

**AAU
Re**

Studio

Joni-Kristian Hämäläinen



signaalinkäsittelyn professori
Tampereen teknillinen yliopisto
joni.kamarainen@tut.fi

Jori Pölkki



lehtori
Kymenlaakson ammattikorkeakoulu
jori.polkki@kyamk.fi

Tommi Tykkälä



tutkija
Lappeenrannan teknillinen yliopisto
tommi.tykkala@lut.fi

Marko Siitonen



pt. tuntiopettaja
Kymenlaakson ammattikorkeakoulu
marko.siitonen@kyamk.fi

Samuli Karevaara



tutkimuspäällikkö
Kymenlaakson ammattikorkeakoulu
samuli.karevaara@kyamk.fi

Suvi Pylvänen



pt. tuntiopettaja
Kymenlaakson ammattikorkeakoulu
suvi.pylvanen@kyamk.fi



KYAMK

University of Applied Sciences

Pääkirjoitus

Vipuvoimaa

EU:lta

2007-2013

Joni-Kristian Kämäräinen

Ensimmäiset siemenet AuReStudio-hankeesta kylvettiin jo vuosina 2009-2010, kun Lappeenrannan teknillisen yliopiston Kouvolan yksikkö (LUT Kouvola) ja Kymenlaakson Ammattikorkeakoulun (KyAMK) Taiteen ja viestinnän yksikkö etsivät yhteisiä kiinnostuksenkohteita tutkimuksessa ja opetuksessa. Hyvin pian löydettiin poikkitaiteellinen (ja poikkitaiteellinen) idea yhdistää KyAMK:n AV-tuotanto-osaaminen sekä loistavat studiotilat Kasarminmäellä ja samanaikaisesti kasvava LUT Kouvolan uusi tutkimusala: tietokonenäkö. Lähtäkohta oli, että KyAMK:n taiteellinen ja tekninen osaaminen elokuva- ja televisiotuotannossa ja LUT Kouvolan konenäkötempu yhdessä loisivat mielenkiintoisia tutkimusideoita ja innostavan oppimisympäristän opiskelijoille. Maakunnan ja Suomen kannalta toivottiin toki myös ristisiitoksen tuottavan uusia tuotteita ja työpaikkoja pitemmällä tähtäimellä.

2011 alkoivat palaset loksahdella paikalleen. Ensin saimme lupauksen merkittävästä rahoituksesta Teknologiateollisuuden 100-vuotissäätiältä ja sitten toisen merkittävän rahoituksen Kymenlaakson liiton hallinnoimasta EU:n EAKR-rahoituksesta. Suuri kiitos rahoittajille luottamuksesta ja

rohkeudesta, joka on jo kantanut hedelmää ja joka toivottavasti kantaa sitä vielä paljon tulevaisuudessa. 2012 starttasi AuReStudio.

Uuden teknologian meille toi Ranskasta päässään nuori ja innokas DI Tommi Tykkälä, ja niin aloimme opettelemaan teknologiaa, joka oli aivan uutta - Tommin sanoin: ”Tiivis RGB-D trakkäys Kinectillä.” 2012-2013 Tommi kehitti menetelmän, joka pystyy haastamaan kaikki vastaavat kilpailijat maailman parhaissa yliopistoissa ja 2014 hyväksyttiin hänen erinomainen väitöskirjansa alan huipputoimesta.

Samaan aikaan KyAMK kävi läpi kovia rakenteellisia muutoksia, joissa painopistettä siirrettiin perinteisemmästä AV-tuotannosta kohti digitaalisia sisältöjä ja pelejä. Ennakoluuloton uuden luominen KyAMK:ssa ja uudet tuulet LUT Kouvolassa saivat aikaan Kasarminmäellä jotain sellaista, joka ei muuten olisi ollut mitenkään mahdollista. Nyt Kasarminmäellä ”kuplii” ja KyAMK:lla on mahdollisuus tarjota opetusta ja tutkimusta täysin uusien teknologioiden ja niiden mahdollistamien uusien ideoiden parissa. Tulevaisuutta ei tarvitse yrittää ennustaa jos pystyy itse keksimään sen!



LUT

Lappeenranta

University of Technology

Vipuvoimaa

EU:lta

2007-2013



Euroopan unioni

Euroopan aluekehitysrahasto

Lisätyn todellisuuden teknologiaa TV-studioille

Tommi Tykkälä

AureStudio-hankkeessa on tutkittu tapoja millä digitaalista 3D sisältöä, kuten reaaliaikaista tietokoneanimaatiota, voidaan yhdistää live TV-lähetykseen. Pääongelmina on kameran kolmiulotteinen liike-estimointi suhteessa studion staattiseen 3D-malliin sekä renderöityjen kuvapikslien yhdistäminen oikeassa syvyysjärjestyksessä TV-kuvan kanssa.

”Lisätyllä todellisuudella” tarkoitetaan synteettisen grafiikkasisällön reaaliaikaista piirtämistä kuvatun videomateriaalin päälle. Perinteisessä elokuva- ja TV-tuotannossa on totuttu tekemään yhdistelmä päinvastoin: täysin tietokonegeneroituun kuvaan tuodaan kuvista leikatut ihmiset mukaan. Hankkeessa toteutettiin lisätyn todellisuuden teknologia, joka myöhemmin modifioitiin yhteensopivaksi perinteisemmän greenscreen-käytön kanssa.

Tyypillisesti Hollywood-elokuvissa tietokonegrafiikan yhteensovittaminen tehdään

jälkikäsitteilynä, mikä vaatii paljon käsityötä. Kameranliike estimoidaan ensin joko manuaalisesti mittaamalla tai puoliautomaattisesti kuvasisällön muutoksia analysoimalla. Tämän jälkeen sisältöä työestetään valmisohjelmilla, kunnes ihmissilmä ei havaitse enää artefakteja. Tavanomaisia artefakteja ovat kappaleiden välinen epänormaali liike, toisiaan leikkaavat kappaleet ja peitossa olevat kappaleet etualalla. Tietokoneella generoitu 3D grafiikka sovitetaan myös kuvissa näkyvään valaistukseen, jotta liitos vaikuttaa mahdollisimman uskottavalta. Prosessi on kallis ja hidas siitä syystä, että se vaatii tarkkaa käsityötä ja ongelmat jälkikäsitteilyvaiheessa vaativat toisinaan koko materiaalin uudelleenkuvaamista.

Kun vastaava prosessi sovitetaan live-lähetykseen, syntyy uudenlaisia haasteita. Teknologian täytyy toimia reaaliaikaisesti, mutta samanaikaisesti sen pitää olla käyttövarma ja edullinen. Taideteollisen korkeakoulun yhteistyönä on rakennettu LuMe-studio, jossa

”Lisätyllä todellisuudella” tarkoitetaan synteettisen grafiikkasisällön reaaliaikaista piirtämistä kuvatun videomateriaalin päälle.

kuvataan myös lisätyn todellisuuden TV-lähetyksiä. Järjestelmä täyttää käyttökriteerit, mutta sen hinta on satojatuhansia euroja. Kallis hinta muodostuu pääosin siitä, että TV-kameroiden liike estimoidaan käyttämällä ulkoista kamerajärjestelmää, mikä on kiinnitetty studion kattoon. Kameran seuravat 60-100Hz taajuudella TV-kameroihin kiinnitettyjä heijastavia markkeripalloja ja niistä kasattuja referenssiobjekteja, joita seurataan infrapuna-aallonpituuksilla.

Infrapunataajuudet ovat siinä mielessä hyvä valinta, että studiovalaistusta ei tarvitse millään tavalla muokata. Kun kamerajärjestelmä pysyy kiinteästi paikallaan ja kohteena oleva TV-kamera liikkuu, voidaan markkeripallojen liikkeitä seuraamalla estimoida TV-kameran liike suhteessa huoneen koordinaatistoon. Vastaavaa teknologiaa hyödynnetään esimerkiksi peliteollisuudessa animaatiohahmojen liikkeenkaappauksessa (Vicon- ja NaturalPoint-järjestelmät).

Kouvolan AuReStudio-hankkeessa pääideana oli kehittää uutta teknologiaa ja julkaista tiedefoorumeilla kehityksestä syntyneet menetelmät osana väitöskirjatyötä. Järjestelmän pääkriteerit olivat reaaliaikaisuus ja alhainen hinta. Edullisuus on tärkeää, sillä hankkeen yhdeksi päämääräksi asetettiin myös uusi

startup-yritys, joka voisi kilpailla olemassa olevien hintavien järjestelmien kanssa.

Pelikonsoleissa käytetty teknologia on edullista ja siitä syystä Nintendo Wii:n ohjainkapula (”wiimote”) sekä Xbox360 mukana levitetty Kinect-sensori olivat vahvoja kandidaatteja liikkeen mittaamiseen. Hankkeessa päädyttiin Kinect-pohjaiseen teknologiaan, sillä sensorista voidaan lukea RGB-videon lisäksi syvyyskuvaa (D), mitä puolestaan mahdollistaa saumattomasti tietokonegrafiikan yhdistämisen. Kinect-sensorin käyttö on teknisesti haastavaa, sillä liikkeenestimointi voi käyttää ainoastaan raakaa RGB-D kuvadataa kolmiulotteisen kameranradan tuottamiseen. Lisäksi Kinect-sensorin käyttöalue on puolesta metrillä n. kuuteen metriin mikä asettaa haasteita käyttöetäisyyksille.

Tieteessä kuvapohjaista liike-estimointia tutkittu paljon ja ratkaisut nojaavat harvojen kiintopisteiden irrottamiseen ja määrittämiseen eri ajanhetkiltä otettujen kuvien kanssa. Kun tunnetaan joukko 2D pisteitä ja niille vastinparit useassa eri näkymässä, voidaan päätellä kameran 3D liike. Koska kiintopisteitä ei aina ole (esimerkiksi valkoinen seinä), on eräänä mahdollisuutena kiinnittää infrapuna-LEDEjä studion takaseinämään, ja varmistua siitä että liike on aina ratkaistavissa. LEDien pysyvä kiinnitys ja uudelleenkiinnitys on kuitenkin hieman hankalaa etenkin Meduusa-TV:n tiloissa, jossa lavasteita muutetaan usein.

Kuvasta irrotetut kiintopisteet ovat toisaalta alttiita virheellisille määrittämisille mikäli samannäköinen kuvan nurkkapiste löytyy monesta paikasta. Ranskan Sophia-Antipoliksessa, RGB-D kuvapohjaisen kameraseurannan ongelmaa on lähestytty hyödyntämällä tiheää pikselidataa. RGB-D kuva kiinnitetään referenssinäkymäksi ja pyritään

matemaattisen mallintamisen kautta selittämään uusin RGB-D kuva pikselitasolla.

Oletuksen on väriakioisuus, mikä tarkoittaa sitä, että värien oletetaan pysyvän samoina riippumatta kuvanottosuunnasta. Tämä oletus pitää studioympäristössä hyvin paikkansa, kunhan kiiltäviä peilipintoja tai läpinäkyvää lasipintaa ei varta vasten tuoda studiorekvisiitaksi. Menetelmässä minimoidaan epälineaarinen kustannusfunktio, jonka tuloksena siis löytyy se 3D liike, joka parhaiten selittää uuden kuvan suhteessa referenssiin.

Toisinkuin kiintopisteiden seurannassa kysymyksessä on suora menetelmä (direct method), koska kuvat käytetään estimoinnissa sellaisenaan muuttamatta niitä ensin kiintopistejoukoksi. Koska käytettävissä on teoriassa (kuvaleveys) \times (kuvakorkeus) kuvapistettä, liikkeenestimoinnista tulee huomattavasti varmempaa, koska dataa on paljon ja 2D-pisteiden mätsääminen voidaan välttää kokonaan.

Jotta menetelmää voidaan käyttää tosielämän käyttösovelluksissa, kuvasisällön täytyy olla tekstuuriiltaan rikasta eli sisältää kuvareunoja mahdollisimman paljon, mitkä vastaavat staattista taustageometriaa, ja kuvauskohteet ovat Kinect-sensorin kantaman piirissä. Tiheitä ja suoraa liikkeenestimointimenetelmiä on tutkittu ja sovellettu vuodesta 2007 eteenpäin RGB-D sensorien kanssa ja tulokset ovat olleet hyviä.

Ensimmäiset tulokset on saatu katunavigointisovelluksista, missä GPS sijaan kulkuneuvon liikettä seurataan katolle kiinnitetyn stereokameran syötteen avulla. Kinect-sensorien yleistymisen myötä teknologia on siirtynyt nopeasti sisätiloissa tehtäviin ”SLAM-sovelluksiin. SLAM (Simultaneous Localization and Mapping) tar-

Kun tunnetaan joukko 2D pisteitä ja niille vastinparit useassa eri näkymässä, voidaan päätellä kameran 3D liike.

koittaa kameran liikeseuranta kuvadatan perusteella johon on yhdistetty samanaikainen ympäristön 3D-mallinnus.

Tutuimmat SLAM-sovellukset ovat robotipölynimureissa, jotka rakentavat lattiamallin ja navigoivat sen perusteella. Studiovalaistus asettaa hieman haasteita edulliselle Kinect-kameralle, sillä värit palavat hyvin herkästi puhki. Tästä syystä valaistusta jouduttiin testeissä pitämään hieman diffuusimpana.

Videoesimerkkejä:

y2u.be/wALQB3eDbUg

y2u.be/fLC3Mjz2na4

Idealisesti TV-studiolla voitaisiin käyttää tarkkaa CAD 3D-mallia kuvapohjaiseen kameran liikeseurantaan (esim. Metaio-kirjasto). Koska 3D-malli ei kuitenkaan ole helposti käytettävissä muuttuvasta studiostatusta, se pitää mitata Kinect-sensorilla ennen kuvauksia. Joukko RGB-D referenssinäkymiä, joihin on assosioitu kameran tarkka kuvauspaikka ja -suunta, vastaavat värikkäistä pisteistä koostettua 3D-mallia studiosta.

RGB-D kuvat voidaan helposti muuntaa värikkäiksi pistepilviksi, mutta tarkat kameran paikat ja suunnat vaativat erityisen mitausmenetelmän.

Hankkeessa kehitettiin teknologia, jolla referenssinäkymät ja niitä vastaavat kameran liikesuunnat saadaan muodostettua. Pääajatuksena on liikutella Kinect-sensoria siten että kaikki kuvaussuunnat tulevat katetuksi. Liikeradan varrelta poimitaan referenssinäkymät tasaisin välein, ja prosessin aikana ratkaistaan kutakin input kuvaa kohti kameran 3D-liikeinkrementti käyttäen suoraa ja tiheää kuvavirheen minimointitekniikkaa. Kun liikeinkrementit ketjutetaan, voidaan ylläpitää käsitystä kameran senhetkisen asennon ja alkupaikan asennon välillä. Taustan 3D-mallinnus tehdään aina tyhjälle näyttämölle, siitä syystä, että dynaamiset kohteet kuten liikkuvat ihmiset eivät saa tulla mukaan staattiseen referenssimalliin. Ensimmäisistä testeistä ja lähestymistä vasta julkaistiin Barcelonassa järjestettyyn VISAPP-konferenssiin artikkeli [1].

Videoesimerkkejä:

y2u.be/Xnn_06r7tFE

y2u.be/5UIDnCG5E50

y2u.be/9uBS6SSEN5M

SLAM (Simultaneous Localization and Mapping) tarkoittaa kameran liikeseuranta kuvadatan perusteella johon on yhdistetty samanaikainen ympäristön 3D-mallinnus.

Kun liikeinkrementit ketjutetaan, voidaan ylläpitää käsitystä kameran senhetkisen asennon ja alkupaikan asennon välillä.

Kuvauksen alkaessa näyttämölle astelee 1-2 juontajaa, jotka eivät saa peittää koko kuvapinta-alaa. Taustamallista on poimittu kuvareunat, jotka toimivat referenssinä kameran 3D-liikkeenseurannalle. Järjestelmä luottaa siihen, että kuvareunojen värit sekä etäisyydet kamerasta eivät oleellisesti muutu kuvausten aikana. Ne uuden RGB-D kuvan kuvareuna-alueet, jotka eivät selity taustamallin avulla, katsotaan ihmisiksi ja muiksi dynaamisiksi kohteiksi, joita tulee välttää. Koska järjestelmä sietää vain osittain etualalla liikkuvia ihmisiä, se asettaa käyttörajoituksen TV-kuvauksiin. Jos kuitenkin kuvaukset ovat suunniteltu siten, että tekstuuria on paljon taustalla, ja etuala peittää vain osan taustasta, niin järjestelmä toimii hienosti.

Videoesimerkkejä:

y2u.be/iVdYPbHY2ro

Lähetysten aikana käytetään samaa teknologiaa kameran liikkeen estimointiin suhteessa kiinnitettyihin referenssinäkyymiin. Koska viimeistään tässä vaiheessa estimoinnin on pakko toimia reaaliajassa (30Hz), ohjelmakoodin optimointiin sijoitettiin merkittävästi aikaa ja vaivaa. Toteutus tehtiin Ubuntu Linuxille, jonka erittäin kevyt 48-ytiminen GPU valjastettiin laskentakäyttöön. GPU:n käyttö on erittäin oleellista, sillä GPU:ta hyödyntävät ohjelmatoteutukset skaalautuvat näytönohjainkehityksen myötä suoraan nopeammiksi.

GPU:lla laskenta tapahtuu rinnakkaisesti tuhansille toimituksille yhtä aikaa. Nvidian CUDA-tekniikkaa hyödynnettiin toteutuksen tekemiseen. Toteutuksesta julkaistiin Journal of Visual Communication and Image Representation -lehteen 12-sivuinen artikkeli [2].

Videoesimerkkejä:

y2u.be/L_OLnFc7QxU

y2u.be/gE6u51tkfQM

Ensimmäisen testi-lähetysten aiheena olivat kamppailulajit ja studiovieraana oli brasilialaisen ju-jutsun mestari, jota haastateltiin. Lisätyn todellisuuden elementtinä lähetykseen liitettiin interaktiivisesti animoitu Ninja-hahmo, jota ohjattiin studiotilaan asennetun motion capture-järjestelmän avulla. NaturalPointin liikkeenkaappausjärjestelmä ei automaattisesti ollut liitettävissä hankkeessa kehitettyyn kameranseuranta-

moduuliin, joten liiketietojen siirtoa varten rakennettiin C++ kielellä räätälöity client/server-ohjelma, joka kytkeytyy NaturalPointin kaappausohjelmaan, lukee liikettä ohjelmointiasennon ja konvertoi sen Panda3D grafiikkamoottorin ymmärtämään muotoon paikallista visualisaatiota varten. Visualisointi-ikkuna mahdollistaa esikatselun, josta näkee miten liiketieto sovituu TV-lähetyksessä käytettävän 3D-hahmon liikkeiksi.

Testilähetysten 3D-hahmo kehitettiin yhteistyössä Kyamkin kanssa.

Testilähetysten 3D-hahmo kehitettiin yhteistyössä Kyamkin kanssa. Hahmon mallinsi opiskelija Taija Lumitähti Blender-ohjelmalla. Hahmo koostui artikuloidusta luurangosta jonka ympärille sovitettiin teksturoitu pinta, joka pysyi pistepainojen avulla luurangon liikkeessä mukana. Hahmon raajojen mittasuhteita ja skaalan muuttelua varten luotiin tekstitiedostopohjainen säätömahdollisuus. Toisinaan kuvauksissa huomattiin, että hahmon jalat tai kädet voisivat olla ulottuvammat ja niitä muutettiin lennosta.

Ninja-hahmon liiketieto siirrettiin lähiverkon ylitse kannettavalle tietokoneelle, jossa ajettiin myös kameran liike-estimointia Kinect-sensorin avulla. Tietokone luki TV-kameraan kytketyn Kinect-sensorin kuvat sekä

ninjan asennon ja kompositioi 3D ninjasta renderöidyn 2D-kuvan saumattomasti Kinectin RGB-kuvan kanssa. TV-lähetys saatiin onnistuneesti pakettiin, mutta se TV-kamera mihin Kinect-sensori oli kytketty ei liikkunut merkittävästi kuvauksen aikana, joten demonstraatio osoitti enemmänkin reaaliaikaisen 3D animaation yhdistämistä live-lähetykseen, kuin varsinaisesti kameranseurannan.

Ensimmäinen testi osoitti myös muutamia parannuskohteita kameranseurannan käytössä, nimittäin Kinectin kantama oli ääriarjoilla, ja näyttämön taustamalli ei sisältänyt riittävästi kuvareunoja. Testilähetystä aiemmin järjestelmää testattiin mm. kiskokameraan kytkettynä joka tuntui toimivan huomattavasti paremmin. Kiskokameraa käytettäessä taustan 3D-mallinnus on suoraviivaista sillä yksi kameran liike, joka kattaa koko kiskon pituuden riittää.

Videoesimerkkejä:

y2u.be/C5YgGpgdW9l

y2u.be/QhuTg11vLcc

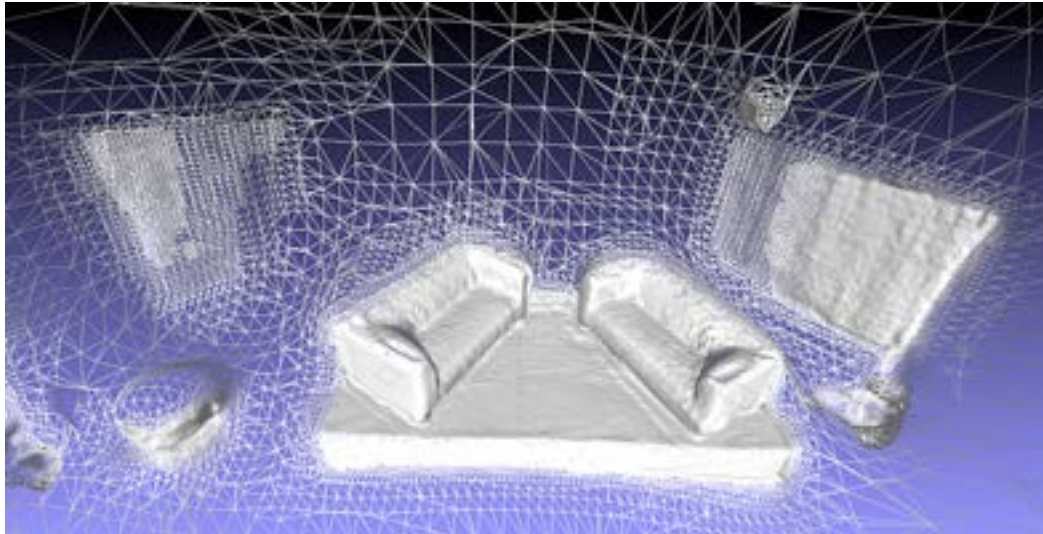
AuReStudiolle oli hankkeen alussa hankittu väriavainin, joka mahdollistaa greenscreenin käytön etualan hahmojen irrotukseen. Tämä mahdollisti järjestelmän jatko-jalostamisen perinteisempään käyttötapaukseen, jossa etualan ihmiset siirretään tietokonegeneroituun ympäristöön. Studiolle hankittiin kuvioitu matto, joka kiinnitettiin TV-kameran alle. Matto oli sen kokoinen, että TV-kamera pystyi liikuttamaan n. metrin joka suuntaan ja kääntelemään joitakin kymmeniä asteita sivu- ja pystysuunnassa. Kun Kinect-sensori suunnattiin alaviistoon

mattoa kohti, itse näyttämön suunnassa oli mahdollista käyttää greenscreeniä, sillä kuvareunat löytyivät jo matosta.

Nyt 3D-hahmon kohdistuksesta tuli kuitenkin hankalampaa, sillä hahmon piirtoon käytetyn virtuaalisen kameran parametrit piti muuntaa mattokameran liikkeistä. TV-kameran ja Kinect-kameran tarkkaa asentoa ei kalibroitu vaikka siihen löytyy myös tieteestä menetelmiä.

Teknologiademoa varten Kinect-kamera suunnattiin 90-asteen kulmassa alaspäin ja oletettiin että 90-astetta on riittävän lähellä todellisuutta. Tuloksena syntyi toinen testilähetys, jossa TV-kameran kuva ohjattiin väriavainimelle, jonka kautta TV-ohjaamon puolelle saatiin taustaerotettua kuvaa näyttämöltä. Taustalle määriteltiin 3D temppele ja sen ympäristö, jota piirrettiin kameramoduulin määräämien suuntien mukaan. Piirretyt kuvat ohjattiin kannettavalta tietokoneelta VGA-kaapelin avulla myös avaintimeen, joka suoritti kuvien yhdistämisen. Koska avainin tekee yhdistämisen puhtaasti avainvärin perusteella, ei ollut mahdollista hyödyntää mitattuja syvyyskuvia toisella Kinect-sensorilla joka olisi suunnattu eteenpäin.

Kannettavalla toimivaan ohjelmaan lisättiin hahmon monistustuki, jonka avulla 3D hahmo voitiin replikoida kokonaiseksi joukoksi. Lopulta kuvattiin demolähetys kasaan jossa esiteltiin taistelukohtaus näyttämöllä taistelevan ihmisen ja 3D-hahmolauman välillä. Prototyypistä tuli hauska ja kiinnostava. Sitä olisi mahdollista soveltaa erilaisten puhetilaisuuksien kokeilemiseen, missä yleisön massakäytöstä voitaisiin mallintaa vapaasti 3D-hahmoilla.



Tommi Tykkälä, Sofa

Järjestelmä toimi huomattavasti paremmin teksturoidun maton kanssa, mutta pari ongelmaa kuitenkin oli myös. TV-kuvan ja Kinect-kuvan välinen latensi aiheutti sen, että nopeissa kameranliikkeissä kesti hetken ennen kuin etualan grafiikka mätsäsi tempelimaismen kanssa.

Toinen ongelma liittyi avaintimen käyttöön: siitä huolimatta, että tietokonegeneroitu kuva oli jo virheetöntä, avainnin teki muunnoksen kuvasignaalille joka heikensi oleellisesti sen laatua. Avaintimeen ei löytynyt lähetyksen aikana konfiguraatiota joka olisi mahdollistanut alkuperäisen virheetömän kuvan näkymisen loppulähetyksessä. Jatkokehittäväksi jäi myös RGB-D sensorin lattiasuunnan ja TV-kameran suunnan tarkka asentoeron estimaatio.

Videoesimerkkejä:

y2u.be/fbyRb8vj1H8
y2u.be/NdL9lXmCeY8
y2u.be/Sna1zIstdnF4

Studioteknologian kehityksestä ja teknisistä yksityiskohdista kirjoitettiin 160 sivuinen väitöskirja ”REAL-TIME IMAGE-BASED RGB-D CAMERA MOTION TRACKING AND ENVIRONMENT MAPPING” [4], johon liittyvät väitöstilaisuudet pidettiin Ranskassa Sophia-Antipoliksessa 3.9.2013 ja Lappeenrannan teknillisessä yliopistossa 21.09.2013. Koska järjestelmän teknisenä pohjana toimi Sophia-Antipolixen taustatutkimus, väitöskirjassa on myös kappaleita, jotka esittelevät projektia edeltävää materiaalia Ranskan tutkimusyksiköstä, jossa alikirjoittanut opiskeli tietokonenäkömenetelmiä 2 vuotta.

Projektin aikana huomattiin, että Kinect-sensoripohjainen sisätilamallinnus on mahdollista studioympäristöjen lisäksi myös asunnoissa. Projektin loppuvaiheilla tutkittiin mitä vaaditaan, jotta RGB-D sensorilla

voidaan tuottaa vesitiiviitä ja teksturoituja 3D-malleja sisätiloista. Kehitetyllä menetelmällä onnistuttiin luomaan komposiittisia 3D-malleja, jotka yhdisteltiin useista näkymistä kasaan.

Pienenä laatuongelmana oli Kinect-sensorin kohina, joka aiheutti n. +5cm poikkeamia pahimmillaan todellisista pinnoista. Laatuongelmaan kehitettiin ratkaisu, jossa useista syvyyskuvista yhdistetään pisteet samaan referenssikoordinaatistoon ja suodatetaan lopulliset pinnat niistä. Lopulta Poisson-triangulaatiomenetelmään sovellettiin tuottamaan vesitiivis geometria. Sisätilamallinnuksesta innostui Tampereen Nokia NRC, jonka kanssa on aloitettu kaksi yhteistyöhanketta jotka kehittävät teknologiaa eteenpäin osaksi Here-kartoitusohjelmistoja. Sisätilamallinnuksesta julkaistiin artikkeli myös Tokion IROS-konferenssissa, joka keskittyy robotiikka-alan tuoreimpiin innovaatioihin [3].

Videoesimerkkejä:

y2u.be/tD3lFxrCHaw
y2u.be/Mhk-fWv27GU

Hankkeen tuloksena syntynyt kameran liikkeen seurantamoduuli täyttää kaikki sille asetetut kriteerit. Se on kuvapohjainen tapa seurata kameran liikkeitä, tarkka kunhan tekstuuria riittää ja käyttöetäisyys sensorin rajoissa, reaaliaikainen, tarvitsee vain hyvin vaatimattoman GPU:n laskutoimituksia varten, toteutus on skaalautuva mikäli syötekuvien kokoa on tarve kasvattaa myöhemmin, ja ratkaisu on ennen kaikkea edullinen, koska tarvitaan ainoastaan Kinect-sensori jonka hinta on ~100€. Kehityksestä on julkaistu 3 tieteellistä artikkelia ja 160s väitöskirja [1,2,3,4], jotka dokumentoivat teknologian tarkasti.

Menetelmää verrattiin myös RGB-D SLAM-alan state-of-the-art teknologiaan, KinectFusioniin, ja todettiin kilpailukykyiseksi (kts. video). KinectFusion tuotti hieman tarkempaa rataa mutta hajoaa helposti tasomaisiin pintoihin joita sisätiloissa paljon (seinät, lattiat). Hanke keskittyi tekniikkakehitykseen, mutta jatkokehityksessä on Tampereen yliopistolla GUI, joka mahdollistaa moduulin helpon käytön erilaisissa sovelluksissa.

Videoesimerkkejä:

y2u.be/tNz1p1sdTrE

Julkaisut:

- [1] RGB-D Tracking and Reconstruction for TV Broadcasts, T. M. Tykkälä, H. Hartikainen, A. I. Comport, and J-K. Kämäräinen, 8th International Conference on Computer Vision Theory and Applications (VISAPP), Barcelona, Spain, Feb 2013.
- [2] ”Live RGB-D Camera Tracking for Television Production Studios” T. M. Tykkälä, A.I. Comport, J-K. Kämäräinen and H. Hartikainen. Journal of Visual Communication and Image Representation, Elsevier, Apr 2013.
- [3] ”Photorealistic 3D Mapping of Indoors by RGB-D Scanning Process” T. M. Tykkälä, A. I. Comport, and J-K. Kämäräinen, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, Nov 2013.
- [3] Real-time Image-based RGB-D Camera Motion Tracking and Environment Mapping, Väitöskirja, 160s, T. M. Tykkälä, 2013

Liikkeenkaappaus

Marko Siitonen

Liikkeenkaappaus (Motion capture) on tekniikka, jolla tarkoitetaan esiintyjän tai erilaisten objektien liikkeiden tallentamista ja muuntamista digitaaliseen muotoon. Liikkeenkaappaustekniikkaa voidaan hyödyntää usealla eri tavalla. Lääketieteessä liikedataa käytetään esimerkiksi kuntoutuksessa liikeratojen tutkimiseen tai erilaisten urheilusuoritusten analysointiin. Tuotesuunnittelussa liikkeenkaappauksesta saatua tietoa voidaan hyödyntää esimerkiksi tuotteiden käytettävyyden ja ergonomian suunnittelussa.

Suuremmalle yleisölle tekniikka on tullut tutuksi viihde- ja peliteollisuuden kautta. Liikkeenkaappausta käytetään paljon erilaisen tehosteiden animaatioiden ja pelihahmojen liikkeiden toteuttamisessa. Tekniikan avulla animaatioiden tuotantoprosessi nopeutuu verrattuna perinteisiin animaatiomenetelmiin. Näyttelijästä tallennetulla liikeinformaatiolla on mahdollista päästä myös animaatioissa luonnollisemman näköiseen lopputulokseen.

Liikkeenkaappaus voidaan toteuttaa useammalla eri tekniikalla. Yleisimmät niistä perustuvat optisiin tai mekaanisiin järjestelmiin. Mekaanisissa järjestelmissä esiintyjä

pukeutuu erikoispukuun, johon kiinnitetyt sensorit tallentavat esiintyjän kehon eri osien liikeratoja ja kiihtyvyyksiä. Sensoreista saatu tieto välitetään tietokoneelle. Tietokoneella oleva ohjelmisto tulkitsee sensorien sijainnit ja niiden liikeradat. Tästä tiedosta voidaan toteuttaa animoitavan hahmon liikuttamisessa tarvittava luuranko.

Suuremmalle yleisölle tekniikka on tullut tutuksi viihde- ja peliteollisuuden kautta.

Mekaaninen järjestelmä ei käytä liikkeen tallentamiseen erillisiä kameroita, joten suoran näköyhteyden puuttuminen sensoreille ei aiheuta liikedatan tallentamiselle ongelmia. Mekaaninen liikkeenkaappausjärjestelmä ei ole riippuvainen tietystä paikasta ja ne ovat yleensä helposti liikuteltavia järjestelmiä.

Optinen liikkeenkaappausjärjestelmä perustuu erikoiskameroiden ja heijastavien merkkipisteiden eli markkereiden käyttämi-

seen. Merkkipisteet voivat olla passiivisia eli heijastavasta materiaalista toteutettuja tai aktiivisia LED – valoihin perustuvia. Passiivisessa järjestelmässä jokainen kamera heijastaa kaappausalueelle infrapunavaloa. Infrapunavalo heijastuu esiintyjään kiinnitetyistä merkkipisteistä takaisin kameroille. Järjestelmään kuuluvalla ohjelmistolla voidaan laskea jokaisen kameralle heijastuvan pisteen sijainti kolmiulotteisessa tilassa. Saadusta informaatiosta toteutetaan animoitavalle hahmolle luuranko, jonka avulla voidaan liikuttaa virtuaalista 3d hahmoa, joko reaaliaikaisesti tai yhdistää taltioitu liikedata 3d hahmoon myöhemmin.

Kameroiden määrää lisäämällä ja kaappausalueen kokoa kasvattamalla voidaan taltioida useiden esiintyjien liikedataa samanaikaisesti. Järjestelmää on mahdollista käyttää myös kamerasijainnin ja liikkeen taltiointiin. Tätä informaatiota voidaan hyödyntää virtuaalisen kamerasijoittamisessa 3d maailmaan.

Optinen liikkeenkaappausjärjestelmä voidaan toteuttaa myös ilman esiintyjään kiinnitettäviä heijastavia merkkipisteitä. Tällöin esiintyjän liike lasketaan ohjelmallisesti kahden tai useammasta kamerasijainnista. Tästä tunnetuimmat esimerkit ovat Microsoftin Kinect, sekä Organic Motionin ja Xsensin kehittämät järjestelmät.

Kymenlaakson ammattikorkeakoulun liikkeenkaappausjärjestelmä on NaturalPointin valmistama OptiTrack ja se perustuu opti-

seen järjestelmään. Järjestelmä koostuu kahdestatoista Flex 13 kamerasta, joka ovat kiinnitetty alumiiniseen trussikehikkoon. Kaappaustila on kooltaan 6 x 6 metriä. Korkeutta kaappaustilalla on 3 metriä. Tila mahdollistaa noin 3 x 3 metrin kokoisen optimaalisen alueen esiintyjän liikkeiden taltioimista varten. Järjestelmän jokainen kamera kuvaa 120 kuvaa sekunnissa, resoluution ollessa 1280 x 1024 pikseliä. Saatut tarkkuus riittää esiintyjän sormien liikkeiden tallentamisesta aina nopeiden taisteluliikkeiden taltiointiin. Käytössä oleva järjestelmä mahdollistaa kahden henkilön samanaikaisesti tapahtuvan liikedatan tallentamisen. OptiTrack liikkeenkaappausjärjestelmään sisältyy myös erillinen kuudesta kamerasta koostuva kasvoanimaatioiden tallentamisjärjestelmä.

Tietokoneella oleva ohjelmisto tulkitsee sensorien sijainnit ja niiden liikeradat.

Lisätty todellisuus aistiemme parantajana

Samuli Karevaara

Tieteellinen kehitys on useita kertoja historian saatossa parantanut ihmisen aisteja. Yksi ensimmäisistä suurenuslasin maininnoista on ajanlaskumme alusta, keisari Neron ajan Roomasta. Neron neuvonantaja, Seneca nuorempi, neuvoi keisaria katsomaan pieniä kirjaimia lasipallon tai vettä täynnä olevan juomalasin lävitse. Runsaat tuhat vuotta myöhemmin lasin-

tekotaito oli kehittynyt jo niin paljon, että 1200-luvun lopun Italiassa valmistettiin ensimmäiset nykyaikaiset silmälasit. Myöhemmin kykymme nähdä sekä kauas että lähelle on parantunut valtavasti, kiitos mikroskooppien ja teleskooppien.

Kuuloaistimme parantamiseen kehitettiin 1600-luvulla erityisiä ”korvatorvia”. Nämä pienivät aikanaan sen verran, ettei niitä tarvinnut enää erikseen pidellä kädellä. Puhelimen keksiminen joudutti äänentoiston



**Voimme
älylaitteiden avulla
muuntaa
esimerkiksi korvalle
kuulumattomia
äänialueita**

elektroniikan kehitystä, ja lopulta päädyttiin kömpelöistä korvatorvista on päästy nykyaikaisiin elektronisiin, korvansisäisiin kuulolaitteisiin.

Niin paljon kuin silmälasit ja kuulolaitteet ovatkin parantaneet nykyaikaisen ihmisen elämänlaatua, elämme valtavirrassa yhä aistien parantamisen kivikautta. Silmälasilla ei voi zoomata kaukaiseen liikennemerkkin teksteihin, laitteet eivät osaa muistuttaa minua siitä, mihin juuri laskin kädestäni kuusiokoloavaimen, minun pitää vilkuilla ajaessani erillistä navigaattorilaitetta ja niin edelleen. Laboratorioissa aistiemme parantamisessa on päästy jo kuitenkin paljon pidemmälle. Niin sanotut älylasit sisällyttävät elektroniikkansa ja äänentoistolaitteensa silmälasien sankoihin. Näyttönä toimivat luonnollisesti silmälasien linssit.

Lisätyssä todellisuudessa, englanniksi Augmented Reality, on kyse siitä, että näkemämme todellisen todellisuuden päälle tuotetaan digitaalisesti lisätietoa meitä ympäröivästä maailmasta. Yksinkertainen esimerkki tästä on vaikkapa keihäänheittokilpailuissa ruohokentälle piirtyvä raja, joka heittäjän pitää ylittää ollakseen mitalitaistelussa mukana. Tuo raja saattaa muuttua jokaisen heiton jälkeen, jolloin sitä ei kannata tuottaa nurmikkoon todellisenä. Riittää, että raja lisätään televisiolähetykseen. Stadionilakin rajan näkee, kun seuraa lähetystä isolta näyttöpäätteeltä.

Lisätyn todellisuuden avulla voisimme esimerkiksi lisätä kuulovammaisille tarkoitettun

tekstityksen aitoon tilanteeseen. Kuuro voi nähdä tekstin puhuvan henkilön pään alapuolella. Voimme siis rakentaa eräänlaisen kuulolaitteen jopa täysin kuuroille. Silmälasien sankoihin upotettava taas optiikka ylittää nykyään jo silmän erottelukyvyn. Täten voimme älylasilla tarkentaa ja suurentaa reaaliajassa kaikkea sitä, mihin laseillamme katselemme.

Voimme älylaitteiden avulla muuntaa esimerkiksi korvalle kuulumattomia äänialueita joko korvallemme soveltuvaan muotoon, tai sitten edellisen esimerkin mukaisesti esimerkiksi tekstittää ne: ”ultraääni-pillin vihellys kuuluu”. Näkyvän valon lisäksi voimme muuntaa vaikkapa ultravioletti- eli uv-säteilyn silmälle nähtävään aallonpituuteen. Tällöin voimme tarkkailla maailmaa esimerkiksi mehiläisen silmin: kukkasissa on kuvioita, jotka erottuvat vain uv-lampulla. Aistien melko suoraviivaisen parantamisen lisäksi voimme kohottaa ymmärryksemme tasoa lisätyn todellisuuden avulla. Opetustilanteessa voimme lisätä esimerkiksi ilmaan heitettyyn palloon siihen kohdistuvat voimat eri värisinä ja paksuisina nuolina. Voimme visualisoida myös lähestyvän auton ääneen liittyvän dopplerin ilmiön äänen aaltoliikkeenä.

Älylasien kehittyessä voimme jopa hidastaa liian nopeita tilanteet. Toistaiseksi tämä kuitenkin tarkoittaisi juuri sekunnin murto-osa sitten nauhoitetun lähetyksen katse-lua hidastettuna, todellisuuden tai reaktio-aikaamme kulkua emme kuitenkaan pysty hidastamaan. Auton alle jäämisen voimme kuitenkin estää siten, että lisätty todellisuus varoittaa meitä kuvan, äänen ja tuntoaistien avulla eli esimerkiksi varoitusnuolella, äänimerkillä ja värisevin silmälasin sangoin lähestyvistä autosta.

Augmentoitu todellisuus ja saavutettavuutta parantavat sovellukset

Lisättyä todellisuutta studioon vai kadulle?

Jori Pölkki

Suvi Pylvänen

Augmentoidun todellisuuden, puheen tunnistuksen ja tekstin tulkkausten yhdistelmä voi tukea huonokuuloisia, sokeita ja tehdä sovellusten käytön helpommaksi henkilöille, jolla on motoristisia vaikeuksia.

Kännykkäsovelluksissa AR teknologian avulla voidaan lisätä audiovisuaalista tietoa todellisuudesta. Esimerkiksi matkustaessa puhe voi ohjata käyttäjää suunnistamaan uudessa paikassa ja kertoa lisätietoa lokaatiosta ja nähtävyyksistä sanallisesti. Esimerkiksi suomalainen BlindSquare sovellus auttaa sokeita löytämään lähistöllä sijaitsevia kohteita kuten kauppoja ja ravintoloita ja italialainen MapAbility auttaa liikuntarajoitteisia reittien suunnittelussa.

Lasien kautta lisättyä todellisuutta muodostavat laitteet kuten Google-lasit puolestaan mahdollistavat tekniikan käytön kokoajan liikuttaessa – ilman tarvetta käyttää käsiä. Tulevaisuudessa esimerkiksi kuuro ja sokea voisivat keskustella keskenään ilman tulkkia: lasit tulkkavat kuurolle puheen tekstiksi ja viittoman sokealle puheeksi, vaikkakin vielä nykyisin viittomakielen tulkkausovellukset ovat kömpelöitä. Kasvojen ilmeiden tunnistusmekanismit voisivat auttaa myös autistisia henkilöitä identifioimalla ihmisten kasvojen ilmeiden perusteella heidän tunteitaan.

Tekniikan kehittyessä näyttöön tai laseihin lisätty puheen tunnistus, ääniopastus, tekstitys ja tulkkaukset sekä visuaalinen informaatio tarjoavat tulevaisuudessa mielenkiintoisia mahdollisuuksia moniin saavutettavuutta parantaviin sovelluksiin.

Kun aloitimme AURE-Studio hankkeen yhdessä LUTin kanssa suurin mielenkiintoni liittyi siihen saammeko toteutettua toimivan ns. "halpis"-version tv-työhön tarkoitettuista virtuaalstudioista, joita maailmalta saa kyllä ostaa, mutta vain paksuhkolla dollarinipulla sekä saammeko tuotua live-lähetykseen lisätyn todellisuuden elementtejä esim. pelimaailmasta hahmojen, rekvisiitan ja muiden tv-työn sisältöä rikastuttavien kerronnallisten elementtien muodossa.

Perinteiset virtuaalstudiot, kuten esim. Brain Storm-järjestelmä, perustuvat kamerakoordinaattien tarkkaan tallentamiseen infrapunasensorien tai laserin avulla sekä ohjelmaan, johon voidaan kalibroida speksit studiokameroissa käytettävistä objekteista, valovoimasta, niiden polttovälimuutoksista ja tarkennuksesta. Kun tähän dataan yhdistetään virtuaalilavasteet saadaan kameroiden

liikkeet, zoomaukset, tarkennusmuutokset ym. näkymään reaaliaikaisesti ja todentuntuisesti myös virtuaalilavastepinnoilla.

Avainnuksen (Keying) avulla tähän virtuaalitudellisuuteen tuodaan esim. juontaja ja muita elementtejä kuten huonekaluja tms. reaali maailmasta kuvaamalla nämä objektit yksiväristä chromasävyistä taustaa vasten, jonka tilalle virtuaalilavasteet avainnetaan reaaliaikaisesti tai jälkityönä editointivaiheessa.

Virtuaalstudioilla on paljonkin käyttöä tv:n ajankohtaisohjelmissa

Ammattikameroiden linssiteknologia on vaativaa ja sen muutoksiin reaaliaikaisessa tv-työssä reagoivalta järjestelmältä edellytetään hyvin tarkkaa kalibrointityötä.

Myös studiotyössä reaaliaikaiseen avainnukseen käytettävät laitteet ovat kalliita (ositain patenttisysteemit) ja lähinnä markkinoita hallitsevan Ultimatten tuottamia. Avainnukseen liittyy lisäksi studiovalojen tarkka hallinta ja rajoitettu käyttö, jotta virtuaalitausta saadaan avainnettua chroma-taustan tilalle ja henkilöt "irrotettua" taustasta.

Hankkeessa lähdettiin hakemaan liikkeen tallentamista optisin ratkaisuin kameroiden infrapuna- tai laser-träkkäyksen sijaan.

Näiden rajoitteiden takia virtuaalstudioita ei juurikaan käytetä fiktiivisessä kerronnassa, joka vaatii tavallisesti draamallisempaa valaistusta sekä näyttelijätyön uskottavuuden vuoksi konkreettisia tiloja sekä rekvisiittaa (action props), jonka kanssa toimia. Hahmojen kanssa on vaikea näyttellä.

Sitä vastoin virtuaalstudioilla on paljonkin käyttöä tv:n ajankohtaisohjelmissä, joiden staattisuutta voidaan elävöittää ja sisältöjä rikastuttaa lisätyn todellisuuden elementeillä.

Aure-Studio hankkeessa tätä reaaliaikaista lisättyä todellisuutta lähdettiin hakemaan halvan uuden teknologian kautta, joka voisi siten laajentaa virtuaalitudion käyttömah-

dollisuuksia esim. peliympäristöihin ja näyttelijän työn harjoituksiin, pois kalliista vuokrattavista studiotiloista.

Hankkeessa lähdettiin hakemaan liikkeen tallentamista optisin ratkaisuin kameroiden infrapuna- tai laser-träkkäyksen sijaan. Tietokonenäköä tutkinut Tommi Tykkälä LU-Tista vastasi teknisestä kehittäytystä väitöskirjansa puitteissa ja hän toi testattavaksi pelikonsoleiden mukana markkinoille saatuneen Kinect-sensorin, jonka myyntihinta markkinoilla on alle 100€.

Kinectia testattiin ensin Meduusa-Studioissa kameraan kiinnitettynä ajoradalla, jonka tulokset olivat yllättävänkin lupaavia. Kameraliikkeen taltiointi onnistui hyvin ja toiston myötä syntyneet erheet Tommi pystyi korjaamaan kalibroinnilla melko helposti. Suurimmat haasteet syntyivät studiovalaistuksesta, lavasteiden pintastruktuurista ja struktuurien edessä liikkuvista henkilöistä. Kinectin heppoisesta optiikasta ongelmat valaistuksen suhteen olivat myös korjattavissa, mutta kiintopisteiden poimiminen lavastepeinoista ja esineistä osoittautui hankalammaksi etenkin kun kiintopisteet

Suurimmat haasteet syntyivät studiovalaistuksesta, lavasteiden pintastruktuurista ja struktuurien edessä liikkuvista henkilöistä.



jäivät välillä edessä liikkuvien juontajien tai vieraiden taakse.

Varsinaiseen Meduusa-tv lähetykseen tuodun virtuaali-ninjan pystymme avantamaan vain yhdessä kamerassa, joka tietysti rajoitti hahmon esiintymistä kolmasosaan kamerakulmista ja mahdollisuudet testata liikkeen seuraamista jäivät myös tuon yhden kameran liikkeen varaan.

Myöhemmin studioissa toteutettiin vielä perinteisempään virtuaalitudiomaailmaan kuuluva testi, jossa yksittäinen henkilö upotettiin virtuaalimaailman hahmojen keskelle

Kinectia testattiin ensin Meduusa-Studioissa kameraan kiinnitettynä ajoradalla, jonka tulokset olivat yllättävänkin lupaavia.

ja kameran liikettä tallennettiin kuvioidun maton kiintopisteiden seurannalla. Tämä testi oli käyttöympäristöltään ja toteutustavaltaan lähinnä kaupallisten järjestelmien toteutusta ja osoitti halpaversionkin pystyvän tuottamaan uskottavan lisätyn todellisuuden maailman tv-työhön.

Positiivisista tuloksista huolimatta oli selvää, että Kinectillä toteutetun AR:n lisääminen tv:n päivittäistuotantoon edellyttää kuitenkin paljon tarkkaa kalibrointia ja käsityötä aina lähetyiskohtaisesti, joka taas syö halvan tekniikan tuomia etuja.

Kaiken kaikkiaan hanke oli positiivinen kokemus, joka tuotti hyviä oppimiskokemuksia ja tuloksia

Sitä vastoin on selvää, että Kinectillä tai vastaavilla optisilla sensoreilla voidaan tuoda lisäarvoa kuluttajien päivittäiseen elämään liittämällä tekniikkaa esim. google laseihin tms. Jolloin voisimme vaikka sisustussuunnittelun yhteydessä kävellä pitkin ostoharkinnassa olevaa asuntoa ja sijoitella sinne haluamiamme virtuaalisia huonekaluja testataksemme kuinka asunto ja huonekalut ovat yhteen sovitettavissa.

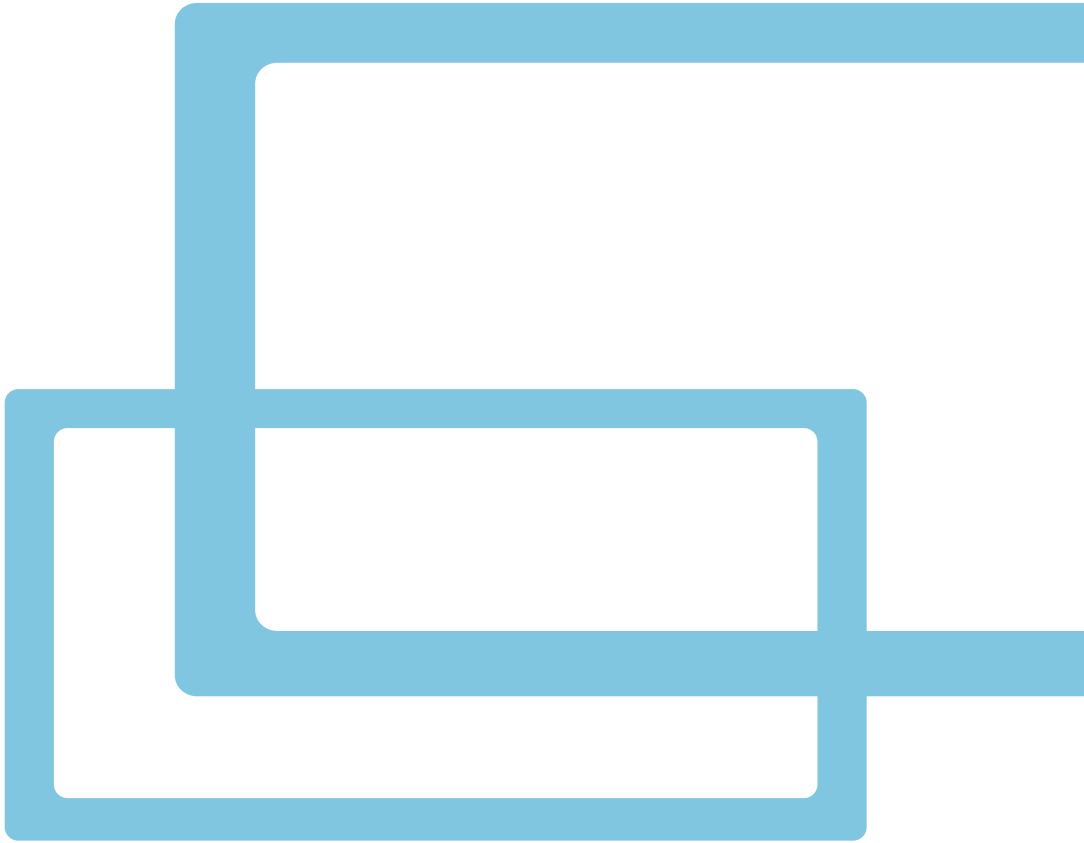
Tämän tyyppisiä kuluttajaratkaisuja AU-RE-Studiossa kehiteltyllä tekniikalla voidaan varmasti jatkossa toteuttaa, joko google-laseihin, verkkopalveluihin tai johonkin vielä

tuntemattomaan lähitulevaisuuden palveluun, joita Tommi Tykkälä onkin jo jatkokehittelemässä Nokian HERE-palvelun yhteydessä.

Kaiken kaikkiaan hanke oli positiivinen kokemus, joka tuotti hyviä oppimiskokemuksia ja tuloksia, joiden työstämistä mielellämme jatkaisimmekin LUTin ja muiden hankkeen yhteistyökumppanien kanssa.



AuRe **Studio**



Teknologiateollisuuden
100-VUOTISSÄÄTIÖ

Vipuvoimaa
EU:lta
2007-2013



KYAMK
University of Applied Sciences



LUT
Lappeenranta
University of Technology



Euroopan unioni
Euroopan aluekehitysrahasto