

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 482

Korespondencijska ugrađivanja za stereoskopsku rekonstrukciju

David Kerman

Zagreb, svibanj 2022.

ZAVRŠNI ZADATAK br. 482

Pristupnik: **David Kerman (0036522014)**

Studij: Elektrotehnika i informacijska tehnologija i Računarstvo

Modul: Računarstvo

Mentor: prof. dr. sc. Siniša Šegvić

Zadatak: **Korespondencijska ugrađivanja za stereoskopsku rekonstrukciju**

Opis zadatka:

Stereoskopska rekonstrukcija važan je zadatak računalnog vida s mnogim zanimljivom primjenama. Tom problemu možemo pristupiti ugrađivanjem piksela slike u visokodimenzionalni metrički prostor koji je pogodan za ostvarivanje korespondencije. Ugrađivanje možemo naučiti na stereoskopskim slikama s poznatom dubinskom informacijom. U okviru rada, potrebno je odabrati okvir za automatsku diferencijaciju te upoznati biblioteke za rukovanje tenzorima i slikama. Proučiti i ukratko opisati postojeće pristupe za stereoskopsku rekonstrukciju koji se temelje na dubokom učenju. Odabrati slobodno dostupni skup slika te oblikovati podskupove za učenje, validaciju i testiranje. Predložiti prikladnu arhitekturu dubokog modela za korespondencijsko ugrađivanje. Uhodati postupke učenja modela i validiranja hiperparametara. Primijeniti naučene modele te prikazati i ocijeniti postignutu točnost. Radu priložiti izvorni i izvršni kod razvijenih postupaka, ispitne slijedove i rezultate, uz potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 10. lipnja 2022.

*Zahvaljujem mentoru prof. dr. sc. Siniši Šegviću, na savjetima i ukazanoj pomoći.
Također, zahvaljujem svojoj obitelji, prijateljima i kolegama na potpori tijekom mojeg
dosadašnjeg obrazovanja.*

SADRŽAJ

1. Uvod	1
2. Stereoskopska rekonstrukcija	2
2.1. Epipolarna geometrija	2
2.2. Pretprocesiranje slika	3
2.2.1. Rektifikacija	4
2.3. Disparitet	4
2.4. Algoritam stereoskopske rekonstrukcije	6
3. Duboko učenje	8
3.1. Osnovni pojmovi u dubokom učenju	8
3.1.1. Perceptron	8
3.1.2. Prijenosne funkcije	8
3.1.3. Konvolucijski slojevi	8
3.1.4. Optimizacija	8
3.1.5. Regularizacija	8
3.2. Učenje dubokih modela	8
3.2.1. Funkcija gubitka	8
3.2.2. Gradijentni spust	8
3.2.3. Memorijska zahtjevnost	8
4. Podatkovni skup za učenje	9
4.1. Podatkovni skup KITTI 2015	9
5. Duboko učenje u kontekstu stereoskopske rekonstrukcije	10
5.1. Ugrađivanje okana u metrički prostor	10
5.2. Učenje modela za ostvarivanje korespondencije	10
6. Eksperimentalni rezultati	11

7. Programska izvedba i vanjske biblioteke	12
8. Zaključak	13

1. Uvod

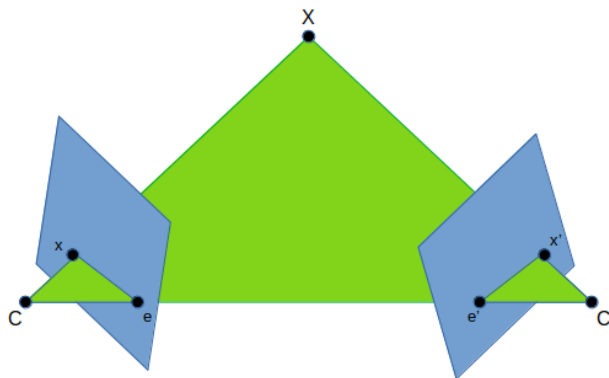
Računalni vid kao jedna od temeljnih grana umjetne inteligencije omogućava računalima da izluče značajnu informaciju iz slika, videa ili sličnih vizualnih ulaza. Područje stereoskopske rekonstrukcije ima za cilj ostvariti stvaran trodimenzionalan položaj točaka koje su promatrane u dvije ili više slika, te kao takvo ima vrlo široke primjene uključujući one za određivanje sličnosti slikovnih okana. Kroz povijest postojale su klasične metode za određivanje korespondencijskih metrika, no pojavom snažnijih računala iskorišteno je duboko učenje kao alat kojim je moguće korespondencijske metrike naučiti na stvarnim podacima. Za potrebe metoda temeljenih na ugrađivanju u visokodimenzionalni prostor potrebne su rektificirane stereoskopske slike s poznatom dubinskom informacijom.

U okviru ovog radu opisana je geometrija stereoskopskog para, kalibracija, rektifikacija i algoritmi korišteni za ostvarivanje korespondencije. Zatim definirani su osnovni pojmovi u dubokom učenju, te postupci korišteni u dubokom učenju uz memorijske zahtjeve. Predstavljen je podatkovni skup na kojem su izvršeni postupci treniranja i validacije. Opisan je model kojim je ostvareno korespondencijsko ugrađivanje slikovnih okana, koji su korišteni u stereoskopskoj rekonstrukciji. Nastavno na to, opisan je postupak učenja takvog modela, te eksperimentalno vrednovanje rekonstrukcijske točnosti.

2. Stereoskopska rekonstrukcija

2.1. Epipolarna geometrija

Epipolarna geometrija je geometrija koja ima primjenu u stereoskopskoj rekonstrukciji, a po definiciji opisuje odnos točaka u slikama dobivenih iz dviju kamera. Također naziva se i epipolarno ograničenje jer uvelike smanjuje broj točaka koje je potrebno provjeriti pri traženju korespodentnih točaka. Točka X koja je prikazana na slici 2.1 prikazuje točku u 3D prostoru koja se snima iz dviju kamera. Točke C i C' prikazuju centre lijeve, odnosno desne kamere. Projekcija točke X na ravninu lijeve kamere je x , a na ravninu desne kamere x' . Važno je napomenuti da su točke X , x i x' u istoj ravnini, te zajedno sa centrima kamera C i C' čine tzv. *epipolarnu ravninu* π .



Slika 2.1: Epipolarna ravnina u kojoj se nalaze točka X , te točke nastale projekcijom x i x'

Nadalje, na slici su označeni *epipolovi* oznakama e i e' te oni prikazuju točke u kojima pravac koji povezuje centre kamera C i C' siječe slikovne ravnine. Najvažniji element su *epipolarne linije*, označene dužinama koje spajaju x i e , te x' i e' , ali su zapravo pravci koji su nastali presjekom epipolarne ravnine sa slikovnom ravninom. Epipol je točka u kojoj se sijeku svi epipolarni pravci.

Poznavajući činjenicu da se točke nalaze u istoj ravnini, te ako znamo položaj točke x možemo odrediti položaj točke x' . Rješenja koja se moraju ispitati da bi saznali položaj točke x' nalaze se na epipolarnoj liniji, što uvelike olakšava pronalaženje korespondentnih točaka.

2.2. Pretprocesiranje slika

Parametre geometrije sustava kamera moguće je podijeliti na dvije vrste: intrinzični i ekstrinzični. Ekstrinzični parametri su oni koji opisuju odnos svojstva koja su zajednička kamerama. Intrinzični pak opisuju svojstva kamera koje se odnose svaku kameru pojedinačno. Kao primjere za intrinzične faktore navodimo svojstva koja utječu na nji, a to su nesavršenost leća (na slici 2.2 prikazano je radijalno izobličenje), pomak senzora od centra leća te slične fizičke karakteristike nesavršenosti kamera.



Slika 2.2: Primjeri radijalnog izobličenja. Izvor:[1]

S druge strane ekstrinzični parametri nas dovode do nužnih transformacija kojima se slike dovode u istu ravninu projekcije te se time postiže da su pikseli duž horizontalnog pravca jedne slike na istoj visini na drugoj slici. Taj pravac je već spomenut, riječ je o epipolarnoj liniji. Ovakvim transformacijama uvelike se olakšava pronalaženje korespondentnih točaka koje u ovom slučaju treba tražiti samo duž jedne linije, epipolarne linije. Problem je sveden na jednu dimenziju. Kalibracijom se rješava problem nesavršenosti leća, odnosno problem izobličenja slika.

2.2.1. Rektifikacija

Za postupak rektifikacije važni su ekstrinzični faktori kamera, dok na njihov izračun utječu intrinzični faktori. Transformacijom slika rektifikacijom dobivamo slike čiji pikseli koji u prostoru odgovaraju točkama na istoj visini, a nalaze se duž iste epipolarne linije. Također, nakon provedenih transformacija intrinzični parametri kamera postaju jednaki.



Slika 2.3: Prikaz epipolarnih linija u rektificiranim slikama skupa KITTI 2015

Postupak se provodi kroz nekoliko koraka, a uključuje šahovsku ploču koja je pogodna zbog svojih svojstava. Prvi korak je traženje šahovskih polja. Postupak se nastavlja određivanjem izobličenja i ostalih parametara uz pomoć šahovskih polja. Na slici 2.2 prikazan je par slika iz stereo kamere koji je rektificiran. Slike su iz skupa KITTI 2015, te se na njima vide epipolarne linije. Lako je primjetiti da iste točke u prostoru nalaze na istoj visini, odnosno na istoj epipolarnoj liniji. Ovakvom transformacijom, korespondentne piksele potrebno je tražiti samo na istoj y -koordinati.

2.3. Disparitet

Prijašnjim postupcima kalibracije i rektifikacije omogućen je postupak računanja dispariteta, odnosno postupak računanja koji za svaki piksel koji je nastao lijevom kamerom određuje piksel slike koja je nastala desnom kamerom, ali pomaknutog za udaljenost d piksela po horizontalnoj osi. *Disparitet* je definiran kao horizontalni pomak d između dvaju korespondentnih piksela u slikama nastalim iz dviju konkurentnih stereo kamera. Kao rezultat računanja dispariteta nastaje mapa dispariteta koja sadrži disparitet za svaki piksel referentne kamere. Odnos piksela slike iz lijeve kamere, koja je u ovom slučaju referentna, i slike iz desne kamere definira se kao:

$$I_L(x, y) = I_D(x - d, y) \quad (2.1)$$

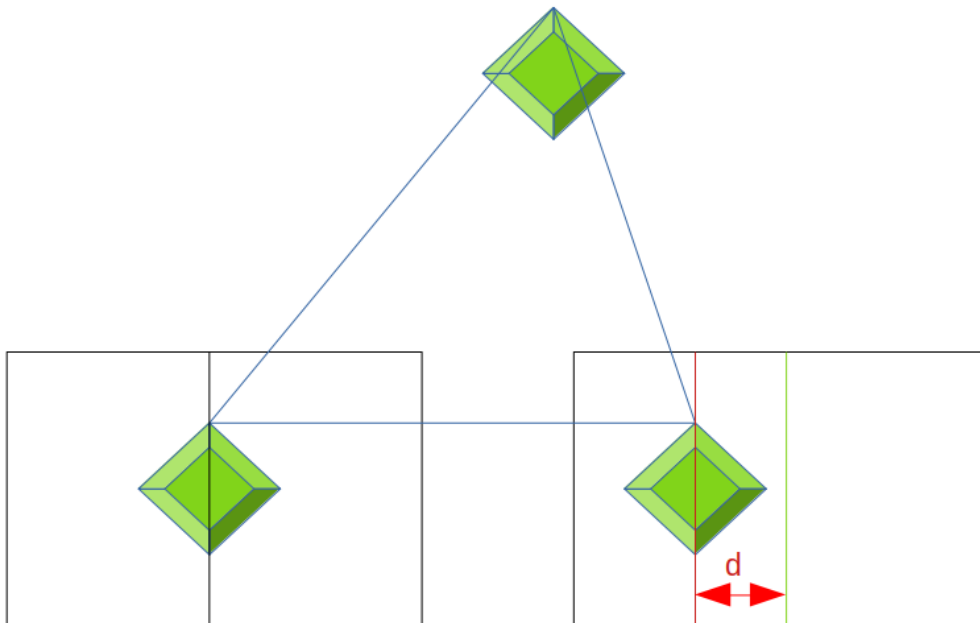
gdje su I_L i I_D slike nastale lijevom, odnosno desnom kamerom, respektivno. Uređenim parom (x, y) opisani su pikseli slike iz lijeve kamere, a njemu korespondentni pikseli

u slikama iz desne kamere označeni su pomakom za disparitet d po horizontalnoj osi $(x - d, y)$. Važno je napomenuti da su slike u skupu KITTI 2015 rektificirane i kalibrirane, što znači da korespondentne piksele tražimo samo po horizontalnoj osi. Pomoću izračunatog dispariteta moguće je izraziti odnos između piksela kamera i dubine scene kao:

$$Z = f \frac{B}{d} \quad (2.2)$$

gdje je dubina scene, odnosno udaljenost objekta od kamere označena sa Z , d je disparitet, B je označena udaljenost između središta kamera, a f je označena fokalna udaljenost kamere. Iz jednadžbe 2.2 je vidljiva obrnuta proporcionalnost između udaljenosti objekta od kamere, te dispariteta s druge strane. Kada je položaj objekta blizu kamere, disparitet će biti velik. A kako se udaljavamo od kamere, disparitet se smanjuje i u beskonačnosti postiže vrijednost 0.

Na slici 2.4. je vidljivo da su epipolarne linije paralelne, te im je y -koordinata za korespondentne piksele jednaka. Disparitet je označen slovom d i prikazuje horizontalni pomak.



Slika 2.4: Prikaz objekta u slikama iz lijeve i desne kamere. Plavom bojom označena je epipolarna ravnina. Okomitim osima označeni su projicirani pikseli u konkurentnim slikama. Disparitet je označen slovom d .

2.4. Algoritam stereoskopske rekonstrukcije

Postupak stereoskopske korespodencije je zapravo postupak pronalaženja istih piksela u međusobno konkurentnim kamerama koji odgovaraju istoj točki u trodimenzionalnom prostoru. Postoje dvije vrste korespodencija, *rijetke* i *guste*. Razlika je u tome što rijetke ne izračunavaju disparitete za sve piksele, dok guste to rade. Guste korespodencije omogućene su današnjom snagom računala, te se rijetke više toliko ne koriste. Postupak guste korespodencije je tim teži što se prilikom izračuna susreće s nekim anomalijama u slikama, primjerice područjima bez tekstone, refleksivnim podlogama na kojima se svjetlost drugačije lomi i odbija, te točkama koje se iz jedne kamere vide, a iz druge ne (*stereoskopska sjena*).

Dolazimo do stereo algoritama, svaki od njih moguće je opisati u četiri koraka ([2],[3]):

1. izračunavanje razlike između dvaju slikovnih okana, odnosno izračun podatkovnih cijena po prije definiranoj mjeri razlike
2. prikupljanje podatkovnih cijena za disparitete koji se razmatraju
3. izračun mape dispariteta, odnosno odabir najmanje prikupljene cijene za svaki piksel, uz optimizaciju
4. zaglađivanje mape dispariteta

Prvi korak ovog generičkog algoritma za stereoskopsku rekonstrukciju je izračun podatkovnih cijena. Izračun podatkovnih cijena može se računati prema pojedinačnim pikselima ili skupini piksela. Izračun nam govori kolika je razlika između dvaju slikovnih okana, a razlika koja se koristi je definirana prije. Mjere razlike mogu biti sljedeće:

- srednja kvadratna razlika
- srednja apsolutna razlika
- kvadratna razlika između intenziteta piksela
- apsolutna razlika između intenziteta piksela

Drugi od koraka generičkog stereo algoritma je prikupljanje podatkovni cijena, tj. agregacija. Na početku je postavljena granica za maksimalan disparitet, pa se tijekom ovog koraka izračunava podatkovna cijena za sve disparitete i skupa od 0 do D , te ćemo tako za svaki piksel imati $D + 1$ podatkovnih cijena.

Treći korak algoritma je izračun mape dispariteta. Nakon što su izračunate podatkovne cijene za razmatran skup dispariteta, treba odabrati metodu za izračun mape dispariteta. U praksi postoje *lokalne* i *globalne*. Lokalne metode prikupljaju cijene nad ograničenim područjem unutar same mape dispariteta, odnosno traže na ograničenom području oko referentne točke. Područje može biti 2D ili 3D, te se prikupljanje može primjerice izvesti konvolucijom. S druge strane, globalne metode preskaču korak prikupljanja podatkovnih cijena, te odmah nakon izračuna podatkovnih cijena vrše optimizaciju. Globalne metode minimiziraju kriterij nad cijelom slikom. Odnos između ovi metoda može se svesti na to da su lokalne brže, no globalne daju puno glađe mape dispariteta, što ih čini kvalitetnijima.

Zadnji korak koji je naveden je zaglađivanje mape dispariteta. Ovaj korak je opcionalan i neke metode ga koriste. Glavni cilj ovog koraka je uklanjanje šuma koji je nastao generiranjem mape.

3. Duboko učenje

3.1. Osnovni pojmovi u dubokom učenju

3.1.1. Perceptron

3.1.2. Prijenosne funkcije

3.1.3. Konvolucijski slojevi

3.1.4. Optimizacija

3.1.5. Regularizacija

3.2. Učenje dubokih modela

3.2.1. Funkcija gubitka

3.2.2. Gradijentni spust

3.2.3. Memorijska zahtjevnost

4. Podatkovni skup za učenje

4.1. Podatkovni skup KITTI 2015

5. Duboko učenje u kontekstu stereoskopske rekonstrukcije

5.1. Ugrađivanje okana u metrički prostor

5.2. Učenje modela za ostvarivanje korespodencije

6. Eksperimentalni rezultati

7. Programska izvedba i vanjske biblioteke

8. Zaključak

Zaključak.

Korespondencijska ugrađivanja za stereoskopsku rekonstrukciju

Sažetak

Sažetak na hrvatskom jeziku.

Ključne riječi: Ključne riječi, odvojene zarezima.

Title

Abstract

Abstract.

Keywords: Keywords.