

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

**Procjena optičkog toka dubokim
povratnim modelima**

Frano Rajić

Voditelj: *prof. dr. sc. Siniša Šegvić*

Zagreb, svibanj 2021.

SADRŽAJ

1. Uvod	1
2. Optički tok	2
2.1. Kategorija KD	3
2.2. Kategorija DD	5
2.3. Evaluacijske mjere	6
2.4. Skupovi podataka	6
2.4.1. MPI-Sintel	7
2.4.2. KITTI-2012	8
2.4.3. KITTI-2015	8
2.4.4. HD1K	9
2.4.5. FlyingChairs	9
2.4.6. FlyingThings	9
3. Duboki model RAFT	12
3.1. Vađenje značajki	12
3.2. Računanje vizualne sličnosti	13
3.3. Iterativna ažuriranja procjene toka	15
3.4. Funkcija gubitka	17
4. Eksperimenti autora	20
4.1. Rezultati	20
4.2. Studija ablacija	20
5. Zaključak	25
6. Literatura	26
7. Sažetak	30

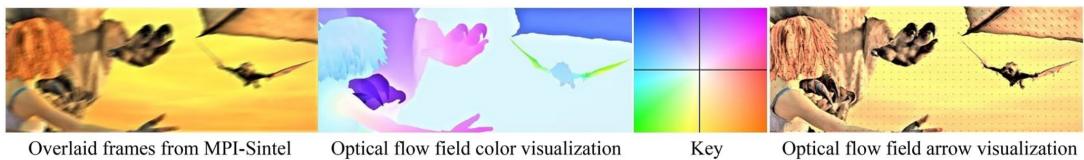
1. Uvod

Optički tok igrao je važnu ulogu u različitim znanstvenim područjima, a među njima su mehanika fluida, solarna fizika, autonomna vožnja, biomedicinske slike, rak dojke, rak mokraćnog mjehura, sigurnosni nadzor, nadzor prometa, virtualna stvarnost, prepoznavanje i praćenje lica i prepoznavanje radnji u videu [22].

Duboko učenje poboljšalo je uspješnost mnogih zadataka računalnog vida. Među tim zadatcima je i procjena optičkog toka. Cilj ovog seminara je uvesti pojam optičkog toka i opisati nadzirani duboki povratni model RAFT [26] (*Recurrent All-Pairs Field Transforms*) koji služi za njegovu procjenu. Optički tok i pregled pristupa kojima se on procjenjuje su opisani u sljedećem poglavlju. U poglavlju *duboki model RAFT* opisana je arhitektura dotične duboke neuronske mreže. U poglavlju *eksperimenti autora* dan je pregled eksperimenata i rezultata autora originalnog rada.

2. Optički tok

Optički tok je vektorsko polje koje opisuje prividno kretanje u slijedu slika neke vizualne scene [10]. Može biti uzrokovani relativnim gibanjem scene u odnosu na kameru ili promjenljivim osvjetljenjem. Svakom pikselu slike $(x, y) \in \mathbb{N}^2$ pridjeljuje vektor $(u, v) \in \mathbb{R}^2$ koji govori gdje se taj piksel pomaknuo u idućoj slici: $(x', y') = (x + u, y + v)$. Ovo se polje može predočiti pomoću strelica i pomoću obojenja. U obojenju, svaka boja predstavlja jedan smjer i orientaciju vektora, a intenzitet predstavlja duljinu vektora, odnosno iznos pomaka. Slika 2.1 prikazuje opisane vizualizacije na danom slijedu slika.



Slika 2.1: Primjer dvaju slika iz MPI-Sintel (postavljenih jedna na drugu) i odgovarajuće polje optičkog toka. [22]

Optički tok dugovječan je problem računalnog vida koji ostaje neriješen [26], ali uspješnost procjene optičkog toka neprestano napreduje [22]. U posljednja četiri desetljeća na ovom se području razvila skupina različitih tehnika kao i novih koncepata. Te su metode u pregledu znanstvenih radova [16] svrstane u 3 kategorije:

1. tradicionalni postupci zasnovani na ekspertnom znanju (KD, *knowledge-driven methods*)
2. postupci zasnovani na učenju iz podataka (DD, *data-driven methods*)
3. postupci koji kombiniraju KD i DD (H, *hybrid methods*)

U narednim potpoglavlјima je detaljniji pregled kategorija KD i DD, a u ostatku poglavlja opisuju se metrike i skupovi podataka korištene za evaluaciju modela

procjene optičkog toka.

2.1. Kategorija KD

Rad [26] opisuje da se procjeni optičkog toka tradicionalno pristupalo kao ručno navođenom optimizacijskom problemu u prostoru gustih polja pomaka između para slika (jer je gusti optički tok definiran kao to vektorsko polje). Općenito je cilj tih optimizacija definiran s dvama elementima koje je potrebno uravnotežiti. Prvi element obuhvaća poravnanje vizualno sličnih dijelova dviju slika. Drugi element ima regularizacijski utjecaj i unosi pretpostavke o izgledu optičkog toka, kao na primjer da je vektorsko polje optičkog toka glatko. Takav je pristup postigao znatan uspjeh, ali daljnji se napredak pokazao izazovnim zbog poteškoća u ručnom dizajniranju cilja optimizacije koji bi bio robustan na rubne slučajevе. Ti rubni slučajevi su: veliki pomak piksela (zbog brzog kretanja predmeta ili kamerе), okluzija (jedan predmet u sceni prekriva drugi), različiti uvjeti osvjetljenja i šum.

Prvi KD pristup i prvi pristup procjene optičkog toka uopće predlažu Horn i Schunck [9] koji koriste varijacijsku metodu za procjenu toka. Nakon [9], Lucas i Kanade [18] uvode pretpostavku lokalnosti optičkog toka i procjenjuju rijetki optički tok (*sparse optical flow*, zadatak u kojem se prate pomaci samo određenog broja zanimljivih piksela). Na slici 2.2 prikazana je primjena njihovog algoritma. Na temelju osnovnih pristupa koje su [9] i [18] uveli, predloženo je mnogo poboljšanja i modifikacija u kategoriji KD pristupa [16].

Znanstveni rad [22] ističe da je izuzetan razvoj u procjeni optičkog toka zabilježen u posljednjem desetljeću. Zahtjevniji skupovi podataka kao što su MPI-Sintel [3], KITTI-2012 [7] i KITTI-2015 [20] predstavili su značajne izazove za algoritme optičkog toka te se pojavilo mnogo novih strategija u KD algoritmima. Te strategije su donijele izvanredne performanse, ali je poboljšanje u točnosti popratilo povećanje vremena izvođenja. Niti jedna od glavnih tradicionalnih metoda trenutno se ne izvodi u stvarnom vremenu. To predstavlja prepreku za usvajanje tih algoritama u područjima primjene.

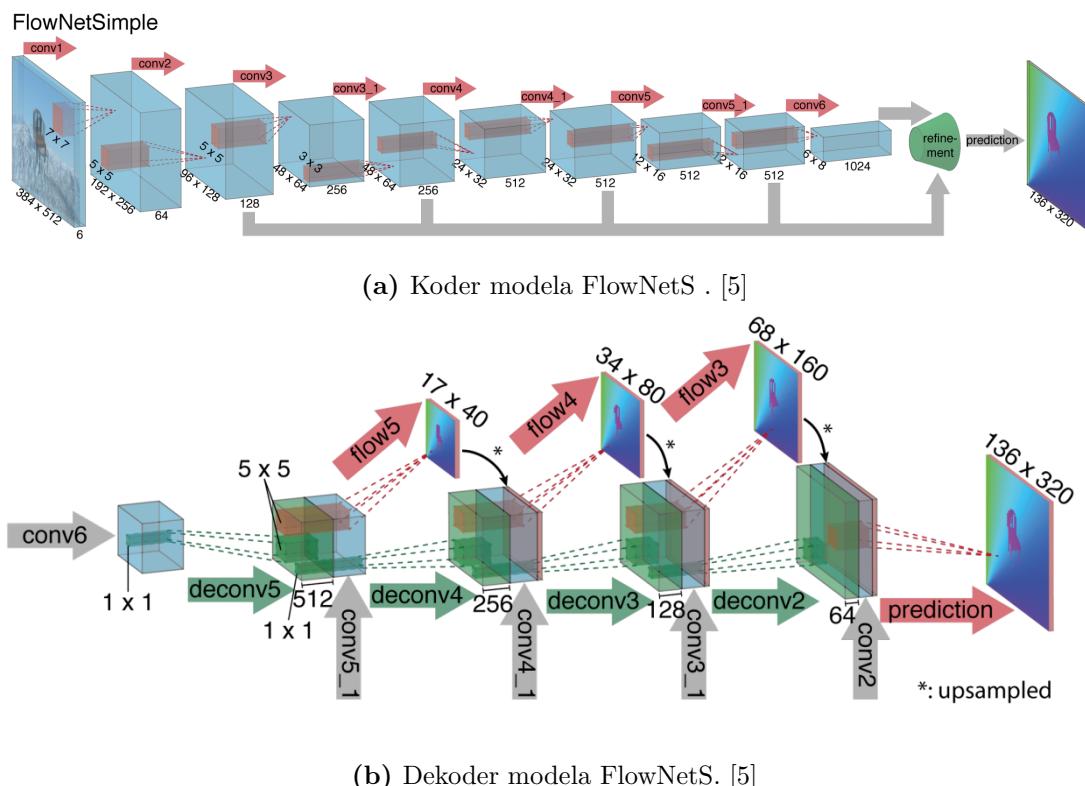


Slika 2.2: Primjena tradicionalnog algoritma za procjenu rijetkog optičkog toka koji su uveli Lucas i Kanade [18] na niz sekvene slika iz sintetičkog skupa MPI-Sintel opisanog u pododjeljku 2.4.1. Prikazana je svaka treća sličica. Istaknute točke su početno odabrane i za njih spomenuti algoritam računa optički tok, odnosno mjesto gdje će se naći na sljedećoj sličici u nizu. Trag koji su točke ostavile prilikom praćenjem svojeg optičkog toka je prikazan bojom jednakoj boji pripadne istaknute točke. Ovaj niz slika narušava mnogo pretpostavki koje ovaj algoritam ima te su rezultati očigledno loši.

2.2. Kategorija DD

Autori pregleda radova [16] tvrde da se konvolucijske neuronske mreže, kao dominirajuća tehnika u dubokom učenju, uspješno koriste za rješavanje problema optičkog toka. Smatraju da se postojeće arhitekture DD modela mogu podjeliti u modele koji koriste *U-Net* i modele koji koriste prostornu piramidalnu mrežu.

U-Net je koder-dekoder arhitektura koju za procjenu optičkog toka prvo koriste u [5], a nakon koje su razvijeni mnogi uspješni duboki modeli. FlowNetS je mreža uvedena u [5] i njena arhitektura je prikazana na slici 2.3. Ulaz mreže čine dvije susjedne slike. Koder se sastoji od uzastopnih slojeva konvolucije, a dekoder od niza slojeva za dekonvoluciju. Nakon dekonvolucije, mreža daje procjenu optičkog toka na izlazu.



Slika 2.3: Arhitektura modela FlowNetS.

S druge strane, glavna prednost korištenja prostorne piramidalne mreže je prema [16] visoka učinkovitost — veličina modela i vrijeme rada prikladni su za praktičnu primjenu. Dodatno, prostorna piramidalna mreža je prilagođenija za procjenu optičkog protoka jer sadrži nekoliko klasičnih principa korištenih za rješavanje

problema optičkog toka, kao što su prostorna piramida, savijanje slike i naknadna obrada. Zahvaljujući korištenju ovih principa, točnost se znatno poboljšava.

Pristupi procjeni optičkog toka temeljeni na dubokom učenju su prema [22] nadmašili u potpunosti tradicionalne metode u točnosti i vremenu izvođenja. Modeli dubokog učenja rade u stvarnom vremenu i s puno većom točnošću. Međutim, metode dubokog učenja oslanjaju se na kvalitetu označenoga skupa podataka za učenje. To je veliki problem jer je za stvarne scene izuzetno nezgodno dobiti oznake optičkog toka.

2.3. Evaluacijske mjere

Dvije najkorištenije mjere pogreške u algoritmima optičkog toka su [22]:

1. *Endpoint error* (EE) – za odabrani piksel mjeri euklidsku udaljenost između predviđenog vektora optičkog toka (u, v) i stvarnog vektora (u_g, v_g) :

$$\text{EE} = \sqrt{(u - u_g)^2 + (v - v_g)^2} \quad (2.1)$$

2. *Angular error* (AE) – neka je $(u, v, 1)$ prošireni trodimenzijski vektor predviđenog optičkog toka za odabrani piksel, a $(u_g, v_g, 1)$ prošireni vektor točnog optičkog toka. AE odgovara kutu između ta dva vektora:

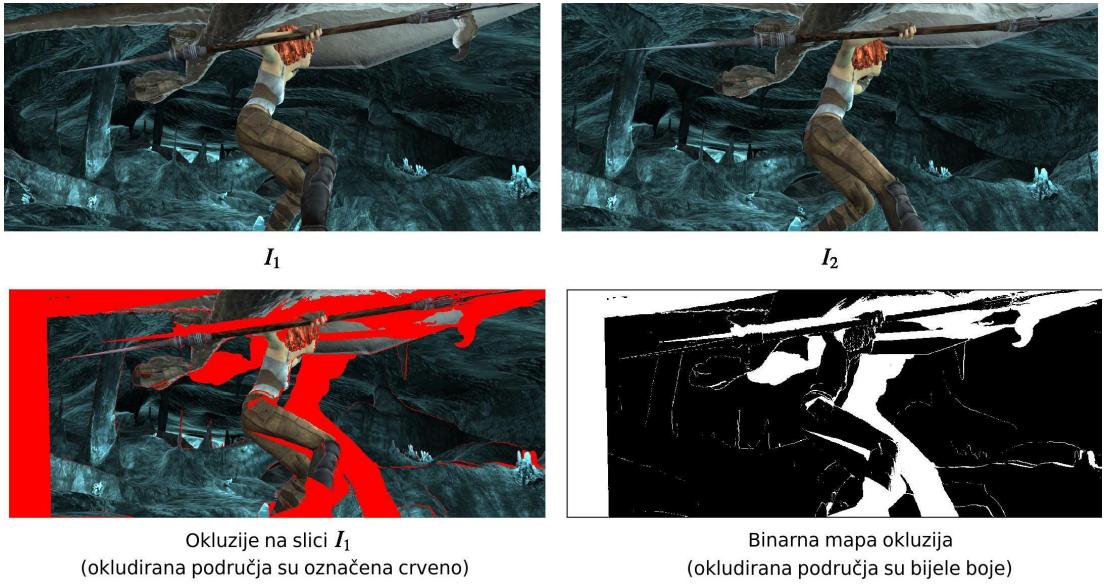
$$\text{AE} = \cos^{-1} \left(\frac{u \cdot u_g + v \cdot v_g + 1.0}{\sqrt{(u^2 + v^2 + 1.0)(u_g^2 + v_g^2 + 1.0)}} \right) \quad (2.2)$$

AE je prikladniji za male pomake i skloniji je podcjenjivanju velikih pomaka piksela, dok je EE prikladniji za velike pomake i zato se uglavnom koristi u radi s naprednim skupovima podataka (poput MPI-Sintela, KITTI-2012 i KITTI-2015) [22]. Prosječne vrijednosti ovih dviju pogrešaka nazivaju se *average endpoint error* (AEE) i *average angular error* (AAE). Iznimno, skup KITTI-2015 koristi drugčiju mjeru pogreške, opisanu u pododjeljku 2.4.3.

2.4. Skupovi podataka

U novijim skupovima podataka, kao što su MPI-Sintel [3], KITTI-2012 [7] i KITTI-2015 [20], pojavljuju se zahtjevniji izazovi za algoritme procjene optičkog toka [16]. Ti izazovi uključuju pojave okluzije, različite uvjete osvjetljenja i velike

pomake piksela. Primjer okludiranih područja iz skupa MPI-Sintel prikazan je na slici 2.4. DD metode su općenito uspješnije od KD metode na ovim skupovima podataka [16]. U nastavku slijedi kratak pregled spomenuta tri skupa podataka te skupova podataka FlyingChairs [5], FlyingThings [19] i HD1K [15].



Slika 2.4: Primjer okludiranih područja u sceni iz skupa podataka MPI-Sintel. [6]

2.4.1. MPI-Sintel

MPI-Sintel [3] je sintetički skup podataka temeljen na animiranom filmu. Sadrži mnogo velikih pomaka piksela, uključujući brze pomake malih predmeta. Scene su slične onima iz stvarnog svijeta jer sadrže kretanje tijela koja nisu kruta, složenu strukturu scene, varijacije osvjetljenja i sjene, složene materijale sa zrcalnim odrazima i atmosferske efekte poput magle. MPI-Sintel sadrži ukupno 1628 slika dimenzija 1024×436 podijeljenih u skup za učenje veličine 1064 slike i skup za testiranje veličine 564. Skup za učenje ima popratne oznake gustog optičkog toka, a oznake za skup za testiranje autori nisu javno objavili nego ih koriste za evaluaciju rješenja koja je moguće predati putem njihove [web stranice](#). Postoje dvije verzije skupa MPI-Sintel: *clean* i *final*. *Final* verzija, za razliku od *clean* verzije, dodaje svjetlosnom i cjelokupnom izgledu scene atmosferske efekte, zamućenje dubine polja (*depth of field blur*), zamućenje pokreta (*motion blur*), korekciju boje i druga umjetnička uljepšavanja. Ta je verzija slična objavljenom filmu. Primjeri iz ovog skupa podataka dani su na slici 2.5.



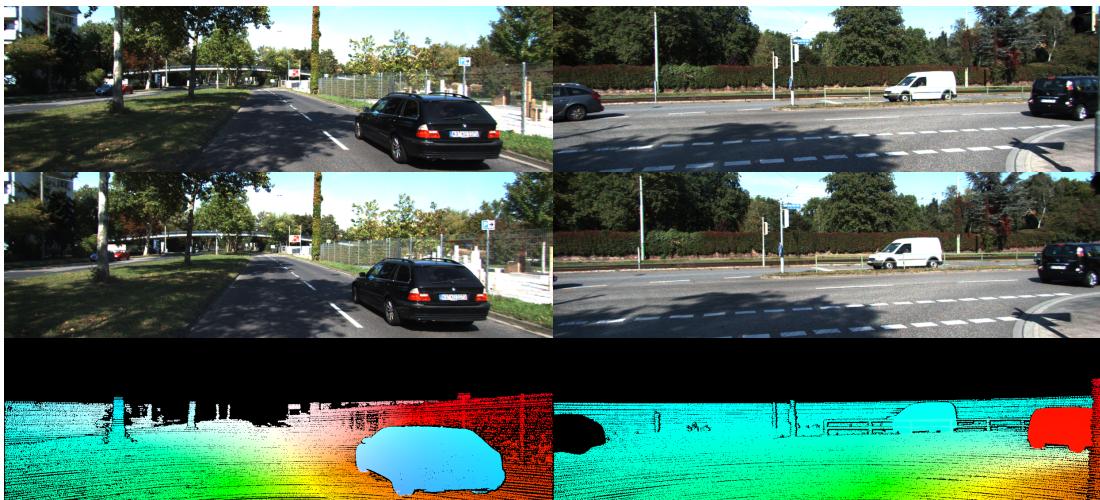
Slika 2.5: Primjeri podataka iz skupa podataka MPI-Sintel. U prvom retku je prva slika iz slijeda, u drugom druga, a u trećem je oznaka gustog optičkog toka koja odgovara tom nizu.

2.4.2. KITTI-2012

KITTI-2012 [7] je skup podataka iz stvarnog svijeta koji je prikupljen snimanjem iz automobila. Sadrži 389 slika veličine 1226×370 od kojih je 194 u skupu za učenje i ima popratne oznake rijetkog optičkog toka te 195 slika u skupu za testiranje.

2.4.3. KITTI-2015

KITTI-2015 [20] je skup koji je također snimljen iz automobila u pokretu. Sadrži 400 slika veličine 1242×375 od kojih je 200 u skupu za učenje i ima popratne oznake rijetkog optičkog toka, a preostalih 200 slika je u skupu za testiranje. Primjeri iz ovog skupa podataka prikazani su na slici 2.6. Ovaj skup podataka koristi drukčiju mjeru pogreške, Fl-all, koja mjeri postotak piksela za koje je procijenjena vrijednost optičkog toka loša. Pritom se procjena smatra dobrom ako je euklidska udaljenost između predviđenog i točnog optičkog toka manja od 3 piksela, odnosno $\text{EE} < 3$, ili ako je pogreška manja od 5% duljine vektora točnog optičkog toka, tj. $\frac{\text{EE}}{\sqrt{u_g^2 + v_g^2 + \epsilon}} < 0.05$ (ϵ je mali broj korišten zbog numeričke stabilnosti odnosno zbog izbjegavanja dijeljenja s nulom). Fl-all se računa samo nad pikselima koji imaju točnu oznaku optičkog toka (jer su popratne oznake dane samo za rijetki optički tok, ne gusti) i uzima se kao prosjek nad svih 200 slika skupa za testiranje.



Slika 2.6: Primjeri podataka iz skupa podataka KITTI-2015. U trećem retku je oznaka gustog optičkog toka određena na temelju slika iz prva dva retka.

2.4.4. HD1K

HD1K [15] je nesintetički skup podataka koji sadrži 28 504 stereo parova slike s popratim oznakama optičkog toka i toka scene.

2.4.5. FlyingChairs

FlyingChairs [5] je sintetički skup podataka koji sadrži veliku količinu generiranih scena u kojima stolice lete nad slučajno odabranim pozadinama s Flickr. Podatci nisu slični stvarnom svijetu, ali ih se može generirati proizvoljno mnogo. Konvolucijske mreže učene na tim podacima iznenađujuće dobro generaliziraju na realnim skupovima podataka, čak i bez dodatnog ugađanja modela (postupka poznatog kao *fine-tuning*). Skup sadrži 22 232 slike za učenje i 640 slika za testiranje. Slike su dimenzija 512×384 i imaju popratne oznake gustog optičkog toka. Primjeri iz ovog skupa podataka su na slici 2.7.

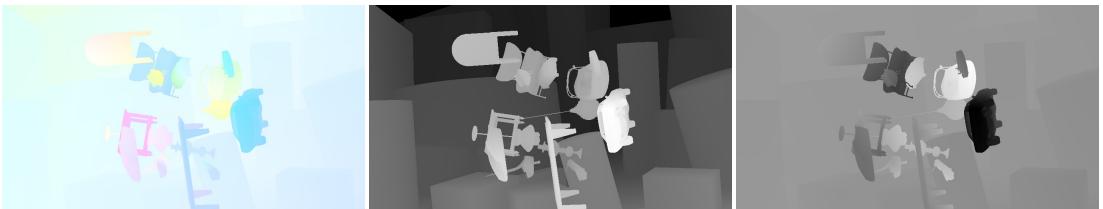
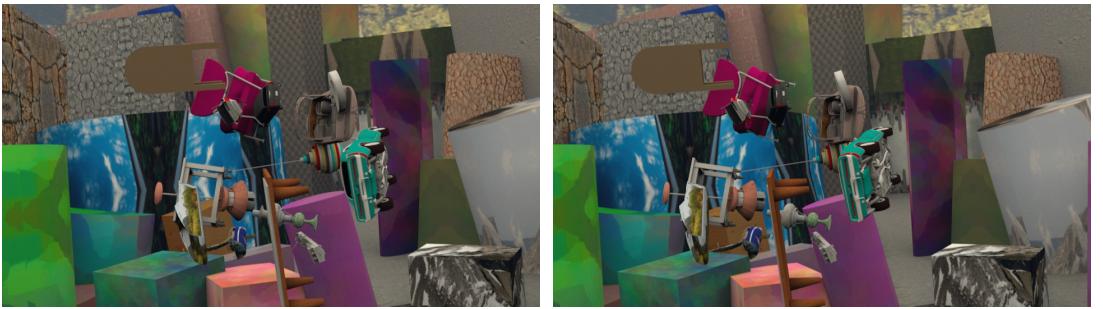
2.4.6. FlyingThings

FlyingThings [5] je također sintetički skup podataka koji sadrži 25 000 stereo parova slike dimenzija 960×540 s pripadnim oznakama gustog toka scene (*scene flow*). Postoji 2247 različitih scena koje su pretežito sačinjene od svakodnevnih objekata poput kauča, stola i stolice koji lete duž slučajne 3D putanje. U skupu za učenje je 21 818 podataka, a preostalih 4248 je u skupu za testiranje. FlyingThings se može koristiti za treniranje konvolucijskih neuronskih mreža koje

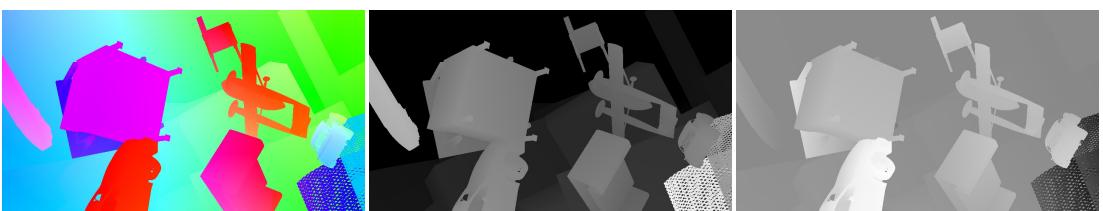


Slika 2.7: Primjeri podataka iz sintetičkog skupa podataka FlyingChairs. U trećem retku je oznaka gustog optičkog toka određena na temelju slika iz prva dva retka.

predviđaju optički tok, tok scene (*scene flow*) i disparitet (*disparity*, zadatak u kojem je potrebno za svaki piksel lijeve slike stereo para slika odrediti horizontalni pomak tog piksela do istovjetnog na desnoj slici). Primjeri stereo parova slika iz ovog skupa prikazani su na slici 2.8.



(a)



(b)

Slika 2.8: Dva primjera (a i b) stereo parova slika iz sintetičkog skupa podataka FlyingThings s popratnim oznakama, preuzeta [19]. Prvi red pokazuje stereo par slika, a u drugom redu su pripadne oznake optičkog toka (lijevo), dispariteta (sredina) i promjene dispariteta (*disparity change image* – mjeri koliko se disparitet promijenio za dva uzastopna para stereo slika; slika desno).

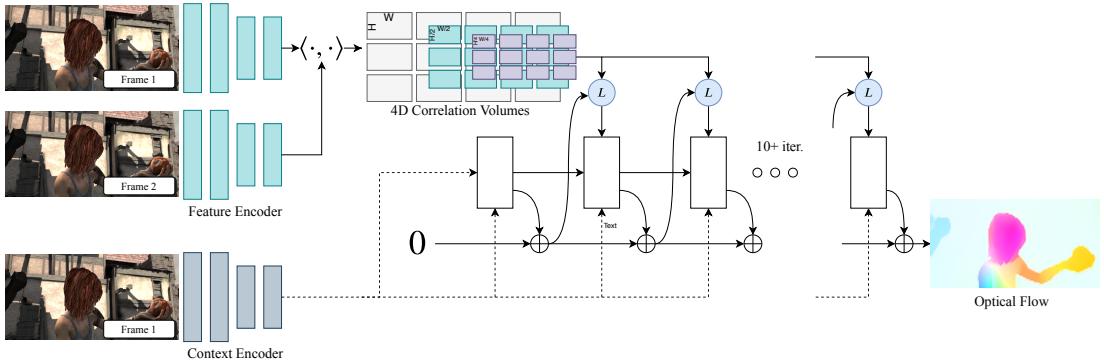
3. Duboki model RAFT

Znanstveni rad u kojem je duboki povratni model RAFT predstavljen odabran je za najbolji rad na konferenciji European Conference on Computer Vision 2020 (ECCV2020). U to vrijeme RAFT je postizao najveću uspješnost u procjeni optičkog toka na skupovima podataka MPI-Sintel i KITTI-2015. Model se odlikuje visokom sposobnošću generalizacije na neviđenim podacima – model koji je predtreniran na sintetičkim podatcima postiže Fl-all od 5.04% na podskupu za treniranje skupa KITTI-2015. Model se odlikuje i vrsnom efikasnošću [26] u vidu brzine izvođenja – na grafičkoj kartici GTX 1080Ti za slike dimenzija 1080×436 brzina predviđanja optičkog toka iznosi 10 sličica u sekundi. Manja verzija modela RAFT (RAFT-S) koja ima pet puta manje parametara obrađuje 20 sličica u sekundi.

Arhitektura modela RAFT je prikazana na slici 3.1, a sastoji se od tri komponente. Svakoj komponenti odgovara jedna faza u radu modela: (1) vađenje značajki iz ulaznih slika, (2) računanje vizualne sličnosti ulaznih slika i (3) iterativna ažuriranja procijenjenog optičkog toka. U narednim odjeljcima dan je detaljniji opis svake faze.

3.1. Vađenje značajki

Za dani par RGB slika I_1, I_2 cilj je previdjeti gusti optički tok $(\mathbf{f}^1, \mathbf{f}^2)$ koji svakom pikselu (u, v) u I_1 pridjeljuje pripadne koordinate $(u', v') = (u + f^1(u), v + f^2(v))$ u I_2 . Prva faza u tome je vađenje značajki iz I_1 i I_2 pomoću konvolucijskih slojeva. Značajke se vade pomoću kodera značajki (*feature encoder*) g_θ koji je izdvojeno prikazan na slici 3.2, a sastoji se od 6 rezidualnih blokova. Prva dva bloka rade sa slikama dva puta smanjene rezolucije, sljedeća dva bloka sa slikama četiri puta manje rezolucije i posljednja dva s osam puta manjom rezolucijom. Prema tome, ekstraktor značajki se može gledati kao preslikavanje $\mathbb{R}^{H \times W \times 3} \mapsto \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times D}$, gdje



Slika 3.1: Arhitektura modela RAFT se sastoji od tri komponente: (1) Koder značajki (*feature encoder*) koji vadi značajke iz svake slike pojedinačno te koder konteksta (*context encoder*) koji zasebno vadi značajke iz prve slike. (2) Korelacijski sloj koji stvara četverodimenzijski $\frac{W}{8} \times \frac{H}{8} \times \frac{W}{8} \times \frac{H}{8}$ tenzor korelacije kao skalarni umnožak piksela prve slike s pikselima druge slike. (3) Komponenta temeljena na modificiranoj verziji povratne neuronske mreže *Gated Recurrent Unit* koja iterativno ažurira procjenu optičkog toka koristeći trenutnu procjenu optičkog toka za pretraživanje tenzora korelacija. [26]

D iznosi 256. Dodatno postoji koder konteksta (*context encoder*) h_θ koji ima istu arhitekturu kao i koder značajki, ali se upotrebljava samo za vađenje značajki iz I_1 i koristi drukčije slojeve normalizacije – koder značajki koristi *instance normalization*, a koder konteksta koristi *batch normalization*.

Koder značajki g_θ i koder konteksta h_θ zajedno čine fazu vađenja značajki. Ova se faza odvija samo jedanput.

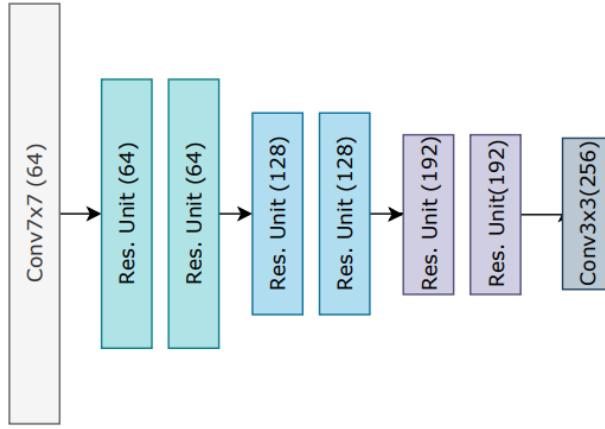
3.2. Računanje vizualne sličnosti

Neka je $H' = \frac{H}{8}$ i $W' = \frac{W}{8}$. U drugoj fazi računaju se skalarni umnošci između svaka dva para iz mapa značajki tako da je prvi element para iz $g_\theta(I_1) \in \mathbb{R}^{H' \times W' \times D}$, a drugi element iz $g_\theta(I_2) \in \mathbb{R}^{H' \times W' \times D}$. Rezultat toga je četverodimenzijski tenzor korelacije \mathbf{C} :

$$\mathbf{C}(g_\theta(I_1), g_\theta(I_2)) \in \mathbb{R}^{H' \times W' \times H' \times W'} \quad (3.1)$$

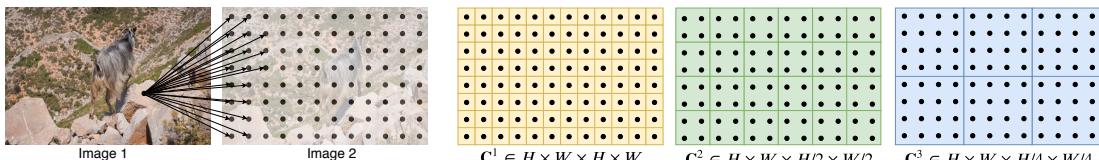
$$C_{ijkl} = \sum_h g_\theta(I_1)_{ijh} \cdot g_\theta(I_2)_{klh} \quad (3.2)$$

Iz \mathbf{C} se stvara četveroslojna piramida $\{\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3, \mathbf{C}^4\}$ tako da se provodi sažimanje prosječnom vrijednošću nad posljednje dvije dimenzije, jezgrama veličine $\{1, 2, 4, 8\}$ i pripadnim pomakom (iznosa jednakog veličini jezgre) kao što je pri-



Slika 3.2: Arhitektura ekstraktora značajki. Koder značajki i koder kontekta imaju ovu arhitekturu. Ukupno je 6 rezidualnih blokova, a svaki rezidualni blok sadrži dva konvolucijska sloja odnosno tri ako dolazi se nakon tog bloka smanjuje rezoluciju. Strelice označavaju da je u posljednjem sloju došlo do smanjenja rezolucije. Rezolucija se smanji tako da se pomak (*stride*) u neposrednom konvolucijskom sloju postavi na 2. [26]

kazano na slici 3.3. \mathbf{C}^k je dakle dimenzija $H' \times W' \times \frac{H'}{2^{k-1}} \times \frac{W'}{2^{k-1}}$. Ova piramida sadrži informacije i o velikim i o malim pomacima piksela. Te informacije su sačuvane na visokoj rezoluciji $H' \times W'$ jer prve dvije dimenzije nisu sudjelovale u sažimanju.



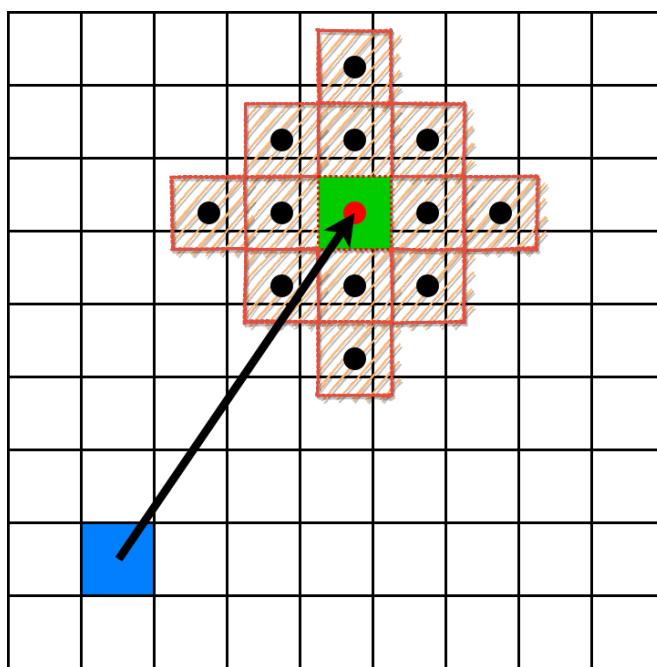
Slika 3.3: Na slici je za jedan odabrani piksel ulazne slike prikazana vizualizacija tensora korelacije odnosno slojeva piramide nastalih sažimanjem prosječnom vrijednošću iz tensora korelacijske. Odabrani piksel se gleda u paru sa svim pikselima iz druge slike te se računa skalarni umnožak svaka dva piksela. Valja primijetiti da se u stvarnom modelu ne gleda korelacija nad pikselima ulaznih slika, nego nad vektorima iz mapa značajki koje se dobiju iz ulaznih slika kao izlaz kodera značajki. [26]

Stvorena piramida će se indeksirati na poseban način. Za pojedini piksel $x = (u, v)$ gledat će se gdje trenutni optički tok \mathbf{f} predviđa poziciju tog piksela $x' = (u + f^1(u, v), v + f^2(v))$. Oko predviđene pozicije x' se gleda lokalno susjedstvo na

cjelobrojnim udaljenostima u zadanom radijusu r , mjereno L1 udaljenošću:

$$\mathcal{N}(x')_r = \{x' + dx \mid dx \in \mathbb{Z}^2, \|dx\|_1 \leq r\} \quad (3.3)$$

Vizualizacija ovog operatora prikazana je na slici 3.4. Slična se susjedstva pro-nalaze i za niže slojeve piramide. Pronađena lokalna susjedstva se koriste za indeksiranje piramide na svim razinama i spajanje dobivenih vrijednosti u kore-lacijske značajke koje su potrebne kao ulaz u svakoj iteraciji sljedeće faze, a za svaku iteraciju se korelacijske značajke računaju iznova jer se procjena optičkog toka nakon svake iteracije ažurira. Autori ovaj operator nazivaju *Lookup* operator L_C .



Slika 3.4: Vizualizacija operatora koji računa lokalno susjedstvo na temelju trenutne procjene optičkog toga. Zeleni piksel x' je procijenjena pozicija plavog piksela x' na temelju trenutne procjene optičkog toka. Označene točke se uzimaju kao lokalno susjedstvo $\mathcal{N}(x')_r$. Na temelju koordinata točaka iz $\mathcal{N}(x')_r$ se bilinearnom interpolacijom vade korelacijske značajke iz korelacijskih tenzora $\{\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3, \mathbf{C}^4\}$.

3.3. Iterativna ažuriranja procjene toka

U posljednjoj fazi se kreće od nekog početno inicijaliziranog optičkog toka \mathbf{f}_0 . Početna vrijednost toka \mathbf{f}_0 su ili nule, $\mathbf{f}_0 = \mathbf{0}$, ili je vrijednost optičkog toka od prethodnog para slika ako je riječ o slijedu slika iz videa. Potonji slučaj autori nazivaju *warm-start*.

Zatim se procijenjeni optički tok iterativno ažurira. Taj iterativni postupak opnaša korake tradicionalnog optimizacijskog algoritma, ali značajke i *a priori* distribucija optičkog toka nisu ručno izrađeni, već ih se uči pomoću kodera značajki i kroz parametre unutar iterativnog postupka. Svaka iteracija t računa $\Delta\mathbf{f}$ i novu procjenu optičkog kao $\mathbf{f}_t = \mathbf{f}_{t-1} + \Delta\mathbf{f}$. U svakoj iteraciji se kao ulaz x_t prima spoj sljedećih dijelova:

1. prethodna procjena optičkog toka \mathbf{f}_{t-1}
2. značajke $h_\theta(I_1)$ koje je izvadio koder konteksta
3. korelacijske značajke izvađene iz piramide $\{\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3, \mathbf{C}^4\}$ na temelju procjene optičkog toka \mathbf{f}_{t-1}

Ove ulaze prima modificirana verzija povratne neuronske mreže *Gated Recurrent Unit* (ConvGRU) prikazana na slici 3.5. Modificirana je zato što tradicionalni GRU koristi potpuno povezane slojeve, a u ConvGRU oni su zamijenjeni konvolucijskim slojevima. ConvGRU ažurira skriveno stanje h_t prema sljedećim formulama:

$$z_t = \sigma(\text{Conv}_{3x3}([h_{t-1}, x_t], W_z)) \quad (3.4)$$

$$r_t = \sigma(\text{Conv}_{3x3}([h_{t-1}, x_t], W_r)) \quad (3.5)$$

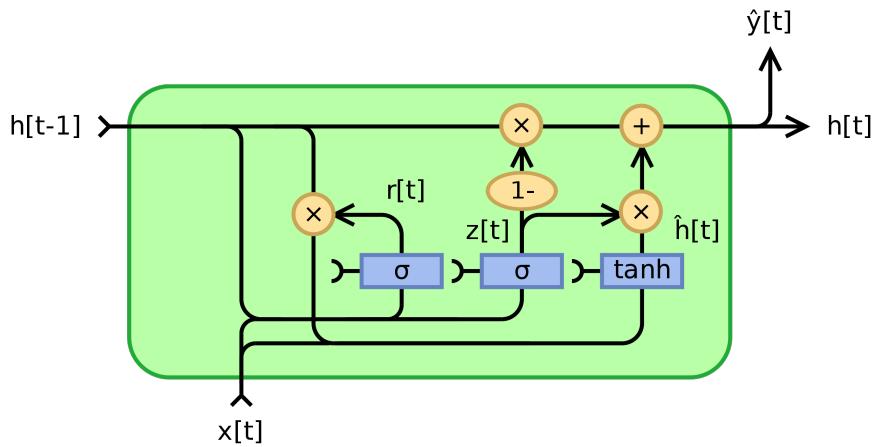
$$\tilde{h}_t = \tanh(\text{Conv}_{3x3}([r_t \odot h_{t-1}, x_t], W_h)) \quad (3.6)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (3.7)$$

Da bi se ažurirao optički tok \mathbf{f}_{t-1} , skriveno stanje h_t se provlači kroz dva konvolucijska sloja da bi se dobio $\Delta\mathbf{f}$, što se je prikazano na slici 3.1. Skriveno stanje h_t inicijalizira se značajkama $h_\theta(I_1)$ kodera konteksta.

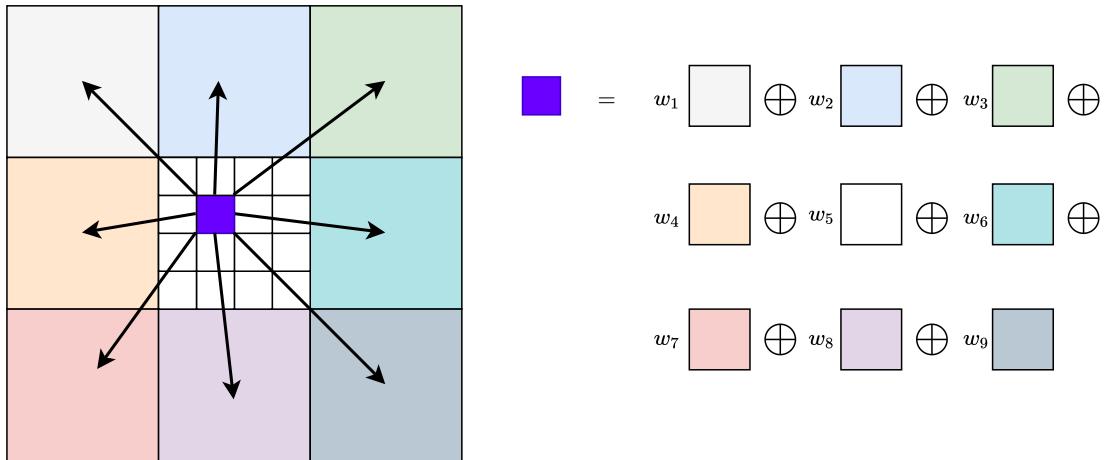
Umjesto ConvGRU koji koristi Conv_{3x3} slojeve, može se koristiti niz od dva ConvGRU bloka tako da prvi blok koristi Conv_{1x5} slojeve, a drugi Conv_{5x1} . Model RAFT koristi ovakav niz od dva bloka, a model RAFT-S koristi jedan blok s Conv_{3x3} slojevima. Detaljan prikaz razlika između arhitekture modela RAFT i manjeg modela RAFT-S je prikazan na slici 3.8.

Procjene optičkog tok $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N\}$ nisu rezolucije $H \times W$ željenog optičkog toka koji je potrebno procijeniti, nego su rezolucije $\frac{H}{8} \times \frac{W}{8}$. Stoga se procjene procesom naduzorkovanja proširuju na potrebnu rezoluciju. U RAFT-u se provodi konveksno naduzorkovanje prikazano na slici 3.6 s težinama koje predviđa posebna konvolucijska neuronska mreža koja na ulaz prima predviđeni optički



Slika 3.5: Na slici [4] je općeniti prikaz *Gated Recurrent Unit* mreže. U modelu RAFT ne postoji skicirani izlaz $\hat{y}[t]$, nego se samo skriveno stanje $h[t]$ koristi. Dodatno, na mjestu potpuno povezanih slojeva su konvolucijski slojevi.

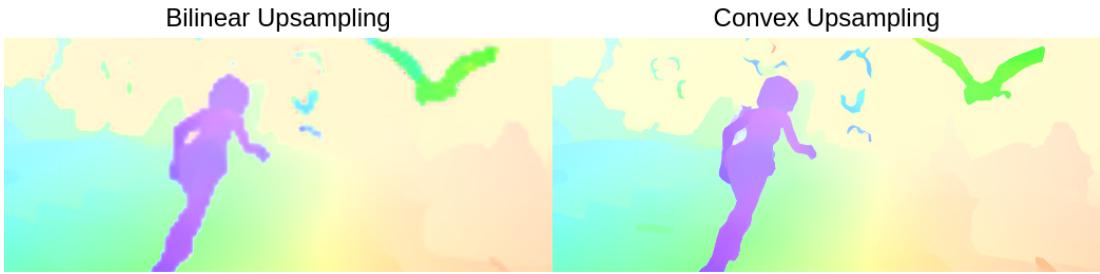
tok. Usporedba rezultata ovakvog konveksnog naduzorkovanja s rezultatima naduzorkovanja bilinearnom interpolacijom dan je na slici 3.7.



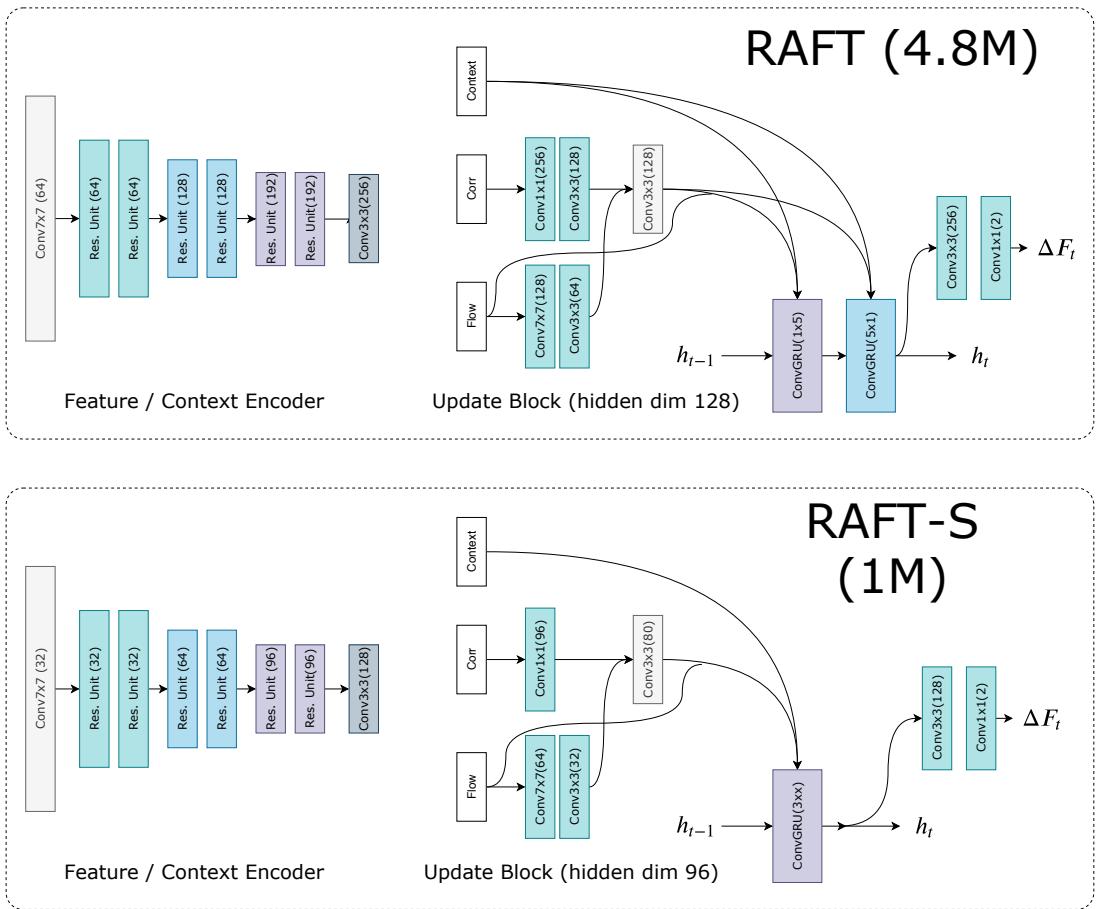
Slika 3.6: Prikaz korištenog konveksnog naduzorkovanja. Svaki piksel u naduzorkovanom optičkom toku visoke rezolucije (slika lijevo) se računa kao konveksna kombinacija 9 susjednih piksela optičkog toka niže rezolucije. Težine koje se u konveksnoj kombinaciji (slika desno) računa posebna konvolucijska neuronska mreža. [26]

3.4. Funkcija gubitka

Za učenje parametara modela, RAFT koristi funkciju gubitka definiranu kao sumu L1 udaljenost između točne oznake optičkog toka \mathbf{f}_g ("g" kao *ground truth*) i svakog procijenjenog optičkog toka $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N\}$, s eksponencijalno rastućom



Slika 3.7: Lijevo je primjer naduzorkovanja bilinearnom interpolacijom, a desno prikaz konveksnog naduzorkovanja koje koristi RAFT. Vidljivo je povećana točnost interpolacije pri rubnim područjima objekata. [26]



Slika 3.8: Arhitektura modela RAFT i RAFT-S. RAFT sadrži 4.8 milijuna (M) parametara (odnosno 5.3M uračunaju li se slojevi za naduzorkovanje), a RAFT-S sadrži 1.0M parametara. RAFT koristi dva ConvGRU bloka – prvi koristi konvolucijske slojeve Conv_{1x5}, a drugi Conv_{5x1}. RAFT-S koristi *bottleneck* rezidualne blokove i koristi samo jedan ConvGRU blok s konvolucijskim slojevima Conv_{3x3}. [26]

težinom γ (koju su autori postavili na 0.8):

$$\mathcal{L} = \sum_{i=1}^N \gamma^{N-i} \|\mathbf{f}_g - \mathbf{f}_i\|_1 \quad (3.8)$$

Rastuća težina γ osigurava da se pogreška u izlaznoj procjeni optičkog toka više kažnjava od privremenih procjena. L1 gubitak je robustan [23].

4. Eksperimenti autora

Autori rada [26] implementiraju RAFT u programskom okviru PyTorch. Tijekom postupka učenja koriste optimizacijski algoritam AdamW [17] (naziv dolazi od *Adam with decoupled weight decay*) i heuristiku odsijecanja gradijenata (*gradient clipping*) na raspon $[-1, 1]$. Pri učenju na skupu MPI-Sintel, provode 32 iteracije ažuriranja procjene optičkog toka, a na skupu KITTI-2015 provode 24 iteracije. Gradijent se u $\Delta\mathbf{f} + \mathbf{f}_t$ pri unatražnom prolazu ne provodi kroz granu \mathbf{f}_t nego samo granom $\Delta\mathbf{f}$.

U nastavku slijedi opis rezultata autora i studija ablacija koju su proveli.

4.1. Rezultati

U [26] se testiranje provodi na skupovima MPI-Sintel i KITTI-2015. Tablica 4.1 je preuzeta iz originalnog rada i prikazuje usporedbu uspješnosti različitih metoda na tim skupovima podataka. Postoje novije metode koje postižu bolju performanse od modela RAFT, ali one nisu prikazane u ovoj tablici jer nisu bile poznate kada se je originalni rad objavio. U svoje vrijeme, RAFT je postizao najveću uspješnost (*state-of-the-art*) u procjeni optičkog toka na ta dva skupa. Neke procjene optičkog toka koje je daje model RAFT su prikazane na slikama 4.1

4.2. Studija ablacija

Da bi se pokazala relativna važnost svake komponente, u radu [26] se provodi niz eksperimenta u kojima se pojedinačna komponenta odstrani te se gleda naknadna uspješnost modela. U svim tim eksperimentima je model učen na skupovima FlyingChairs i FlyingThings (C+T). Rezultati su prikazani u tablici 4.2.

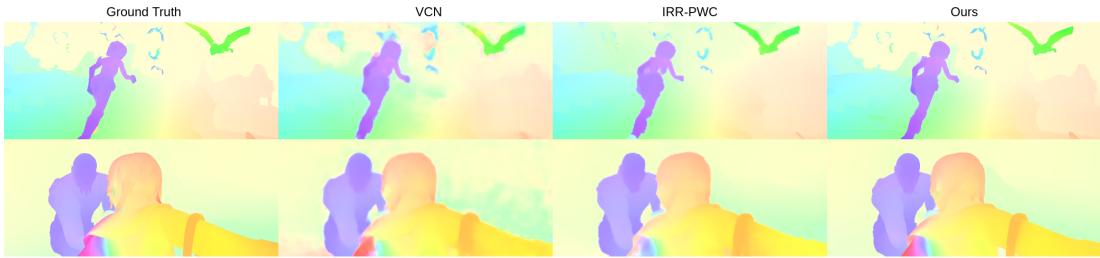
Pokazalo se da korištenje ConvGRU modula s dijeljenim težinama i s kontekstom

Training Data	Method	MPI-Sintel (train)		KITTI-2015 (train)		MPI-Sintel (test)		KITTI-2015 (test)	
		Clean	Final	EE	Fl-all	Clean	Final	Fl-all	
-	FlowFields [1]	-	-	-	-	3.75	5.81	15.31	
-	FlowFields++ [21]	-	-	-	-	2.94	5.49	14.82	
S	DCFlow [28]	-	-	-	-	3.54	5.12	14.86	
S	MRFlow [27]	-	-	-	-	2.53	5.38	12.19	
C + T									
C + T	HD3 [30]	3.84	8.77	13.17	24.0	-	-	-	
	LiteFlowNet [10]	2.48	4.04	10.39	28.5	-	-	-	
	PWC-Net [24]	2.55	3.93	10.35	33.7	-	-	-	
	LiteFlowNet2 [11]	2.24	3.78	8.97	25.9	-	-	-	
	VCN [29]	2.21	3.68	8.36	25.1	-	-	-	
	MaskFlowNet [31]	2.25	3.61	-	<u>23.1</u>	-	-	-	
C+T+S/K									
C+T+S/K	FlowNet2 [13]	(1.45)	(2.01)	(2.30)	(6.8)	4.16	5.74	11.48	
	HD3 [30]	(1.87)	(1.17)	(1.31)	(4.1)	4.79	4.67	6.55	
	IRR-PWC [12]	(1.92)	(2.51)	(1.63)	(5.3)	3.84	4.58	7.65	
	ScopeFlow [2]	-	-	-	-	<u>3.59</u>	<u>4.10</u>	<u>6.82</u>	
	RAFT	(0.77)	(1.20)	(0.64)	(1.5)	2.08	3.41	5.27	
C+T+S+K+H									
C+T+S+K+H	LiteFlowNet2 ¹ [11]	(1.30)	(1.62)	(1.47)	(4.8)	3.48	4.69	7.74	
	PWC-Net+ [25]	(1.71)	(2.34)	(1.50)	(5.3)	3.45	4.60	7.72	
	VCN [29]	(1.66)	(2.24)	(1.16)	(4.1)	2.81	4.40	6.30	
	MaskFlowNet [31]	-	-	-	-	2.52	4.17	<u>6.10</u>	
	RAFT	(0.76)	(1.22)	(0.63)	(1.5)	<u>1.94</u>	<u>3.18</u>	5.10	
	RAFT (<i>warm-start</i>)	(0.77)	(1.27)	-	-	1.61	2.86	-	

Tablica 4.1: Rezultati na skupovima MPI-Sintel i KITTI-2015. Za MPI-Sintel su prikazane mjere EE pogreške. Tablica je preuzeta iz originalnog rada [26] i prikazuje usporedbu uspješnosti različitih metoda na tim skupovima podataka. Postoje novije metode koje postižu bolju performanse od modela RAFT, ali one nisu prikazane u ovoj tablici jer nisu bile poznate kada se je originalni rad objavio. U svoje vrijeme, RAFT je postizao najveću uspješnost (*state-of-the-art*) u procjeni optičkog toka na ta dva skupa. U tablici je prikazana uspješnost za različite skupove na kojima su modeli učili – FlyingChairs (C), FlyingThing (T), MPI-Sintel (S), KITTI-2015 (K) i HD1K (H). Modeli trenirani na C+T nisu dodatno ugađani (*fine-tuning*) na skupovima na kojima se testiraju, pa se na svojevrstan način testira sposobnost generalizacije modela učenog samo na sintetičkim skupovima. Učenje na C+T+S/K znači da se C+T model dodatno ugađao odnosno učio na skupu za učenje onog skupa podataka na kojemu će biti testiran (npr. na MPI-Sintel(train) ako je testiran na MPI-Sintel(test)). Modeli učeni na C+T+S+K+H su koristili svih pet skupova za učenje. (¹ Rezultat je prikazan bez treniranja na HD1K jer se u [11] pokazalo da korištenje skupa HD1K nije puno pomoglo, zbog čega je objavljen rezultat bez njega.)

Experiment	Method	MPI-Sintel (train)		KITTI-2015 (train)		Parameters
		Clean	Final	EE	Fl-all	
<i>Reference Model</i> (bilinear upsampling), Training: 100k(C) → 60k(T)						
Update Op.	<u>ConvGRU</u>	1.63	2.83	5.54	19.8	4.8M
	Conv	2.04	3.21	7.66	26.1	4.1M
Tying	Tied Weights	1.63	2.83	5.54	19.8	4.8M
	Untied Weights	1.96	3.20	7.64	24.1	32.5M
Context	<u>Context</u>	1.63	2.83	5.54	19.8	4.8M
	No Context	1.93	3.06	6.25	23.1	3.3M
Feature Scale	<u>Single-Scale</u>	1.63	2.83	5.54	19.8	4.8M
	Multi-Scale	2.08	3.12	6.91	23.2	6.6M
Lookup Radius	0	3.41	4.53	23.6	44.8	4.7M
	1	1.80	2.99	6.27	21.5	4.7M
	2	1.78	2.82	5.84	21.1	4.8M
	<u>4</u>	1.63	2.83	5.54	19.8	4.8M
Correlation Pooling	No	1.95	3.02	6.07	23.2	4.7M
	<u>Yes</u>	1.63	2.83	5.54	19.8	4.8M
Correlation Range	32px	2.91	4.48	10.4	28.8	4.8M
	64px	2.06	3.16	6.24	20.9	4.8M
	128px	1.64	2.81	6.00	19.9	4.8M
	<u>All-Pairs</u>	1.63	2.83	5.54	19.8	4.8M
Features for Refinement	<u>Correlation</u>	1.63	2.83	5.54	19.8	4.8M
	Warping	2.27	3.73	11.83	32.1	2.8M
<i>Reference Model</i> (convex upsampling), Training: 100k(C) → 100k(T)						
Upsampling	<u>Convex</u>	1.43	2.71	5.04	17.4	5.3M
	Bilinear	1.60	2.79	5.17	19.2	4.8M
Inference Updates	1	4.04	5.45	15.30	44.5	5.3M
	3	2.14	3.52	8.98	29.9	5.3M
	8	1.61	2.88	5.99	19.6	5.3M
	<u>32</u>	1.43	2.71	5.00	17.4	5.3M
	100	1.41	2.72	4.95	17.4	5.3M
	200	1.40	2.73	4.94	17.4	5.3M

Tablica 4.2: Rezultati eksperimenata u provedenoj studiji ablacije, preuzeti iz [26]. Konfiguracija korištena u konačnom RAFT modelu je podcrtana. Za MPI-Sintel su prikazane mjere EE pogreške. Eksperimenti su detaljnije opisani u poglavljju 4.2



Slika 4.1: Procjene optičkog toka za različite modele na dvama primjerima iz skupa MPI-Sintel. U prvom stupcu je točna oznaka optičkog toka, a u sljedeća tri su procjene koje daju redom modeli VCN [29], IRR-PWC [12] i RAFT. [26]



Slika 4.2: Procjene optičkog toka koje je dao RAFT na primjerima iz skupa podataka KITTI-2015. [26]

daje bolje rezultate od modela koji:

1. ne koristi koder konteksta (i ne dovodi dodatne značajke izvadene kao kontekst iz prve slike) – eksperiment *Context* u tablici 4.2
2. koristi tri konvolucijska sloja i ReLU aktivacijski sloj umjesto ConvGRU modula – eksperiment *Update Op.*
3. ne koristi dijeljene težine u iterativnom ažuriranju nego ima toliko puta više parametara koliko se iteracija provodi – eksperiment *Tying*

Uloga radijusa korištenog u računanju lokalnog susjedstva $\mathcal{N}(x')_r$ je također bitna. Za radijus jednak 0 su rezultati značajno lošiji, a u testiranim konfiguracijama je radijus jednak 4 dao najbolje rezultate – eksperiment *Lookup Radius*.

Korišteni način računanja korelacijskih tenzora se pokazao bolji od konfiguracija u kojima se:

1. umjesto korelacija svih parova ulaznih značajki računaju korelacije za ograničen broj susjednih piksela – 32 piksela, 64 piksela i 128 piksela, eksperiment *Correlation Range*
2. umjesto jednog skupa korelacijskih tenzora izračunatog na temelju značajki

jedne rezolucije (a to je rezolucija $\frac{H}{8} \times \frac{W}{8}$ izlaznih značajki kodera konteksta) da se izgradi više skupova korelacijskih tenzora izračunatih na temelju različitih rezolucija – eksperiment *Feature Scale*

3. ukloni sažimanje kojima se stvore $\{\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3, \mathbf{C}^4\}$, nego se koristi samo \mathbf{C}^1 – eksperiment *Correlation Pooling*
4. umjesto korelacije koristi *warping* u kojem se značajke slike I_2 izobliče na temelju trenutne procjene toka i postave nad značajke slike I_1 da bi se onda usporedile razlike između njih – eksperiment *Features for Refinement*

Konveksno naduzorkovanje se pokazalo značajno boljim od bilinearog naduzorkovanja – eksperiment *Upsampling*. U eksperiment *Inference Updates* se gleda utjecaj korištenog broja iteracija u modulu ConvGRU i pokazuje da procjena optičkog toka ne divergira ni za konfiguracije s velikim brojem od 200 iteracija.

5. Zaključak

Duboko učenje poboljšalo je uspješnost mnogih zadataka računalnog vida među kojima je i procjena optičkog toka. Noviji skupovi podataka poput MPI-Sintel [3], KITTI-2012 [7] i KITTI-2015 [20] donose zahtjevnije izazove za algoritme procjene optičkog toka – pojave okluzije, različite uvjete osvjetljenja i velike pomake piksela [16]. U okviru ovih izazova, veću uspješnost postižu postupci procjene optičkog toka zasnovani na korištenju dubokog učenja [16] koji su u potpunosti nadmašili tradicionalne metode jer procjene provode u stvarnom vremenu i s puno većom točnošću [22].

RAFT [26] je duboki povratni model koji uvodi nekoliko noviteta među duboke modele za procjenu optičkog toka da bi postigao u svoje vrijeme najvišu uspješnost na skupovima MPI-Sintel i KITTI-2015. Ti noviteti uključuju (1) interni rad s jednom procjenom optičkog toka koja je visoke i fiksne rezulocije, (2) iterativni operator koji dijeli mal broj parametara (2.7M) kroz proizvoljan broj iteracija i (3) korištenje ConvGRU modula u kombinaciji s *Lookup* operatorm nad 4D tenzorom korelacija svih parova značajki ulaznih slika. Autori modela RAFT provode studiju ablacji koja opravdava konačnu konfiguraciju modela RAFT.

Optički tok igrao je važnu ulogu u različitim znanstvenim područjima [22], ali je i dalje otvoren problem [8]. Budući da je optički protok koristan za procjenu kretanja, dispariteta i semantičke korespondencije (*semantic correspondence*), poboljšanja u procjeni optičkog toka imaju izravnu korist za zadatke poput vizualne odometrije, stereo procjene dubine (*stereo depth estimation*) i praćenja objekta [14]. Daljnji razvoj metoda za procjenu optičkog toka može dodatno poboljšati praktičnu uporabljivost te uspješnost drugih zadataka računalnog vida. [8]

6. Literatura

- [1] Christian Bailer, Bertram Taetz, i Didier Stricker. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. U *Proceedings of the IEEE international conference on computer vision*, stranice 4015–4023, 2015.
- [2] Aviram Bar-Haim i Lior Wolf. Scopeflow: Dynamic scene scoping for optical flow. U *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, stranice 7998–8007, 2020.
- [3] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, i Michael J. Black. A naturalistic open source movie for optical flow evaluation. *Computer Vision – ECCV 2012*, stranice 611–625, 2012. doi: 10.1007/978-3-642-33783-3_44. URL https://link.springer.com/chapter/10.1007/978-3-642-33783-3_44.
- [4] Wikimedia Commons. Gradient recurrent unit, fully gated version, 2018. URL https://en.wikipedia.org/wiki/File:Gated_Recurrent_Unit,_base_type.svg.
- [5] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, i Thomas Brox. Flownet: Learning optical flow with convolutional networks. U *Proceedings of the IEEE international conference on computer vision*, stranice 2758–2766, 2015.
- [6] Denis Fortun, Patrick Bouthemy, i Charles Kervrann. Optical flow modeling and computation: A survey. *Computer Vision and Image Understanding*, 134:1–21, 2015. ISSN 1077-3142. doi: <https://doi.org/10.1016/j.cviu.2015.02.008>. URL <https://www.sciencedirect.com/science/article/pii/S1077314215000429>. Image Understanding for Real-world Distributed Video Networks.

- [7] A Geiger, P Lenz, C Stiller, i R Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32:1231–1237, 08 2013. doi: 10.1177/0278364913491297. URL <https://journals.sagepub.com/doi/full/10.1177/0278364913491297>.
- [8] Fatma Guney, Laura Sevilla-Lara, Deqing Sun, i Jonas Wulff. " what is optical flow for?": Workshop results and summary. U *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018.
- [9] Berthold K.P. Horn i Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 08 1981. doi: 10.1016/0004-3702(81)90024-2. URL <https://www.sciencedirect.com/science/article/pii/0004370281900242>.
- [10] Tak-Wai Hui, Xiaoou Tang, i Chen Change Loy. Liteflownet: A lightweight convolutional neural network for optical flow estimation. U *Proceedings of the IEEE conference on computer vision and pattern recognition*, stranice 8981–8989, 2018.
- [11] Tak-Wai Hui, Xiaoou Tang, i Chen Change Loy. A lightweight optical flow cnn—revisiting data fidelity and regularization. *arXiv preprint arXiv:1903.07414*, 2019.
- [12] Junhwa Hur i Stefan Roth. Iterative residual refinement for joint optical flow and occlusion estimation. U *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, stranice 5754–5763, 2019.
- [13] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, i Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. U *Proceedings of the IEEE conference on computer vision and pattern recognition*, stranice 2462–2470, 2017.
- [14] Rico Jonschkowski, Austin Stone, Jon Barron, Ariel Gordon, Kurt Konolige, i Anelia Angelova. What matters in unsupervised optical flow. *ECCV*, 2020. URL <https://arxiv.org/pdf/2006.04902.pdf>.
- [15] Daniel Kondermann, Rahul Nair, Katrin Honauer, Karsten Krispin, Jonas Andrulis, Alexander Brock, Burkhard Gussefeld, Mohsen Rahimimoghadam, Sabine Hofmann, Claus Brenner, et al. The hci benchmark suite: Stereo and flow ground truth with uncertainties for urban autonomous driving.

U *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, stranice 19–28, 2016.

- [16] C. Lopez-Molina, C. Marco-Detchart, H. Bustince, i B. De Baets. A survey on matching strategies for boundary image comparison and evaluation. *Pattern Recognition*, 115:107883, Jul 2021. doi: 10.1016/j.patcog.2021.107883. URL <https://www.sciencedirect.com/science/article/pii/S0031320321000704>.
- [17] Ilya Loshchilov i Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [18] Bruce Lucas i Takeo Kanade. An iterative image registration technique with an application to stereo vision (ijcai). *svezak* 81, 04 1981.
- [19] Nikolaus Mayer, Eddy Ilg, Philip Häusser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, i Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. U *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, stranice 4040–4048, 2016. doi: 10.1109/CVPR.2016.438.
- [20] Moritz Menze i Andreas Geiger. Object scene flow for autonomous vehicles. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 06 2015. doi: 10.1109/cvpr.2015.7298925. URL <https://ieeexplore.ieee.org/document/7298925>.
- [21] René Schuster, Christian Bailer, Oliver Wasenmüller, i Didier Stricker. Flowfields++: Accurate optical flow correspondences meet robust interpolation. U *2018 25th IEEE International Conference on Image Processing (ICIP)*, stranice 1463–1467. IEEE, 2018.
- [22] Syed Tafseer Haider Shah i Xiang Xuezhi. Traditional and modern strategies for optical flow: an investigation. *SN Applied Sciences*, 3, 02 2021. doi: 10.1007/s42452-021-04227-x. URL <https://link.springer.com/article/10.1007/s42452-021-04227-x>.
- [23] Deqing Sun, Stefan Roth, i Michael J. Black. Secrets of optical flow estimation and their principles. U *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, stranice 2432–2439, 2010. doi: 10.1109/CVPR.2010.5539939.

- [24] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, i Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. U *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, stranice 8934–8943, 2018.
- [25] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, i Jan Kautz. Models matter, so does training: An empirical study of cnns for optical flow estimation. *arXiv preprint arXiv:1809.05571*, 2018.
- [26] Zachary Teed i Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow, 2020.
- [27] Jonas Wulff, Laura Sevilla-Lara, i Michael J Black. Optical flow in mostly rigid scenes. U *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, stranice 4671–4680, 2017.
- [28] Jia Xu, René Ranftl, i Vladlen Koltun. Accurate optical flow via direct cost volume processing. U *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, stranice 1289–1297, 2017.
- [29] Gengshan Yang i Deva Ramanan. Volumetric correspondence networks for optical flow. U *Advances in Neural Information Processing Systems*, stranice 793–803, 2019.
- [30] Zhichao Yin, Trevor Darrell, i Fisher Yu. Hierarchical discrete distribution decomposition for match density estimation. U *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, stranice 6044–6053, 2019.
- [31] Shengyu Zhao, Yilun Sheng, Yue Dong, Eric I Chang, Yan Xu, et al. Maskflownet: Asymmetric feature matching with learnable occlusion mask. U *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, stranice 6278–6287, 2020.

7. Sažetak

Ovaj seminar uvodi pojam optičkog toka, daje kratak pregled pristupa za procjenu optičkog toka i opisuje duboki povratni model RAFT koji služi za procjenu optičkog toka. Posebna pažnja posvećena je arhitekturi modela RAFT, a uz to su opisani eksperimenti originalnih autora. Znanstveni rad u kojem je duboki povratni model RAFT predstavljen odabran je za najbolji rad na konferenciji European Conference on Computer Vision 2020 (ECCV2020). U to vrijeme je RAFT postizao najveću uspješnost (*state-of-the-art*) u procjeni optičkog toka na skupovima podataka MPI-Sintel i KITTI-2015. Postignutu uspješnost duguje novitetima koje je uveo u duboke modele za procjenu optičkog toka.