ELEKTROTEHNIŠKI VESTNIK 83(5): 217–224, 2016
IZVIRNI ZNANSTVENI ČLANEK

Hierarhični modeli videza v vizualnem sledenju

Luka Čehovin Zajc, Aleš Leonardis, Matej Kristan

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana, Slovenija
E-pošta: luka.cehovin@fri.uni-lj.si

Povzetek. V članku obravnavamo problem kratkoročnega vizualnega sledenja, v okviru katerega predstavljamo koncept hierarhičnih modelov opisa videza objektov. Hierarhični modeli opis videza strukturirajo v več plasti. Najnižja plast vsebuje najbolj specifične informacije o videzu, ki se hitro spreminjajo, višje plasti pa opisujejo videz v trajnejši, posplošeni, obliki. Hierarhična urejenost se odraža tudi v posodabljanju vizualnega modela, kjer višje plasti vodijo posodabljanje nižjih plasti, te pa so v primeru lastne zanesljivosti vir informacij za osveževanje višjih plasti. Koristi hierarhičnega modela sta predstavljeni s povzetkom dveh izpeljank modela v okviru dveh sledilnikov, ki sta namenjeni predvsem sledenju ne-togih in artikuliranih objektov, saj so ti še poseben izziv za večino obstoječih sledilnikov. Prva implementacija je sestavljena iz dveh plasti, druga pa doda še tretjo plast kot odgovor na nekatere pomanjkljivosti prve implementacije. Predstavljena eksperimentalna analiza na obstoječih primerjalnih zbirkah podatkov pokaže, da opisana sledilnika spadata v sam vrh raziskav na področju kratkoročnega vizualnega sledenja ter se še posebej odlikujeta v sledenju netogih objektov.

Ključne besede: računalniški vid, vizualno sledenje, model videza, hierarhije

Hierarchical appearance models in visual tracking

The paper addresses the problem of short-term visual tracking in the scope of which we present a concept of hierarchical models to describe the appearance of an object. The key property of these models is that they structure the appearance description into multiple layers. The lowers layers contain the most specific information that can change quickly, while the higher layers contain the appearance information in a more general and lasting form. The structure is also reflected in the update process where the higher layers are guiding the update process of the lower layers, while the lower layers provide a reliable information for updating the higher layers. The benefits of this hierarchical organization are presented with a summary of two shuch models in two visual trackers that are primarily designed for tracking articulated and non-rigid objects, which present a difficulty for many tracking approaches. The first implementation is composed of two layers, while the second one adds another layer to address several shortcomings of the first implementation. The presented experimental analysis on several established benchmarks shows that the described trackers are comparable to the state-of-the-art and excel in tracking non-rigid objects.

Keywords: computer vision, visual tracking, appearance model, hierarchy

1 UVOD

Vizualno sledenje je pomembno raziskovalno področje v okviru računalniškega vida, katerega glavni cilj vizualnega sledenja je določitev stanja enega ali več objektov v toku slik ob upoštevanju časovne soslednosti le-teh. Algoritme, ki opravljajo nalogo vizualnega sledenja, imenujemo *vizualni sledilniki*, in jih lahko uporabimo na številnih, tako novih kot tudi že uveljavljenih, tehnoloških področij, kot so npr. robotika [35], videonadzorni sistemi [40], [20], interakcija med človekom in računalnikom [5], [22], [18], avtonomna vozila in analiza športa [25]. Zaradi široke palete možnosti uporabe vizualnega sledenja se je razvilo veliko podvrst formalizacije problema, vsaka s svojimi izzivi in predpostavkami. V tem članku obravnavamo tip vizualnega sledenja, kjer sledimo samo enemu objektu v enem samem toku slik, geometrijskih lastnosti objekta ne poznamo vnaprej, predpostavljamo pa tudi, da objekt ne bo nikoli izginil iz opazovanega območja v sliki. Takemu sledenju pravimo *kratkoročno sledenje*. Poleg tega predpostavljamo, da je tok slik potencialno neskončen in ga torej ne moremo shraniti in nato obdelati v celoti z naključnim dostopom do slik. Vizualni sledilniki za dosego cilja naloge uporabljajo različne *modele videza*, ki na različne načine opisujejo videz objekta. Ker se ta tekom sekvence spreminja, je treba model videza posodabljati, to pa je pogosto problem, saj neuspešna posodobitev, ki je lahko rezultat netočne lokalizacije ali toge zasnove vizualnega modela, vodi v počasno spiralo odklona opisa videza objekta od realnega stanja, to pa pripelje do odpovedi sledilnika oziroma *zdrsa*.

V tem članku predstavljamo napredni koncept konstrukcije vizualnega modela, ki temelji na hierarhičnem združevanju vizualnih informacij. Tak način opisa videza daje možnosti za uspešno sledenje v številnih zahtevnih scenarijih, še zlasti pa je primeren za sledenje netogih in artikuliranih objektov. Uporabo hierarhičnega vizualnega modela smo potrdili z razvojem dveh sledilni-

Prejet 18. avgust, 2016
Odobren 12. oktober, 2016

kov [6], [7], ki se glede na empirične primerjave uvrščata v sam vrh raziskav na tem področju. V članku predstavljamo enovit okvir formalizacije hiearhičnih modelov, kamor spadata [6], [7] in eksperimentalno analizo obeh izpeljank. V poglavju 2 najprej predstavljamo raziskovalno področje ter motiviramo naše delo. V poglavju 3 opišemo idejo hierarhičnih modelov videza ter povzamemo podrobnosti obeh izpeljanih modelov videza. V poglavju 4 predstaviljamo eksperimentalne rezultate, v poglavju 5 pa sklepne ugotovitve in ideje za nadaljnje delo.

2 PREGLED PODROČJA

Modele videza lahko razvrstimo glede na tip uporabljenih vizualnih značilnic za opis objekta in glede na način hranjenja ter obdelave informacij o videzu. Najbolj razširjena vrsta modelov so *holistični* modeli videza, ki hranijo monolitno reprezentacijo videza objekta. Taki modeli videz objekta največkrat opisujejo z barvnimi histogrami [9], [23], slikovnimi predlogami [39], [33], [4], [43], obrisi [19] in teksturami [38]. Pogosto uporabljene metode iskanja maksimalnega ujemanja vizualnega modela s sliko uporabljajo sekvenčno jedrno [9] ter optimizacijo Monte-Carlo [36], [23]. V zadnjih dveh desetletjih je postalo popularno sledenje z uporabo diskriminativnih modelov, kar pomeni, da model videza vsebuje klasifikator, ki določi, ali določena regija vsebuje objekt ali ne. Ta klasifikator mora biti med sledenjem sproti osveževan, kar je eden izmed večjih problemov takih pristopov. Ena izmed prvih uspešnih implementacij sledenja z uporabo detekcije je uporabljal kaskadni ojačevalni (*boosting*) klasifikator, prirejen za sprotno osveževanje [14]. Pristop je bil kasneje večkrat razširjen [15], [1], navdihnil pa je tudi druge prostope k integraciji diskriminativne informacije, npr. uporabo strukturiranih podpornih vektorjev [16] in naključnih projekcij [45]. Kljub očitnemu uspehu holističnih modelov videza pa so hitre spremembe strukture objekta še vedno velik izziv. Pri holističnih modelih je namreč celotna reprezentacija videza objekta osvežena naenkrat, kar povečuje verjetnost, da bo pravilen del vizualne informacije pokvarjen z novo informacijo. To se lahko zgodi, ker sledilniku ne uspe določiti pravilnega položaja objekta, kar pomeni, da bo model osvežen z informacijami, ki ne pripadajo objektu, ali ker sledilnik ne uporablja značilnic, ki bi bile v danem scenariju zmožne razločevati objekt od ozadja. Drugi problem holističnih modelov videza je predpostavka, da objekt lahko opišemo s pravokotno regijo v sliki. Čeprav je to smiselna predpostavka v številnih praktičnih primerih (npr. sledenje obrazov ali avtomobilov), obstaja veliko scenarijev, kjer ta predpostavka ne drži, npr. pri netogih in artikuliranih objektih. Vse geometrijske deformacije tarče, ki bi jih lahko upoštevali v geometrijskem okviru, morajo biti v holističnem vizualnem modelu obdelane s korakom osveževanja, kar povečuje možnost zdrsa. En od načinov obravnave nekaterih pomanjkljivosti posameznih holističnih sledilnikov je njihovo združevanje [41], [29], [3], ki izvira iz opažanja, da se posamezni sledilniki v določenih okoliščinah obnašajo dobro in da lahko s pametnim preklapljanjem med njimi izboljšamo njihovo skupno delovanje. A tudi ta pristop dejansko ne naslavlja sledenja netogim objektom, ki se deformirajo in spreminjajo obliko.

Po drugi strani pa je glavna ideja modelov videza, ki temeljijo na več delih, da je videz razdeljen na več lokalnih modelov in povezav med njimi. Vrste lokalnih modelov in oblike povezav se lahko med modeli videza zelo razlikujejo. Primer te vrste modelov videza temelji na množici lokalnih značilnic, ki sledijo z ocenjevanjem optičnega toka [22]. Optični tok je bil uporabljen tudi v [21], kjer se robustne ocene lokalnih premikov združijo v oceno premika z uporabo mediane. Drugi pristop k sledenju z več deli je uporaba stabilnih regij, npr. v [44] avtorji zaznajo stabilne dele in s predpostavljanjem globalne afine transformacije omejijo iskanje ujemanj ter se izognejo zdrsu. V [13] avtorji za sledenje predlagajo uporabo posplošene Houghove transformacije, ta pristop pa je bil kasneje razširjen v [10]. V [37] so uporabljene značilnice SIFT[31], videz objekta pa je predstavljen kot množica značilnic, ki se pogosto pojavijo skupaj. Na splošno je število stabilnih regij odvisno od vizualnih lastnosti specifičnega objekta (npr. jasnosti teksture), to pa neposredno vpliva na uspešnost sledilnika, saj je leta odvisna od števila in ponovljivosti stabilnih regij. Če imamo opravka z barvno homogenimi objekti, značilnice SIFT ne bodo številne in ponovljive, sledilnik pa bo zato neuspešen.

V [11] avtorji obravnavajo problem postavitve delov v sliko kot optimizacijski problem in predlagajo sledenje objektu s pomočjo množice lokalnih jeder, ki so med seboj povezana prek omejitev v obliki afine transformacije. V [32] je globalna afina transformacija razbita na lokalne afine transformacije trojic delov, v [2] pa je polno povezan graf omejitev rešen z uporabo filtra z delci za manjše število delov. V [8] avtorji za zapis prostorskih omejitev med deli uporabijo markovska slučajna polja. Problem vseh omenjenih pristopov je, da morajo biti omejitve ročno nastavljene glede na strukturne lastnosti objekta, čemur pa je v številnih scenarijih sledenja nemogoče zadostiti. Poleg tega je množica delov v teh modelih fiksna in se ne more prilagajati večjim spremembam v videzu objekta. V [34] avtorji predlagajo sledenje artikuliranim objektom s požrešnim deljenjem segmentacijske maske objekta na več delov. Bolj prilagodljiv geometrijski model, ki omogoča dolgoročno osveževanje, je predstavljen v [28]. Preprost zvezdast model povezuje posamezne dele, le-te pa lahko s časom dodajamo in odvzemamo. Novi deli so v model dodani z uporabo globalnega barvnega modela, ki je kombiniran z detektorjem stabilnih regij, kar pomeni, da je postopek omejen na teksturirane objekte. Naslednji model, ki

uporablja višjenivojski globalni videz za postavljanje delov, je predstavljen v [13]. Segmentacijski algoritem, inicializiran z uporabo najdenih ujemanj lokalnih značilnic, rezultat segmentacije pa je nato uporabljen za učenje novih značilnic. Uspeh tega pristopa je neposredno odvisen od robustnosti segmentacije, ki je pri zamegljenih ali šumnih scenah dokaj nizka. Preprostejša, hitrejša, a tudi manj zanesljiva segmentacija je uporabljena v [10]. Uspešnost vseh teh pristopov kaže na uporabnost visokonivojske informacije, saj ta omogoča daljšo življenjsko dobo sledilnikov, ki temeljijo na kombinaciji lokalnih opisov v scenarijih, kjer se videz objekta spreminja. Kljub temu pa ostaja mehanizem integracije globalne in lokalne informacije o videzu objekta le delno raziskan.

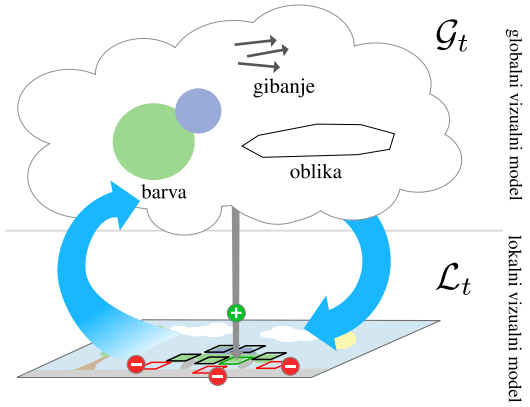
3 HIERARHIČNI MODEL VIDEZA

Kot smo omenili v prejšnjem poglavju, holistični modeli niso primerni za vse scenarije sledenja. Zato predstavljamo novo formalizacijo modela videza, ki mu pravimo *hierarhični model videza*. Ta temelji na združevanju obeh glavnih paradigem zasnove modelov videza, torej holističnega načina opisa videza v kombinaciji z opisom z deli. Motivacija za hierarhični opis videza objekta izhaja iz potrebe po prostorskem in časovnem strukturiranju teh podatkov, rezultat pa je vizualni model, ki je dovolj specifičen za učinkovito lokalizacijo objekta v sliki ter dovolj prožen in prilagodljiv glede na spremembe v videzu objekta. Konceptualno je hierarhični model definiran kot množica plasti, vsaka od njih opisuje videz na specifičen način. Spodnja plast vsebuje najbolj jasno informacijo o trenutnem videzu objekta, višje plasti pa informacijo o splošnejšem, časovno manj spremenljivem videzu. Funkcija posameznih plasti se odraža tudi v osveževanju vizualnega modela. Spodnje plasti pri osveževanju vodijo višje-ležeče plasti, višje plasti pa so osveževane z izluščeno in posplošeno vizualno informacijo spodnjih plasti, če je le-ta dovolj zanesljiva. Če informacija v nekem trenutku ni zanesljiva, se osveževanje višjih plasti ustavi, plasti pa so tako zaščitene pred drsenjem in lahko z vodenjem osveževanja spodnjih plasti pripomorejo k okrevanju celotnega vizualnega modela.

Hierarhični model videza ponuja odprt in prožen teoretični okvir, ki je lahko vodilo za razvoj bolj robustnih sledilnikov. Spodnja plast je najbliže videzu objekta v danem trenutku, vendar se mora nenehno spreminjati in prilagajati spremembam v sliki. To lahko dosežemo z uporabo vizualnega modela z visoko stopnjo prostih parametrov, npr. prožna konstelacija delov, vendar pa lahko pri taki predstavitvi na dolgi rok hitro nastanejo problemi pri iskanju optimalnega nabora vrednosti parametrov zaradi velikega števila lokalnih maksimumov. Prav pri tem pridejo do izraza višje plasti vizualnega modela, ki omogočajo spodnji plasti vodenje, na primer z odvzemanjem zastarelih delov ter dodajanjem novih, s čimer se spodnja plast prilagaja spremembam in ohranja jasnost opisa. V nadaljevanju bomo povzeli dva modela videza, ki ju lahko obravnavamo kot instanco predstavljenega splošnega hierarhičnega koncepta modeliranja videza.

3.1 Model dveh plasti

V tem članku kot prvi model, ki sledi ideji hierarhične organizacije vizualne informacije. povzemamo idejo t. i. *sklopljenega modela videza*, ki je bil podrobneje predstavljen v [6]. Gre za model, ki videz objekta hrani v dveh plasteh, v njih pa združuje lokalno in globalno predstavitev videza objekta, kot je to prikazano na sliki 1.



Slika 1: Shematični prikaz dvoplastnega modela videza

Spodnja plast modela sestavlja množica delov, ki opisujejo lokalne lastnosti videza,

$$\mathcal{L}_t = \{\langle \mathbf{x}_t^{(i)}, \mathbf{h}^{(i)}, w_t^{(i)} \rangle\}_{i=1:N_t}, \quad (1)$$

kjer je $\mathbf{x}_t^{(i)}$ položaj i-tega dela, $\mathbf{h}^{(i)}$ njegov model videza, gre za sivinski histogram iz lokalnega območja, ki je zajet iz slike ob postavitvi dela, $w_t^{(i)}$ pa je utež, ki označuje pomembnost dela znotraj modela. Za primerjavo posameznega dela s sliko smo uporabili razdaljo Bhattacharyja. Iskanje prileganja cele množice delov z novo sliko v zaporedju je formalizirano kot sikanje maksimuma verjetnostne porazdelitve nad položaji delov in v odvisnosti od vizualne informacije in geometrijskih omejitev,

$$p(\mathbf{Y}_t, \mathbf{X}_t | \mathbf{X}_{t-1}) = \sum_{i=1}^{N_t} w_t^{(i)} p(\mathbf{Y}_t, \mathbf{x}_t^{(i)} | \varepsilon_t^{(i)}, \mathbf{z}^{(i)}), \quad (2)$$

kjer $\varepsilon_t^{(i)}$ označuje okolico i-tega dela, torej množico delov, s katerimi je del i povezan. Če privzamemo neodvisnost geometrijskih omejitev in vizualne podobnosti dela, lahko ujemanje posameznega dela opišemo kot

$$p(\mathbf{Y}_t, \mathbf{x}_t^{(i)} | \varepsilon_t^{(i)}, \mathbf{z}^{(i)}) = p(\mathbf{Y}_t | \mathbf{x}_t^{(i)}, \mathbf{z}^{(i)}) p(\mathbf{x}_t^{(i)} | \varepsilon_t^{(i)}). \quad (3)$$

Pri tem je člen $p(\mathbf{Y}_t|\mathbf{x}_t^{(i)}, \mathbf{z}^{(i)})$ definiran kot vizualno ujemanje prek razdalje Bhattacharyja, $p(\mathbf{x}_t^{(i)}|\varepsilon_t^{(i)})$ pa kot geometrijsko ujemanje prek odstopanja od položaja, ki ga za del i predlagajo njegovi sosedi. Iskanje optimuma take funkcije je problematično zaradi visoke dimenzionalnosti in kompleksnosti prostora z veliko lokalnimi optimumi. Algoritem, ki smo ga uporabili za hitro in robustno reševanje problema, se opira na idejo o postopni nekonveksnosti in razdeli iskanje optimuma na dva koraka: globalno optimizacijo toge konstelacije in residualne popravke posameznih delov. Podrobneje je algoritem opisan v [6]. Poleg prilagajanja položajev delov, kar zagotavlja kratkoročno točnost opisa, se mora množica delov med sledenjem ustrezno prilagajati tudi večjim spremembam videza, kar dosežemo z dodajanjem novih in odvzemanjem starih delov. Kriterij za odstranjevanje starih delov je njihov majhen pomen, torej utež $w_t^{(i)}$. Ta se spreminja na podlagi trenutnega ujemanja posameznega dela s sliko in njegove oddaljenosti od drugih delov. Pri dodajanju novih delov igra zelo pomembno vlogo zgornja plast, ki vsebuje globalni opis objekta v treh vizualnih modalnostih: barvi (C_t), gibanju (M_t) in obliki (S_t),

$$\mathcal{G}_t = \{C_t, M_t, S_t\}. \tag{4}$$

Vse tri modalnosti hranijo informacije na njim lasten način, ki je podrobneje opisan v [6], barva je predstavljena z barvnim histogramom, gibanje z vektorjem premika, oblika pa z množico poligonov. Vsem trem modalnostim je skupno, da lahko za dano sliko generirajo verjetnostno porazdelitev, da posamezni slikovni element $\mathbf{x}$ pripada objektu. Taka porazdelitev lahko nato služi za vzorčenje območja, ki je primerno za postavitev novega elementa. Ob predpostavki, da so vse tri modalnosti med seboj neodvisne, lahko skupno verjetnostno porazdelitev zapišemo kot

$$p(\mathbf{x}|C_t, M_t, S_t) \propto p(C_t|\mathbf{x})p(M_t|\mathbf{x})p(S_t|\mathbf{x}). \tag{5}$$

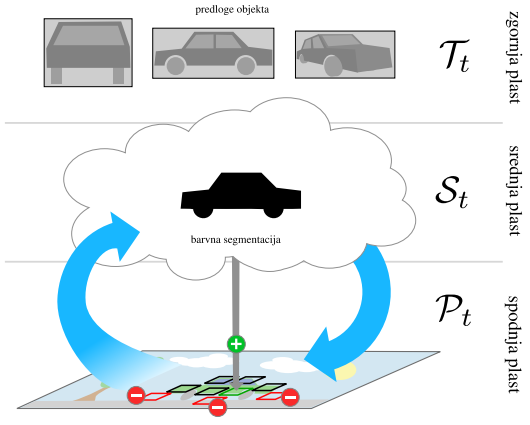
V vsakem koraku se vse tri plasti posodobijo. Pri tem s svojo informacijo o položaju sodelujejo zgolj deli lokalne plasti z dovolj veliko pomembnostjo (utežjo), kar zagotavlja višjo robustnost posodabljanja.

Pomembna lastnost modela videza je tudi začetni zajem vizualne informacije, scenariji sledenja namreč predpostavljajo, da je sledilnemu algoritmu ob njegovi inicializaciji poleg začetne slike podana tudi bolj ali manj natančna regija celotnega objekta. V našem primeru predpostavljamo, da gre za pravokotnik, ki je tudi edina informacija, ki jo imamo o objektu, zato dele spodnje plasti razporedimo v mrežo znotraj podanih mej, v katerih pridobimo tudi začetne informacije za posamezne globalne modalnosti.

Kot je razvidno iz rezultatov primerjalnih poskusov v poglavju 4, je predlagana kombinacija lokalne in globalne informacije smiselna, tak opis videza objekta sledilniku omogoča robustno in računsko učinkovito sledenje v raznolikih pogojih, še posebno pa se model videza izkaže pri sledenju objektov, ki se ne-togo deformirajo. Kljub temu pa prinaša tak način opisa videza tudi probleme. Vizualna informacija na obeh plasteh modela se spreminja dokaj hitro, ob vdoru vizualne informacije ozadja v model pa le-ta nima mehanizma, da bi si od tega opomogel. To se v rezultatih kaže v slabi natančnosti sledenja, še posebno v primerih, ki so dokaj preprosti za sledilnike, ki objekt ocenjujejo z manj parametri, lete pa lahko tako ocenijo bolj natančno. Primeri takih scenarijev so sledenje pretežno togih objektov v primeru zakrivanja, kjer se zaradi prilagodljive spodnje plasti pokvari geometrija delov, pa tudi v primeru vizualne podobnosti med objektom in ozadjem, ko odpove globalna plast. Da bi hierarhični model videza lahko kljuboval tudi takim primerom, je treba uvesti novo plast, ki date časovno stabilnejšo informacijo o videzu objekta in omogoča hitro okrevanje v primeru, ko ima objekt na sliki podoben videz, kot ga je imel že v preteklosti.

3.2 Model treh plasti

Drugi prestavljeni model videza razširja hierarhijo s tretjo plastjo, kot je prikazano na sliki 2, prvi dve plasti drugega vizualnega modela pa sta konceptualno zelo podobni prvemu vizualnemu modelu in ju bomo opisali pozneje.



Slika 2: Shematični prikaz troplastnega modela videza

Najvišja plast modela je še bolj trajna glede na sekvenco slik ter uvaja koncept *sidrnih predlog*. To je pomnilniški sistem, ki vsebuje množico holističnih predlog videza objekta, pridobljenih čez zaporedje slik ob različnih trenutkih $\mathcal{T}_t = \{T_1, T_2, \ldots\}$. V našem primeru smo za opis zaplat uporabili korelacijske filtre nad opisniki HOG [17]. V novi sliki zaporedja iščemo predlogo, ki najbolje pojasni vizualno informacijo t. j.

$$\hat{T}_t = \arg\max_{T \in \mathcal{T}_{t-1}} d(T, \mathbf{Y}_t), \tag{6}$$

kjer $d(\cdot,\cdot)$ označuje funkcijo ujemanja, ki nam vrne najboljši odziv v okolici, $\mathbf{Y}_t$ trenutno sliko, $\hat{T}_t$ pa

označuje *sidrno predlogo*, torej predlogo z najboljšim odzivom. Kakovost ujemanja predloge določa nadaljnji način uporabe. Predloga je lahko pri visokem ujemanju obravnavana kot *detekcija*, pri srednjem ujemanju je položaj njene detekcije *vodilo* za nižje plasti modela videza, pri slabem ujemanju pa se predloga ne uporablja. Prednost tega mehanizma je, da se zaplate uporabijo samo pri zanesljivi uporabi, lahko tudi samo za krajša obdobja v sekvenci slik. To nam posledično dovoljuje tudi uporabo zelo konservativnega mehanizma za posodabljanje množice predlog z novimi predlogami, ki se dodajo samo ob soglašanju spodnjih plasti modela o velikosti in položaju objekta. Več podrobnosti o opisanem mehanizmu je na voljo v [7].

Spodnja plast modela je enaka spodnji plasti modela dveh plasti, vendar pa v tem primeru namesto stohastične optimizacije za iskanje prileganja uporabimo deterministični algoritem, ki najprej poizkuša za posamezni del oceniti optični tok, če ocena ni zanesljiva, pa preide najprej na globalno oceno premika z uporabo posplošene Houghove transformacije, nato pa na izboljšavo z uporabo algoritma Iterated Conditional Modes (ICM). Podrobneje je algoritem opisan v [7]. Tudi v tem primeru se množica delov osvežuje še z odstranjevanjem nepomembnih delov in dodajanjem novih. Deli, ki niso pomembni, so določeni na podlagi trenutnega ujemanja posameznega dela s sliko in algoritma prilagajanja srednje vrednosti (*mean-shift*) [12] nad položaji posameznih delov. V algoritmu se uporabi uniformno jedro, ki je po velikosti enako trenutni ocenjeni velikosti objekta. Deli, ki so za model nepomembni, ali pa celo škodljivi, so tisti, ki jih algoritem izloči iz območja jedra ob končni konvergenci.

Novi deli so v množico dodani na podlagi segmentacije objekta na podlagi barve, kar bomo opisali v naslednjem odstavku, pri določitvi natančnega položaja začetne postavitve pa se upoštevajo tudi lastnosti slike, ki bi zagotovile čim bolj kakovostno oceno optičnega toka, s tem pa hitro določitev prilagajanja množice delov v novi sliki. Za vzorčenje položajev se uporabi funkcija $q(\mathbf{x}) = H(\mathbf{x}) + \alpha_U U(\mathbf{x})$, kjer je $H(\mathbf{x})$ Harrisova ocena kotov za točko $\mathbf{x}$, $U(\mathbf{x})$ periodična funkcija, v našem primeru gre za kosinusni signal v dveh dimenzijah, ki naredi mrežast vzorec, α_U pa konstanta. Funkcija $q(\mathbf{x})$ nam za teksturirane regije zagotavlja postavitev delov na kotih, kjer je ocena optičnega toka bolj kakovostna, na območjih z manj teksture pa postavitev delov v mrežnem vzorcu.

Srednja plast je, podobno kot pri prvem vizualnem modelu, namenjena določanju območij v sliki, ki pripadajo objektu. To se doseže s preprosto in hitro segmentacijo na podlagi barvnega modela. Plast vsebuje informacijo o barvi objekta, s katero za dano sliko generiramo segmentacijsko masko objekta, na podlagi te pa določimo območja, primerna za postavitev novih delov na spodnji plasti. Konceptualno gre za podoben pristop kot pri zgornji plasti modela dveh plasti, s tem, da tu algoritem glede na predznanje o velikosti objekta samodejno določi prag nad verjetnostno porazdelitvijo in jo razdeli v binarno masko. Če prag ne more biti določen, barvna predstavitev velja za nezanesljivo in se v danem časovnem koraku ne uporabi. Algoritem za določitev praga je opisan v [7].

V nasprotju z modelom dveh plasti se tudi začetni zajem vizualne informacije zanaša na segmentacijo in oceno kotov. Razporeditev delov je zato bolj naravna in se prilagaja strukturi objekta. Poleg tega začetni položaj objekta služi tudi za pridobitev prve predloge objekta v tretji plasti.

Najpomembnejša lastnost izboljšav modela videza je tretja plast. Ta daje spodnjim plastem zanesljivo informacijo o položaju in velikosti objekta pri dobrem ujemanju ene izmed predlog s sliko, sicer pa delovanju modela ne škoduje. Tako torej tretja plast pripomore k hitremu okrevanju celotnega modela, če ta delno zdrsne na ozadje ali zgolj na del objekta. Eksperimentalna analiza, predstavljena v naslednjemu poglavju, potrdi koristi tretje plasti in predlaganega mehanizma, saj sledilnik s takim modelom videza izboljša natančnost, pa tudi splošno kakovost sledenja.

4 REZULTATI

Opisana sledilnka, ki ju bomo označili z oznakama LGT [6] in ANT [7], kljub svoji konceptualni podobnosti do zdaj nista bila eksperimentalno ovrednotena z enako metodologijo, zato bomo v okviru članka naredili poenoteno analizo iz vidika hiearhičnih modelov videza ter ju testirali z uporabo dveh primerjalnih testov za vizualne sledilnke, VOT2013 [27] in VOT2014 [26]. Oba testa določata zbirko sekvenc z ročnimi anotacijami objektov in protokol izvedbe eksperimentov ter obdelave rezultatov, obenem pa so na voljo tudi okolje za izvedbo eksperimentov in rezultati za veliko sledilnih algoritmov, kar omogoča primerjavo z referenčnimi algoritmi.

Podatki in eksperimentalni protokol. Zbirki sekvenc sta sestavljeni iz 16 (VOT2013) in 25 (VOT2014) ročno anotiranih sekvenc, ki vsebujejo različne scenarije, ki so iz različnih razlogov zahtevni za nalogo sledenja, npr. sprememba osvetlitve, deformacije objekta, hitre spremembe gibanja, sprememba velikosti, rotacija, zakrivanje. Kot je opisano v [27], [26], so bile sekvence izbrane iz večjega nabora z metodo gručenja po lastnostih z namenom, da dobimo reprezentativno množico sekvenc obvladljive velikosti.

V primerjavi sledimo uradni metodologiji VOT, ki predpisuje, da je za vsako sekvenco sledilnik najprej postavljen na položaj, ki ga določa zlati standard v prvi sliki sekvence. Nato sledilnik sledi objektu, dokler ne zdrsne z njega; v tem primeru sledilnik ponastavimo na pravilen položaj in si zapolnimo položaj odpovedi. Končni rezultati so povzeti v obliki natančnosti (povprečno prekrivanje regije zlatega standarda in trajektorije sledilnka) in robustnosti (število zdrsov). Poskusi so bili

opravljeni z uporabo uradnega okolja za izvedbo poskusov, ki omogoča tudi izvedbo analize z razvrščanjem. V tem primeru se, da se izognemo pristranskosti, upoštevata tudi statistična in praktična razlika v natančnosti in robustnosti. Podrobnosti metodologije so opisane v [27], [26], [24].

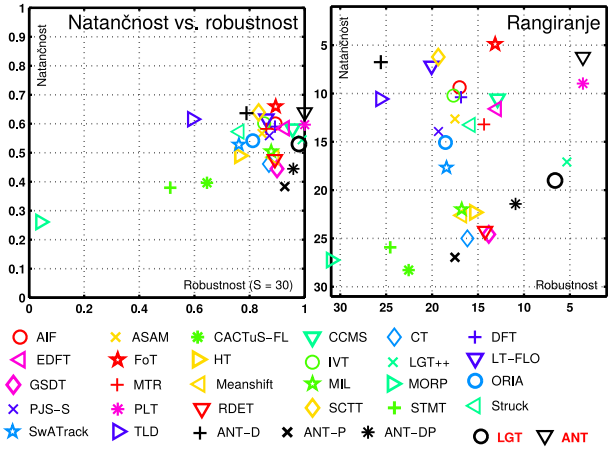
Implementacija in nabor parametrov. Oba predlagana sledilnika sta bila v okviru raziskav implementirana v jeziku Matlab, bolj kompleksni deli algoritma pa so implementirani v jeziku C++. Čeprav je implementacija namenjena razvoju in razumevanju delovanja in se deli algoritma izvedejo večkrat zavoljo jasnosti, celoten algoritem pa teče zaporedno kljub velikim potencialom paralelizacije, se algoritem LGT izvaja s hitrostjo tri slike na sekundo, algoritem ANT pa celo s petimi slikami na sekundo na računalniku s procesorjem AMD Opteron 6238. Algoritem LGT je bil v okrnjeni obliki implementiran tudi v jeziku C++, ki deluje s hitrostjo 30 slik na sekundo, zato lahko trdimo, da sta oba algoritma primerna za procesiranje slikovnih tokov v realnem času.

Kot je to predvideno v metodologiji VOT, so parametri obeh sledilnikov fiksni v vseh sekvencah obeh primerjalnih testov. Pri sledilniku LGT so podrobni parametri objavljeni v [6], parametri sledilnika ANT pa so navedeni v [7].

4.1 VOT2013

Rezultate primerjave na zbirki VOT2013 povzema slika 3, na kateri so prikazane surove vrednosti za natančnost in robustnost ter razvrščanje z upoštevanjem statistične enakovrednosti. Poleg primerjave z referenčnimi sledilniki, katerih rezultati so javno dostopni v okviru zbirke, smo v okviru zbirke raziskali tudi pomen tretje plasti modela videza, ki je vključen v sledilnik ANT. Z manipulacijo parametrov, ki vplivajo na funkcijo tretje plasti, smo poleg glavnega sledilnika ustvarili še tri dodatne: sledilnik ANT-D sledi samo z eno predlogo, ki jo dobi na začetku sekvence, in ji vedno brezpogojno zaupa, sledilnik ANT-P uporablja samo spodnji dve plasti in ignorira vpliv sidrne predloge, sledilnik ANT-DP pa uporablja vse tri plasti, vendar zgornja plast deluje le v načinu detekcije, ne upošteva pa se vmesni način vodenja spodnjih plasti pri manj zanesljivemu ujemanju.

Iz rezultatov lahko vidimo, da sta oba sledilnika, tako LGT kot ANT, v samem vrhu kar se tiče robustnosti, sledilnik ANT pa celo premaga vse sledilnike, če upoštevamo kombinacijo natančnosti in robustnosti. Ob tem je treba poudariti, da so nekateri sledilniki opazno boljši v natančnosti, npr. FoT [42] in LT-FLO [30], vendar je razlog za to v pogostih zdrsih sledilnika, ki jim sledijo ponastavitve. To se seveda odraža v nizki robustnosti takih sledilnikov. Analiza je pokazala tudi, da je večina holističnih sledilnkov, npr. IVT, Struck in EDFT, manj robustnih, ko govorimo o sledenju netogih objektov, odrežejo pa se bolje pri sledenju togim objektom, pri katerih so uspešni tudi pri zakrivanju in spremembah osvetlitve. Po drugi strani pa sledilnik ANT



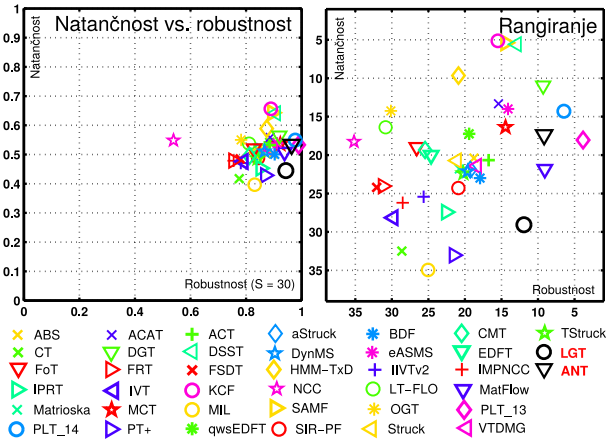
Slika 3: Rezultati na zbirki VOT2013

združuje kvalitete holističnega opisa in opisa z deli, kar privede do izboljšav na obeh področjih. To je potrdila tudi analiza rezultatov treh dodatnih sledilnikov. Iz nje je razvidno, da ANT-D doseže dobro natančnost, predvsem zato, ker uporablja zglolj začetno predlogo, ki ne more upoštevati sprememb videza, zato tudi hitro zdrsne z objekta. Po drugi strani sledilnik ANT-P doseže dobro robustnost, vendar dokaj slabo natančnost, saj gre za samonadzorovano osveževanje spodnjih dveh plasti brez dodatnega nadzora in možnosti okrevanja, ki ga prinaša sistem sidrnih predlog. Sledilnik ANT-DP integrira lastnosti ANT-D in ANT-P in tako izboljša rezultat s preklapljanjem med detekcijo s predlogami in sledenjem z množico delov, vendar pa ne vključuje mehanizma, pri katerem lahko predloge ob nepopolnem ujemanju še vedno sodelujejo pri osveževanju spodnjih plasti. S tem mehanizmom sledilnik ANT opazno izboljša delovanje, s tem pa se potrdi tudi naša hipoteza, da sidrne predloge v opisanem načinu delovanja izboljšajo robustnost modela videza in posledično kakovost sledenja.

4.2 VOT2014

Druga zbirka, ki smo jo uporabili za analizo je VOT2014, rezultate pa povzema slika 4. Čeprav gre za zahtevnejšo zbirko z novejšimi sledilniki, sta LGT in ANT glede robustnosti še vedno v vrhu. V natančnosti so rezultati nekoliko slabši še posebej pri sledilniku LGT, medtem ko se sledilnk ANT odreže primerljivo z večino drugih sledilnikov. Primerljivi sledilnik DGT se odreže bolje v natančnosti z uporabo računsko potratne segmentacije, holistični sledilniki, npr. DSST, KCF in SAMF, pa se v natančnosti odrežejo bolje, vendar ob opazno večjem številu zdrsov.

Kot je jasno razvidno iz slike, sledilnik ANT z uporabo treh plasti opazno izboljša natančnost glede na sledilnik LGT, obenem pa izboljša tudi robustnost. To pomeni, da izboljšana natančnost ni zgolj rezultat kompromisa med dvema pogledoma na sledenje, ampak gre za izboljšavo modela videza.



Slika 4: Rezultati na zbirki VOT2014

5 Sklep

V članku smo opisali problem kratkoročnega vizualnega sledenja in predstavili koncept hierarhičnega modela videza. Tak način opisa vizualne informacije nam omogoča, da se po eni strani osredotočimo na trenutni videz objekta, vendar pa ohranimo dovolj splošne informacije, ki se uporabi kot vodilo pri posodabljanju modela. V članku smo povzeli teorijo dveh modelov videza, ki izpolnjujeta merila hierarhične ureditve, in predstavili eksperimentalne rezultate, ki kažejo na velik potencial ideje, še zlasti pri sledenju netogih objektov.

Na koncu je treba poudariti, da je definicija kratkoročnega sledenja v trenutni obliki dokaj problematična, saj sledenje stanju poljubnega objekta zahteva integracijo veliko večje količine znanja, kot je samo trenutni videz objekta. Da bi lahko poljuben objekt zanesljivo sledili v poljubni situaciji, bi moral sistem integrirati algoritme z več področij računalniškega vida in sklepanja, kar daleč presega trenutno stanje na tem raziskovalnem področju. Po drugi strani pa že zdaj obstaja veliko možnosti za uporabo vizualnega sledenja v okviru določenih aplikacij, kjer je scenarij sledenja bolj definiran in omejen. Prav med tema dvema pogledoma vidimo veliko priložnost hierarhičnih modelov videza, saj dajejo teoretični okvir, ki omogoča po eni strani postopen prehod s problema sledenja na druge domene računalniškega vida, kot sta kategorizacija in detekcija, po drugi strani pa na podoben način omogoča tudi intuitivno uvajanje omejitev, ki izvirajo iz aplikacije. To so zato tudi naše smernice za nadaljnje raziskovanje in delo.

Literatura

[1] B. Babenko, M.-H. Yang, in S. Belongie. Robust Object Tracking with Online Multiple Instance Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1619–1632, aug 2011.

[2] V. Badrinarayanan, F. Le Clerc, L. Oisel, in P. Perez. Geometric Layout Based Graphical Model for Multi-Part Object Tracking. Objavljeno v *International Workshop on Visual Surveillance*, 2008.

[3] V. Badrinarayanan, P. Perez, F. Le Clerc, in L. Oisel. Probabilistic Color and Adaptive Multi-Feature Tracking with Dynamically Switched Priority Between Cues. Objavljeno v *IEEE International Conference on Computer Vision*, strani 1–8, 2007.

[4] Chenglong Bao, Yi Wu, Haibin Ling, in Hui Ji. Real time robust L1 tracker using accelerated proximal gradient approach. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, strani 1830–1837. IEEE, jun 2012.

[5] Gary R. Bradski. Real Time Face and Object Tracking as a Component of a Perceptual User Interface. Objavljeno v *Winter Conference on Applications of Computer Vision*, stran 214. IEEE Computer Society, 1998.

[6] Luka Čehovin, Matej Kristan, in Aleš Leonardis. Robust Visual Tracking using an Adaptive Coupled-layer Visual Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):941–953, apr 2013.

[7] Luka Čehovin, Aleš Leonardis, in Matej Kristan. Robust visual tracking using template anchors. Objavljeno v *WACV*. IEEE, mar 2016.

[8] W.-Y. Chang, C.-S. Chen, in Y.-P. Hung. Tracking by Parts: A Bayesian Approach With Component Collaboration. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(2):375–388, 2009.

[9] D. Comaniciu, V. Ramesh, in P. Meer. Kernel-Based Object Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–575, 2003.

[10] Stefan Duffner in Christophe Garcia. PixelTrack: A Fast Adaptive Algorithm for Tracking Non-rigid Objects. Objavljeno v *IEEE International Conference on Computer Vision*, dec 2013.

[11] Z. Fan, M. Yang, in Y. Wu. Multiple Collaborative Kernel Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1268–1273, 2007.

[12] Keinosuke Fukunaga in Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40, 1975.

[13] Martin Godec, Peter M. Roth, in Horst Bischof. Hough-based tracking of non-rigid objects. Objavljeno v *IEEE International Conference on Computer Vision*, strani 81–88, Barcelona, nov 2011. IEEE.

[14] H. Grabner, M. Grabner, in H. Bischof. Real-Time Tracking via On-line Boosting. Objavljeno v *British Machine Vision Conference*, strani 47–56, 2006.

[15] Helmut Grabner, Christian Leistner, in Horst Bischof. Semi-supervised on-line boosting for robust tracking. Objavljeno v *European Conference on Computer Vision*, strani 234–247. Springer, 2008.

[16] Sam Hare, Amir Saffari, in Philip H. S. Torr. Struck: Structured output tracking with kernels. Objavljeno v *IEEE International Conference on Computer Vision*, strani 263–270. IEEE, nov 2011.

[17] J F Henriques, R Caseiro, P Martins, in J Batista. High-Speed Tracking with Kernelized Correlation Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.

[18] Jesse Hoey, A. von Bertoldi, P. Poupart, in A. Mihailidis. Tracking using flocks of features, with application to assisted handwashing. Objavljeno v *British Machine Vision Conference*, strani 367–376, 2006.

[19] Michael Isard in Andrew Blake. Contour tracking by stochastic propagation of conditional density. Objavljeno v Bernard Buxton in Roberto Cipolla, editors, *European Conference on Computer Vision*, del 1064 of *Lecture Notes in Computer Science*, strani 343–356. Springer Berlin Heidelberg, 1996.

[20] Pakorn KaewTrakulPong in Richard Bowden. A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes. *Image and Vision Computing*, 21(10):913–929, sep 2003.

[21] Zdenek Kalal, Krystian Mikolajczyk, in Jiri Matas. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1409–1422, 2012.

[22] M. Kölsch in M. Turk. Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition Workshops, del 10, stran 158, Washington, DC, USA, 2004. IEEE Computer Society.

[23] M. Kristan, J. Perš, S. Kovačič, in A. Leonardis. A Local-motion-based probabilistic model for visual tracking. *Pattern Recognition*, 2008.

[24] Matej Kristan, Jiri Matas, Ales Leonardis, Tomas Vojir, Roman Pflugfelder, Gustavo Fernandez, Georg Nebehay, Fatih Porikli, in Luka Čehovin. A Novel Performance Evaluation Methodology for Single-Target Trackers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2016.

[25] Matej Kristan, Janez Perš, Matej Perše, in Stanislav Kovačič. Closed-world tracking of multiple interacting targets for indoor-sports applications. *Computer Vision and Image Understanding*, 113(5):598–611, may 2009.

[26] Matej Kristan, Roman Pflugfelder, Aleš Leonardis, Jiri Matas, Luka Čehovin, Georg Nebehay, Tomáš Vojir, Gustavo Fernández, et al. The Visual Object Tracking VOT2014 challenge results. Objavljeno v *European Conference on Computer Vision Workshops*, 2014.

[27] Matej Kristan, Roman Pflugfelder, Aleš Leonardis, Jiri Matas, Fatih Porikli, Luka Čehovin, Georg Nebehay, Gustavo Fernandez, Tomáš Vojir, et al. The Visual Object Tracking VOT2013 challenge results. Objavljeno v *IEEE International Conference on Computer Vision Workshops*, strani 98–111, 2013.

[28] J. S. Kwon in K. M. Lee. Tracking of a non-rigid object via patch-based dynamic appearance modeling and adaptive Basin Hopping Monte Carlo sampling. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, strani 1208–1215, 2009.

[29] Junseok Kwon in Kyoung M. Lee. Visual tracking decomposition. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, strani 1269–1276. IEEE, jun 2010.

[30] Karel Lebeda, Simon Hadfield, Jiri Matas, in Richard Bowden. Long-Term Tracking Through Failure Cases. Objavljeno v *IEEE International Conference on Computer Vision Workshops*, 2013.

[31] David G Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[32] B. Martinez in X. Binefa. Piecewise affine kernel tracking for non-planar targets. *Pattern Recognition*, 41(12):3682–3691, 2008.

[33] Xue Mei in Haibin Ling. Robust visual tracking and vehicle classification via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2259–72, nov 2011.

[34] S.M. Shahed Nejhum, Jeffrey Ho, in Ming-Hsuan Yang. Online visual tracking with histograms and articulating blocks. *Computer Vision and Image Understanding*, 114(8):901–914, aug 2010.

[35] N.P. Papanikolopoulos, P.K. Khosla, in T. Kanade. Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision. *IEEE Transactions on Robotics and Automation*, 9(1):14–35, 1993.

[36] P. Pérez, C. Hue, J. Vermaak, in M. Gangnet. Color-Based Probabilistic Tracking. Objavljeno v *European Conference on Computer Vision*, del 1, strani 661–675. Springer-Verlag, 2002.

[37] Federico Pernici, Alberto Del Bimbo, in Alberto Del Bimbo. Object Tracking by Oversampling Local Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(12):2538–2551, 2014.

[38] F. Porikli, O. Tuzel, in P. Meer. Covariance Tracking using Model Update Based on Means on Riemannian Manifolds. Technical report, Mitsubishi Electric Research Laboratories, 2006.

[39] David A Ross, Jongwoo Lim, Ruei-Sung Lin, in Ming-Hsuan Yang. Incremental Learning for Robust Visual Tracking. *International Journal on Computer Vision*, 77(1-3):125–141, may 2008.

[40] C. Stauffer in W.E.L. Grimson. Adaptive background mixture models for real-time tracking. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, del 2, strani 246–252. IEEE Comput. Soc, 1999.

[41] B. Stenger, T. Woodley, in R. Cipolla. Learning to track with multiple observers. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, strani 2647–2654. IEEE, jun 2009.

[42] Tomáš Vojír in Jirí Matas. Robustifying the Flock of Trackers. Objavljeno v Andreas Wendel, Sabine Sternig, in Martin Godec, editors, *Computer Vision Winter Workshop*, strani 91–97, Inffeldgasse 16/II, Graz, Austria, 2011. Graz University of Technology.

[43] Yi Wu, Bin Shen, in Haibin Ling. Online robust image alignment via iterative convex optimization. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, strani 1808–1814, 2012.

[44] Z. Yin in R. Collins. On-the-fly Object Modeling while Tracking. Objavljeno v *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, strani 1–8, 2007.

[45] Kaihua Zhang, Lei Zhang, in Ming-Hsuan Yang. Real-time Compressive Tracking. Objavljeno v *European Conference on Computer Vision*, 2012.

Luka Čehovin Zajc je leta 2015 doktoriral na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Zaposlen je v Laboratoriju za umetne vizualne spoznavne sisteme na Fakulteti za računalništvo in informatiko kot asistent, njegova raziskovalna področja pa so računalniški vid, interakcija med človekom in računalnikom, mobilna robotika in spletne tehnologije.

Aleš Leonardis je profesor na *School of Computer Science, University of Birmingham* in direktor Centra za računsko nevroznanost in kognitivno robotiko na *University of Birmingham*. Je tudi profesor na Fakulteti za računalništvo in informatiko Univerze v Ljubljani ter gostujoči profesor na Fakulteti za računalništvo na Tehniški univerzi v Gradcu. Njegova raziskovalna področja so robustne in prilagodljive metode v računalniškem vidu, razpoznava in kategorizacija predmetov, statistično učenje v računalniškem vidu, 3-D modeliranje objektov in biološko motiviran računalniški vid.

Matej Kristan je leta 2008 doktoriral na Fakulteti za elektrotehniko Univerze v Ljubljani. Zaposlen je kot docent v Laboratoriju za umetne vizualne spoznavne sisteme na Fakulteti za računalništvo in informatiko, poleg tega pa je docent tudi na Fakulteti za elektrotehniko. Njegova raziskovalna področja so verjetnostni modeli v računalniškem vidu s poudarkom na vizualnem sledenju, dinamični modeli ter sprotno učenje v računalniškem vidu in mobilni robotiki.