilno poveže več zaporednih detekcij. V primeru odpovedi algoritma je potreben ročen poseg operaterja, zaradi česar mora ta stalno nadzirati potek sledenja.

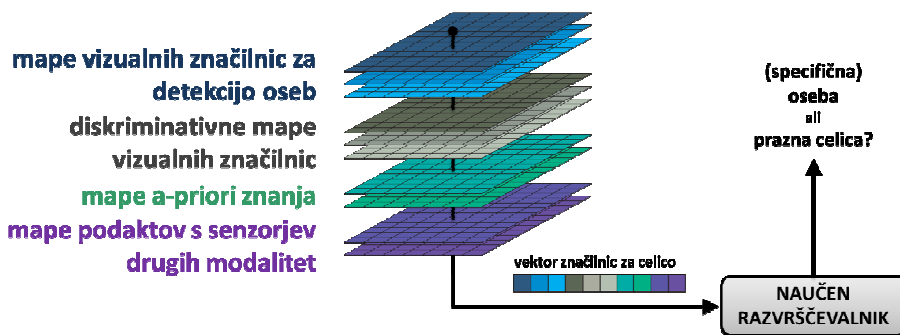
Kot odgovor na pomanjkljivosti sekvenčnih metod so se v zadnjem času uveljavile metode, ki temeljijo izključno na detekciji (npr. [2,1,6]); detekcija oseb je v vsakem trenutku izvedena neodvisno od prejšnjih trenutkov, dobljeni položaji pa so nato povezane s pomočjo robustnih metod iz teorije grafov. Tako posamezne napake v detekcijah niso več katastrofalne. Po drugi strani pa te metode temeljijo na širjenju

identitet posameznikov skozi čas, zato v primeru, ko ljudje pridejo blizu skupaj, lahko pride do zamenjave njihovih identitet. Posledica tega je spet katastrofalna napaka, ki ponovno zahteva poseg operaterja. Ker obstoječe metode ne omogočajo samodejne *detekcije pojava napak*, morajo biti algoritmi stalno pod nadzorom operaterja, kar močno omejuje njihovo učinkovitost.

V Laboratoriju za strojni vid smo zato začeli razvijati sistem, ki omogoča sočasno detekcijo, lokalizacijo in identifikacijo posameznikov v posameznih trenutkih, in tako naslavlja prej omenjene omejitve obstoječih pristopov. Oddaljitev od tradicionalnega sledenja tako po eni strani odpravlja možnost pojava in širjenja katastrofalnih napak, vendar pa nas po drugi strani sili, da uporabimo precej več razpoložljive informacije oziroma značilnic, ki jih lahko pridobimo s slik. Združevanje razpoložljive informacije poteka znotraj ogrodja, ki ga bomo predstavili v nadaljevanju.

2. OPIS SISTEMA

Naš sistem uporablja več video kamer in temelji na uporabi številnih šibkih značilnic, s pomočjo katerih detektiramo in identificiramo posameznike. Zaradi uporabe več kamer razpoložljive značilnice zapišemo v obliki *map značilnic* (angl. »feature maps«), ki jih, kot je prikazano na Sliki 1, združimo s pomočjo naučenega razvrščevalnika – v našem primeru z metodo podpornih vektorjev (angl. »Support Vector Machine – SVM«).



Slika 1: shematičen prikaz delovanja našega sistema

2.1 Šibke značilnice

Za razliko od obstoječih metod računalniškega vida smo ljudje sposobni razpoznati ljudi tudi v težavnih okoljih in razlikovati med posamezniki celo v uniformiranih okoljih, kot je na primer šport.

V sodelovanju z raziskovalci Fakultete za šport Univerze v Ljubljani, ki so glavni uporabniki metod sledenja oseb (igralcev) z video kamerami, smo ugotovili, da si pri razpoznavanju igralcev pomagajo s številnimi lastnostmi in podrobnostmi, kot so barva

oblačil, telesna višina in širina, barva kože, barva las, frizura, obrazne poteze, način gibanja, položaj na igrišču, itd. Znano je, da se ljudje tudi pri sami detekciji oseb zanašamo na različne značilnice, npr. izstopanje iz ozadja, gibanje in obliko.

Tovrstne značilnice so *šibke*, saj same po sebi ne omogočajo detekcije oziroma razločevanja med posamezniki; slednje je možno šele s kombinacijo večjega števila šibkih značilnic.

2.2 Mape značilnic

Koncept *map značilnic* predstavlja posplošitev koncepta *map zasedenosti* (angl. »occupancy map«). *Mape zasedenosti* uporabimo tako, da opazovano območje diskretiziramo v mrežo in nato za vsako od celic ocenjujemo verjetnost, da je zasedena (torej, da je se v njej nahaja oseba). To storimo na podlagi informacije z več kamer, običajno v obliki binarnih slik, ki jih pridobimo s postopkom segmentacije ozadja.

Najnaprednejša metoda za ocenitev *map zasedenosti* je t.i. algoritem POM (»*Probabilistic Occupancy Map*« – *verjetnostna mapa zasedenosti*) [2]. Postopek je iterativen, in uporablja vzvratno projekcijo z modelom, ki aproksimira silhuete posameznikov s pravokotniki fiksne višine; za vsako zasedeno celico na ustrezna mesta v vhodnih slikah projicira pravokotnike in optimira mapo zasedenosti tako, da minimizira razliko med vhodnimi binarnimi slikami in reprojiciranimi pravokotniki.

Predlagane *mape značilnic* so podobne mapam zasedenosti, le da vsaka celica namesto verjetnosti zasedenosti vsebuje vrednost značilnice. Takšno kodiranje značilnic omogoča enostavno združevanje informacije iz večih kamer, saj lahko z upoštevanjem geometrijskih omejitev občutno zmanjšamo šum.

Znotraj našega sistema predvidevamo uporabo različnih vrst map značilnic:

- mape vizualnih značilnic za detekcijo oseb
- diskriminativne mape vizualnih značilnic
- mape a-priori znanja
- mape z informacijo s senzorjev drugih modalitet

Mape vizualnih značilnic za detekcijo oseb služijo razlikovanju oseb od ozadja. Pravzaprav gre za več map zasedenosti, pridobljenih z algoritmom POM in različnimi postopki obdelave vhodnih slik, s katerimi izluščimo različne značilnice – segmentacija ozadja [8], gosti optični tok [7] in detektor zgornjega dela telesa [5]. Za vsako od modalitet konstruiramo več map z različnimi višinami pravokotnikov v algoritmu POM (npr. 1.60 m, 1.75 m in 1.85 m), s čimer implicitno zakodiramo tudi diskriminativno informacijo o višinah posameznikov.

Diskriminativne mape vizualnih značilnic so namenjene razločevanju med posamezniki; posledično ima vsak posameznik lasten nabor takšnih map. Diskriminativne značilnice so npr. barva majice, barva hlač, barva las, tekstura las, obrazne poteze, telesna višina in

širina, itd. Trenutno uporabljamo samo barvo majic, ki jo za vsako osebo zapišemo v obliki mape razdalj do referenčnega barvnega histograma.

V nekaterih primerih imamo na voljo tudi a-priori znanje, ki ga lahko uporabimo pri razločevanju posameznikov. Primer je npr. znanje o verjetnih položajih igralcev, ki so v nekaterih športih tesno povezani z njihovimi vlogami. Tudi to informacijo lahko zapišemo v obliki mape značilnic in uporabimo znotraj našega sistema.

Prek zapisa v mapo značilnic naš sistem omogoča tudi vključitev informacije s senzorjev drugih modalitet, npr. radijskega sistema za sledenje [9].

2.3 Zlivanje (fuzija) map značilnic

Bistvena ideja našega pristopa je, da zberemo čim večje število map šibkih značilnic. Uporabnost posameznih map je odvisna od situacije – nekatere mape so v nekaterih situacijah le delno uporabne, nekatere pa so povsem neuporabne. Zato se sistem v fazi učenja sam nauči, katere mape so pomembne in katere ne. To dosežemo z učenjem razvrščevalnika, ki ga nato uporabimo za zlivanje map. Zlivanje izvedemo tako, da za določeno celico v prostoru zberemo vrednosti značilnic iz posameznih map, ki jih združimo v en vektor značilnic, za razvrščanje pa uporabimo metodo podpornih vektorjev (Slika 1).

Za potrebe učenja mora operater označiti učni del posnetkov, ki naj bo ustrezno reprezentativen. Če želimo izvesti samo detekcijo, na učnih posnetkih učimo en sam razvrščevalnik, če pa želimo izvesti še identifikacijo, pa naučimo en razvrščevalnik za vsako osebo. Tako naučene razvrščevalnike nato uporabimo za obdelavo preostalega posnetka. S tem razbremenimo operaterja, ki mora označiti samo del posnetka, preostanek pa lahko obdelamo brez nadzora, saj napaka v določenem trenutku nima nobenega vpliva na delovanje v naslednjih trenutkih.

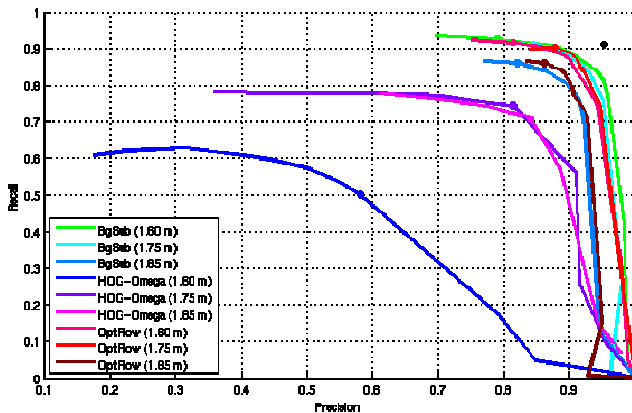
3. REZULTATI

Delovanje našega sistema smo preverili s postavitvijo štirih kamer v našem laboratoriju in snemanjem petih oseb, ki so hkrati hodile po prostoru.

V prvem delu eksperimenta smo z uporabo večih map za detekcijo in enega samega razvrščevalnika poskusili izboljšati detekcijo oseb, brez identifikacije. Slika 2 prikazuje rezultat; prikazane so krivulje preciznosti-priklica (angl. »*precision-recall*«) posameznih map, pridobljenih z algoritmom POM in spreminjanjem njegovih parametrov. Pike na krivuljah predstavljajo delovne točke, ki smo jih izbrali za zlivanje map, črna pika pa predstavlja rezultat zlivanja – kot lahko vidimo, z uporabo večih značilnic za detekcijo občutno izboljšamo uspešnost detekcije.

V drugem delu smo za detekcijo in identifikacijo uporabili tako mape za detekcijo kot tudi diskriminativno mapo razdalj do referenčnega barvnega histograma ter zelo šibko mapo a-priori znanja o položajih (dve od oseb sta se gibali samo po vnaprej določenem

delu prostora). Rezultat je prikazan v Tabeli 1; ker tri osebe nosijo črne majice, dve pa sivi majici (Slika 3), uporabljene diskriminativne značilnice ne omogočajo povsem zanesljivega razlikovanja med njimi. Vendar pa zaradi uporabe šibke mape a-priori znanja ter implicitno kodirane višine ljudi v mapah za detekcijo dosežemo večjo preciznost kot v primeru, ko bi ne bi imeli na voljo nobenih diskriminativnih značilnic oziroma če bi uporabili samo barvo majic, kar je razvidno iz vrednosti preciznosti v Tabeli 1.



Slika 2: rezultat detekcije z zlivanjem večih map vizualnih značilnic za detekcijo. Prikazane so krivulje preciznosti-priklica za posamezne mape, pridobljene z algoritmom POM in različnimi vrstami informacije (segmentacija ozadja – *BgSub*, optični tok – *OptFlow* in detekcija zgornjega dela telesa – *HOG-Omega*) ter različnimi višinami pravokotnikov. Črna pika prikazuje rezultat zlivanja, ki je boljši od vseh posameznih značilnic.

Tabela 1: vrednosti priklica in preciznosti za vsako od oseb, ki so sodelovale v eksperimentu. Za vsako od oseb sta poleg doseženih vrednosti navedeni še pričakovani vrednosti preciznosti v primeru, ko ne bi uporabili nobene diskriminativne značilnice ter v primeru, ko bi uporabili samo barvo majic ob predpostavki, da ni napak zaradi zakrivanja med posamezniki. Predvsem v primeru osebe #1 s pomočjo dodatnih šibkih značilnic dosežemo znatno višjo preciznost.

Oseba	Priklic [%]	Preciznost [%]	Naključna preciznost (brez disk. zn.) [%]	Naključna preciznost (samo barve) [%]
#1	76,2	80,3	20	33,3
#2	76,2	33,0	20	33,3
#3	71,4	58,7	20	50
#4	93,9	55,2	20	33,3
#5	76,2	46,0	20	50



Slika 3: osebe, ki so sodelovale v eksperimentu. Razločevanje med njimi samo na osnovi barve majic je težavno, saj tri osebe nosijo črne, dve pa sivi majici.

4. ZAKLJUČEK

V članku smo predstavili sistem za sočasno detekcijo, lokalizacijo in identifikacijo oseb z več kamerami. Posebnost predstavljenega sistema je, da ne deluje na principu sledenja (časovnega filtriranja), pač pa sta tako detekcija kot identifikacija izvedeni v vsakem trenutku neodvisno od prejšnjih trenutkov. S tem preprečimo pojav katastrofalnih napak, ki so prisotne v tradicionalnih sistemih za sledenje; tovrstne napake zaradi širjenja od trenutka nastopa dalje predstavljajo glavno oviro pri izvedbi sistemov, ki bi omogočali povsem avtonomno obdelavo realistično dolgih video posnetkov (npr. celotne tekme).

Slabost predstavljenega pristopa je, da zaradi odsotnosti komponente sledenja ne zagotavlja povsem zanesljive identifikacije oseb. Ker pa je njegova napaka omejena na posamezne trenutke, ga lahko uporabimo v kombinaciji s tradicionalnimi metodami sledenja, in sicer za preverjanje in ponovno vzpostavitev identitet.

LITERATURA

1. J. Berclaz, F. Fleuret, E. Turetken, P. Fua (2011), Multiple object tracking using K-Shortest paths optimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, str. 1806-1819.
2. F. Fleuret, J. Berclaz, R. Lengagne, P. Fua (2007), Multicamera people tracking with a probabilistic occupancy map, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, str. 267-282.
3. M. Kristan, S. Kovačič, A. Leonardis, J. Perš (2010), A Two-Stage Dynamic Model for Visual Tracking, *Systems, Man, and Cybernetics: Part B*, vol. 40, str. 1505-1520

4. M. Kristan, J. Perš, M. Perše, S. Kovačič (2009), Closed-world tracking of multiple interacting targets for indoor-sports applications, *Computer Vision and Image Understanding*, vol. 113, str. 598-611.
5. M. Li, Z. Zhang, K. Huang, T. Tan (2008), Estimating the number of people in crowded scenes by MID-based foreground segmentation and head-shoulder detection, *Proceedings ICPR 2008*, Tampa, Florida: IEEE, str. 1-4.
6. H. Shitrit, J. Berclaz, F. Fleuret, P. Fua (2011), Tracking multiple people under global appearance constraints, *Proceedings ICCV 2011*, Barcelona: IEEE, str. 137-144.
7. M. Werlberger, T. Pock, H. Bischof (2010), Motion estimation with non-local total variation regularization, *Proceedings CVPR 2010*, San Francisco: IEEE, str. 2464-2471.
8. Z. Zivkovic, F. van der Heijden (2006), Efficient adaptive density estimation per image pixel for the task of background subtraction, *Pattern Recognition Letters*, vol. 27, str. 773-780.
9. <http://www.ubisense.net>
Ubisense Real-Time Location Systems – radijski sistem za sledenje.