

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1149

**Gusta stereoskopska
rekonstrukcija scene
predstavljene ravninskim
odsječcima**

Dino Kovač

Zagreb, lipanj 2015.

Umjesto ove stranice umetnite izvornik Vašeg rada.
Da bi ste uklonili ovu stranicu obrišite naredbu \izvornik.

Veliko hvala profesor Siniši Šegviću te Ivanu Kreši za pruženu pomoć i vodstvo.

SADRŽAJ

1. Uvod	1
2. Kalibracija stereo para	3
2.1. Intrinzični parametri	4
2.2. Ekstrinzični parametri	5
2.2.1. Ispravljanje redoslijeda značajki	5
2.3. Geometrija kalibriranog stereoskopskog para	7
2.3.1. Epipolarna ograničenja i triangulacija	8
2.3.2. Ispravljanje slike	11
2.3.3. Ovisnost dubine o disparitetu	12
3. Stereoskopska rekonstrukcija	14
3.1. Podatkovna cijena	14
3.1.1. Srednja kvadratna pogreška	14
3.1.2. Srednja apsolutna pogreška	14
3.1.3. Census transformacija	15
3.1.4. Podudaranje neovisno o uzorkovanju	15
3.2. Lokalne metode	16
3.3. Globalne metode	16
4. Poluglobalno podudaranje	18
4.1. Izračun podudaranja elemenata	18
4.2. Agregacija cijene	18
4.2.1. Optimizacija potrebnog vremena i prostora	21
4.3. Izračun mape dispariteta	22
5. Poluglobalno podudaranje s geometrijskom regularizacijom	23
5.1. SGM Flow	23
5.1.1. Prilagodba SGM-a	25

5.2. SGM StereoFlow	25
6. Zaglađivanje tehnikom ravninske dekompozicije	27
6.1. Definicija energije	27
6.2. Efikasni koordinatni spust po blokovima	30
7. Eksperimentalni rezultati	33
7.1. Mapa dispariteta	37
7.2. Brzina izvođenja	37
8. Zaključak	38
Literatura	39

1. Uvod

Računalni vid je područje računalne znanosti koje radi na razvoju tehnologije kojom računala razumijevaju i upotrebljavaju slike. Slike predstavljaju projekciju stvarnog, trodimenzionalnog svijeta. Računalni vid se bavi rekonstrukcijom scene i raspoznavanjem objekata na slici. Rekonstrukcijski pristupi pokušavaju otkriti geometrijske odnose među elementima scene. To obično implicira određivanje dubine slikovnih elemenata, odnosno određivanje njihove udaljenosti od otvora kamere.

Postupci stereoskopske rekonstrukcije primjenjuju se u robotici i automobilizmu za autonomno navigiranje, izbjegavanje prepreka, procjenu kretanja kamere, i sl.

Kako bismo rekonstruirali scenu s dvodimenzionalne slike obično su potrebne dodatne informacije. Jedan od načina da se dobije više informacija s ciljem rekonstrukcije prostora jest snimanje prostora s dvije horizontalno razmaknute kamere. Na taj način se dobivaju dva pogleda na sliku, odnosno dobiva se pogled iz dva različita kuta – tzv. par stereo slika. Iz para stereo slika moguće je dobiti informacije o dubini pojedinih objekata, njihovim konturama, orijentaciji površina na slici te naborima na objektima u slici. Postupak izvlačenja dubine scene iz para stereo slika naziva se stereoskopska rekonstrukcija.

Stereoskopska rekonstrukcija uključuje sljedeće korake obrade parova stereo slika:

1. kalibracija kamera
2. ispravljanje slika (eng. rectification)
3. uparivanje dijelova slika
4. izračun mape dispariteta

U koraku uparivanja dijelova slika stereoskopske rekonstrukcije mogu se koristiti lokalne ili globalne metode. Lokalne metode podudaranje bloka s referentne slike traže u prozoru konačne veličine na drugoj slici. U pravilu su brze i daju dobre rezultate.

Nedostatak lokalnih metoda je što odnos između značajki slike prve i druge kamere nije bijektivan. Značajka na slici druge kamere može biti pridružena većem broju značajki s prve (referentne) slike.

Metode globalne optimizacije pokušavaju optimizirati pridruživanje globalno. Veći je naglasak na postizanje globalnog optimuma nego na optimalno pridruživanje pojedinih značajki.

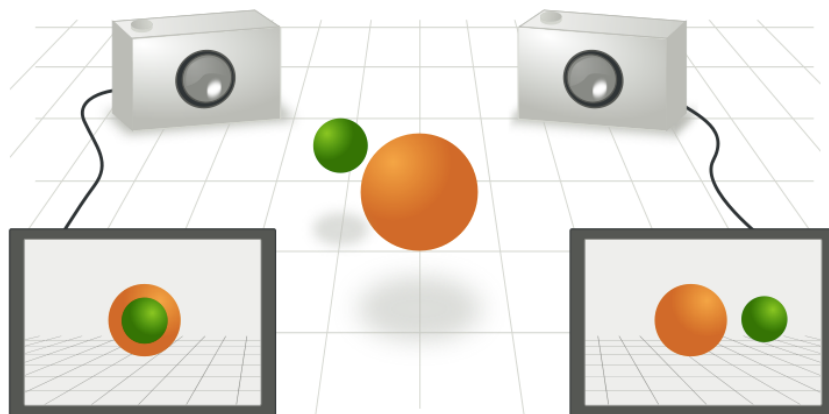
Globalne metode daju vrlo precizne rezultate u usporedbi lokalnim metodama. Cilj je pronaći optimalnu funkciju dispariteta $d(x, y)$, koja minimizira globalni kriterij E koji kombinira podatkovni član i član koji opisuje glatkoću.

$$E(d) = E_{podaci}(d) + \lambda E_{glatkost}(d) \quad (1.1)$$

Globalni kriterij favorizira rješenje koje je *glatko* zato jer susjedni slikovni elementi često imaju sličnu vrijednost jer su dio istog objekta na slici.

2. Kalibracija stereo para

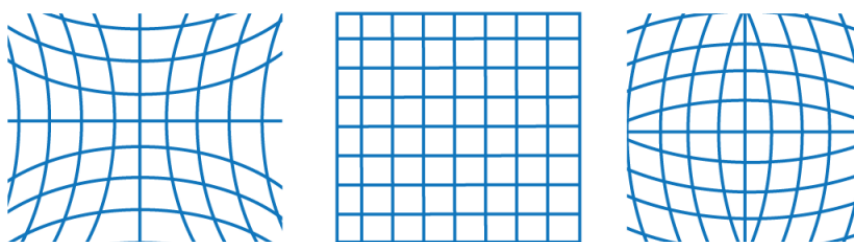
Kako bismo mogli iskoristiti informacije sadržane u slikama dobivenim pomoću stereoskopskog para kamera, najprije moramo znati geometrijski odnos između kamera. Također moramo kompenzirati za izobličenje slike koje nastaje zbog nesavršenosti izrade leća kamere.



Slika 2.1: Prikaz stereo para kamera u prostoru s pogledima na scenu s obje kamere

Kalibracija stereoskopskog para se vrši u 2 koraka:

1. Izračun intrinzičnih parametara za svaku kameru
2. Izračun ekstrinzičnih parametara za stereo par kamera

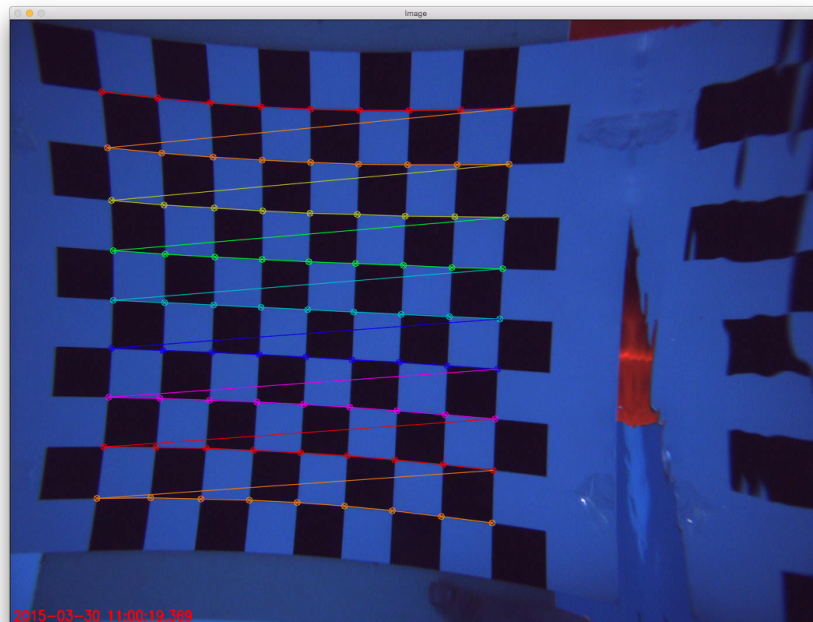


Slika 2.2: Grafički prikaz izobličenja nastalog zbog nesavršenosti leće: negativno radijalno izobličenje (lijevo), bez izobličenja (sredina), pozitivno radijalno izobličenje (desno)

Za tu primjenu može se iskoristiti uzorak šahovske ploče. Uzorak šahovske ploče je pogodan za kalibraciju kamere jer sadrži veliki broj značajki koje se mogu detektirati jednostavnim metodama. Kalibracijski program može iskoristiti činjenicu da su značajke, odnosno kutovi šahovskih polja, pravilno raspoređeni po pravcima.

2.1. Intrinzični parametri

Intrinzični parametri kamere se računaju tako da se s kamerom koju kalibriramo snimi barem desetak fotografija scene na kojoj se nalazi šahovska ploča. Pritom je bitno da je šahovska ploča u cijelosti vidljiva na fotografijama i da pokriva veliki dio fotografije.

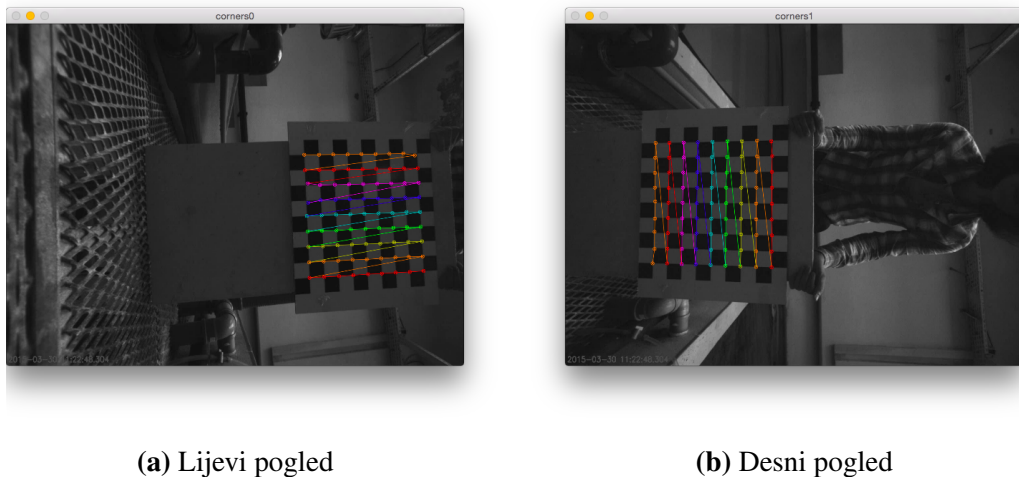


Slika 2.3: Uzorak šahovske ploče jasno otkriva negativno radijalno izobličenje fotografije

Tako pribavljen niz slika za pojedinu kameru se koristi kao ulazni skup u program za kalibraciju intrinzičnih parametara. Kalibracijski program koristi ulazni skup slika i ulazne podatke o dimenzijama polja šahovske ploče da bi pronašao dodirne točke svih crnih polja na šahovskoj ploči. Zatim se pomoću pozicija slikovnih elemenata gdje su pronađene dodirne točke računaju koeficijenti distorzije slike koji nam govore kako pozicije točaka odstupaju od pozicija istih točaka na kameri sa savršenim lećama.

2.2. Ekstrinzični parametri

Ekstrinzični parametri opisuju geometrijski odnos između stereo kamera, odnosno međusobni položaj i rotaciju kamera (Slika 2.1). Za izračun ekstrinzičnih parametara potrebni su intrinzični parametri obje kamera te skup stereo parova u kojem je na svakom stereo paru cijela šahovska ploča vidljiva na obje slike.



Slika 2.4: Stereo par za kalibraciju ekstrinzičnih parametara

Tako pripremljen ulazni skup koristi se u programu za ekstrinzičnu kalibraciju. Kalibracijski program najprije na svakoj slici posebno pronalazi sve značajke šahovskog polja. Pronađene lokacije značajki se u memoriju računala spremaju kao jednodimenzionalno polje. Zatim se uz pomoć lokacija slikovnih elemenata gdje su pronađene značajke šahovskog polja na lijevoj i desnoj slici računa razmak između kamera u prostoru te međusobna rotacija kamera.

2.2.1. Ispravljanje redoslijeda značajki

Kako se detekcija značajki šahovskog polja odvija zasebno za lijevu i desnu sliku scene, može se dogoditi da poredak elemenata u polju značajki za lijevu sliku nije jednak poretку značajki u polju za desnu sliku. Primjer takvog slučaja prikazan je na slici 2.4. U tom slučaju potrebno je ispraviti poredak značajki za desnu sliku.

Ako sve točke u polju redom povežemo u retke kao što prikazuje slika 2.4 onda postoji 8 mogućih odnosa između poretka značajki s lijeve i desne slike s obzirom na smjer pružanja redaka i stupaca. Kako bi se poredak značajki s desne slike ispravio, najprije je potrebno identificirati koji od 8 mogućih slučajeva je primjenjiv na stereo par koji se obrađuje.

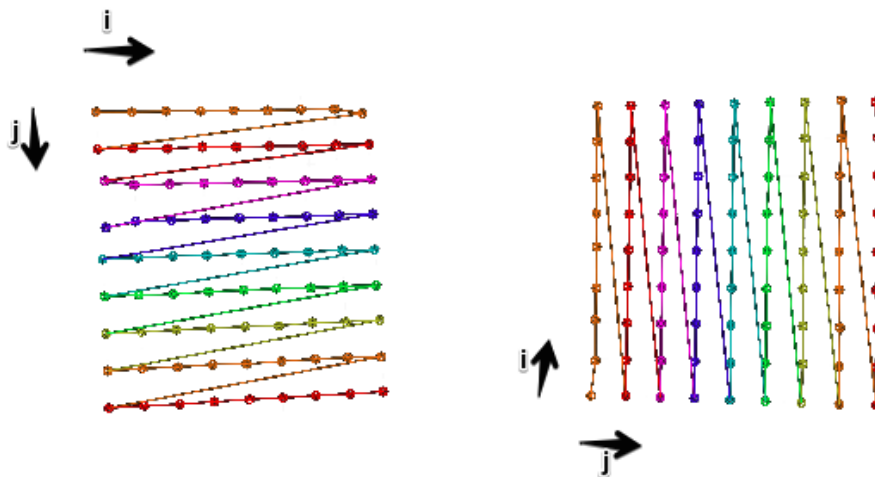
Označimo s C_L i C_R polje značajki za lijevu, odnosno desnu sliku scene te s n broj značajki u jednom retku šahovske ploče. Definirajmo sljedeće vektore pomoću točaka iz polja značajki:

$$\mathbf{i}_L = C_L[1] - C_L[0] \quad (2.1)$$

$$\mathbf{i}_R = C_R[n] - C_R[0] \quad (2.2)$$

$$\mathbf{j}_L = C_L[1] - C_L[0] \quad (2.3)$$

$$\mathbf{j}_R = C_R[n] - C_R[0] \quad (2.4)$$



Slika 2.5: Grafički prikaz vektora $\mathbf{i}$ i $\mathbf{j}$ za primjer na slici 2.4

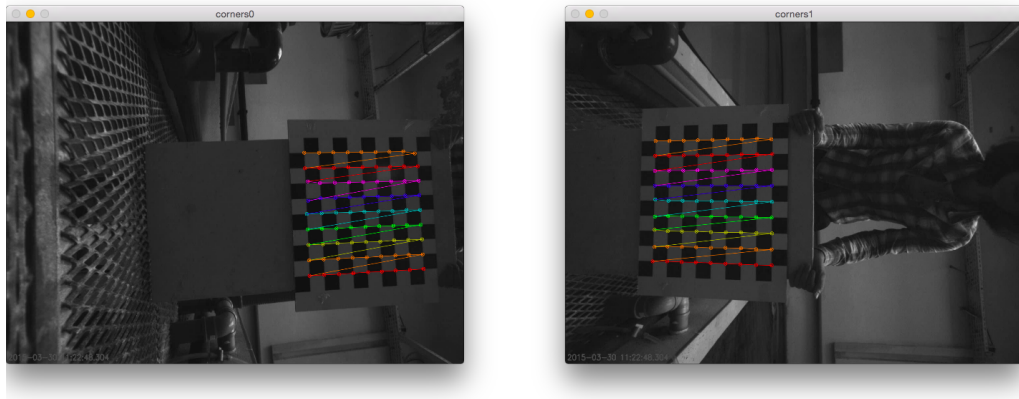
Tada iz odnosa smjerova vektora $\mathbf{i}_L$ i $\mathbf{i}_R$ te odnosa smjerova vektora $\mathbf{j}_L$ i $\mathbf{j}_R$ možemo odabrati jedan od mogućih slučajeva te na osnovi prepoznatog slučaja ispraviti poredak značajki u polju C_R kako bi se vektori $\mathbf{i}$ i $\mathbf{j}$ podudarali na obje slike. Odnos dvaju vektora kvantiziramo s korakom od 90 stupnjeva. S obzirom da će vektori $\mathbf{i}$ i $\mathbf{j}$ uvijek biti razmaknuti za otprilike 90 stupnjeva, vektori pružanja polja na lijevoj i desnoj slici mogu biti u 8 različitih odnosa.

Ovisno o pronađenom slučaju, nad matricom značajki koja je u memoriji spremjena kao jednodimanzionalno polje potrebno je izvršiti nijednu, jednu ili više transformacija:

- Obrtanje redoslijeda redaka

- Obrtanje redoslijeda stupaca
- Transponiranje matrice

Rezultat ispravljanja za primjer na slici 2.4 vidljiv je na slici 2.6.



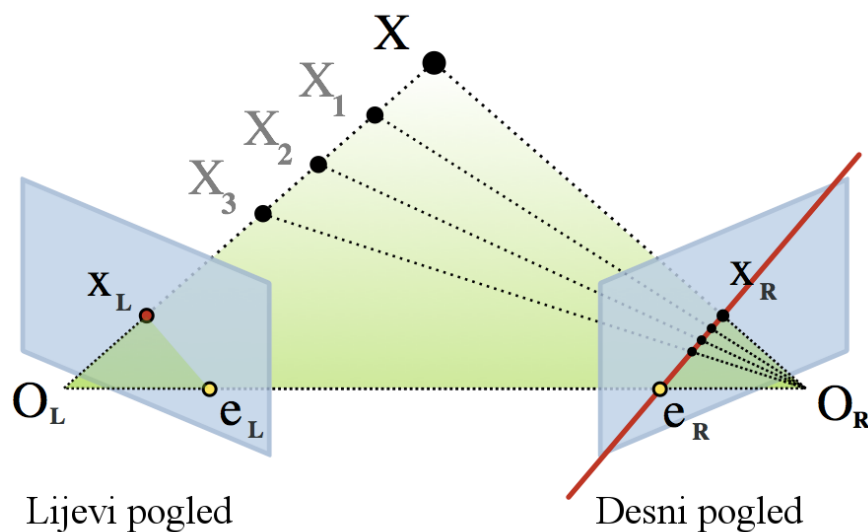
(a) Lijevi pogled

(b) Desni pogled

Slika 2.6: Stereo par za kalibraciju ekstrinzičnih parametara nakon korekcije redoslijeda značajki

2.3. Geometrija kalibriranog stereoskopskog para

Kalibrirani stereoskopski par kamera koji snima trodimenzionalnu scenu stvara dvije dvodimenzionalne slike koje predstavljaju projekcije te scene na dvodimenzionalnu površinu. Točke na dobivenim slikama imaju definiran geometrijski odnos. Taj odnos definira *epipolarna geometrija*.



Slika 2.7: Epipolarna geometrija stereo sustava

Svaka kamera ima projekcijski centar, O_L i O_R . Slika projekcijskog centra uparene kamere naziva se epipolarna točka ili epipol te se označava s e_L za lijevu, odnosno e_R za desnu kameru. Projekcijski centri i epipolovi obje kamere nalaze se na zajedničkom pravcu. Ako se promatranu točku u trodimenzionalnom prostoru označi s X , a projekcije te točke na slike kamera s X_L i X_R , može se definirati epipolarna ravnina; to je ravnina na kojoj leže točke O_L , O_R i X . Sjecište epipolarne ravnine i slike kamere definira epipolarni pravac.

2.3.1. Epipolarna ograničenja i triangulacija

Ako je poznata translacija i rotacija jedne kamere u odnosu na drugu, odgovarajuća epipolarna geometrija omogućuje dva bitna zaključka:

1. Ako je poznat položaj projekcije točke iz trodimenzionalnog prostora X_L , onda je epipolarni pravac $e_R - X_R$ također poznat te se točka X_R nalazi na tom epipolarnom pravcu. To znači da za svaku točku na slici kamere, odgovarajuća točka na slici uparene kamere mora ležati na poznatom epipolarnom pravcu. To predstavlja *epipolarno ograničenje* koje odgovarajuće točke moraju zadovoljiti što nam omogućava provjeru odgovaraju li dvije točke na projiciranim slikama istoj točki u trodimenzionalnom prostoru.
2. Ako su poznati položaji točaka X_L i X_R , njihovi projekcijski pravci su također poznati. Ako dvije točke na projiciranim slikama odgovaraju istoj točki u trodimenzionalnom prostoru, projekcijski pravci moraju se sjeći točnu u toj točki koju

označavamo s X . To znači da se položaj točke X može izračunati iz koordinata točaka X_L i X_R . Taj izračun naziva se *triangulacija*.

Da bismo matematički opisali epipolarno ograničenje, najprije treba uvesti pojmove *esencijalne* i *fundamentalne matrice*. Esencijalna matrica E sadrži informacije o translaciji i rotaciji koje vežu uparene kamere u prostoru, odnosno opisuju odnos između točaka X_L i X_R , u koordinatnom sustavu slike. Fundamentalna matrica F sadrži, uz informacije sadržane u esencijalnoj matrici, intrinzične parametre obje kamere. Ona povezuje točke X_L i X_R iz lijeve i desne slike. Slijedi izvod esencijalne matrice.

Pretpostavimo dvije normalizirane kamere koje projiciraju istu točku X iz stvarnog svijeta na svoje ravnine slike. Koordinate točke X su $X_L = (x_1, y_1, z_1)$ u koordinatnom sustavu prve kamere i $X_R = (x_2, y_2, z_2)$ u koordinatnom sustavu druge kamere. Kamere su normalizirane pa su korespondentne koordinate točke u slikama nakon perspektivne projekcije:

$$\mathbf{y}_1 = \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \frac{1}{z_1} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \quad (2.5)$$

$$\mathbf{y}_2 = \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \frac{1}{z_2} \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \quad (2.6)$$

Još jedno svojstvo normaliziranih kamera jest činjenica da su njihovi koordinatni sustavi u relaciji preko translacije i rotacije pa se točke jednostavnom transformacijom mogu prebaciti iz jednog u drugi sustav:

$$\mathbf{x}_2 = \mathbf{R}(\mathbf{x}_1 - \mathbf{t}) \quad (2.7)$$

gdje je $\mathbf{R}$ rotacijska matrica dimenzije 3×3 , a $\mathbf{t}$ translacijski vektor dimenzije 3. Esencijalna matrica je time definirana kao:

$$\mathbf{E} = \mathbf{r}[\mathbf{t}]_x \quad (2.8)$$

gdje je $[\mathbf{t}]_x$ matrična reprezentacija vektorskog produkta s vektorom $\mathbf{t}$. Matrična reprezentacija vektorskog produkta je uvijek antisimetrična matrica.

$$\mathbf{t} \times \mathbf{a} = [\mathbf{t}]_x \mathbf{a} = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad (2.9)$$

Slijedi dokaz.

$$\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = (\mathbf{R}(\mathbf{x}_1 - \mathbf{t}))^T \mathbf{R} [\mathbf{t}]_x \mathbf{x}_1 \quad (2.10)$$

$$\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = (\mathbf{x}_1 - \mathbf{t})^T \mathbf{R}^T \mathbf{R} [\mathbf{t}]_x \mathbf{x}_1 \quad (2.11)$$

$$\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = (\mathbf{x}_1 - \mathbf{t})^T [\mathbf{t}]_x \mathbf{x}_1 \quad (2.12)$$

$$\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = 0 \quad (2.13)$$

Korištene relacije i svojstva:

1. $\mathbf{x}_2 = \mathbf{R}(\mathbf{x}_1 - \mathbf{t})$
2. $\mathbf{R}^T \mathbf{R} = \mathbf{I}$ jer je $\mathbf{R}$ ortogonalna matrica
3. Svojstva matrične reprezentacije vektorskog produkta:

$$(a) \mathbf{a}^T [\mathbf{b}]_x \mathbf{a} = 0$$

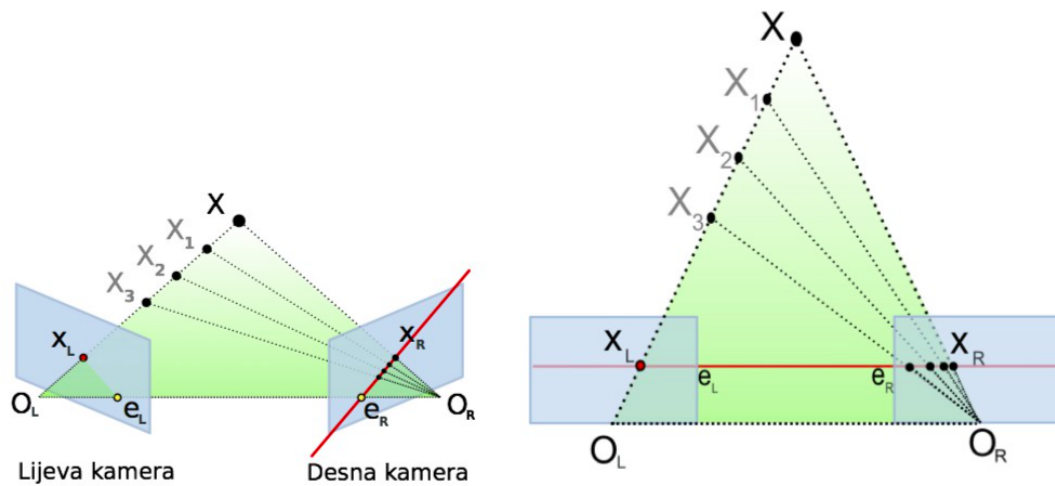
$$(b) \mathbf{b}^T [\mathbf{b}]_x \mathbf{a} = 0$$

Također možemo pretpostaviti da su z_1 i z_2 uvijek veći od nule jer bi točke inače bile iza otvora kamere pa ne bi bile vidljive. Uz tu pretpostavku vrijedi sljedeće:

$$\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = \frac{1}{z_2} \mathbf{x}_2^T \mathbf{E} \frac{1}{z_1} \mathbf{x}_1 = \mathbf{y}_2^T \mathbf{E} \mathbf{y}_1 = 0 \quad (2.14)$$

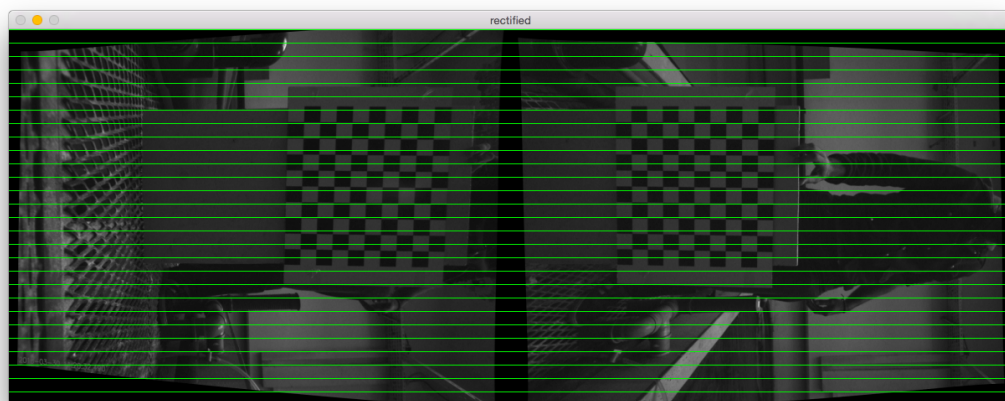
gdje je $\mathbf{y}_2^T \mathbf{E} \mathbf{y}_1 = 0$ epipolarno ograničenje.

2.3.2. Ispravljanje slike



Slika 2.8: Stereo sustav prije i poslije ispravljanja slika

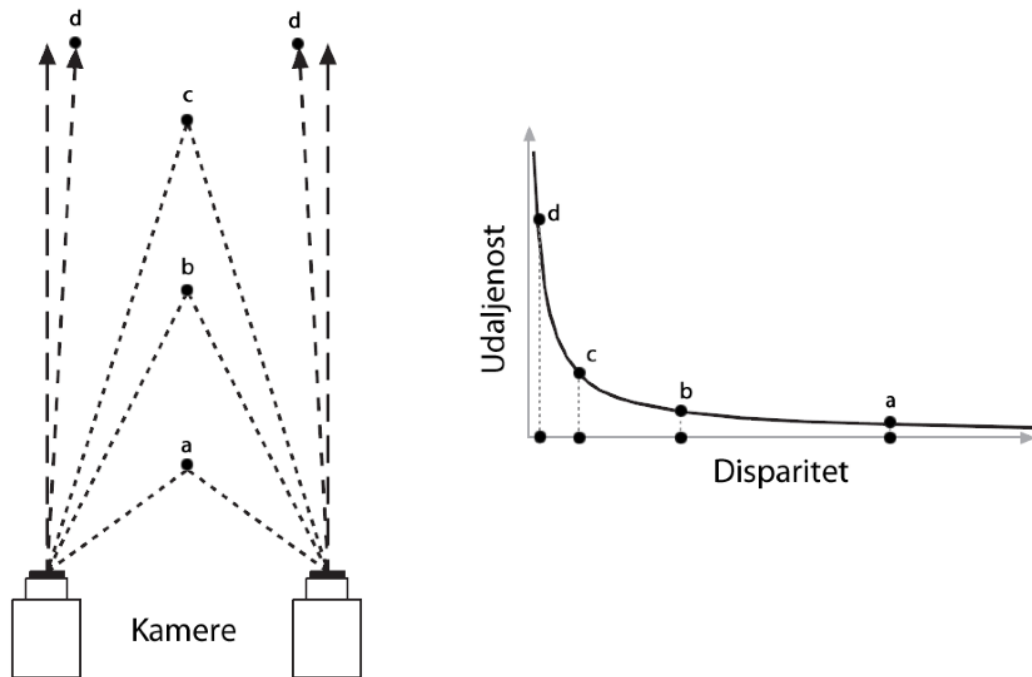
Nakon što smo kalibracijom otkrili parametre lijeve i desne kamere te njihov međusobni geometrijski odnos, možemo provesti postupak ispravljanja slika. Da bi se slike ispravile, potrebno je najprije ukloniti izobličenje nastalo zbog nesavršenosti leća te zatim odrediti transformaciju slike koja bi se dobila ako bi projekcijske ravnine kamera ležale u istoj ravnini jedna uz drugu na istim visinama. To se postiže postavljanjem epipolova lijeve i desne kamere u beskonačnost na X osi.



Slika 2.9: Ispravljene slike s iscrtanim epipolarnim linijama

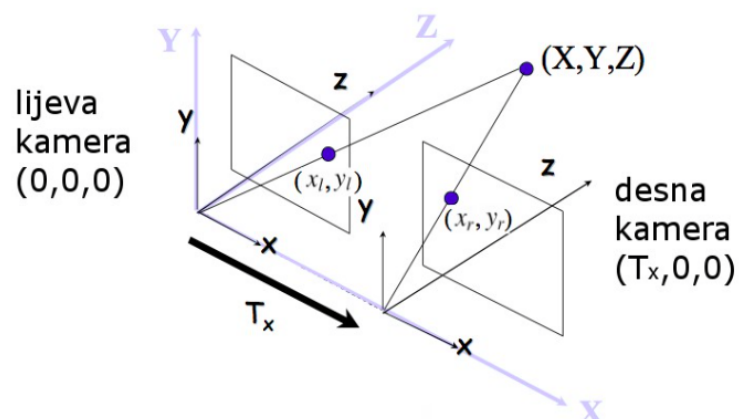
2.3.3. Ovisnost dubine o disparitetu

Što je razmak između projekcije točke na jednoj i drugoj slici veći, to je točka manje udaljena od ravnine na kojoj leže kamere. To je intuitivno jasno jer na jednak način funkcioniraju ljudske oči.



Slika 2.10: Odnos dubine i disparitetu

Promotrimo sustav na slici 2.11.



Slika 2.11: Stereo sustav

Označimo s f žarišnu duljinu kamere, pod pretpostavkom da obje kamere imaju

jednaku žarišnu duljinu. Desna kamera je od lijeve razmaknuta za T_x . Tada se koordinate projekcije promatrane točke (X, Y, Z) na slikama računaju na sljedeći način.

Za lijevu kameru:

$$x_l = f \frac{X}{Z} \quad (2.15)$$

$$y_l = f \frac{Y}{Z} \quad (2.16)$$

Za desnu kameru:

$$x_r = f \frac{X - T_x}{Z} \quad (2.17)$$

$$y_r = f \frac{Y}{Z} \quad (2.18)$$

Disparitet točke je tada:

$$d = x_l - x_r = f \frac{X}{Z} - \left(f \frac{X - T_x}{Z} \right) \quad (2.19)$$

odnosno

$$d = f \frac{T_x}{Z} \quad (2.20)$$

Tako dobivamo sljedeću formulu za izračun dubine točke:

$$Z = f \frac{T_x}{d} \quad (2.21)$$

gdje je f žarišna duljina kamere, T_x horizontalni odmak kamera, a d disparitet točke. Time smo dokazali da su dubina točke i njen disparitet u obrnuto proporcionalnom odnosu.

3. Stereoskopska rekonstrukcija

Stereoskopska rekonstrukcija podrazumijeva rekonstrukciju scene iz dvodimenzionalnih projekcija scene, odnosno iz dva pogleda na istu scenu. Ovaj postupak pretpostavlja prethodnu kalibraciju kamera i ispravljanje slika.

3.1. Podatkovna cijena

Kako bi se rekonstruirala dubina scene, najprije je potrebno upariti značajke sa slike koju vidi jedna kamera s odgovarajućim značajkama sa slike koju vidi druga kamera. Uparivanje značajki obavlja se tako što se za projekciju značajke na lijevi pogled traži projekcija iste značajke na desni pogled scene uspoređivanjem razlika. Podatkovna cijena je mjera razlike koja se koristi prilikom uspoređivanja dvaju slikovnih okana.

3.1.1. Srednja kvadratna pogreška

Srednja kvadratna pogreška (eng. Mean squared error) je jedna od dostupnih mjera razlike. Rezultat funkcije je zbroj kvadrata razlike vrijednosti elementa prvog bloka i vrijednosti odgovarajućeg elementa drugog bloka podijeljen s brojem elemenata u bloku.

$$MSE(x_1, y_1, x_2, y_2) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (slika_1[x_1 + i][y_1 + j] - slika_2[x_2 + i][y_2 + j])^2 \quad (3.1)$$

3.1.2. Srednja apsolutna pogreška

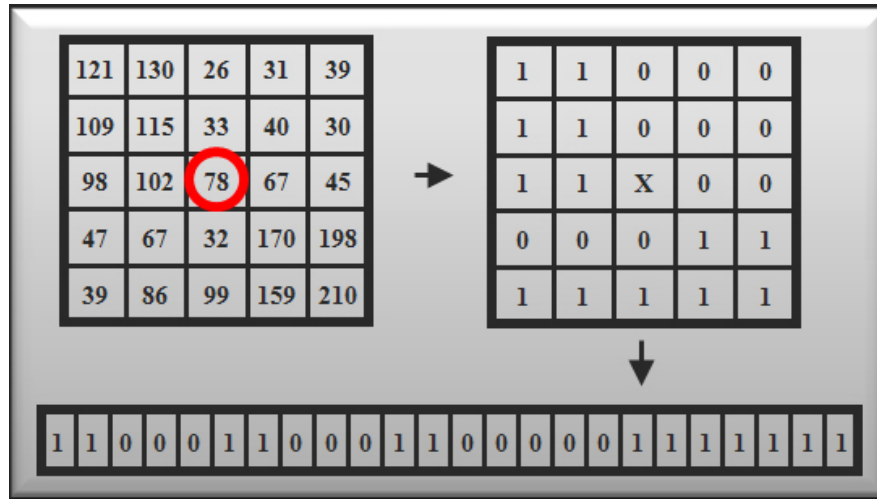
Srednja apsolutna pogreška (eng. Mean absolute difference) računa prosjek apsolutne razlike između dva slikovna okna.

$$MAD(x_1, y_1, x_2, y_2) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |slika_1[x_1 + i][y_1 + j] - slika_2[x_2 + i][y_2 + j]| \quad (3.2)$$

3.1.3. Census transformacija

Dosad navedene mjere razlike su osjetljive na varijacije u vrijednosti slikovnih elemenata zbog položaja i svojstava kamere. Mjera razlike zasnovana na Census transformaciji se ne oslanja na konkretnu vrijednost elemenata slike, već na lokalni raspored vrijednosti.

Census transformacija mijenja vrijednost središnjeg elementa slikovnog okna s nizom bitova koji označavaju odnos vrijednosti okolnih slikovnih elemenata i središnjeg slikovnog elementa. Na odgovarajuće mjesto u nizu bitova stavlja se 0 ako je vrijednost promatranog elementa manja od vrijednosti središnjeg elementa, 1 inače.



Slika 3.1: Grafički prikaz census transformacije

3.1.4. Podudaranje neovisno o uzorkovanju

Birchfield i Tomasi [3] opisali su funkciju podudaranja neovisnu o uzorkovanju. Cijena C_{BT} za slikovni element i određeni disparitet je minimalna apsolutna razlika intenziteta slikovnih elemenata sa obje slike u rasponu od polovice slikovnog elementa u oba smjera duž X osi, odnosno duž epipolarne linije.

Cijena C_{BT} računa se kao minimum dvaju simetričnih funkcija:

$$C_{BT} = \min(C_{BT}(x_i, y_i, I_L, I_R), C_{BT}(y_i, x_i, I_R, I_L)) \quad (3.3)$$

gdje su x_1 i y_i slikovni lementi na lijevoj odnosno desno slici, a I_L i I_R lijeva odnosno desna slika.

Funkcija $C_{BT}(x_i, y_i, I_L, I_R)$ definirana je sljedećim izrazima:

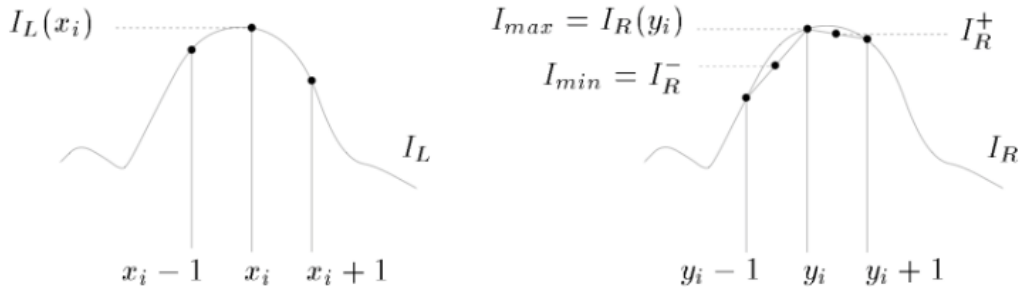
$$I^+ = \frac{1}{2}(I_R(y_i) + I_R(y_i + 1)) \quad (3.4)$$

$$I^- = \frac{1}{2}(I_R(y_i) + I_R(y_i - 1)) \quad (3.5)$$

$$I_{min} = \min(I_R^-, I_R^+, I_R(y_i)) \quad (3.6)$$

$$I_{max} = \max(I_R^-, I_R^+, I_R(y_i)) \quad (3.7)$$

$$C_{BT}(x_i, y_i, I_L, I_R) = \max(0, I_L(x_i - I_{max}, I_{min} - I_L(x_i)) \quad (3.8)$$



Slika 3.2: Grafički prikaz izračuna maksimalnog i minimalnog intenziteta u susjedstvu od polovice slikovnog elementa

3.2. Lokalne metode

Metode lokalne podudarnosti uzimaju značajku sa slike jedne kamere te traže tu istu značajku na slici druge kamere unutar konačnog prozora oko položaja značajke. Prozor unutar kojeg se traži podudaranje značajke jest konačan, no nema nužno konstantnu veličinu.

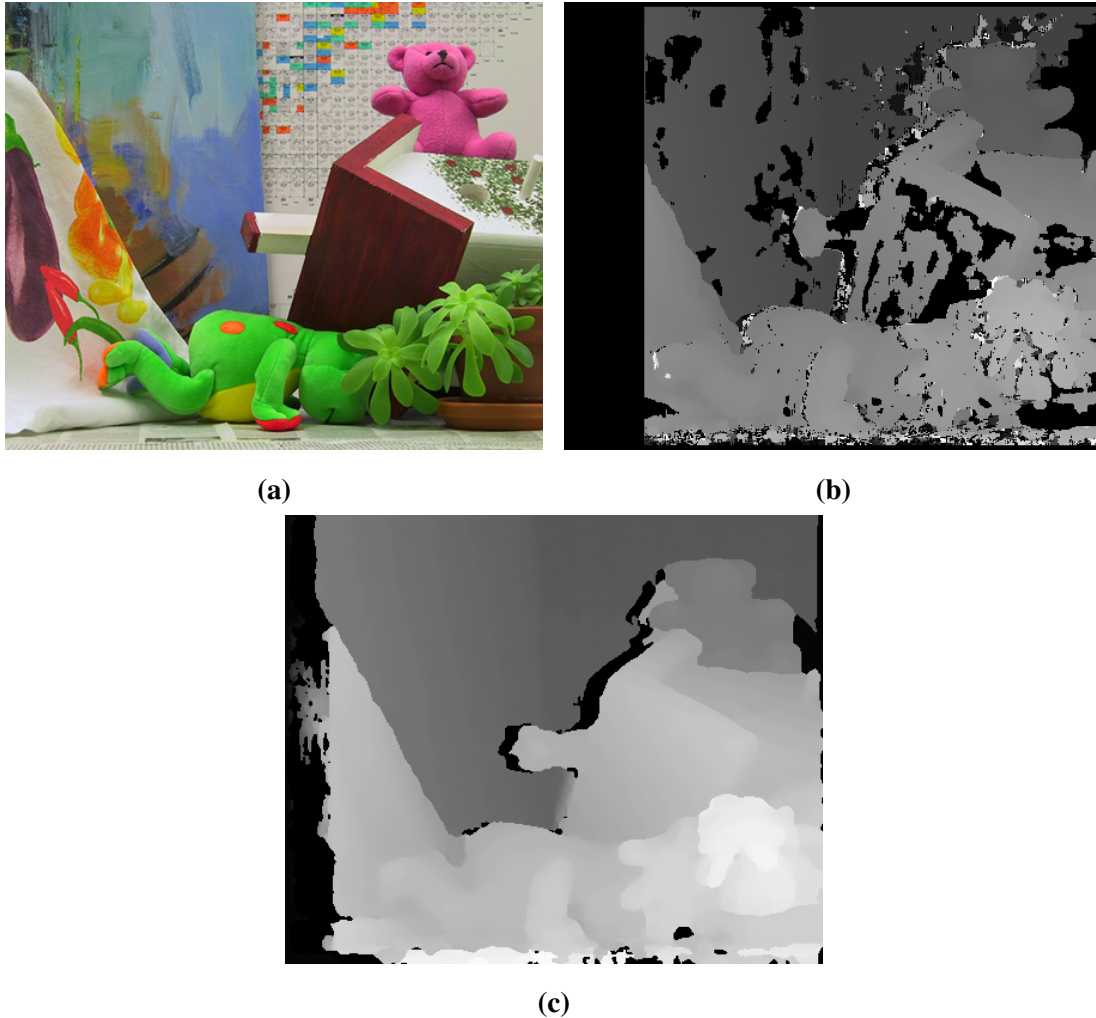
Lokalne metode su u pravilu brze te daju dobre rezultate.

3.3. Globalne metode

Globalne metode daju vrlo precizne rezultate u usporedbi lokalnim metodama. Cilj je pronaći optimalnu funkciju dispariteta $d(x, y)$, koja minimizira globalni kriterij E koji kombinira podatkovni član i član koji opisuje glatkoću mape dispariteta.

$$E(d) = E_{podaci}(d) + \lambda E_{glatkost}(d) \quad (3.9)$$

Globalni kriterij favorizira rješenje koje je *glatko* zato jer susjedni slikovni elementi često imaju sličnu vrijednost jer su dio istog objekta na slici.



Slika 3.3: Usporedba rezultata lokalne i globalne metode: (a) Ispitna slika 'teddy' iz middlebury stereo skupa, (b) Mapa dispariteta dobivena lokalnom metodom, (c) Mapa dispariteta dobivena SGM metodom

Globalne metode uključuju sljedeće korake:

1. Izračun podudaranja elemenata
2. Agregacija cijene
3. Izračun mape dispariteta

4. Poluglobalno podudaranje

Algoritam Poluglobalnog podudaranja (eng. Semi-Global Matching), u daljnjem tekstu SGM, je algoritam za stereoskopsku rekonstrukciju koji daje dobre rezultate uz relativno dobre performanse.

Slijedi opis algoritma kroz tri koraka. Podrazumijeva se da su ulazne slike ispravljene pomoću ranije izračunatih intrinzičnih i ekstrinzičnih parametara stereo para kamera.

4.1. Izračun podudaranja elemenata

U ovom koraku iterativno se izračunava podudaranje vrijednosti parova slikovnih elemenata iz lijeve i desne slike za svaki disparitet unutar $d \in [0, maksDisparitet]$. Pritom se *maksDisparitet* predaje implementaciji algoritma prije pokretanja te ovisi o konkretnoj primjeni.

Razultat ovog koraka je trodimenzionalno polje C koje sadrži vrijednost funkcije podudaranja za svaki redak, stupac i disparitet. Primjerice, $C[5][7][3]$ sadrži vrijednost funkcije podudaranja slikovnog elementa na lokaciji (5, 7) lijeve slike te slikovnog elementa na lokaciji (5, 4) desne slike.

Jedna od funkcija podudaranja koje daju dobre rezultate jest podudaranje neovisno o uzorkovanju (poglavlje 3.1.4). Odabir funkcije podudaranja ovisi o konkretnoj primjeni.

4.2. Agregacija cijene

Korištenje samo funkcije podudaranja slikovnih elemenata ne daje dovoljnu točnost zbog smetnji, slikovnih elemenata istih intenziteta u okolini i sl. Kako bi se povećala točnost rješenja uvodi se dodatno ograničenje koje poboljšava glatkoću rješenja tako što penalizira velike promjene u disparitetu susjednih elemenata u mapi dispariteta.

Podudaranje i glatkoća cijele slike objedinjeni su u energiju $E(D)$ koja je za mapu dispariteta D definirana kao:

$$E(D) = \sum_p (C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1]) \quad (4.1)$$

gdje q, p, D_p, D_q označavaju redom slikovni element na lijevoj i desnoj slici te odgovarajuće disparitete tih slikovnih elemenata, a funkcija T vraća 1 ako je izraz istinit, 0 u suprotnom.

Energija je suma tri podizraza po svim elementima slike. Prvi podizraz $C(p, D_p)$ označava podudaranje slikovnog elementa. Drugi podizraz dodaje fiksnu cijenu P_1 za sve elemente q iz susjedstva elementa p za koje je razlika u disparitetu jednaka 1. Treći podizraz dodaje fiksnu cijenu P_2 za sve elemente q iz susjedstva elementa p za koje je razlika u disparitetu veća od 1.

S tako definiranim izrazom, problem izračuna mape dispariteta svodi se na pronalazak mape dispariteta D za koju je energija $E(D)$ najniža. S obzirom na vrlo veliku složenost takvog iscrpnog pretraživanja, u praksi se koriste pristupi koji koriste dinamičko programiranje te aproksimiraju minimizaciju navedene energije.

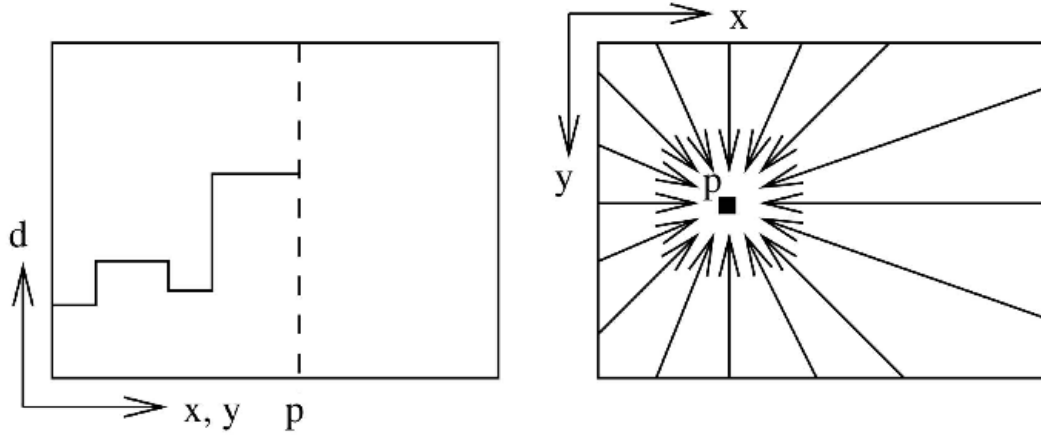


Slika 4.1: Mapa dispariteta dobivena minimizacijom po retcima

Minimizacija energije samo po retcima se može izvesti efikasno u polinomijalnoj složenosti, no to uzrokuje pojavu pruga na mapi dispariteta (slika 4.1).

Problem s takvim jednodimenzionalnim pristupom jest što se snažna ograničenja po retcima kombiniraju s slabim ili nikakvim ograničenjima po stupcima. Taj problem

je doveo do ideje agregiranja cijene po svim smjerovima jednako. Agregirana cijena $S(p, d)$ računa se zbrajanjem minimalnih cijena puteva iz svih smjerova koji završavaju u promatranom slikovnom elementu p .



(a) Put s minimalnom cijenom $L_r(p, d)$

(b) 16 puteva po svim smjerovima r

Slika 4.2: Agregacija puteva po svim smjerovima

Ti putevi se projiciraju kao ravne crte na referentnu (lijevu) sliku te kao "vijugave" crte na desnu sliku. Same puteve kroz desnu sliku nije potrebno pamtit, već samo minimalnu cijenu puta. Put $L'(p, d)$ definiran je rekurzivno sljedećim izrazom:

$$L'_r(p, d) = C(p, d) + \min(L'_r(p - r, d), \\ L'_r(p - r, d - 1) + P_1, \\ L'_r(p - r, d + 1) + P_1, \\ \min_i L'_r(p - r, i) + P_2)$$

gdje su P_1 i P_2 cijene za manju, odnosno veću razliku dispariteta, a r je smjer puta.

Koristeći gornji izraz, vrijednost $L'(p, d)$ se povećava prolaskom po nekom od puteva što predstavlja problem kod implementacije. Kako bi se taj problem riješio, od izraza se oduzima minimalna cijena puta slikovnog elementa koji prethodi u zadanom smjeru:

$$\begin{aligned}
L_r(p, d) = C(p, d) + \min(&L_r(p - r, d), \\
&L_r(p - r, d - 1) + P_1, \\
&L_r(p - r, d + 1) + P_1, \\
&\min_i L_r(p - r, i) + P_2) \\
&- \min_k L_r(p - r, k)
\end{aligned}$$

Modificirani izraz $L_r(p, d)$ ima gornju granicu vrijednosti od $L_r(p, d) \leq C_{max} + P_2$, neovisno o duljini puta.

Cijene puteva L_r se sumiraju za svaki slikovni element po svim putevima:

$$S(p, d) = \sum_r L_r(p, d) \quad (4.2)$$

Da bi se ostvarila dobra pokrivenost 2D slike putevima, broj smjerova treba biti 8 ili 16. Smjer se definira kao pomak u smjeru X osi te pomak u smjeru Y osi.

Korak agregacije cijena je vremenski najzahtjevniji dio SGM algoritma. Vremenska složenost mu je $O(WHD)$, gdje je W broj stupaca, H broj redaka, a D maksimalni mogući disparitet.

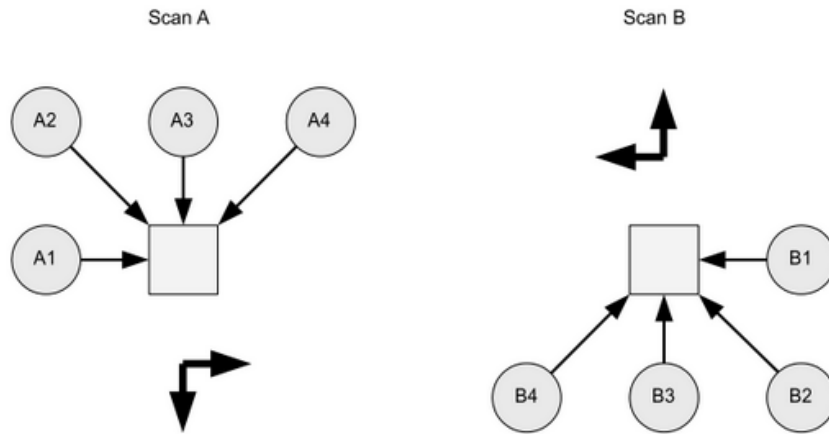
Zbog jednostavnosti operacija i podatkovnih struktura, ovaj korak se može relativno jednostavno ubrzati korištenjem vektorskih naredbi procesora za paralelno izvođenje koda nad više setova podataka.

4.2.1. Optimizacija potrebnog vremena i prostora

Mnoge primjene imaju specifične zahtjeve na brzinu izvođenja algoritma te je potrebno dodatno optimizirati implementaciju.

Ukoliko specifična primjena zahtijeva veliku paralelnost uz dostupne velike memorijske resurse, moguće je za svaki od r puteva alocirati posebno trodimenzionalno polje veličine $W \times H \times D$ i paralelno agregirati cijene te ih na kraju sumirati po putevima u polje S . Memorijska složenost takvog pristupa je $O(r \times W \times H \times D)$, ali će izvođenje biti relativno brzo zbog lokalnosti podataka s kojima se računa.

Ukoliko primjena ima strože zahtjeve na memorijsku potrošnju, a vrijeme izvođenja nije toliko bitno može se koristiti samo jedno polje veličine $W \times H \times D$ za pohranu cijene puta te sekvencijalno izračunati cijene za sve puteve r . Vrijeme izvođenja će bit osjetno dulje, ali će potrebna memorija za pohranu cijena puteva biti r puta manja.



Slika 4.3: Agregacija cijene u dva prolaza, A i B

U stvarnim primjenama prisutna su ograničenja i na vrijeme izvođenja i na memorijsku potrošnju. Uravnotežen pristup koristi $r/2$ polja veličine $W \times H \times D$ za pohranu cijene puteva te paralelizira izračun tako da se agregiraju cijene za r smjerova u dva prolaza. Na taj način se dobiva paralelnost od $r/2$ i iskorištava se lokalnost podataka kod izračuna. Slika 4.3 ilustrira navedeni pristup za $r = 8$.

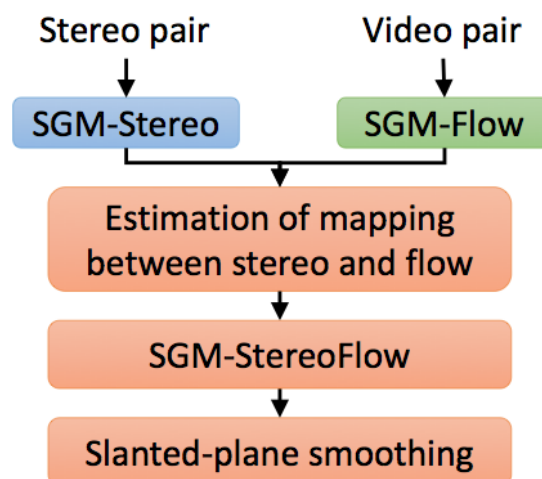
4.3. Izračun mape dispariteta

Rezultat koraka agregacije cijena je trodimenzionalno polje S koje sadrži cijenu odabira svake razine dispariteta za svaki element referentne slike. Primjerice, element polja $S[3][5][7]$ sadrži cijenu odabira vrijednosti dispariteta 7 za slikovni element na lokaciji (3, 5).

Kako cijena u svakom elementu polja S već sadrži cijenu susjedstva po svim putevima, izračun mape dispariteta je jednostavan. Disparitet u nekom slikovnom elementu referentne slike dobiva se iteriranjem kroz cijene u polju S za taj slikovni element po svim mogućim vrijednostima dispariteta s ciljem nalaženja minimalne cijene. Indeks koji indeksira minimalnu cijenu jest vrijednost dispariteta tog slikovnog elementa.

5. Poluglobalno podudaranje s geometrijskom regularizacijom

U [12] opisana je metoda stereoskopske rekonstrukcije koja koristi tri ulazne slike: lijevi pogled na scenu, desni pogled na scenu i lijevi pogled na scenu u kasnijem trenutku, odnosno koristi se stereo par (lijevi i desni pogled na scenu) te video par (lijevi pogled na scenu u dva različita trenutka u vremenu).



Slika 5.1: Koraci obrade za algoritam poluglobalnog podudaranja s geometrijskom regularizacijom

5.1. SGM Flow

SGM Flow je prilagodba SGM algoritma za estimaciju pokreta kamere iz ulaznog video para slika. SGM flow svodi dvodimenzionalni problem pretraživanja na jednodimenzionalno pretraživanje po epipolarnim linijama koje prate kretanje kamere.

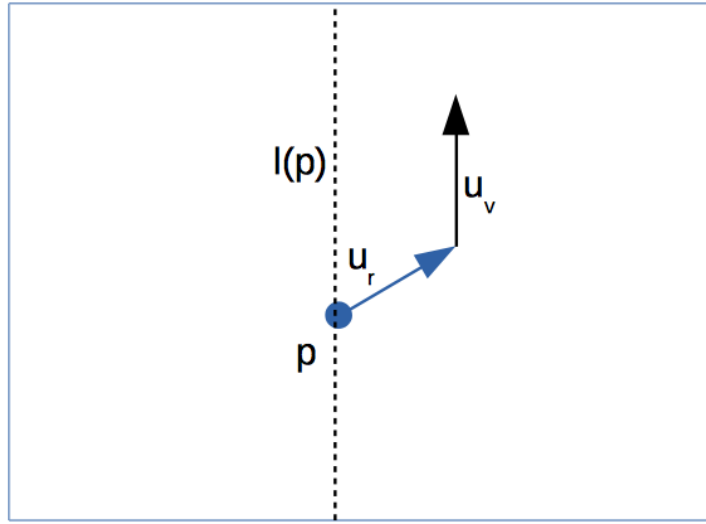
U prvom koraku algoritma potrebno je estimirati fundamentalnu matricu koja defi-

nira epipolarne linije. Pomoću pronađenih *SIFT* značajki obiju slika video para dobivaju se parametri kretanja koji definiraju translaciju i rotaciju kamere posebno.

Označimo s $\mathbf{r} = (r_x, r_y, r_z)$, a s $\mathbf{v} = (v_x, v_y, v_z)$ translaciju kamere između vremena t i vremena $t + 1$. Pod pretpostavkom da je scena statična, vektor kretanja $\mathbf{u} = (u_x, u_y)$ za slikovni element $\mathbf{p} = (x, y)$ na slici u vrijeme t se definira na sljedeći način:

$$\mathbf{u} = u_r(\mathbf{p}) + u_v(\mathbf{p}, Z_p) \quad (5.1)$$

gdje su $u_r(\mathbf{p})$, $u_v(\mathbf{p}, Z_p)$ redom komponente kretanja zbog rotacije i translacije kamere, a Z_p je dubina slikovnog elementa $\mathbf{p}$.



Slika 5.2: Komponente optičkog toka - rotacijska komponenta u_r i translacijska komponenta u_v

Pod pretpostavkom da je rotacija kamere između dvije slike mala, $u_r(\mathbf{p})$ se može izraziti na sljedeći način [10]:

$$u_r(\mathbf{p}) = \begin{pmatrix} fr_y - r_z\bar{y} + \frac{r_y}{f}\bar{x}^2 - \frac{r_x}{f}\bar{x}\bar{y} \\ -fr_x + r_z\bar{x} + \frac{r_y}{f}\bar{x}\bar{y} - \frac{r_x}{f}\bar{y}^2 \end{pmatrix} \quad (5.2)$$

gdje je f žarišna duljina kamere i te vrijedi $\bar{x} = x - c_x$, $\bar{y} = y - c_y$ uz središnju točku (c_x, c_y) . Onda možemo $u_r(\mathbf{p})$ izraziti kao 5-parametarski model.

$$u_r(\mathbf{p}) = \begin{pmatrix} a_1 - a_3\bar{y} + a_4\bar{x}^2 + a_5\bar{x}\bar{y} \\ a_2 + a_3\bar{x} + a_4\bar{x}\bar{y} + a_5\bar{y}^2 \end{pmatrix} \quad (5.3)$$

Dodatno ograničenje koje se može iskoristiti za estimaciju rotacijske komponente kretanja jest činjenica da je $u_v(\mathbf{p}, Z_p)$ paralelna epipolarnoj liniji $l(\mathbf{p}')$ koja prolazi kroz slikovni element na slici u vremenu $t + 1$. Epipolarna linija je definirana izrazom $l(\mathbf{p}') = \mathbf{F}\tilde{\mathbf{p}}$ gdje je $\tilde{\mathbf{p}}$ slikovni element $\mathbf{p}$ izražen u homogenim koordinatama.

Ako je $u_v(\mathbf{p}, Z_p)$ paralelan epipolarnoj liniji $l(\mathbf{p}')$, onda $\mathbf{p} + u_r(\mathbf{p})$ mora biti na $l(\mathbf{p}')$ pa se može napisati sljedeći uvjet:

$$l(\mathbf{p}')^T(\tilde{\mathbf{p}} + \tilde{u}_r(\mathbf{p})) = 0 \quad (5.4)$$

Mogu se estimirati parametri rotacije kamere, $\mathbf{a} = (a_1, \dots, a_5)$ tako da se minimizira suma za sve slikovne elemente s lijeve strane izraza 5.4. Nakon toga je potrebno estimirati kretanje kamere samo u smjeru epipolarnih linija. To je jednodimenzionalni izračun koji nije računski zahtjevan te daje precizne rezultate jer nameće snažnu regularizaciju.

5.1.1. Prilagodba SGM-a

SGM algoritam radi direktno s disparitetima. U slučaju kretanja kamere, korištenje video dispariteta direktno dovodi do međudjelovanja između epipolarne geometrije video para i dubine scene jer je disparitet u svakoj točki kompleksna nelinearna funkcija dubine Z_p . Potrebna je bolja parametrizacija koja je približno linearna.

Translacijsku komponentu možemo izraziti na sljedeći način:

$$u_v(\mathbf{p}, Z_p) = e'(\mathbf{p}) \cdot d(\mathbf{p}, Z_p) \quad (5.5)$$

gdje je $e'(\mathbf{p})$ jedinični vektor u smjeru epipolarne linije, a $d(\mathbf{p}, Z_p)$ je disparitet na epipolarnoj liniji.

[11] definira tzv. *VZ-omjer* kao $\omega_p = \frac{v_z}{Z_p}$ te koristi taj omjer umjesto mjere dispariteta u SGM algoritmu. Tako se dobiva procjena kretanja kamere između vremena t i $t + 1$.

5.2. SGM StereoFlow

Kad je scena statična dobiva se konstantan odnos između dobivenog dispariteta i dobivenog VZ-omjera $\alpha = \omega_p/d_p$. No, zbog pogrešaka u kalibraciji i uparivanju, α je definiran kao linearna funkcija koordinata slike:

$$\omega_p = \alpha(\mathbf{p})d_p = (\alpha_x p_x + \alpha_y p_y + \alpha_c) \quad (5.6)$$

U praksi se $\alpha = (\alpha_x, \alpha_y, \alpha_z)$ robustno estimira korištenjem RANSAC metode nad setom koji sadrži samo slikovne elemente za koje postoji i estimacija dispariteta d_p i estimacija VZ-omjera ω_p .

Koristeći estimirani $(\alpha_x, \alpha_y, \alpha_z)$, formulira se SGM algoritam za zajedničku estimaciju dubine i kretanja pomoću izraza 5.6. Za taj SGM algoritam definiramo energiju sustava kao zbroj podatkovne energije C_{sf} i energije glatkoće S_{sf} :

$$E_{sf}(\mathbf{d}) = \sum_p C_{sf}(\mathbf{p}, d_p) + \sum_{\mathbf{p}, \mathbf{q} \in N} S_{sf}(d_p, d_q) \quad (5.7)$$

gdje je $\mathbf{d}$ polje koje svakom slikovnom elementu pridjeljuje disparitet.

Kaže se da je slikovni element zaklonjen u polju kretanja (odnosno dispariteta) ako SGM-Flow (odnosno SGM-Stereo) ne može estimirati vrijednost za taj slikovni element. Unarna cijena dubine za slikovni element je prosječna vrijednost cijena vrijednosti za kretanje i disparitet. Kad polje dispariteta ili polje kretanja označi slikovni element kao anomalnu vrijednost, cijena se računa koristeći samo vrijednost iz drugog polja. Ostatak postupka je analogan SGM algoritmu za disparitet opisanom u poglavlju 4.

6. Zaglađivanje tehnikom ravninske dekompozicije

Zaglađivanje tehnikom kose ravnine gradi podjelu slike na *segmente* s ravinom za svaki segment, zastavicom anomalne vrijednosti za svaki slikovni element i oznakom granice za svaki par susjednih segmenata.

6.1. Definicija energije

Ukupna energija je označena kao $E(s, \theta, f, o, I, d)$ gdje je s segmentacija, θ pridjeljuje ravninu svakom segmentu, f pridjeljuje zastavicu anomalne vrijednosti svakom slikovnom elementu, o označava vrstu granice za svaki par susjednih segmenata. Neka je s_p indeks segmenta kojega segmentacija s pridjeljuje slikovnom elementu $\mathbf{p}$, a μ_i i c_i srednji položaj i boja segmenta i . Srednji položaj i boja se održavaju inkrementalno u svrhu računske efikasnosti. Neka je $\theta_i = (\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i)$ ravnina dispariteta koju θ pridjeljuje segmentu i . Za svaki slikovni element možemo izračunati disparitet:

$$\hat{d}(\mathbf{p}, \theta_i) = \mathbf{A}_i p_x + \mathbf{B}_i p_y + \mathbf{C}_i \quad (6.1)$$

gdje su (p_x, p_y) koordinate slikovnog elementa $\mathbf{p}$. Neka je $f_p \in \{0, 1\}$ zastavica anomalne vrijednosti slikovnog elementa $\mathbf{p}$.

[12] definira energiju sustava kao zbroj energija za izgled, položaj, disparitet, glatkoću i energije segmentnih granica:

$$\begin{aligned}
E(s, \theta, f, o, I, d) = & \underbrace{\sum_{\mathbf{p}} E_{col}(\mathbf{p}, c_{s_p})}_{\text{boja}} + \underbrace{\lambda_{pos} \sum_{\mathbf{p}} E_{pos}(\mathbf{p}, \mu_{s_p})}_{\text{položaj}} + \underbrace{\lambda_{depth} \sum_{\mathbf{p}} E_{depth}(\mathbf{p}, \theta_{s_p}, f_p)}_{\text{dubina}} \\
& + \underbrace{\lambda_{smo} \sum_{\{i,j\} \in N_{seg}} E_{smo}(\theta_i, \theta_j, o_{i,j})}_{\text{glatkoća}} + \underbrace{\lambda_{com} \sum_{\{i,j\} \in N_{seg}} E_{prior}(o_{i,j})}_{\text{apriorna oznaka}} \\
& + \underbrace{\lambda_{bou} \sum_{\{\mathbf{p}, \mathbf{q}\} \in N_8} E_{bou}(s_p, s_q)}_{\text{duljina granice}} \tag{6.2}
\end{aligned}$$

gdje je ovisnost o I i d implicirana, a N_8 je set parova iz okoline od 8 susjednih slikovnih elemenata. Slijedi detaljniji opis svakog dijela izraza.

Položaj Ovim izrazom zaglađivanje preferira segment pravilnog oblika.

$$E_{pos} = \|\mathbf{p} - \mu_{s_p}\|_2^2 \tag{6.3}$$

Izgled Ovaj izraz potiče slikovne elemente da budu dio segmenta ako su slične boje kao i segment.

$$E_{col} = \|I_{L, c_{s_p}}(\mathbf{p}) - c_{s_p}\|_2^2 \tag{6.4}$$

Disparitet Ovaj izraz potiče estimiranu ravninu da se poklapa s podacima dobivenih iz SGM-StereoFlow koraka. Ako je slikovni element anomalne vrijednosti dodaje se cijena λ_d što sprječava trivijalno rješenje gdje su svi slikovni elementi anomalne vrijednosti.

$$E_{depth}(\mathbf{p}, \theta_{s_p}, f_p) = \begin{cases} (d(\mathbf{p} - \hat{d}(\mathbf{p}, \theta_{s_p})))^2 & \text{ako } f_p = 0 \\ \lambda_d & \text{inače} \end{cases} \tag{6.5}$$

gdje je λ_d konstanta.

Kompleksnost Preferiraju se jednostavna rješenja, odnosno segmenti koji se nalaze na istoj ravnini.

$$E_{prior}(o_{i,j}) = \begin{cases} \lambda_{occ} & \text{ako } o_{i,j} = lo \vee o_{i,j} = ro \\ \lambda_{hinge} & \text{ako } o_{i,j} = hi \\ 0 & \text{ako } o_{i,j} = co \end{cases} \quad (6.6)$$

gdje su λ_{occ} , λ_{hinge} konstante za koje vrijedi $\lambda_{occ} > \lambda_{hinge} > 0$, odnosno za energiju je najpovoljnije ako su segmenti koplanarni, nešto manje povoljno ako segmenti dijele granicu, a najnepovoljnije ako segmenti ne dodiruju uopće.

Podudaranje granica Ovaj izraz potiče ravnine susjednih segmenata da budu slične ako pripadaju istom objektu. Energija između dvije susjedne ravnine ovisi o granici između njih: ako su susjedni segmenti na istoj ravnini njihove ravnine bi se trebale u potpunosti poklapati, ako segmenti zatvaraju kut njihova granica se treba poklapati, a ako se segmenti ne dodiruju onda segment koji je ispred treba biti bliže kameri, odnosno treba imati manji disparitet. Sukladno tome definira se energija glatkoće:

$$E_{smo}(\theta_i, \theta_j, o_{i,j}) = \begin{cases} \phi_{occ}(\theta_i, \theta_j) & \text{ako } o_{i,j} = lo \\ \phi_{occ}(\theta_j, \theta_i) & \text{ako } o_{i,j} = ro \\ \frac{1}{B_{i,j}} \sum_{\mathbf{p} \in B_{i,j}} (\hat{d}(\mathbf{p}, \theta_i) - \hat{d}(\mathbf{p}, \theta_j))^2 & \text{ako } o_{i,j} = hi \\ \frac{1}{S_i \cup S_j} \sum_{\mathbf{p} \in S_i \cup S_j} (\hat{d}(\mathbf{p}, \theta_i) - \hat{d}(\mathbf{p}, \theta_j))^2 & \text{ako } o_{i,j} = co \end{cases} \quad (6.7)$$

gdje je $B_{i,j}$ skup slikovnih elemenata na granici segmenata i i j , S_i skup slikovnih elemenata u segmentu i , a $\phi_{occ}(\theta_{ispred}, \theta_{iza})$ je funkcija koja penalizira granice susjednih segmenata koji se ne dodiruju ako parametri ravnina nisu odgovarajući.

$$\phi_{occ}(\theta_{ispred}, \theta_{iza}) = \begin{cases} \lambda_{pen} & \text{ako } \sum_{\mathbf{p} \in B_{ispred,iza}} (\hat{d}(\mathbf{p}, \theta_{ispred}) - \hat{d}(\mathbf{p}, \theta_{iza})) < 0 \\ 0 & \text{inače} \end{cases} \quad (6.8)$$

gdje je λ_{pen} konstanta.

Duljina granica Segmenti trebaju biti pravilni, odnosno preferiraju se ravne granice

$$E_{bou}(s_p, s_q) = \begin{cases} 0 & \text{ako } s_p = s_q \\ 1 & \text{inače} \end{cases} \quad (6.9)$$

6.2. Efikasni koordinatni spust po blokovima

Minimizacija izraza 6.2 je NP-teška, sadrži velik broj diskretnih (s, f, o) te kontinuiranih (θ, μ, c) varijabli. [12] opisuje efikasan algoritam koordinatnog spusta po blokovima koji je nekoliko redova veličine brži od prijašnjih pristupa.

Algoritam naizmjenično koristi tri koraka:

1. zajednički se određuju anomalne zastavice f za sve slikovne elemente, svaki slikovni elementu se pridjeljuje nekom segmentu s_i , određuju se izgled c i lokacija μ svakog segmenta
2. procjenjuju se granične oznake o
3. procjenjuju se parametri ravnina θ

Slijedi detaljniji opis algoritma.

Algoritam 1 Algoritam koordinatnog spusta po blokovima

Inicijaliziraj segmentaciju na pravilnu mrežu.

Izračunaj μ_i i c_i za svaki segment i .

Inicijaliziraj rješenje pomoću SOT (algoritam 3)

za svaki segment i napravi

Inicijaliziraj θ_i koristeći RANSAC približno minimizirajući

$$\sum_{\mathbf{p}|s_p=1} E_{depth}(\mathbf{p}, \theta_i)$$

kraj za svaki

za svaki $k = 1$ do $vanskih\ iteracija$ napravi

Izračunaj s, f koristeći PSOT (algoritam 2)

za svaki $j = 1$ do $unutarnjih\ iteracija$ napravi

za sve granice (i, j) napravi

$$o_{i,j} = \operatorname{argmin}_{o_{i,j}} E(s, \mu, c, \theta, o, f)$$

kraj za svaki

za svaki segment i napravi

$$\theta_i = \operatorname{argmin}_{\theta_i} E(s, \mu, c, \theta, o, f)$$

kraj za svaki

kraj za svaki

kraj za svaki

Algoritam 2 PSOT: Proširena segmentacija s očuvanjem topologije

Inicijaliziraj stog tako da sadrži sve granične slikovne elemente

sve dok stog nije prazan **napravi**

 Uzmi slikovni element p sa stoga

ako $valjana_povezanost(p) = 0$ **onda**

 nastavi

kraj ako

$\{f_p, s_p\} = \underset{f_p, s_p \in \cup \{s_{N_4(p)}}}{\operatorname{argmin}} E(s, \mu, c, \theta, o, f)$

ako s_p promijenjen **onda**

 Incrementalno ažuriraj μ i c za dva promijenjena segmenta

 Dodaj granične slikovne elemente iz $N_4(p)$ na stog

kraj ako

kraj sve dok

Prvi korak zajednički optimizira segmentaciju, zastavice anomalne vrijednosti, te lokaciju i izgled segmenata uz uvjet da je svaki segment skup povezanih slikovnih elemenata koji nema rupa. U tu svrhu [12] opisuje algoritam za proširenu segmentaciju s očuvanjem topologije (algoritam 2) koji radi na sljedeći način.

Najprije se inicijalizira stog tako da se popuni svim slikovnim elementima koji predstavljaju granice između segmenata. Sve dok stog nije prazan, uzima se slikovni element sa stoga te se vrši provjera da li će promjena segmenta za taj slikovni element prekinuti povezanost njegovog trenutnog segmenta. Ako to nije slučaj, ponovno se postavlja segment za taj slikovni element, njegova zastavica anomalne vrijednosti te lokacija i izgled dva segmenta koja se mijenjaju promjenom pridruženog segmenta, odnosno za segment kojemu je slikovni element bio pridružen u prošloj iteraciji i novi pridruženi segment. Zatim se na stog dodaju novi granični slikovni elementi koristeći susjedstvo od 4 slikovna elementa oko p jer se granica promijenila promjenom pridruženog segmenta.

Algoritam 3 SOT: Segmentacija s očuvanjem topologije

Inicijaliziraj stog tako da sadrži sve granične slikovne elemente

sve dok stog nije prazan **napravi**

Uzmi slikovni element $\mathbf{p}$ sa stoga

ako $valjana_povezanost(\mathbf{p}) = 0$ **onda**

nastavi

kraj ako

$$s_p = \underset{s_p \in \cup \{s_{N_4(\mathbf{p})}\}}{\operatorname{argmin}} \{E_{col}(\mathbf{p}, c_{s_p}) + \lambda_{pos} E_{pos}(\mathbf{p}, \mu_{s_p}) + \lambda_{bou} \sum_{p,q \in N_8}(s_p, s_q)\}$$

ako s_p promijenjen **onda**

Incrementalno ažuriraj μ i c za dva promijenjena segmenta

Dodaj granične slikovne elemente iz $N_4(\mathbf{p})$ na stog

kraj ako

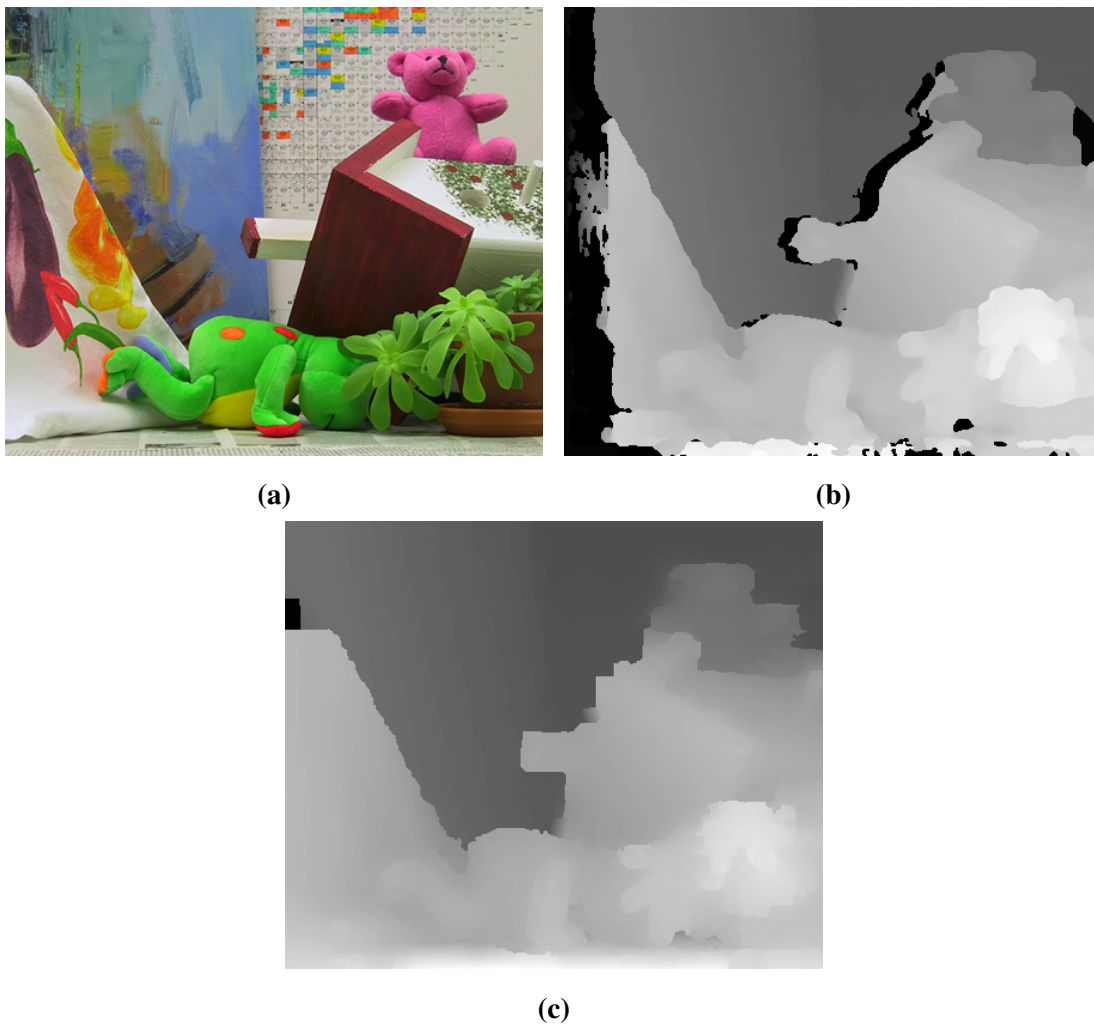
kraj sve dok

7. Eksperimentalni rezultati

[12] uz članak ima pripadajuću implementaciju SGM-a s geometrijskom regularizacijom. Navedena implementacija ne sadrži kod koji koristi procjenu kretanja, već samo SGM koji koristi stereo par te kod za zaglađivanje. Taj kod je korišten za dobivanje rezultata opisanih u nastavku.

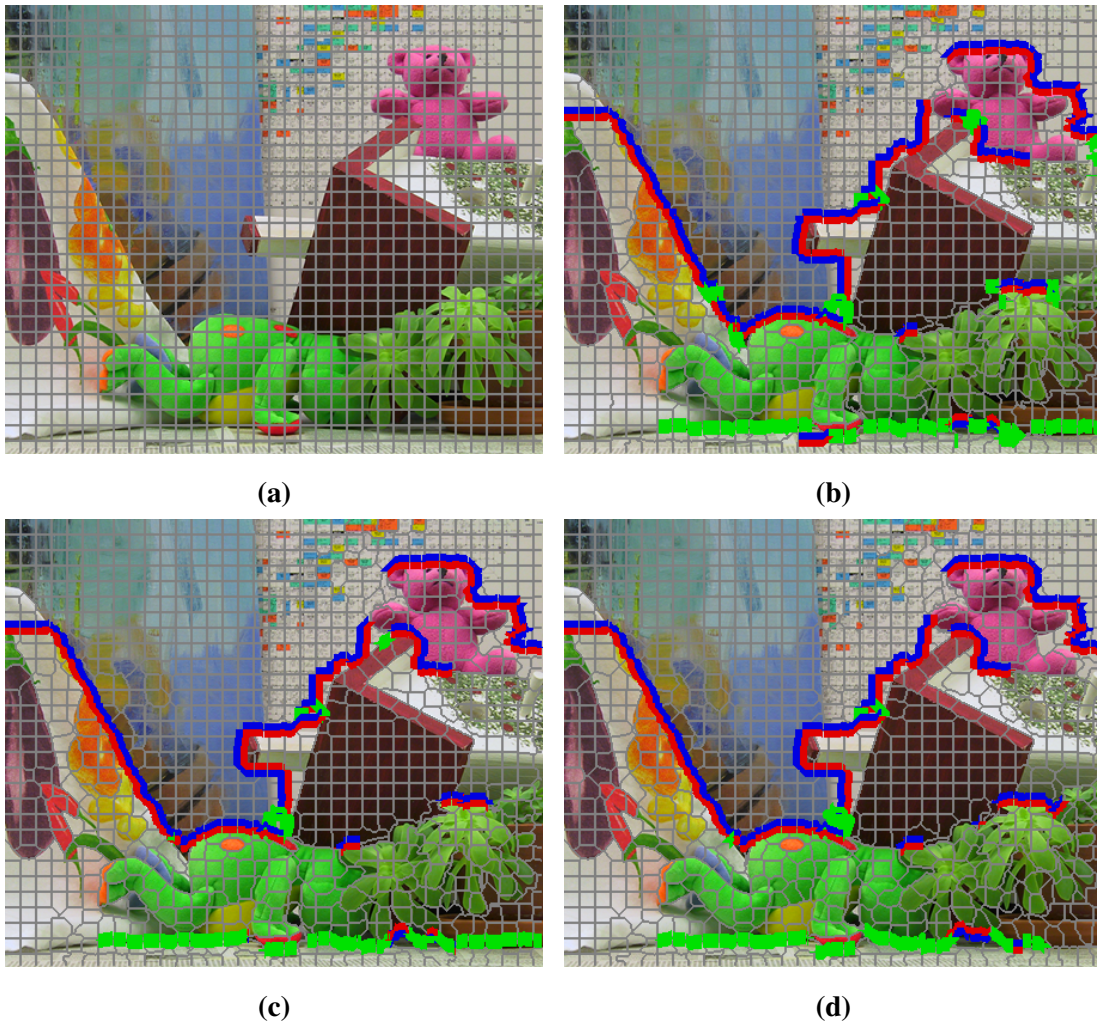
Ako usporedimo mapu dispariteta dobivenu SGM metodom i mapu dispariteta dobivenu SGM metodom s geometrijskom regularizacijom, vidljivo je da geometrijska regularizacija daje "glađu" mapu dispariteta koja sadrži vrlo malo slikovnih elemenata za koje nemamo procjenu dubine (označeno crnom bojom, slika 7.1).

Dodatna svojstvo mape dispariteta dobivene uz geometrijsku regularizaciju jesu i mekše, "razlivenne" granice između objekata - kao što je vidljivo na krovu kućice s ispitne slike teddy (Slika 7.1).



Slika 7.1: Usporedba rezultata lokalne i globalne metode: (a) Ispitna slika 'teddy' iz middle-bury stereo skupa, (b) Mapa dispariteta dobivena SGM metodom, (c) Mapa dispariteta dobivena SGM-om s geometrijskom regularizacijom

Na slici 7.2 vidljivo je da algoritam zaglađivanja brzo konvergira prema rješenje; već nakon prvog prolaza vanske petlje (algoritam 1) dobiva se segmentacija slična konačnoj segmentaciji.



Slika 7.2: Promjena segmentacije kroz deset prolaza zaglađivanja: (a) inicijalna pravilna mreža, (b) segmentacija nakon prvog prolaza, (c) segmentacija nakon trećeg prolaza, (d) segmentacija nakon zadnjeg prolaza

Slika 7.3 uspoređuje stereoskopsku rekonstrukciju na uzorku iz KITTI seta stereo slika [[5]]. Tu je također vidljivo da SGM s geometrijskom regularizacijom daje glađu mapu dispariteta, no još uvijek sadrži greške na rubu slike.



(a)



(b)



(c)



(d)

Slika 7.3: Usporedba stereoskopske rekonstrukcije na uzorku iz KITTI seta: (a) originalna slika, (b) rezultat lokalnog algoritma, (c) rezultat SGM-a, (d) rezultat SGM-a s geometrijskom regularizacijom

7.1. Mapa dispariteta

Tablica 7.1 prikazuje rezultate usporedbe mape dispariteta s idealnom mapom dispariteta za SGM sa i bez geometrijske regularizacije. Prikazani rezultati su dobiveni zbrojem poklapanja vrijednosti za svaki slikovni element mape dispariteta s slikovnim elementom idealne mape dispariteta za iste koordinate.

Algoritam	Podudaranje
SGM (stereoSGBM)	0.79
SGM s geometrijskom regularizacijom	0.88

Tablica 7.1: Usporedba podudaranja dobivene mape dispariteta s idealnom mapom dispariteta, Middlebury skup slika [1]

[12] sadrži usporedbu mapa dispariteta za KITTI skup slika s većim brojem poznatih rješenja za stereoskopsku rekonstrukciju.

7.2. Brzina izvođenja

Na modernom procesoru, SGM s geometrijskom regularizacijom je za red veličine sporiji od SGM-a bez geometrijske regularizacije (Tablica 7.2). Prilikom testiranja korištena je implementacija SGM-a *stereoSGBM* iz biblioteke *OpenCV* [8].

Algoritam	Vrijeme izvođenja (s)
SGM (stereoSGBM)	0.26
SGM s geometrijskom regularizacijom	3.16

Tablica 7.2: Brzina izvođenja stereoskopske rekonstrukcije nad slikama veličine 1241 x 376 na modernom procesoru

Brzina SGM-a s geometrijskom regularizacijom na modernom procesoru nije dovoljno velika za upotrebu u stvarnom vremenu, no na specijaliziranom sklopovlju se takva upotreba gotovo sigurno može realizirati.

8. Zaključak

SGM je poluglobalna metoda za stereoskopsku rekonstrukciju scene koja omogućuju dobivanje dobrih rezultata uz manji trošak procesorskog vremena nego druge globalne metode.

SGM s geometrijskom regularizacijom nadograđuje SGM s dodatnim korakom zaglađivanja kako bi rezultatna mapa dispariteta bila kvalitetnija. Zaglađivanje optimizira segmentaciju slike uzimajući u obzir izgled, položaj i oblik segmenata te podudaranje i duljinu granica između segmenata. Minimizacija ukupne energije po mnogo različitih varijabli se vrši efikasno koristeći koordinatni spust po blokovima.

Geometrijska regularizacija omogućuje kvalitetniju stereoskopsku rekonstrukciju s manje grešaka uz dodatni utrošak procesorskog vremena.

[12] uspoređuje opisani pristup s drugim suvremenim pristupima i dobiva najbolje rezultate dosad na KITTI skupu slika.

Opisani pristup je prikladan za primjene gdje je bitna kvaliteta stereoskopske rekonstrukcije, a vrijeme izračuna nije kritično. Kako je SGM s geometrijskom regularizacijom za red veličine sporiji od običnog SGM-a, ovaj pristup nije prikladan za primjene gdje je vrlo bitna velika brzina izračuna.

LITERATURA

- [1] Middlebury stereo datasets. URL <http://vision.middlebury.edu>.
- [2] Efficient joint segmentation, occlusion labeling, stereo and flow estimation. URL <http://ttic.uchicago.edu/~dmcalister/SPS/>.
- [3] Stan Birchfield i Carlo Tomasi. Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35(3):269–293, 1999.
- [4] Stefan K Gehrig i Clemens Rabe. Real-time semi-global matching on the cpu. U *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, stranice 85–92. IEEE, 2010.
- [5] Andreas Geiger, Philip Lenz, Christoph Stiller, i Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.
- [6] Heiko Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. U *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, svezak 2, stranice 807–814. IEEE, 2005.
- [7] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):328–341, 2008.
- [8] *The OpenCV Reference Manual*. Itseez, 2014. URL <http://docs.opencv.org/opencv2refman.pdf>. Release 2.4.9.0.
- [9] Jean-Marc Lavest, Gérard Rives, i Jean-Thierry Lapresté. Dry camera calibration for underwater applications. *Machine Vision and Applications*, 13(5-6):245–253, 2003.

- [10] H Christopher Longuet-Higgins i Kvetoslav Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B: Biological Sciences*, 208(1173):385–397, 1980.
- [11] Koichiro Yamaguchi, David McAllester, i Raquel Urtasun. Robust monocular epipolar flow estimation. U *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, stranice 1862–1869. IEEE, 2013.
- [12] Koichiro Yamaguchi, David McAllester, i Raquel Urtasun. Efficient joint segmentation, occlusion labeling, stereo and flow estimation. U *Computer Vision–ECCV 2014*, stranice 756–771. Springer, 2014.

Gusta stereoskopska rekonstrukcija scene predstavljene ravninskim odsječcima

Sažetak

Stereoskopska rekonstrukcija koristi dvije slike iste scene za rekonstrukciju dubine objekata u sceni. Pomoću niza stereo slika šahovske ploče računaju se intrinzični i ekstrinzični parametri kamere. Pomoću parametara kamere slike se ispravljaju i rektificiraju. Rektificirane slike poštuju pravila epipolarne geometrije što znači da su značajke na obje slike vertikalno poravnate pa je dovoljno tražiti podudarajuće značajke uzduž vodoravne linije.

Razlikujemo lokalne i globalne metode stereoskopske rekonstrukcije. Lokalne metode traže podudarajuće značajke unutar prozora fiksne veličine. Globalne metode su sporije, ali preciznije jer uzimaju u obzir glatkoću slike. SGM je poluglobalna metoda stereoskopske rekonstrukcije koja optimizira energiju definiranu kao funkcija podudaranja značajki i glatkoće slike.

Geometrijska regularizacija je dodatni korak zaglađivanja koji se primjenjuje nakon SGM-a kako bi se dobila kvalitetnija rekonstrukcija scene. Zaglađivanje dijeli sliku na segmente te optimizira segmentaciju po svojstvima segmenata i ranije izračunatom disparitetu. SGM s geometrijskom regularizacijom je za red veličine sporiji od običnog SGM-a, ali daje značajno bolje rezultate.

Ključne riječi: računalni vid, kalibracija, rektifikacija, stereoskopska rekonstrukcija, gusti stereo, poluglobalne metode, sgm, geometrijska regularizacija

Dense stereoscopic reconstruction of a piecewise planar scene

Abstract

Stereoscopic reconstruction uses two images of the same scene to reconstruct the scene depth. A set of stereo images of a chessboard is used to calculate the extrinsic and intrinsic parameters of the cameras. These parameters are then used to undistort and rectify the images. Rectified images are subject to the rules of epipolar geometry which means that image features can be matched along a horizontal line.

The two types of stereoscopic reconstruction methods are local and global. Local methods match features within a fixed window on the image. Global methods are slower, but more accurate because they take into account the image smoothness. SGM is a semi-global method which optimizes the energy defined as a function of feature matching and image smoothness.

Geometric regularization is an additional smoothing step performed after SGM to get a more accurate scene reconstruction. Smoothing divides the image into segments and optimizes the segmentation in regard to properties of the segments and the disparities calculated by SGM earlier. SGM with geometric regularization is an order of magnitude slower than a regular SGM, but it produces significantly better results.

Keywords: computer vision, calibration, rectification, stereoscopic reconstruction, dense stereo, semi-global methods, sgm, plane segments