Srđan Radović i Valentina Njaradi

Praćenje šake korišćenjem dubinske kamere

U ovom radu analizirane su performanse sistema za određivanje pozicije svih zglobova šake korišćenjem dubinske kamere. Dobijeni sistem može da rekonstruiše pozu šake u kompleksnim položajima, kao i u položajima gde nisu svi prsti vidljivi. Pozicije i rotacije zglobova koje ovaj sistem određuje sa slike šake zajedno čine vektor položaja. Prvi korak sistema čini inicijalizator koji se sastoji od nekoliko šuma odluke (random decision forest) i određuje skup vektora položaja koji približno odgovaraju ulaznoj slici. Inicijalizator je treniran na bazi sintetisanih dubinskih slika šake u različitim pozama. Polazeći od skupa vektora položaja dobijenih od inicijalizatora, do tačnijeg vektora položaja dolazi se optimizacijom rojeva (particle swarm optimization). Određivanje tačnosti vektora položaja vrši se direktnim poređenjem ulazne slike sa sintetisanom dubinskom slikom 3D modela šake koji odgovara datom vektoru. Sistem je na sintetičkim slikama u stanju da odredi 78% pozicija zglobova šake sa greškom manjom od 3 cm, što je uporedivo sa drugim sistemima za te svrhe.

Srđan Radović (1999), Bačko Dobro Polje, Vojvođanska 56, učenik 4. razreda Gimnazije „J. Jovanović Zmaj" u Novom Sadu

Valentina Njaradi (2001), Beograd, 11. Krajiške divizije 38, učenica 3. razreda Matematičke gimnazije u Beogradu

MENTORI:

Ratko Amanović, student Elektrotehničkog fakulteta u Beogradu

Damjan Dakić, Majkrosoft razvojni centar Srbije

Uvod

Određivanje tačne pozicije i poze šake je veoma bitno za interakciju korisnika i kompjutera. Istraživanja u ovoj oblasti se vrše već decenijama, ali precizno praćenje šake i dalje ostaje veoma složen problem usled kompleksnosti i raznovrsnosti pokreta. Rešenje ovog problema moglo bi da promeni način upravljanja kompjuterskim procesima, sa širokim primenama u proširenoj stvarnosti (augmented reality), robotici i industriji kompjuterskih igara (Sharp *et al*. 2015).

Neka od postojećih rešenja zahtevaju specijalne senzorske rukavice (Dipietro *et al*. 2008), šarene rukavice (Wang i Popović 2009) ili reflektujuće markere (Zhao *et al*. 2012). Međutim, dodatan hardver na ruci u manjoj ili većoj meri ograničava njene pokrete. U radovima koji ne koriste dodatan hardver, glavni fokus je na metodama sa različitom konfiguracijom kamera. Međutim, postoje određeni problemi pri oslanjanju samo na slike

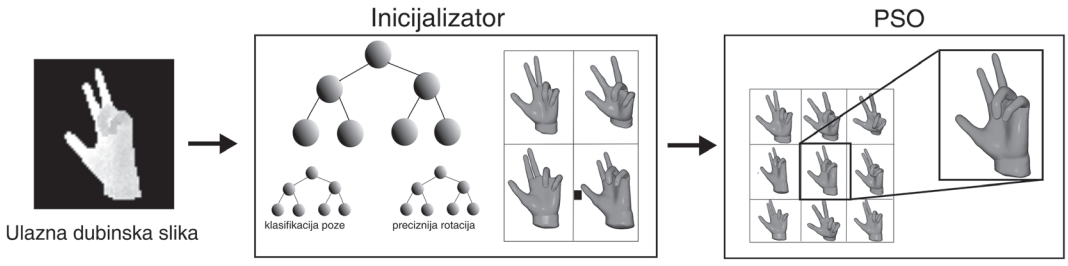
sa kamera, na primer, nisu svi prsti vidljivi u svakoj poziciji šake, vidljivost na slici RGB kamere se značajno pogoršava pri niskom osvetljenju. Zbog ovakvih problema često dolazi do grešaka pri određivanju tačne pozicije šake. Ovo dovodi do značajnih ograničenja u sistemima praćenja šake, kao što je mala dozvoljena udaljenost šake od kamere ili potreba da se koristi veći broj kamera. U radu Šrinata Šridra i saradnika (Sridhar *et al*. 2013) predložena je metoda sa pet RGB kamera i dubinskim senzorom, i dobijeni su vizuelno dobri rezultati, sa prosečnom greškom pozicije vrhova prstiju od 13.24 mm. Glavni nedostatak tog rada jeste komplikovana konfiguracija kamera, ali je problem rešen istraživanjem Majkrosoftovog tima (Sharp *et al*. 2015), gde se koristi samo jedna dubinska kamera, a dobijeni su rezultati približni rezultatima rada Šridra i saradnika (Sridhar *et al*. 2013).

Cilj ovog rada je ispitivanje mogućnosti određivanja pozicije šake i svih zglobova prstiju pomoću jedne dubinske kamere, kao i performansi takvog sistema. Značajna prednost korišćenja dubinskih u odnosu na RGB kamere je u njihovoj invarijantnosti na osvetljenje i različite pozadine. Ova metoda omogućava veliku fleksibilnost, odnosno mogućnost da se šaka nalazi u bilo kojoj poziciji i na bilo kojoj udaljenosti od kamere, a pritom ne zahteva ni statičnost kamere.

Metod

Na slici 1 prikazana je šema postupka. Proces određivanja optimalnog vektora položaja podeljen je u dva koraka. U prvom koraku se inicijalizatorom pozicija koji implementira algoritam mašinskog učenja određuje skup približno tačnih vektora. Zatim se, polazeći od tog skupa vektora određuje optimalan vektor položaja, korišćenjem algoritma optimizacije rojeva.

Ulaz opisanog algoritma je dubinska slika šake. Dubinska kamera vraća 2D sliku kod koje vrednosti piksela predstavljaju udaljenost između kamere i objekta. Sa dubinske slike šake potrebno je odrediti odgovarajuće pozicije i rotacije svih 20 zglobova prstiju, kao i globalnu rotaciju šake. Ove

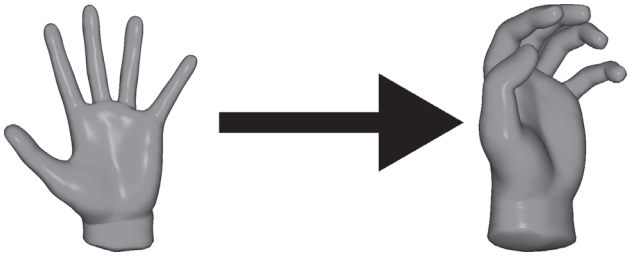


Slika 1. Šematski prikaz algoritma (PSO – algoritam optimizacije rojeva)

Figure 1. Algorithm pipeline. From left to right: Input depth image; Random decision forest initializer for finding approximate pose vectors; Particle swarm optimization algorithm for optimizing pose vectors.

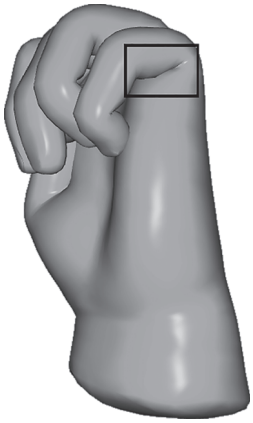
rotacije zajedno, predstavljene u obliku kvaterniona, čine vektor položaja. S obzirom na to da vektor položaja u potpunosti opisuje položaj šake u datom trenutku, pomoću njega je moguće manipulisati i menjati izgled 3D modela šake. Koliko je neki vektor položaja tačan proverava se tako što se 3D model transformiše dobijenim vektorom položaja, sintetiše se dubinska slika modela i uporedi sa ulaznom slikom sa kamere.

3D model. Za generisanje baze podataka korišćen je 3D model iz biblioteke libigl (Jacobson *et al*. 2018). Da bi model mogao da se pomera i menja oblik u zavisnosti od zadate poze, neophodan je i kinematski skelet vezan za model. Skelet se sastoji od 20 kostiju prstiju i podlaktice. Menjanjem vektora položaja koji opisuje skelet, postiže se menjanje oblika celog modela, kao što je prikazano na slici 2. Ovo se postiže LBS algoritmom (Linear Blend Skinning) implementiranim u biblioteci. Ovaj algoritam za jednu tačku mreže modela računa transformacije koje potiču od rotacija njenih okolnih zglobova. Težinskim usrednjavanjem tih transformacija dobija se konačna transformacija za datu tačku. Primenom ovog postupka na sve tačke modela dobija se transformisani model koji odgovara datom vektoru položaja. Nedostatak opisanog algoritma je gubitak zapremine pri uglu rotacije zgloba većem od 90°, takozvani „candy-wrapper" artefakt (Abu Rumman i Fratarcangeli 2017; slika 3).



Slika 2. Menjanje 3D modela šake pod uticajem vektora položaja

Figure 2. Transformation of 3D hand model by pose vector



Slika 3. Gubitak zapremine kod LBS algoritma

Figure 3. Volume loss with linear blend skinning

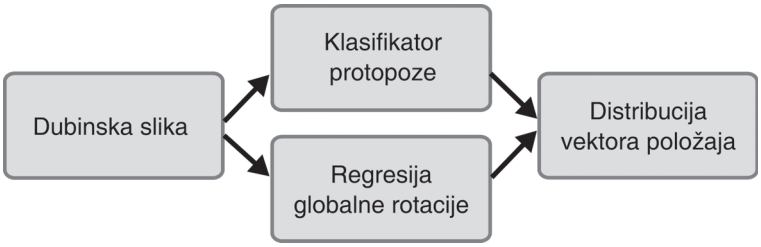
Sintetička baza podataka. Nakon transformacije 3D modela vektorom položaja (slika 2), moguće je sintetisati dubinsku sliku. Parametri vektora položaja su varirani u određenim granicama, za svaki od njih je sintetisana dubinska slika modela šake i dobijena je baza labeliranih dubinskih slika. Ograničenja globalne rotacije određena su empirijski. Rotacije zglobova prstiju su međusobno zavisne, i uvodi se nekoliko različitih poza za prste i šaku, u daljem tekstu protopoza. Korišćeno je 6 protopoza: otvorena (open), zatvorena (closed), poluotvorena (halfopen), ravna (flat), pokazivanje (pointing), i „štipanje" (pinching) (Sharp *et al*. 2015). Za svaku od protopoza varirane su rotacije zglobova sa odgovarajućim ograničenjima, i za svaku od njih je generisano 12 500 slika. Diskretizacijom globalnih rotacija slika, baza je podeljena na 125 grupa (Sharp *et al*. 2015). Zatim je slikama dodat Gausov šum radi simuliranja šuma realnog dubinskog senzora. Sintetička baza sadrži ukupno 75 000 slika veličine 64×64 piksela. Labele slika se sastoje od vektora položaja, protopoza i grupe globalne rotacije. Na slici 4 je prikazano nekoliko primera iz baze.

Slika 4.
Primeri slika iz baze

Figure 4.
Sample images from the database

Inicijalizator pozicija. Ulaz inicijalizatora je dubinska slika šake, a izlaz distribucija vektora položaja koji približno odgovaraju ulaznoj slici. Hipoteza je da globalna rotacija najviše utiče na izgled 3D modela, tj. da će rezultati biti bolji ako se ona tačnije odredi. Zato su upoređivani jednoslojni i dvoslojni inicijalizator. Jednoslojni inicijalizator u jednom koraku odredi globalnu rotaciju i protopozu. Kod dvoslojnog inicijalizatora prvi sloj grubo određuje globalnu rotaciju, dok drugi sloj određuje protopozu i precizniju globalnu rotaciju. Nijedan od konačno dobijenih vektora nije potpuno odgovarajuć ulaznoj slici, ali su približni, što i jeste uloga inicijalizatora. Zbog velike količine raznovrsnih podataka koji su potrebni za treniranje inicijalizatora, napravljena je sintetička baza podataka.

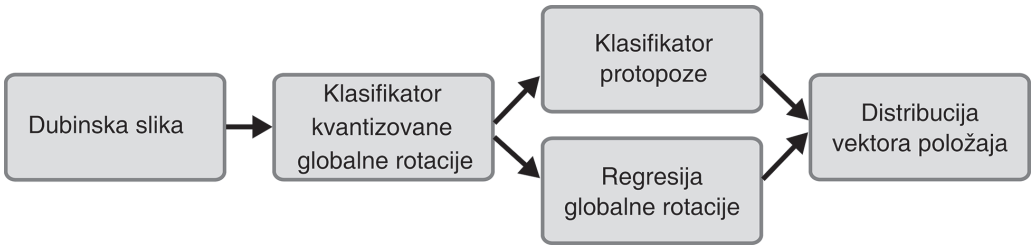
Struktura jednoslojnog inicijalizatora. Šema jednoslojnog incijalizatora je prikazana na slici 5. Sastoji se iz jedne šume odluka za regresiju globalne rotacije i jedne za klasifikaciju protopoze. Ove dve šume se treniraju na celom trening setu. Za datu ulaznu sliku regresorom je određena globalna rotacija i dobijena je distribucija protopoza. Iz te distribucije biraju se protopoze i generišu se nasumični vektori položaja koji odgovaraju datim protopozama i globalnoj rotaciji.



Slika 5.
Šematski prikaz strukture jednoslojnog inicijalizatora

Figure 5.
One-layer initializer pipeline

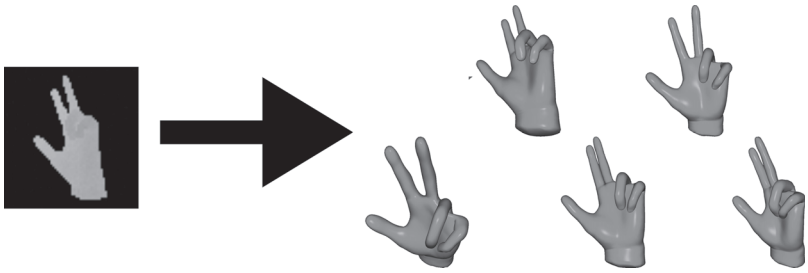
Struktura dvoslojnog inicijalizatora. Na slici 6 je prikazana šema dvoslojnog inicijalizatora. Baza je podeljena na 125 delova prema grupama rotacije. Na ovako podeljenoj bazi istrenirana je šuma odluke (decision forest) (Criminisi i Shotton 2013). Ona klasifikuje u kom delu spektra rotacije se nalazi šaka, tj. kojoj grupi pripada. Zatim su za svaku grupu rotacije istrenirane po još dve šume odluke – jedna za klasifikaciju protopoze i druga za precizniju regresiju globalne rotacije. Istrenirano je ukupno 250 šuma odluke za protopozu i regresiju rotacije i jedna za klasifikaciju rotacije.



Slika 6. Šematski prikaz dvoslojnog inicijalizatora

Figure 6. Two-layer initializer pipeline

Za datu ulaznu sliku prvo se primenjuje klasifikator globalne rotacije čime se dobija distribucija po grupama. Zatim se iz dobijene distribucije bira 10 grupa i za njih se određuju preostali parametri. Za svaku grupu biraju se protopoze iz distribucije dobijene klasifikatorom poza i generišu se vektori položaja. Broj izabranih protopoza zavisi od traženog broja vektora položaja. Primer je prikazan na slici 7.



Slika 7.
Primer izlaza dvoslojnog inicijalizatora

Figure 7.
Example outputs from two-layer initializer

Algoritam optimizacije rojeva (particle swarm optimization). Na osnovu predloga vektora položaja iz inicijalizatora, pomoću algoritma optimizacije rojeva određuje se traženi vektor položaja.

Optimizacija rojeva je algoritam koji se koristi za traženje minimuma funkcije. Sastoji se od određenog broja „čestica", koje se kreću po prostoru pretrage. Svaka čestica sadrži vektor brzine i položaj njenog minimuma. Algoritam pamti i položaj minimuma celog roja (globalni minimum roja, nije isto što i globalni minimum funkcije). Ažuriranje položaja čestice u prostoru zavisi od sopstvenog minimuma, vektora brzine i globalnog minimuma.

Početne položaje čestica čini izlaz inicijalizatora. Potrebna je funkcija koja određuje koliko precizno vektor položaja odgovara ulaznoj slici, i ta funkcija se minimizuje optimizacijom rojeva. Prema tome, prostor pretrage je iste dimenzionalnosti kao i vektor položaja. Upoređivanje se vrši tako što se na osnovu vektora položaja i 3D mreže sintetiše dubinska slika, a zatim se koristi funkcija energije za poređenje sintetisane i ulazne slike. Što je manja vrednost funkcije energije, to ispitivani vektor bolje odgovara ulaznoj slici. Za funkciju energije se koristi sledeća relacija (Sharp *et al*. 2015):

$$E(Z, R) = \sum_{ij} \rho(z_{ij} - r_{ij}) \tag{1}$$

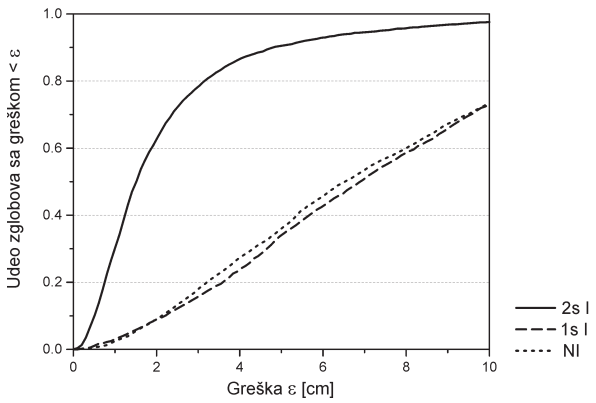
gde je Z ulazna slika, R je sintetisana slika i $\rho(x) = \min(|x|, \tau)$. Vrednost τ se bira tako da funkcija ρ za veliko $|x|$ uvek daje τ. Iz definicije funkcije ρ očigledno je da je ona minimalna ako pikseli z_{ij} i r_{ij} imaju iste vrednosti.

Rezultati i diskusija

Dobijeni rezultati su prikazani u nastavku. Na svim graficima na x-osi je udaljenost ε između stvarne i dobijene pozicije zglobova (u centimetrima), a na y-osi je procentualni udeo zglobova čija je pozicija dobijena sa greškom manjom od ε. Prema tome, linije koje su bliže gornjem levom uglu grafika prikazuju tačnije određene pozicije zglobova.

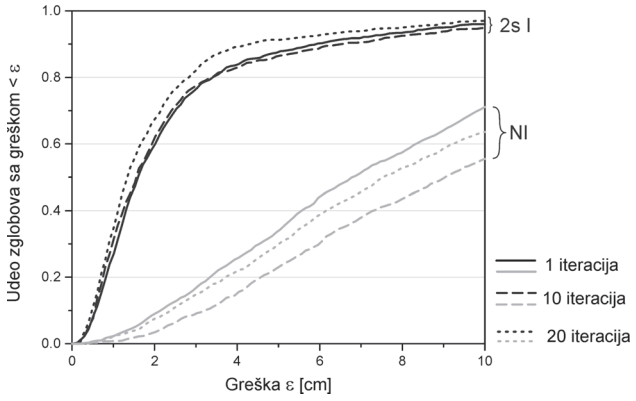
Sa grafika na slici 8 vidi se da korišćenje dvoslojnog inicijalizatora daje značajno veću tačnost od nasumične inicijalizacije. Međutim, jednoslojni inicijalizator daje približno istu tačnost kao i nasumični. Razlog tome je što regresor globalnih rotacija ima malu tačnost kada se trenira na velikom spektru rotacija, kao što je slučaj sa jednoslojnim inicijalizatorom. Rezultati su uporedivi sa onima iz referentnog rada (Sharp *et al*. 2015), sa približno 78% dobro određenih zglobova sa greškom manjom od 3 cm.

Sa grafika na slici 9 se vidi da povećavanje broja iteracija daje bolje rezultate za dvoslojni inicijalizator, dok to nije slučaj sa nasumičnom inicijalizacijom, gde su rezultati bolji sa 1 iteracijom (što znači da se od čestica dobijenih od inicijalizatora za krajni rezultat uzme ona sa najmanjom cost funkcijom) nego sa 10 ili 20. Iz toga se zaključuje da optimizacija rojeva nije odgovarajući algoritam za ovaj problem ako se koristi bez inicijalizatora. S druge strane, povećavanje broja čestica poboljšava rezultate i za dvoslojni inicijalizator i za nasumičnu inicijalizaciju (slika 10).



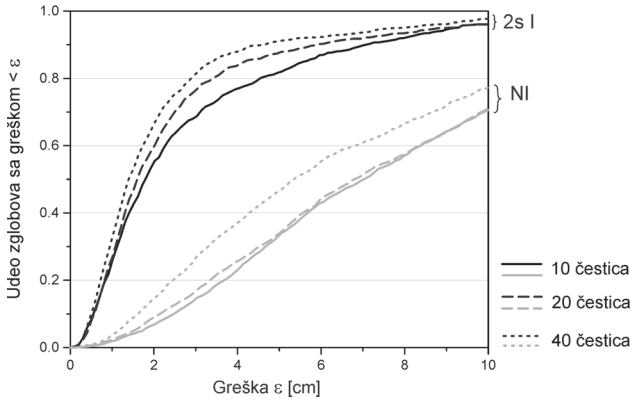
Slika 8. Udeo zglobova određenih sa greškom manjom od ε u zavisnosti od vrednosti ε za jednoslojni (1s I), dvoslojni (2s I) i nasumični (NI) inicijalizator. Grafik ilustruje situaciju sa 20 čestica i jednom iteracijom.

Figure 8. Dependency of percentage of joint positions predicted with an error less than ε on error ε, for two-layer (2s I) and one-layer initializer (1s I), and random initializer, with 20 particles and 1 iteration.



Slika 9. Udeo zglobova određenih sa greškom manjom od greške ε u zavisnosti od vrednosti ε za različiti broj iteracija kod dvoslojnog (2s I) i nasumičnog inicijalizatora (NI). Grafik je dobijen za 20 čestica.

Figure 9. Dependency of percentage of joint positions predicted with an error less than ε on error ε, for different number of iterations and 20 particles in PSO for two-layer initializer (2s I) and random initializer (NI).



Slika 10. Udeo zglobova određenih sa greškom manjom od ε u zavisnosti od ε, za različiti broj čestica pri jednoj iteraciji, za dvoslojni (2s I) i nasumični inicijalizator (NI)

Figure 10. Dependency of percentage of joint positions predicted with an error less than ε on error ε, for different number of particles and 1 iteration in PSO for two-layer initializer (2s I) and random initializer (NI)

Zaključak

U ovom radu predstavljen je fleksibilan sistem koji uspešno određuje poziciju i pozu šake na dubinskoj slici. Dobijeni rezultati pokazuju da dvoslojni inicijalizator daje najtačniji rezultat. Takođe, dobijeno je da veći broj čestica u algoritmu optimizacije rojeva znači i bolje rezultate, ali da povećavanje broja iteracija poboljšava tačnost samo kada se koristi inicijalizator. Vizuelno dobri rezultati, odnosno rezultati gde dobijeni 3D modeli šake vizuelno odgovaraju ulaznim slikama, dobijeni su sa dvoslojnim inicijalizatorom u kombinaciji sa optimizacijom rojeva sa 20 čestica i jednom iteracijom Moguće je dobiti i bolje rezultate sa većim brojem čestica, ali to značajno usporava program. Sa ovakvim sistemom dobija se 75% pozicija zglobova određenih sa greškom manjom od 3 cm. Rezultati su uporedivi sa referentnim radom (Sharp *et al*. 2015), uprkos tome što je ovaj sistem jednostavniji. Nedostatak sistema je što funkcioniše na slikama na kojima se vidi šaka bez ostatka tela ili objekata u pozadini. Moguće poboljšanje sistema bi bio dodatak detekcije šake, čime bi se spomenuti nedostatak izbegao.

Literatura

Abu Rumman N., Fratarcangeli M. 2017. Skin Deformation Methods for Interactive Character Animation. *Communications in Computer and Information Science*, **693**: 153.

Criminisi A., Shotton J. 2013. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer

Dipietro L., Sabatini, A. M., Dario P. 2008. A survey of glove-based systems and their applications. *IEEE Transactions on Systems, Man, and Cybernetics*, C **38** (4): 461.

Jacobson A., Panozzo D., Schüller C., Diamanti O., Zhou Q., SKoch C. *et al*. 2018. Libigl: A simple {C++} geometry processing library. Dostupno na: http://libigl.github.io/libigl/

Sharp T., Keskin C., Robertson D., Taylor J., Shotton J., Kim D., *et al*. 2015. Accurate, Robust and Flexible Real-Time Hand Tracking. U *CHI '15 – Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Seoul: ACM, str. 3633-3642.

Sridhar S., Oulasvirta A., Theobalt C. 2013. Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data. U *Proceedings of the IEEE International Conference on Computer Vision*. Sydney: ICCV, str. 2456-2463

Wang R. Y., Popović J. 2009. Real-time hand-tracking with a color glove. *IACM Transactions on Graphics*, **28** (63): 1.

Zhao W., Chai J., Xu Y.-Q. 2012. Combining marker-based mocap and RGB-D camera for acquiring high-fidelity hand motion data. U *Proc. Eurographics Symposium on Computer Animation*. Eurographics / ACM SIGGRAPH, str. 33-42.

Srđan Radović and Valentina Njaradi

Hand-Tracking Using a Single Depth Camera

Accurate and fast hand-tracking could make a huge impact on human-computer interaction. This is a very complex problem, due to the hand's remarkable dexterity and high degree of freedom. A solution to this problem would have a wide range of applications, including augmented reality, robotics and the gaming industry. This paper focuses on hand-tracking using a single depth camera and the goal is to determine the 3D positions and rotations of all joints from a depth image.

Global hand rotation and relative joint angles together form a pose vector. All joint positions can then be computed from this pose vector. The algorithm pipeline is presented in Figure 1. The first step is a machine learning pose initializer which provides a set of approximate pose vectors. The next step is a particle swarm optimization (PSO) algorithm which optimizes these pose vectors. Pose vectors are then rendered as depth images and compared to the input image, and the best pose is output.

The pose initializer consists of several random decision forests (RDFs) trained on a synthetic database. The database was made by a variation of pose vector parameters within specific limits. Since the joint angles are relative to each other and cannot be varied independently, seven hand poses were introduced. Within each hand pose the joint angle limits were determined separately (Sharp *et al*. 2015). Two initializer architectures were compared. Their pipelines are presented in figures 5 and 6. The one-layer initializer consists of an RDF for predicting global hand rotation and an RDF for classifying the hand pose. The two-layer initializer consists of an RDF classifier for classifying global rotation to one of 125 discrete bins, and for each bin two additional RDFs, a regressor for predicting precise global rotation and a classifier for hand poses. The two-layer initializer consists of 251 random decision forests in total. The advantage of the two-layer compared to the one-layer architecture is in predicting the global rotation in two steps, thus resulting in more precise predictions.

The next part of the system is the particle swarm optimization algorithm. Each particle's movement is influenced by its local best known position, but is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions. In this case, the particles are pose vectors and their starting positions are given by the

initializer. The optimization function proposed in literature (Sharp *et al*. 2015) is used.

A new synthetic dataset with a wide range of hand poses was made for testing the presented system. The results showed that the most critical part of the pipeline for achieving high accuracy is the two-layer initializer. A higher number of particles in PSO achieves better results, however, increasing the number of iterations after a certain limit does not significantly add to the accuracy. This is due to the fact that the initializer predicts pose vectors very close to the optimal solution. Finally, the best results in terms of accuracy/execution time are obtained by using the two-layer initializer in combination with PSO with 40 particles and 10 iterations. A total of 75% of joint positions are predicted with an error less than 3 cm, which corresponds to a visually good result and is comparable to results in previous research (Sharp *et al*. 2015).