

REKONSTRUKCIJA KORISTEĆI SLIKE SA INTERNETA I ALGORITAM POSTEPENOG RASTA REGIONA

Andrija Vučinić, Vladimir Božović, Univerzitet Crne Gore, Prirodno-matematički fakultet
Dubravko Čulibrk, Vladimir Crnojević, Univerzitet Novi Sad, Fakultet tehničkih nauka

Sadržaj - Savremeni pristupi rekonstrukciji 3D modela na bazi fotografija su doživjeli revoluciju posljednjih godina. Stereo algoritmi koji koriste više pogleda danas mogu biti suparnici laserskim skenerima po pitanju preciznosti. Algoritmi postepenog rasta regiona zauzimaju jedno od vodećih mjesta u kreiranju gustih modela. Ideja ovog rada je da opiše osnovne korake u rekonstrukciji 3D modela koristeći slike sa interneta i algoritam postepenog rasta regiona kao i probleme koji se mogu susresti u ovom pristupu. U radu je data i sistematizacija postojećih algoritama sa opisom prednosti i nedostataka kao i pregled nekih perspektivnih pravaca istraživanja u ovoj oblasti. Izvršena je eksperimentalna evaluacija algoritma na skupovima slika koje su prikupljene na internetu, kao i na nekim standardnim skupovima za probleme 3D rekonstrukcije. Za implementaciju su korištena rješenja otvorenog koda.

Ključne riječi: Računarska vizija, Struktura iz pokreta, 3D/stereo analiza scene, Stereo algoritmi sa više pogleda, modeliranje fizičkih atributa

1. UVOD

Cilj stereo algoritama sa više pogleda je da rekonstruiše 3D model objekta iz kolekcije slika koje su napravljene po poznatim parametrima (pozicije kamera, žižne daljine). Posljednjih godina su razvijeni veoma kvalitetni algoritmi, i state-of-art ubrzano napreduje [1, 2, 3]. Istovremeno, digitalna fotografija zajedno sa internetom je stvorila veliku bazu fotografija objekata slikanih sa bezbroj pozicija, različitim osvjetljenjem, nivoom detalja, kao i različitim vremenskih razdoblja. Na primjer, Google pretraga za pojmom "Eiffel Tower" vraće na hiljade slika istog.

Za očekivati je da se ove slike iskoriste za rekonstrukciju i da se napravi sistem koji bi vratio vjerodostojan 3D model za upotrebu. U ovom radu je predstavljen sistem koji iz kolekcije slika prikupljenih na internetu rekonstruiše 3D model kroz 3 koraka, koristeći komponente otvorenog koda. Proizvedeni model je mrežasta mapa¹.

U Sekciji 2 će biti dat pregled relevantnih algoritama na polju 3D rekonstrukcije, kao i njihova podjela po fundamentalnim osobinama.

Algoritam i njegovi koraci su opisani u Sekcijama 3 i 4 sa komentarom na prednosti i mane ovakvog pristupa, kao i detalje na koje treba obratiti pažnju ako se koriste stereo algoritmi sa više pogleda.

Rezultati predstavljeni u Sekciji 5 su dobijeni na kolekcijama slika sa interneta, kao i na provjerenim skupovima za testiranje algoritama za rekonstrukciju.

2. PREGLED RELEVANTNE LITERATURE

Algoritmi za rekonstrukciju značajno variraju po pitanju osnovnih pretpostavki, domenu rada i ponašanja.

Kategorišemo algoritme u 6 grupa: po *reprezentaciji scene*, *mjeri fotokonzistentnosti*, *modelu vidljivosti*,

početnom obliku, *algoritmu za rekonstrukciju*, i *inicijalnim parametrima*.

Geometrija objekta ili scene se može predstaviti na više načina. Većina koristi voksele [5], nivoe [6], poligonalne mreže [7] ili dubinske mape [8].

Mjera fotokonzistentnosti upoređuje piksele jedne slike sa ostalima, i daje nam mjeru korelacije. Može se definisati u *domenu scene* [9] ili *domenu slike* [10].

Modeli vidljivosti specificiraju kakav pogled da koristimo kad evaluiramo mjeru fotokonzistentnosti. Postoje *geometrijski* [11], *kvazi-geometrijski* [12] i *pristupi bazirani na izuzecima* [13].

Nekad fotokonzistentnost nije dovoljna za preciznu rekonstrukciju, pa se zato uvodi pretpostavka *početnog oblika* kako bi "usmjerili" algoritam i dobili željene osobine. [14]

Algoritmi za *rekonstrukciju* se mogu podijeliti u 2 grupe. Prvoj grupi pripadaju algoritmi koji iterativno mijenjaju površ minimizirajući funkciju cijene [15, 16]. Druga grupa su algoritmi koji koriste dubinske mape kako bi rekonstruisali model, bilo da se mjeri konzistentnost između mapa ili se one spajaju kako bi se dobila 3D scena [17, 18].

Svi algoritmi za rekonstrukciju, pored parametara kamera zahtijevaju i neke *inicijalne parametre*. Uglavnom algoritmi zahtijevaju grub granični okvir [9, 11], dok je nekim algoritmima potrebna segmentacija prednjeg plana ili pozadine za svaku sliku.

3. ALGORITAM

Sistemu se predaje kolekcija slika iz koje je potrebno rekonstruisati model. Sistem pokušava da dobije potrebne parametre za svaku sliku. Slike za koje ne može izračunati parametre odbacuje.

Zatim se minimizuje broj slika koje ćemo koristiti za rekonstrukciju, jer se veliki broj slika preklapa, te je

¹ Meshmap

nepotrebno koristiti "slične" slike. Ovim ubrzavamo izvršavanje rekonstrukcije. Sama rekonstrukcija se vrši iterativno, dodajući sliku po sliku, i poklapajući rekonstruisane regione i postepeno ih povećavajući dodavanjem slika.

Dakle, algoritam se izvršava u 3 koraka:

1. *ekstrakcija parametara kamera i rekonstrukcija rijetkog 3D modela*
2. *grupisanje ulaznih fotografija kako bi dobili minimalni dovoljni skup za rekonstrukciju*
3. *rekonstrukcija gustog 3D modela*

3.1. EKSTRAKCIJA PARAMETARA KAMERA I REKONSTRUKCIJA RIJETKOG 3D MODELA

Za rekonstrukciju su nam potrebni precizni parametri o relativnoj lokaciji, orijentaciji kao i podaci o žižnoj daljini za svaku fotografiju u kolekciji. Mnoge digitalne kamere ugrađuju podatak o žižnoj daljini kao i još neke korisne informacije u EXIF² oznaci slike, ali su često ove informacije neprecizne.

Korišćeni algoritam ne koristi ove parametre već upotrebljava robustan algoritam za strukturu-iz-pokreta³.

Prvi korak je nalaženje svih ključnih tačaka⁴. Koristi se SIFT [19] algoritam za nalaženje ovih tačaka, zato što je invarijantan na transformacije nad slikom. Tipična slika sadrži oko par stotina ključnih tačaka.

Za svaki par slika poklapamo njihove ključne tačke koristeći algoritam najbližeg susjeda [20]. Zatim računamo fundamentalnu matricu za svaki par koristeći RANSAC [21]. Na kraju odbacujemo poklapanja koja su izuzeci za dobijenu fundamentalnu matricu. Ako je broj preostalih poklapanja manji od 20, odbacujemo ih sva, u suprotnom ih zadržavamo za razmatranje.

Nakon što nadjemo skup geometrijski konzistentnih poklapanja, organizujemo ih u *trake*, gdje svaka traka sadrži poklopljene tačke iz slika. Ako traka sadrži 2 tačke iz iste slike, odbacujemo je.

Drugi korak je dobijanje parametara kamera i 3D lokacija za svaku traku. Parametri su dobijeni minimizujući rastojanje između projekcija svake trake i korespondirajućih slika.

Problem se rješava iterativno, dodajući kameru po kameru u razmatranje. Inicijalni par kamera bi trebalo da sadrži veliki broj poklapanja. Sledeću kameru biramo po broju već rekonstruisanih tačaka koje vidi, i inicijalizujemo njene parametre koristeći direktnu linearnu transformaciju (DLT) [22] unutar RANSAC procedure.

Na kraju, dodajemo trake koje nova kamera vidi procesu optimizacije. Traku dodajemo ako je vidi makar jedna kamera za koju smo dobili lokaciju. Procedura se ponavlja sve dok ima kamera koje vide rekonstruisane tačke.

Proces se optimizuje tako što nakon svake iteracije

nađemo traku koja sadrži tačku sa najvećom greškom projekcije i uklonimo je.



Slika 1: Vizuelizacija ekstraktovanih parametara kamera

Trajanje izvršavanja na uzorku od 2.600 slika je oko 2 nedjelje, dok je na uzorku od 120 slika do 10 sati.

3.2. GRUPISANJE ULAZNIH FOTOGRAFIJA

Ideja ovog koraka je da početni skup slika razbijemo u manje grupe slika sa velikim stepenom preklapanja kako bi ubrzali korak rekonstrukcije gustog modela. Moramo uzeti u obzir da imamo 3 ograničenja:

1. minimizujemo broj slika koje ćemo koristiti u koraku rekonstrukcije
2. veličina svake grupe je manja od proizvoljnog broja koji određuju računarski resursi (*ograničenje veličine*)
3. novodobijeni skup mora sadržati bar 70% ključnih tačkaka nađenih u prethodnom koraku (*ograničenje pokrivenosti*)

Uklanjanjamo slike iz početnog skupa sve dok je ograničenje pokrivenosti zadovoljeno. Slike su poredane u rastućem poretku po rezoluciji (broj piksela), tako da će prvo biti uklonjene slike sa manjom rezolucijom.

1. Novodobijeni skup dijelimo u manje grupe, tako da zadovoljimo ograničenje veličine. Dijeljenje grupa se vrši algoritmom normalizovanog podrezivanja [23] grafa vidljivosti u kome su čvorovi slike. Težina grane između čvorova I_i i I_j , mjeri koliko ta dva čvora doprinose rekonstrukciji. Mjera doprinosa je broj ključnih tačaka koje čvorovi sadrže. Jasno je da će čvorovi sa velikim doprinosom biti kasnije odsječeni. Korak se ponavlja dok ne zadovoljimo ograničenje veličine za sve grupe.

2. U slučaju da smo u koraku 2 prekršili ograničenje pokrivenosti trebamo svakoj grupi dodati slike kako bi ponovo uspostavili pokrivenost. Prvo napravimo listu mogućih dodavanja, sortirano po efikasnosti koju definišemo kao broj nepokrivenih ključnih tačaka.

Nakon dodavanja slika, moguće je da smo prekršili ograničenje veličine, te ćemo ponavljati korake 1 i 2 sve dok ne zadovoljimo oba ograničenja.

3.3. REKONSTRUKCIJA GUSTOG 3D MODELA

Na novom skupu za rekonstrukciju detektujemo ivice⁵ i

2 Exchangeable image file format (EXIF) specificira format slike za digitalne kamere kao i dodatne meta oznake (na primjer: proizvođač i model kamere, žižna daljina, datum i vrijeme)

3 Structure-from-motion (SfM)

4 Feature point

5 Harris detektor

blob⁶ osobine slika. Da bi smo garantovali uniformnu pokrivenost svaku sliku dijelimo u ćelije veličine $B \times B$ i označavamo sa $C(i, j)$ ćeliju u i -tom redu, j -oj koloni, i za svaku nalazimo N^7 lokalnih maksimuma dva detektora osobina. Nakon nalaženja osobina, sparujemo tačke na više slika da bi smo rekonstruisali rijedak skup regiona koji će kasnije biti prošireni. Razmotrimo sliku I sa optičkim centrom O odgovarajuće kamere. Za svaku osobinu f nađenu na I , tražimo skup osobina F' koje leže unutar 2 piksela od odgovarajućih epipolarnih linija i trijangulacijom dobijamo 3D tačke za parove $(f, f \in F')$. Svaku od ovih tačaka razmatramo kao potencijalni centar regiona u rastućem poretku u odnosu na rastojanje od O i uzimamo prvi fotokonzistentni region u najmanje M^8 slika. Ovim dobijamo skup tačaka koje ćemo proširivati da bi dobili gusti model.

Sada iterativno dodajemo nove susjede postojećim regionima dok ne pokriju površine vidljive u sceni. Dva regiona p i p' su susjedni ako se nalaze u susjednim ćelijama $C(i, j)$ i $C(i', j')$ na istoj slici I . Svaka ćelija $C(i, j)$ u sebi sadrži i 2 skupa: Q_t - rekonstruisani regioni koji su vidljivi u ćeliji $C(i, j)$ i Q_f - rekonstruisani regioni potencijalno vidljivi u ćeliji $C(i, j)$. Pokušavamo dodati novi region samo ako je $Q_t(i', j')$ prazan i nijedan element $Q_f(i', j')$ nije n -susjedan [2]. Kad su oba uslova zadovoljena, dodajemo novi region p' tako da njegov centar bude presjek ose koja prolazi kroz centar ćelije $C(i', j')$ i ravan koja sadrži region p .

Nakon svake iteracije filtriramo regione, tako da nijedan region ne zaklanja bilo koji drugi region. Ovo garantuje da kod kolekcija slika koje imaju puno zaklonjenih djelova rekonstruisemo samo regione koji su vidljivi na većini slika.

4. REZULTATI



Slika 2: Neke od slika iz skupa "Notre Dame" i rekonstruisani model

Za poređenje rezultata smo koristili skupove podataka iz [4]. Naglašavamo da su to skupovi sa datim parametrima kamera, te je za očekivati da sistem koji sam procjenjuje ove

6 Difference-of-Gaussian

7 Koristimo $N = 4$

8 Koristimo $M = 3$

parametre da nešto lošije rezultate.

	"Dino" skup	"Temple" skup
Kompletnost	55.00%	88.00%

Na ovim skupovima se mogu dobro uočiti i prednosti i mane sistema.

Model za Hram skup je veoma dobro rekonstruisan, i ovo je očekivani rezultat sistema za proizvoljan skup slika sa interneta.

Dinosaurius skup je rekonstruisan sa manjom kompletnošću zato što SIFT algoritam za nalaženje ključnih tačaka, kao blob detektori i detektori ivica daju loše rezultate za slabo teksturirane slike. Ovo je jedan od generalnih problema na polju rekonstrukcije.

Nažalost, ne postoje načini za evaluaciju rekonstrukcija koristeći skupove slika sa interneta. Jasno je da je veoma teško dobiti adekvatan model, npr. Katedrale Notre-Dame, te ove rezultate prezentujemo vizuelno bez statističkih podataka.



Slika 3: Neke od slika iz skupa "Hall" i rekonstruisani model

5. ZAKLJUČAK

Kao što je predstavljeno, sve je lakše napraviti sistem od komponenti otvorenog koda koji daje dobre rezultate. Algoritam se može poboljšati ubrzavajući fazu ekstrakcije parametara kamera, obzirom da on postaje veoma spor kako broj fotografija raste. Takođe je moguće popraviti fazu rekonstrukcije, tako što ćemo dopuniti "praznine" u modelu da na najjednostavniji način spajaju piksele koji ih okružuju.

Budući korak našeg istraživanja će biti pokušaj da se

napravi interaktivni 3D model. Sistemu bi se predale fotografije konkretnog objekta, a kao izlaz dobijamo 3D rekonstrukciju u kojoj možemo vidjeti svaku sliku ponaosob i kretati se kroz model.

Ovakav sistem može naći primjenu u marketinške svrhe jer se kreiraju vizuelni modeli koji se mogu koristiti za demonstracije, kao i u medicini za pripremanje komplikovanih zahvata.

LITERATURA

- [1] Noah Snavely, Steven M. Seitz, Richard Szeliski. Photo Tourism: Exploring image collections in 3D. *ACM Transactions on Graphics*, 2006.
- [2] Yasutaka Furukawa, Jean Ponce. Accurate, Dense, and Robust Multi-View Stereopsis, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, July 2007.
- [3] Yasutaka Furukawa, Brian Curless, Steven M. Seitz, and Richard Szeliski. Towards Internet-scale Multi-view Stereo, *CVPR* 2010.
- [4] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. *CVPR*, 1, 2006.
- [5] T. Fromherz and M. Bichsel. Shape from multiple cues: Integrating local brightness information. *In ICYCS*, 1995.
- [6] H. Jin, S. Soatto, and A. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *IJCV*, 63(3):175–189, 2005.
- [7] C. Hernandez and F. Schmitt. Silhouette and stereo fusion for 3D object modeling. *CVIU*, 96(3):367–392, 2004.
- [8] R. Szeliski. A multi-view approach to motion and stereo. *In CVPR*, vol. 1, pp. 157–163, 1999.
- [9] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. *IJCV*, 35(2):151–173, 1999.
- [10] S. Savarese, H. Rushmeier, F. Bernardini, and P. Perona. Shadow carving. *In ICCV*, pp. 190–197, 2001.
- [11] K. Kutulakos and S. Seitz. A theory of shape by space carving. *IJCV*, 38(3):199–218, 2000.
- [12] P. Narayanan, P. Rander, and T. Kanade. Constructing virtual worlds using dense stereo. *In ICCV*, pp. 3–10, 1998.
- [13] M.-A. Drouin, M. Trudeau, and S. Roy. Geo-consistency for wide multi-camera stereo. *In CVPR*, vol. 1, pp. 351–358, 2005.
- [14] G. Vogiatzis, P. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. *In CVPR*, pp. 391–398, 2005.
- [15] G. Zeng, S. Paris, L. Quan, and F. Sillion. Progressive surface reconstruction from images using a local prior. *In ICCV*, pp. 1230–1237, 2005.
- [16] H. Saito and T. Kanade. Shape reconstruction in projective grid space from large number of images. *In CVPR*, vol. 2, pp. 49–54, 1999.
- [17] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. *In ECCV*, vol. III, pp. 82–96, 2002.
- [18] C. Zitnick, S.-B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. on Graphics*, 23(3):600–608, 2004.
- [19] Lowe, D. G., “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2004.
- [20] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman and A. Wu, An optimal algorithm for approximate nearest neighbor searching, *Journal of the ACM*, 45(6):891-923, 1998.
- [21] M. A. Fischler, R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, Vol 24, pp 381-395, 1981.
- [22] Hartley R., Zisserman A. - Multiple View Geometry in Computer Vision, 2004.
- [23] J. Shi, J. Malik. Normalized cuts and image segmentation. *PAMI*, 22(8):888-905, 2000.