

hyvymisp arvosana

arvostelija

Ihmishahmojen kaappaaminen ja animointi

Juho Heikkurinen

Helsinki 16.12.2007

Kandidaatintutkielma

HELSINGIN YLIOPISTO

Tietojenkttelytieteen laitos

Tiedekunta — Fakultet — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Juho Heikkurinen			
Työn nimi — Arbetets titel — Title			
Ihmishahmojen kaappaaminen ja animointi			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Kandidaatintutkielma		16.12.2007	0 sivua
Tiivistelmä — Referat — Abstract			
<p>Tutkielma pyrkii antamaan yleiskuvan eri menetelmistä ihmisten ulkomuodon kaappaamiseen animointiin kelpaavaksi tietokonemalliksi ja animaatioiden luomista näiden mallien liikkeen parametrisoimisella joko ihmisnäyttelijältä tai täysin automatisoidusti.</p>			
Avainsanat — Nyckelord — Keywords			
ihmiskasvot, animointi, kasvoanimointi, lihsmallinnus, fysiologinen mallinnus, kaappaus, stereokamera, fotogrammetri			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — övriga uppgifter — Additional information			

Sisältö

1	Motivaatio	1
2	Historiakatsaus	2
3	Ihmisen ulkonäön kaappaus	2
3.1	Geometrian ja tekstuurin kaappaus	3
3.1.1	Stereokamerakaappaus	3
3.1.2	3-D skannaus	3
3.1.3	Kaappaus yksittäisestä valokuvasta	4
3.2	Heijastavuuden kaappaus	4
4	Mallit	5
5	Animaatioiden syntetisointi	8
5.1	Fysiikkapohjainen animaatioparametrien generointi	8
5.2	Animaatioparametrien hankkiminen liikkeenkaappauksella	9
5.3	Fysiikka/liikkeenkaappaus hybridi	9
5.4	Ihotekstuuri-analysointi, syntetisointi ja renderointi	9
6	Esimerkki: Puhuva lihas-ohjattu pää	10
6.1	Johdanto	11
6.2	Datan hankinta	11
6.3	Animointi ja renderointi	12
6.4	Puheen synkronointi	13
6.5	Tulokset ja päätelmät	14
7	Esimerkki: Matrix-elokuvan jatko-osien digitaaliset kaksoisolennot	14
7.1	Näyttelijän suorituksen ja ulkonäön kaappaus	14
7.2	Heijastavuuden kaappaaminen ja simuloiminen	15
7.3	Renderointi	15

	iii
8 Konklusio	16
9 Kuvien lähteet ja kopio-oikeusilmoitukset	16
Leet	18

1 Motivaatio

Tietokoneella luodut ihmistä muistuttavat kuvat lumoavat katsojansa oli kyse sitten virtuaalimaailmojen 3-D avatareista tai elokuvakankaan synteettisistä sankareista ja konnaista. Tieteiselokuvat ovat jo pitkään olleet täynnä futuristisia käyttöliittymiä, joissa tietokone on saanut ihmisen hahmon ja äänen. Toki puhuvista päistä on konkreettista hyötyäkin: niillä on sovelluksia eri tavoin kognitiivisesti tai aistillisesti vammautuneiden ihmisten terapiassa ja päivittäisessä elämässä. Osittain kuulokykynsä menettäneet ihmiset voivat lukea puhuvan pään huulilta.

Ihmisten ulkonäön kaappaamiessa käytetyillä tekniikoilla on myös sovelluksia kameravalvonnan automatisoimisessa. Vanhenemissimulaatioilla on käyttöä poliiseille etsittäessä kauan sitten kadonneita ihmisiä ja ihmisen ulkonäön konstruomisesta kalon rakenteen perusteella on hyötyä oikeuslääketieteessä. Kirurgit pystyvät pystyvät hyödyntämään ulkonäkösimulaatioita suunnitellessaan leikkauksia ja ihon tekstuurin analysoinnista on hyötyä dermatologisten sairauksien automaattisessa tunnistamisessa.

2000-luvun alussa ihmishahmoiset tietokoneanimaatiot ovat kehittyneet niin uskottaviksi, että ne ovat saavuttaneet asteen, että ne ovat *digitaalisia kaksoisolentoja*, joiden erottaminen oikeasta ihmisestä on usein mahdotonta, mikä avaa yhä visuaalisesti näyttävämmän ja näyttävämmän viihteen lisäksi paljon mahdollisuuksia myös pahan tekoon. Ihmiskunta on jakautunut kolmeen ryhmään: 1) Ihmiset jotka eivät tiedä mitään digitaalisten kaksoisolentojen olemassaolosta. Tämä tietämättömyys on sekä siunaus että kirous. He eivät huolehdi sellaisesta mitä eivät tiedä olevan olemassa, mutta jos joku näistä joutuu digitaalisen kaksoisolento-hyökkäyksen kohteeksi on hänen ja hänen rakkaitensa elämä todennäköisesti melko täydellisesti pilalla hyvin nopeasti 2) Ihmiset jotka tietävät, että niitä on, mutta eivät itse tai heidän lähipiirinsä omaa kykyjä tehdä niitä. Nämä ihmiset eivät enään aivan pureskelematta niele kaikkea mitä näkevät internetissä tai kuulevat nähdyn siellä, mutta toisaalta voivat olla hyvinkin huolissaan potentiaalisesta vaarasta, että joku tai jotkin saattaisivat käyttää näitä aseena heitä tai heidän rakkaitaan kohtaan mikäli he astuvat väärin ihmisten varpaille. 3) Digikraattinen eliitti, jolla on käytettävissään digitaaliset kaksoisolennot mahdollistava teknologia. Heille avautuu laajat mahdollisuudet kohdistaa hirmuvaltaa muihin ihmisiin ja kierouttaa ihmiskunnan ihmisyyksäkäsitys esim. levittämällä tietokonegeneroitua lapsi- tai muuta pornoa sisältäen toimintaa, joka ei ole mahdollista ihmisnäyttelijöihin.

2 Historiakatsaus

Vuonna 1971 Henri Gouraud valmisti ensimmäisen ihmisen kasvokuvan tietokoneella. Mallina tässä ¹ kuvassa toimi hänen vaimonsa Sylvie Gouraud. Alaviitteestä näkyvästä kuvasta päätellen Gouraud käytti hahmonkaappaukseen manuaalista kamerakaappausta ottaen kaksi valokuvaa kasvoista hyödyntäen kasvoon liimattuja merkkejä. Tämä photometrinen kaappausmenetelmä, jossa useammasta kuvasta yhdistelemällä saadaan kaapattua hahmoja automaattisesti tai semi-automaattisesti on voimissaan yhä tänäkin päivänä siitä huolimatta, että saatavilla on 3-D skannereita. Stereokamera- tai jopa multikamera-installaatiot ovat huomattavan halpoja verrattuna 3-D skannereihin ja kehitys hahmontunnistuksen alueella on johtanut siihen, että monissa kaappauksissa ei usein tarvita ihoon kiinnitettyjä merkkejä vaan ihon pinnanmuodon ja eleiden kaappaus voidaan tehdä konenäkö-tekniikoita soveltamalla.

Vuonna 1972 Frederik I. Parke, DARPA:n osittain rahoittamassa projektissaan, loi ensimmäisen tietokoneanimaation ihmiskasvoista käyttäen Gouraudin menetelmää, jossa kaarevia pintoja approksimoidaan polygoneilla, sekä myöskin Gouraudin kehittämää shaderiä[Par72].

Vuonna 1981 Platt ja Badler rakensivat ensimmäisen mallin, joka ei ollut täysin manuaalisesti animoitava vaan sisälsi jotain logiikkaa itsessään. Heidän mallinsa sisälsi sisälsi jousia ja massoja lihaksien aiheuttamien voimien simuloimiseksi ja hyödynsi Facial Action Coding System:iä, jonka Ekman ja Friesen kehittivät vuonna 1978. [Sif05]

Vuonna 1995 Waters ja Frisbie rakensivat lihasmallin puheanimaatiota varten painottaen, että lihakset ohjaavat kasvoja ja siten on hyödyllisempää keskittyä lihasmalliin, eikä pintamalliin[Sif05].

3 Ihmisen ulkonäön kaappaus

Hahmon kaappauksessa on saatava vähintään kaksi asiaa kaapattua kohteesta: Topologia ja tekstuurit. Fotorealistisen renderoinnin saavuttamiseksi pitää kaapata myös kohteen heijastavuus[PiL05]. Esittelen kunkin menetelmän lyhyesti tässä kappaleessa.

¹<http://helios.univ-reims.fr/Labos/LERI/Afig99/biographie.html>

Erityisen vaikeiksi kohteiksi on havaittu korvat niiden monimutkaisen topologian takia, sekä hiukset sillä ymmärrettävästi jokaista yksittäistä hiusta ei voi mallintaa[HaK01]. Myös hampaat ja kieli ovat ongelmallisia alueita koska niitä ei voida perinteisillä menetelmillä kaapata ymmärrettävistä syistä.

3.1 Geometrian ja tekstuurin kaappaus

Geometria ja tekstuuri voidaan kaapata stereokameralla, joka koostuu useammasta samanaikaisesti laukaistavasta kamerasta, 3-D skannerilla, joka on syvyysmittauslaite tai jopa yhdestä valokuvasta sovittamalla muovailtavaa 3-D päämallia siihen.

3.1.1 Stereokamerakaappaus

Fotogrammetria on tekniikka jossa esineen geometrisia ominaisuuksia päätellään valokuvasta tai valokuvista. Käytettäessä kahta tai useampaa digitaalista kameraa jotka laukaistaan samanaikaisesti voidaan kuvia koneellisesti vertaamalla saada geometriatieto sisältäen syvyystiedon käyttäen fotogrammetrisiä algoritmeja.

Stereokamerakaappausta voidaan käyttää sekä hahmon kaappaamiseen että liikkeen kaappaamiseen.

3.1.2 3-D skannaus

Cyberware-skannerit käyttävät lasermittausta topologian kaappaamiseen ja erillistä mittaussysteemiä tekstuuriin kaappaamiseen ja kykenevät skannaamaan 30.000 pistettä sekunnissa X,Y,Z ja R,G,B komponentteihin. Cyberware skannerit saavuttavat nykyisin vain 400 mikronin tarkkuuden Y-akselille ja 50-150 mikronin tarkkuuden Z-akselille. Cyberware on yksityisesti rahoitettu amerikkalainen yritys[Cyberw99].

Arius3d-skannerit ovat moniväri-laser-skannereita, jotka kaappaavat sekä pisteen värin, että sijainnin lähettämällä kolme eri laservalon aallonpituutta samassa keskitetyssä säteessä. Väri saadaan mittaamalla palaavien lasersäteiden voimakkuus. X-akselin tieto saadaan laskemalla skannauspeilin sijainti ja kulma skannauskameran suhteen. Y-koordinaatti lasketaan kameran liikuttelujärjestelmän sijaintitiedosta. Z-koordinaatti eli etäisyys saadaan kameran sisäisellä laserkolmiomittauksella, jossa hyödynnetään galvanometri-ohjatun kaksipuolisen peilin molemmat puolet seuraavasti: ensimmäinen puoli heijastaa säteen kohteen pintaan ja paluusäde ohjataan peilin toiseen puoleen, jolloin kulmanmuutos peiliä käännettäessä ei siirrä paluusä-

teen osumiskohtaa CCD-kennossa vaan ainoastaan skannattavan pisteen etäisyys. Arius3d on kanadalainen yksityisessä omistuksessa oleva yritys, joka on saanut rahoitusta Kanadan hallitukselta skannerinsa kehittämiseen[arius3d]. Arius3d-skannerit pystyvät 100 mikronin tarkkuuteen, joka on George Borshukovin (sivu 106) (väärin kirjoittivat muuten kalvoihin ”Aruis3d”)[HaD04] mukaan juuri pienimpien ihmiskasvoissa erottuvien kohteiden, kuten ihohuokosten ja ohuiden ryppyjen kokoluokka.

3.1.3 Kaappaus yksittäisestä valokuvasta

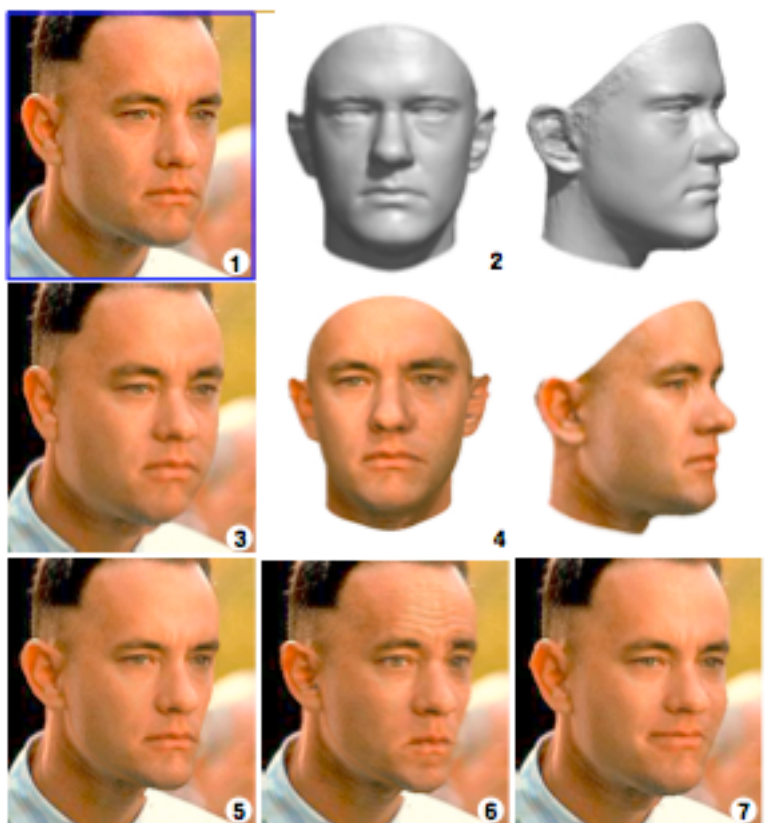
Volker Blanz ja Thomas Vetter ovat kehittäneet muovattavan päämallin käyttäen suurta tietokantaa 3-D skannatuista päistä. Heidän järjestelmänsä kykenee valmistamaan näköisen 3-D mallin ihmispäästä jopa vain *yhden valokuvan perusteella* (Kuva 1) vaatien vain vähän interaktiivista työskentelyä. Ensin käyttäjä määrittelee karkeasti pään asennon kuvassa. Tämän jälkeen järjestelmä automaattisesti sovittaa muovattavan mallin annettuun kuvaan ”analyysi syntetisoimalla”-silmukassa, jossa alkuperäisen kuvan ja syntetisoidun tekstuurikartoitetun 3-D mallin tuottaman kuvan eroista saadaan seuraavan iteraation parametrit. Mallinnetun ihmisen ulkonäköä voidaan muokata esim. laihduttaa tai lihottaa.[BIV99]

3.2 Heijastavuuden kaappaus

Pinnan heijastavuus on eräs materiaalin ominaisuus, jossa havaittu värisävy ja kirkkaus on riippuvainen sekä valaistuksen suunnasta että katseen suunnasta. Sitä voidaan kuvata kaksisuuntaisella heijastavuusjakauma-funktiolla (Bidirectional reflectance distribution function, BRDF), joka siis kuvaa materiaalin annetun pisteen heijastavuuden.

Kaksisuuntainen tekstuurifunktio (Bidirectional texture function, BTF) on 6-ulotteinen funktio joka sisältää tekstuurin tasokoordinaatit ja katselu- ja valaistuskulman. Kaksisuuntainen tekstuurifunktio muodostetaan ottamalla tuhansia kuvia kohteesta eri kamera- ja katselukulmista. Se on siis joukko BRDF:iä jonkin alueen yli.

Heijastavuuden kaappaukseen käytetään valolavaa[Dev00] (kuva 2), joka luo erilaisia valaistusolosuhteita, joiden perusteella pystytään rakentamaan analyttinen BRDF, mikä mahdollistaa 3-D mallin renderoinnin erilaisiin joko tietokonegeneroituihin tai todellisiin (sämplättyihin) ympäristöihin ja valaistusolosuhteisiin mistä tahansa kuvakulmasta. Tuloksia esitellään kuvassa 3[HaD04].



Kuva 1: Muovattavan mallin sovittaminen yhteen ainoaan kuvaan (1) tuottaa 3-D aproksimaation (2) ja tekstuurikaappauksen (4) 3-D malli renderoidaan takaisin kuvaan lihonneena (3) laihtuneena (4) ilmeensä nyrpistäneenä (5) ja pakoitettuna hymyilemään (6)

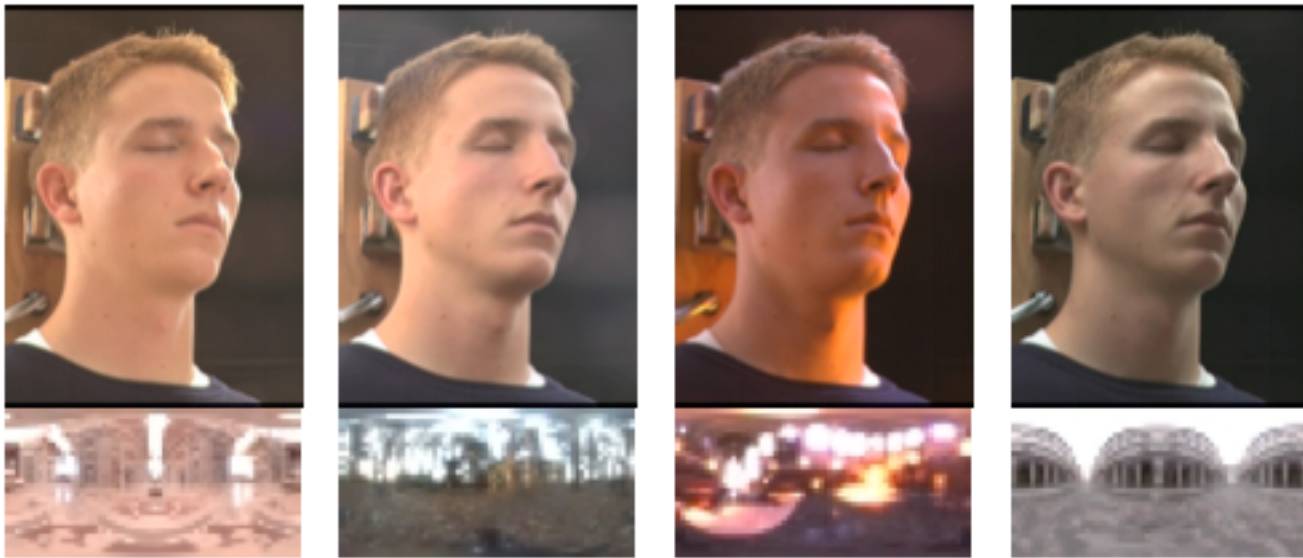
Heijastunut valo voidaan erotella spekulaariseen valoon, joka heijastuu suoraan ihon rasvasta ja diffusoituneeseen valoon, joka on kulkenut ihon sisällä ennen kulkeutumistaan ulos sijoittamalla sekä valolähteeseen, että kameraan lineaarinen polarisoiija (kuva 3). Suoraan pinnasta heijastunut spekulaarinen valo säilyttää polarisaationsa kun taas ihokerroksista siroutuva valo menettää polarisaationsa [Dev00].

4 Mallit

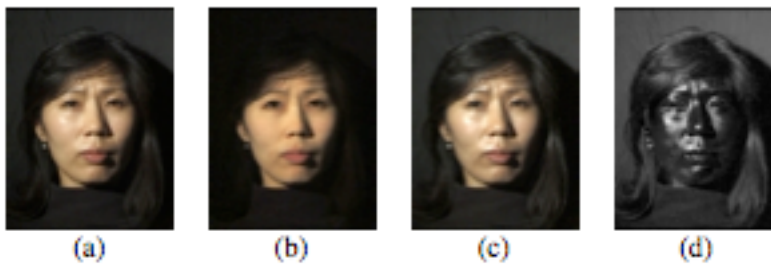
Gouraudin ja Parken mallit olivat monikulmiorepresentaatioita, jotka sisälsivät vain pinnan eli ihon. Nykyaikaiset mallit sisältävät luurangon, lihaksiston, kudokset ja ihon. Nivelet määrittävät luiden liikeratojen maksimipisteet ja näiden määrittely empirisesti on yksinkertaista ja suoraviivaista. Tavalliset lihakset (ts. kahdesta



Kuva 2: Kuvassa on valolava, jota käytetään BRDF:n muodostamiseen tarvittavien mittausten suorittamiseen. Pikkukuvassa pitkällä laukaisujalla saatu kuva valolavan toiminnasta



Kuva 3: Kasvot renderoituna sämplätyssä valaistuksessa. Jokainen kuva kuvaa kasvoja syntetisoidussa valaistuksessa. Alakuvat esittävät kaapattua valokarttaa. Kuvat luodaan ottamalla pistetulo jokaisen pikselin heijastavuusfunktioista valokartan kanssa.



Kuva 4: Spekulaarisen ja diffuusin valon erittely. (a) Normaali kuva pistevalaistuksessa. (b) Kuva diffuusista heijastuksesta, joka saadaan sijoittamalla vertikaalinen polarisoija valolähteen eteen ja horisontaalinen kameran eteen. (c) Kuva korostetusta spekulaarisesta heijastuksesta, joka saadaan asettamalla molemmat polarisoijat vertikaalisesti. (d) c:n ja b:n eroitus, joka antaa spekulaarisen komponentin. Kuvat on skaalattu näyttämään yhtä kirkkailta

pisteestä tukirankaan kiinnittyvät lihakset) voidaan hyvin tuloksin approksimoida ellipsoideiksi, jotka paksunevat, lyhenevät ja suoristuvat lihaksten jännittyessä. Eri-tyistapauksiin kuten esim. orbicularis oris (suun kehän muodostava lihasryhmä) ja kieli (kiinni vain yhdestä päästä) tämä ei tietenkään päde. Jännittyessään ja rentoutuessaan lihakset vaikuttavat muihin niihin liittyviin lihaksiin. Tämä lihasten vuorovaikutus voidaan ottaa malleihin mukaan sitomalla ideaalijousia lihaksien välille [HaK01] Anatomisesti korrekteissa malleissa lihaksia ja muita kudoksia voidaan mallintaa tetrahedronitilavuuksina[Sif05].

5 Animaatioiden syntetisointi

Ensimmäiset ihmiskasvojen animaatiot tehtiin interpolaatiolla, eli käsin määrittelemällä kunkin verteksin sijainti avainkehyksissä, joiden väliset kehykset generoitiin interpoloimalla avainkehyksistä[Par72].

Tämän jälkeen siirryttiin parametriseen animaatioon, jossa kasvojen verteksit on ryhmitelty toiminnallisiin ryhmiin joille on määritelty muokkausoperaatioita. Näitä malleja pystytään ohjaamaan korkeamman abstraktiotason komennoilla esimerkiksi käsketään mallia ”hymyilemään” tai sanomaan ’A’. Tämä vähentää huomattavasti siirrettävän datan määrää mallin animoimisessa.

Nykyaikaisessa ihmishahmojen animoinnissa käytettävien parametrien tuottamiseen on eriytynyt kaksi eri suuntausta, fysiikka- ja liikkeenkaappauspohjainen, jotka tosin yhdistelevät tuloksia toisistaan.

5.1 Fysiikkapohjainen animaatioparametrien generointi

Fysiikkapohjaisella animaatioparametrien generoinnilla pyritään täysin automatisoituun animointiin sovellusalueena esim. virtuaalistan ympäristöjen avatarit ja erityistarpeisiin rakennettavat puhuvat päät. Virtuaalisten hahmojen interaktiot, eli tapahtumapohjainen animaatio on tämän tutkielman ulkopuolella ja keskityn ker- tomaan visuaalisen puhesynteesin tuottamisesta automaattisesti esittelemällä yhden puhuvan pään kohdassa ”Esimerkki: Puhuva lihas-ohjattu pää”.

Fysiikkapohjaisessa visuaalisessa puhesynteesissä ihon ja erityisesti puheelinten muodonmuutokset aikaansaadaan liikuttamalla valikoituja kasvojen lihaksia, leukaluuta ja kieltä oikean suun asennon aikaansaamiseksi. Ylähampaiden sijainti luonnollisesti on kiinnitetty kallon sijaintiin. Sen lisäksi, että alahampaat kiertyvät leukaluun

mukana leukanivelen muodostaman akselin suhteen, ne pystyvät myös työntymään eteen ja taakse ja sivuttain. Kieli pystyy liikkumaan leukaluun mukana liikkumisen lisäksi itsenäisesti.[HaK01].

5.2 Animaatioparametrien hankkiminen liikkeenkaappauksella

Harvassa kaappauksessa käytetään kasvoihin kiinnitettäviä merkkejä, joiden liikettä seuraamalla saadaan harvaa tietoa geometrian muutoksista seuraamalla merkkien liikettä kuvasarjassa.

Tiheässä kaappauksessa ei käytetä erillisiä merkkejä vaan geometrian muutokset saadaan käyttämällä optista virtausta.

Optinen virtaus approksimoi kohteiden liikettä visuaalisessa representaatiossa. Tosi maailman esineissä fyysinen liike ei välttämättä johda väri- tai harmaasävyjen muutoksiin ja kaikki muutokset eivät johdu liikkeestä, mikä vaikeuttaa approksimaation toteuttamista. Useimmiten liike kuvataan vektoreina jotka alkavat tai päättyvät pikseleihin digitaalisessa kuvasarjassa.

5.3 Fysiikka/liikkeenkaappaus hybridi

Sifakis et al. esittivät vuonna 2005 lähestymistavan joka yhdistää fysiikka-pohjaisen ja liikkeenkaappaus-pohjaisen animointiparametrien tuottamisen.[Sif05] Tämä kehityssuunta on jo melko pelottava, sillä fysiikkapohjainen malli jota parametrisoidaan/opetetaan liikkeenkaappausdatalla voidaan takaisinsyöttää liikkeenkaappauksella ohjattavaan animaatioon, mikä vähentää interaktiivisen työn tarvetta ja mataltaa kustannuksia fotorealististen tai ainakin uskottavien animaatioiden tuottamisessa. Esim. Matrix-elokuvan jatko-osien tuotannossa käytettyjen digitaalisten kaksoisolentojen valmistamisessa käsityönä tehty optisessa virtauksessa tapahtuva virheiden kumuloitumisen eliminointi voitaisiin mahdollisesti automatisoida käyttämällä hybridistä lähestymistapaa.

5.4 Ihotekstuurien analysointi, syntetisointi ja renderointi

Tekstuurien tuottaminen synteettisesti on ongelma, jota on tutkittu paljon. Kun tavoitteena on tuottaa ihmisen ihoa muistuttavia tekstuureja, luonnollisten tekstuu-

rien tuottamiseen erikoistuneet menetelmät eivät välttämättä toimi tyydyttävästi ja tarvitaan menetelmiä, jotka liittyvät erityisesti ihon tekstuurin analyysiin, kuten Cula et. al.:in kuvapohjainen menetelmä ihotekstuurin kuvaamiseen.

Iho on monimutkainen maasto, jonka mallintaminen on hyvin vaikeaa monesta syystä. Valon heijastukseen ja kulkeutumiseen vaikuttavat ihokerrosten monimutkaiset optiset ominaisuudet ja myös pinnan mikrogeometria; ihohuokokset ja rypyt. Kun valolähde tulee kohtisuoraan tarkasteltavaan ihonkohtaan tai sitten sivusta on iho aivan erinäköistä samasta tarkastelukulmasta katsottuna.

Cula toteaa, että ihon simuloiminen on vaikea haaste, johon ei ole toistaiseksi löydetty kovin uskottavia tuloksia tuottavaa menetelmää, mutta arvelee että lähestymistapa, joka yhdistää geometriset transformaatiot ja kuva-pohjaisen renderoinnin voisi tarjota tyydyttävän tuloksen[Cu05].

Lihasten ja ihon pehmeys, läpinäkyvyys ja ihon venyvyyden epälineaarisuus ja sen aiheuttamat värin muutokset asettavat melkoisia haasteita uskottavan lopputuloksen saavuttamiseen. Ihon läpinäkyvyys lisää renderoinnin haastavuutta, sillä osa valosta läpäisee ihon ja siroaa sieltä takaisin ulos[Ha04].

Renderoitaessa staattiset rypyt ja ihohuokokset kuvataan staattisena töyssykarttina (bump map) kun taas dynaamiset rypyt, jotka ilmestyvät ja katoavat ilmeiden vaihtumisen myötä täytyy identifioida ja eristää tekstuurikartoista ja kerrostaa tössykartan päälle.[Ha04]

Töyssykarttaistaminen on tietokonegrafiikkateknikka missä jokaisen pikselin kohdalla poikkeama pinnan tasosta katsotaan korkeuskartasta ja tämä otetaan huomioon laskettaessa pikselin valaistusta.

Viime aikojen merkittävimmät edistykset tällä saralla on tehty viihdeteollisuuden sisällä akateemisen maailman roolin jäädessä melko vähäiseksi. ESC Entertainment-yhtiö teki Matrix-elokuvien tuotannossa käytetyt digitaaliset kaksoisolennot ja Sony Pictures Imageworks oli vastuussa Spiderman-elokuvan jatko-osien digitaalisista kaksoisolunnoista.[Sag04]

6 Esimerkki: Puhuva lihas-ohjattu pää

Tässä esimerkissä käydään läpi Haber et al.:in kehittämä puhuva pää artikkelin ”Face to Face: From Real Humans to Realistic Facial Animation” pohjalta. Pää osaa foneemit ja niiden ajoitustiedon sisältävän transkriptin perusteella tuottaa visuaalista

puhetta.[HaK01]

6.1 Johdanto

Uskottavien ihmiskasvojen mallintaminen ja animointi on hyvin vaikeata, koska ihmisaivot ovat erikoistuneet havaitsemaan hyvinkin hienojakoisia ilmeiden vivahteita. Kasvojen rakenteen (kallo, lihakset, iho) monimutkaisuus lisäävät haastavuutta. Yksilön kasvojen geometria ja tekstuurit täytyy kaapata koneelle, valita sopivat animointi-parametrit ja näistä seuraavat muotojen muutokset (deformations) täytyy laskea ja renderoida. Tämän perustan päälle rakentuu ilmeiden ja puheliikkeiden muodostaminen.

Ryhmä esittelee tämän prosessin vaihe vaiheelta joka pääpiirteittäin menee seuraavasti:

1. Data (geometria ja tekstuurit) kaapataan ja muunnetaan polygoni-geometriaksi.
2. Matalan tason animointi perustuu fysiikka-pohjaiseen lähestymiseen, missä lihasten supistumiset antavat animointi-parametrit
3. Virtuaalisen pään rakentaminen perustuu ihmisen anatomiaan ja he ovat kehittäneet työkalun jolla malli saadaan sovitetuksi ja linkitettyksi kasvojen geometriaan
4. Korkeamman tason animaatiosynkronointi tuottaa puheliikkeet

6.2 Datan hankinta

Paperissa esitellyssä menetelmässä geometria ja tekstuurit kaapataan erikseen.

Data hankitaan kolmiomittaus-periaatteella etäisyysmittauksia käyttäen. Etäisyysmittaukset muodostavat pistepilven, mistä muodostetaan polygoni-pinta pinnan etsimisalgoritmilla. Geometriaa kaapattaessa ihmismallilla on suu kiinni ja neutraali ilme, joten suu joudutaan leikkaamaan auki (polygoni-pinnasta) toimivan mallin aikaansaamiseksi. Lisäksi pinnan osa, joka esittää silmiä litistetään, jotta mallille voidaan upottaa synteettiset silmät.

Tutkimusryhmä havaitsi, että tyydyttäviin tuloksiin päästään monikulmioverkolla, jossa on 3000-4000 kolmiota koko päälle, tosin ihmiskasvojen kriittisimmille alueille, jotka ovat silmät ja suu pitää käyttää korkeampaa resoluutiota kuin muille alueille.

Tekstuurit kaapataan ottaen korkea-resoluutioisella digitaalikameralla 4-11 kuvaa hyvin hajaantuneessa valaistuksessa. Valokuvat rekisteröidään ja yhdistetään yhdeksi OpenGL-renderointiin sopivaksi tekstuuriksi. Jokaselle solmulle ei tarvitse lukea omaa tekstuurisyötettä (erityisesti korvien sisällä ja takana) vaan sopiva tekstuuri voidaan interpoloida ympäröivistä solmuista.

Ryhmä havaitsi, että silmät, hampaat ja kieli ovat kohteita joista on vaikeaa saada dataa kaapatuksi ja päätyi käyttämään geneerisiä malleja näistä ihmisruumiin osista. Silmien tekstuurista he hankkivat kolme versiota, sinisen, vihreän ja ruskean.

Ryhmän geneerinen silmämalli tarjoaa seuraavat animointimahdollisuudet:

1. Katseen suunta
2. Pupillin koko
3. Silmäluomien ulkonäkö

Silmäluomien asento vaikuttaa silmien kirkkauteen.

Ylähampaiden sijainti luonnollisesti on kiinnitetty kallon sijaintiin. Sen lisäksi, että alahampaat kiertyvät leukaluun mukana leukanivelen muodostaman akselin suhteen, ne pystyvät myös työntymään eteen ja taakse ja sivuttain.

Kieli pystyy liikkumaan leukaluun mukana liikkumisen lisäksi itsenäisesti.

Käytetyssä fysiikkapohjaisessa simuloinnissa ihon pinnanmuodon muutokset laskeaan lihasten supistumisesta. Puheanimointia varten ryhmä rakensi 12 lihaksen ryhmän ala-kasvojen tärkeimmistä lihaksista ryhmän mallintamalle ne interaktiivisesti.

6.3 Animointi ja renderointi

Perinteisten (kahdesta pisteestä kiinnittyvien) lihaksien fysiologia approksimoitiin ellipsoidiksi, joka paksunee, lyhenee ja suoristuu lihaksen jännittyessä.

Orbicularis oris (suun kehän muodostava lihasryhmä) approksimoitiin renkaaksi. Puhuttaessa suun muodon määrittää lähinnä orbicularis orisiin liittyvät lihakset. Suun tapauksessa jännityksen keskipisteen määrittäminen ei ole toimiva ratkaisu, sillä se mahdollistaisi vain suuaukon koon muuttamisen ja tämän takia ryhmä päätyi suun animoimisessa käyttämään jännitysakselia joka kulkee nielusta, hampaiden välistä ulos. Tämä mahdollistaa suun koon ja muodon (edestäpäin katsottuna) muuttamisen lisäksi huulien liikuttelun eteen ja taakse.

Koska orbicularis oris ei työnny esiin homogeenisesti, vaan lähinnä pään keskivivaa olevat osat työntyvät ja vetäytyvät kokomääräisesti, kun taas suupielet eivät juuri ollenkaan (muut kasvolihakset hallitsevat suunpieliä) he käyttivät työntymisgradienttia, jossa näiden ääripäiden työntyminen/vetäytyminen interpoloitiin lineaarisesti.

Lihasten yhdistyminen siten, että yhden lihaksen jännitystila ja sijainti vaikuttavat toiseen lihakseen approksimoitiin sitomalla ne kiinni toisiinsa ideaalijousilla ja näiden voimia iteroimalla saavutettiin tyydyttävä lopputulos.

Orbicularis oris jaettiin kahteen osaan sekä ylä- että alahuulen osalta tarkemman ulostyöntymisen ja sisäänvetäytymisen aikaansaamiseksi.

Pään luonnollisuutta saatiin aikaan lisäämällä satunnaisia silmänräpäytyksiä ja satunnaista siniaallon mukaista pään kääntelyä x-, y- ja z-akselien ympäri.

Järjestelmä käyttää kahta prosessoria tehokkaasti koska simulointi ja renderointi ovat omia säikeitään. Renderointi saa avainruudut rengaspuskurista johon simulointi kirjoittaa niitä ja sen jälkeen interpoloi avainruutujen väliset muutokset.

6.4 Puheen synkronointi

Tärkein asia puheen synkronoinnissa on ”koartikulointi”. Tämä viittaa siihen, että suun ja kielen asento eri foneemien yhdistelmissä voi olla voimakkaastikin eri. Esim. ”kone” ja ”karitsa” sanoissa alkavan k:n aiheuttama suun muoto on eri, sillä k-kirjaimnta seuraava a-kirjain ei muuta k:n lausunta-asentoa, kun taas o-kirjain muuttaa. Vaikuttavat foneemit voivat olla hyvinkin kaukana ja ylittää sanarajoja joten on erittäin vaikeaa tietää mistä jokin akustinen segmentti alkaa ja mihin se päättyy.

Käytetty puhesynkronointi vaatii transkriptin joka sisältää sekä foneemit että niiden ajoitustiedon. Koartikulointi mallinnetaan dominanssifunktioina. Ne kuvaavat puhesegmentin vaikutuksen ääntämiselinten muotoon ajan suhteen. Ryhmän fysiikkapohjaisessa mallissa he käyttävät supistumis- ja työntymisparametreja viiden avainlihaksen, sekä leukaluun kiertymistä nivelensä suhteen mallintaakseen artikulointia. Jokaista parametriä kontrolloi erillinen funktio joka lasketaan dominanssifunktion painotettuna keskiarvona kaikkien segmenttien yli kerrottuna kohteina olevilla foneemeilla. Lihaksen kohde tietyille foneemille saadaan lihasten tilasta kun foneemi lausuttaisiin erillään muusta puheesta.

6.5 Tulokset ja päätelmät

Ryhmä esitti järjestelmän kasvojen mallinnukseen ja animointiin. Selkeästi eniten ihmistyötä (4-7 h) vaativa osuus kaappauksessa ja mallinnuksessa oli etäisyysmittauksista saatavan pistedatan kokoaminen toimivaksi polygonimalliksi. Ryhmä tutkii geneerisen päämallin sovittamista mittauksesta saatuun pistedataan interaktiivisen työskentelyn vähentämiseksi. Korvien ja huulien sisäpintojen topologia aiheuttivat ongelmia (bugeja ulkonäössä) mutta approksimoinnilla päästiin kuitenkin tyydyttävään tulokseen. Ryhmä oli luottavainen siihen, että järjestelmään pystyttäisiin liittämään myös hiukset etsimällä tukan geometria ja tekstuuri valokuvista.

Puhesynteesin kontekstissa heidän käyttämänsä menetelmä oli suoraviivainen ja kun dominanssifunktion parametrit oli määritelty ja jokaisen lihaksen kohdesuipustusparametrit oli interaktiivisesti määritelty jokaiselle foneemille avainkehukset pystyttiin laskemaan automaattisesti.

7 Esimerkki: Matrix-elokuvan jatko-osien digitaaliset kaksoisolennot

Matrix elokuvan jatko-osia varten tarvittiin fotorealistisia digitaalisia kaksoisolennoja tunnetuista näyttelijöistä ja George Borshukovin johtama ryhmä kehitti tätä tehtävää varten menetelmän, jota he kutsuvat univeraaliksi kaappaukseksi ja se yhdistää kaksi hyvin voimakasta konenäkötekniikkaa: optisen virtauksen ja fotogrammetrian.

7.1 Näyttelijän suorituksen ja ulkonäön kaappaus

Näyttelijän suoritus kaapattiin viidellä synkronoidulla kameralla hajaantuneessa valaistuksessa. Tämän jälkeen kunkin pikselin (sic.) liike ajan suhteen jokaisessa kamerassa mitattiin käyttäen optista virtausta ja tämän prosessin tulos yhdistettiin neutraalin ilmeen omaavaan 3-D malliin näyttelijästä ja kameroiden sijaintien fotogrammetriseen rekonstruktioon. Käytetty algoritmi toimii projisoimalla mallin verteksin jokaiseen kameraan ja sitten seuraamalla tämän verteksin liikettä 2-D näkymissä optisella virtauksella ja joka kehyksessä verteksin 3-D sijainti aproksimoidaan kolmiomittauksella.

Optisessa virtauksessa virheet voivat kumuloitua ajan myötä aiheuttaen ei-toivottua

ajelehtimistä 3-D rekonstruktiossa. Tässä tuotannossa tämän eliminoimiseksi prosessiin otettiin mukaan manuaalinen avainkehysten korjaus, jossa geometrian virheet korjataan käsin kun virhe kasvaa häiritsevän suureksi ja korjaus takaisinsyötetään edellisiin kehyksiin käyttäen käänteistä optista virtausta.

Uskottavaa kasvojen renderointia ei voi saavuttaa ottamatta huomioon, että tekstuuri vaihtelee ilmeiden mukaan. Koska ryhmä ei käyttänyt kaappauksessa kasvoihin lisättyjä merkkejä pystyivät he yhdistämään kuvat useammasta kamerasta ajan suhteen ja näin tuottamaan saumattoman animoidun UV-värikartan, josta saatiin eristettyä olennaisia teksturaalisia muutoksia kuten hienojen juonteiden ja ryppyjen ilmestymistä ja ihon venymisen aiheuttamia värimuutoksia. UV-kartoituksella kuvataan kolmiulotteinen esine kaksiulotteiseksi tekstuurikartaksi.

Vaikka näyttelijän suorituksesta eristetty animaatio sisälsi useimmat liikenyanssit siitä puuttui mittakaavaltaan pienet yksityiskohdat, kuten huokokset ja hienot juonteet. Dynaamiset rypyt identifioitiin ja eristettiin tekstuurikartoista ja kerrostettiin staattisen töyssykartan päälle.

7.2 Heijastavuuden kaappaaminen ja simuloiminen

Elokuvan tuotannossa kasvojen BRDF eristettiin kuvapohjaisesti. Näyttelijät kuvattiin usealla samanaikaisella kameralla useissa eri valaistusolosuhteissa käyttäen valolavaa. Kaappausasetelma värikalibroitiin tarkasti ja fotogrammetrialla rekonstruoidtiin kameroiden sijainnit ja pään sijainti kaikkien kameroiden suhteen. Kerätty kuvamateriaali tuotiin yhteiseen UV-avaruuteen uudelleenprojisoimalla se käyttäen näyttelijän cyberscan-mallia. Tätä avaruutta käyttäen ryhmä pystyi analysoimaan ihon heijastavuusominaisuuksia lukuisissa valaistusolosuhteissa ja valon ulostulokulmissa ja tämän tiedon pohjalta ryhmä eristi parametrit aproksimaaliseen analyttiseen BRDF:ään, joka koostui lambertialaisesta diffuusio-komponentista ja modifioidusta Phong-tyyppisestä kirkas-komponentista, jossa oli Fresnel-tyyppinen efekti.

7.3 Renderointi

Tuotannon edistyessä kävi selväksi, että realistinen ihon renderointi ei ole mahdollista ilman pinnanalaisen hajonnan simuloimista. Olemassaolevat menetelmät läpinäkyvien materiaalien renderointiin havaittiin konetehovaatimuksiltaan kalliiksi ja tulokset epätyytyttäväiksi, joten ryhmä kehitti oman menetelmänsä ihonalaisen si-

ronnan tuottamiseksi, joka oli melko helppo toteuttaa. Hajaantuneen valon heijastuminen kameran suuntaan tallennettiin 2-D valokarttaan ja sitten aproksimaalisesti simuloidaan valon hajaantuminen kuvan alueella. Tuloksen parantamiseksi he muokkasivat diffuusio-parametrejä eri aallonpituuksille. Kohteisiin joiden läpi valo voi kulkea suoraan, kuten korvat, he käyttivät traditionaalisempaa ray tracing-lähestymistapaa halutun läpinäkyvyyssefektin tuottamiseen. [HaD04]

8 Konklusio

Tutkielma esitteli ihmishahmojen kaappausta tietokonemalleiksi ja näiden mallien animointia eri menetelmin ja eri tarkoituksiin. Alan kehittyessä selkeästi koneen näköisistä hahmoista kohti digitaalisia kaksoisolentoja, joiden erottaminen oikeista ihmisistä on erittäin vaikeaa tai jopa mahdotonta on alan kehitystyö siirtynyt lähestulkoon kokonaan akateemiselta puolelta kaupalliselle puolelle, mikä vaikeuttaa suuresti tiedon saantia uusimmista tekniikoista kun ne tapaavat olla liikesalaisuuksia.

Tutkielma toi esiin nykyisen tilanteen, jossa yhdestä ainoasta tavallisesta valokuvasta voidaan luoda hyvin näköinen kasvomalli pitkälti automatisoidusti. Tietojenkäsittelytieteellisestä näkökulmasta asiat käsiteltiin melko pinnallisesti, mutta näiden asioiden saattaminen populaariin tietoisuuteen on tärkeä asia, jotta digitaalisten kaksoisolentojen aiheuttama tuho ja sen uhka ja tapainturmellus saataisiin jonain päivänä lainsäätäjien asialistalle tai edes julkiseen keskusteluun. Asian käsittely oikeustieteellisenä ongelmana olisi mitä tervetulleinta. Teologisessa tarkastelussa on erittäin vaikeaa olla huomaamatta lukuisia yhtymäkohtia ja samankaltaisuuksia ilmestyskirjassa (kirjoitettu n. 200 jkr) kuvattujen petojen ja digitaalisten kaksoisolentojen välillä. Ihmiskunnan jakaantuminen yli-ihmisiin, ali-ihmisiin ja sinne välille on filosofien, psykologien ja yhteiskuntatieteilijöiden näkökulmasta erittäin polttava aihe.

9 Kuvien lähteet ja kopio-oikeusilmoitukset

- Kuva 1. Copyright ACM 1999 - <http://doi.acm.org/10.1145/311535.311556> - Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or

distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

- Kuvat 2, 3 ja 4 Copyright ACM 2000 - <http://doi.acm.org/10.1145/344779.344855>
- Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

Leet

- arius3d Arius3d, Arius3d digital imaging solutions - 3d color laser scanner. <http://www.arius3d.com/main.html?contentId=1>. [16.12.2007]
- BIV99 Blanz, V. ja Vetter, T., A morphable model for the synthesis of 3d faces. *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1999, ACM Press/Addison-Wesley Publishing Co., sivut 187–194.
- Cyberw99 Cyberware, Head face scanner specifications. <http://www.cyberware.com/products/scanners/pxSpecs.html>. [16.12.2007]
- Dev00 Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W. ja Sagar, M., Acquiring the reflectance field of a human face. *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 2000, ACM Press/Addison-Wesley Publishing Co., sivut 145–156.
- HaK01 Haber, J., Kähler, K., Albrecht, I., Yamauchi, H. ja Seidel, H.-P., Face to Face: From Real Humans to Realistic Facial Animation. *Proceedings of the 3rd Israel-Korea Binational Conference on Geometrical Modeling and Computer Graphics*, Seoul, Korea, October 2001, Kyung Moon, sivut 73–82.
- HaD04 Haber, J. ja Terzopoulos, D., Facial modeling and animation. *SIGGRAPH '04: ACM SIGGRAPH 2004 Course Notes*, New York, NY, USA, 2004, ACM Press, sivu 6.
- CuD05 Oana G. Cula, Kristin J. Dana, F. P. M. ja Rao, B. K., Skin texture modeling. *International Journal of Computer Vision*, 62,1(2005).
- Par72 Parke, F. I., Computer generated animation of faces. *ACM'72: Proceedings of the ACM annual conference*, New York, NY, USA, 1972, ACM Press, sivut 451–457.
- PiL05 Pighin, F. ja Lewis, J. P., Introduction. *SIGGRAPH '05: ACM SIGGRAPH 2005 Courses*, New York, NY, USA, 2005, ACM, sivu 1.

- Sag04 Sag04, Reflectance field rendering of human faces for "spider-man 2". *SIGGRAPH '04: ACM SIGGRAPH 2004 Sketches*, New York, NY, USA, 2004, ACM, sivu 118.
- Sif05 Sifakis, E., Neverov, I. ja Fedkiw, R., Automatic determination of facial muscle activations from sparse motion capture marker data. *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, New York, NY, USA, 2005, ACM Press, sivut 417–425.