

Aalto-yliopisto  
Perustieteiden korkeakoulu  
Tietotekniikan koulutusohjelma

# **Eleentunnistus Kinect sensorilla**

**Kandidaatintyö**

**28. marraskuuta 2011**

**Liisa Saileranta**

<b>Tekijä:</b>	Liisa Sailaranta
<b>Työn nimi:</b>	L <sup>A</sup> T <sub>E</sub> X-pohja kandidaatintyötä varten ohjeiden kera ja varuilla keillaan vähän ylipitkää otsikkoa
<b>Päiväys:</b>	28. marraskuuta 2011
<b>Sivumäärä:</b>	Kirjoita tähän oikea määrä, tässä esimerkissä 23
<b>Pääaine:</b>	Tähän sinun pääaineesi nimi, kts main.tex
<b>Koodi:</b>	Txxxx tai Ilyyyy
<b>Vastuupettaja:</b>	Ma professori Tomi Janhunen
<b>Työn ohjaaja(t):</b>	Ohjaajantitteli Sinun Ohjaajasi (Poimi tähän ohjaajasi laitos, DEPT, main.tex)
<p>Tiivistelmä on muusta työstä täysin irrallinen teksti, joka kirjoitetaan tiivistelmälo- makkeelle vasta, kun koko työ on valmis. Se on suppea ja itsenäinen teksti, joka ku- vaa olennaisen opinnäytteen sisällöstä. Tavoitteena selvittää työn merkitys lukijalle ja antaa yleiskuva työstä. Tiivistelmä markkinoi työtäsi potentiaalisille lukijoille, sik- si tutkimusongelman ja tärkeimmät tulokset kannattaa kertoa selkeästi ja napakasti. Tiivistelmä kirjoitetaan hieman yleistajuisemmin kuin itse työ, koska teksti palvelee tiedonvälitystarkoituksessa laajaa yleisöä.</p> <p>Tiivistelmän rakenne: teksti jäsennetään kappaleisiin (3–5 kappaletta); ei väliotsikkoja; ei mitään työn ulkopuolelta; ei tekstiviitteitä tai lainauksia; vähän tai ei ollenkaan viittauksia työhön (ei ollenkaan: “luvussa 3” tms., mutta koko työhön voi viitata esim. sanalla “kandidaatintyössä”; ei kuvia ja taulukoita.</p> <p>Tiivistelmässä otetaan “löysät pois”: ei työn rakenteen esittelyä; ei itsestäänselvyys- kiä; ei turhaa toistoa; älä jätä lukijaa nälkäiseksi, eli kerro asiasisältö, älä vihjaa, että työssä kerrotaan se.</p> <p>Tiivistelmän tyypillinen rakenne: (1) aihe, tavoite ja raja- us (heti alkuun, selkeästi ja napakasti, ei johdattelua); (2) aineisto ja menetelmät (erittäin lyhyesti); (3) tulokset (tälle enemmän painoarvoa); (4) johtopäätökset (tälle enemmän painoarvoa).</p>	
<b>Avainsanat:</b>	avain, sanoja, niitäkin, tähän, vielä, useampi, vaikkei, niitä, niin, montaa, oikeasti, tarvitse
<b>Kieli:</b>	Suomi

# Sisältö

<b>Käytetyt symbolit ja lyhenteet</b>	<b>4</b>
<b>1 Johdanto</b>	<b>5</b>
1.1 Johdantoluku . . . . .	5
<b>2 Eletunnistus videokuvalta</b>	<b>5</b>
<b>3 ChaLearn Gesture Challenge -kilpailu</b>	<b>6</b>
3.1 Kilpailun esittely . . . . .	6
3.2 Katsaus kilpailutöihin . . . . .	7
<b>4 Loppuluku</b>	<b>8</b>
<b>Lähteet</b>	<b>9</b>
<b>A Esimerkkiliite</b>	<b>10</b>

## Käytetyt symbolit ja lyhenteet

3GPP	3rd Generation Partnership Project; Kolmannen sukupolven matkapuhelupalvelu
ESP	Encapsulating Security Payload; Yksi IPsec-tietoturvaprotokolla
$\Omega_i$	hilavitkuttimen kulmataajuus
$\mathbf{m}_{ic}$	hilavitkutinjärjestelmän $i$ painokertoimet

Tähän voidaan listata kaikki työssä käytetyt lyhenteet. Lyhenteistä annetaan selityksenä sekä alkukielinen termi kokonaisuudessaan (esim. englanninkielinen lyhenne avattuna sanoiksi) että sama suomeksi. Jos suoraa käännöstä ei ole tai sellaisesta on vaikea saada sujuvaa, voi käännöksen sijaan antaa selityksen siitä, mitä kyseinen käsite tarkoittaa. Jos lyhenteitä ei esiinny työssä paljon, ei tätä osiota tarvita ollenkaan. Yleensä luettelo tehdään, kun lyhenteitä on 10–20 tai enemmän. Vaikka lyhenteet annettaisiinkin tässä keskitetysti, ne pitää silti avata sekä suomeksi että alkukielellä myös itse tekstissä, kun ne esiintyvät siellä ensi kertaa. Käytetyt lyhenteet -osion voi nimetä myös “Käytetyt lyhenteet ja termit”, jos luettelossa on sekä lyhenteitä että muuta käsitteenmäärittelyä.

**TIK.kand suositus: Lisää lyhenne- tai symbolisivu, kun se näyttää luontevalta ja järkevältä. (Käytä vasta kun lyhenteitä yli 10.)**

# 1 Johdanto

## 1.1 Johdantoluku

Tämä kandidaatintyö käsittelee eletunnistusta Kinect syvyyskameran avulla. Työn tarkoitus on tutustua erilaisiin eletunnistusmenetelmiin ja tutkia niitä käytännönsovellusten kuten elekäyttöliittymien näkökulmasta. Tutkimuksessa kiinnitetäänkin huomioita erityisesti ”yhdestä otoksesta oppimiseen” ja transfer-oppimiseen eli perehdytään siihen miten eletunnistusmenetelmät oppivat tunnistamaan liikkeitä mahdollisimman pienellä määrällä annettua dataa. Tämä on kuluttujaystävällistä, loppukäyttäjät eivät usein halua treenata kameraa tuhansilla toistoilla.

One Shot Learning Gesture Recognition with Kinect Sensor, Di Wu, Fan Zhu, Ling Shao  
Kuvaa ja videokuvaa on tutkittu paljon mutta eletunnistus on vielä melko uusi haaste. Elentunnistuksen mahdollistavat uudet kehittyneet sensorit kuten Microsoftin Kinect-sensori ja kehittynyt laskentateho. Lisätäkseen alan kiinnostusta järjestettiin ChaLearn challenge eletunnistuskilpailu, joissa kilpailijat kehittivät Kinect sensorin datalle suunnattuja eletunnistus menetelmiä. Tämä kandidaatin työ tutustuu kilpailutöihin ja kartoittaa niiden avulla alan uusinta tutkimusta.

Aaron F. Bobick, Member, IEEE Computer Society, and James W. Davis, Member, IEEE Computer Society  
The Recognition of Human Movement Using Temporal Templates  
Kinect sensori tarjoaa ns. 3D -videokuvaa eli tavallisen värikuvan lisäksi Kinect -kamera antaa infrapunakameralla mitattua syvyyskuvaa.

## 2 Eletunnistus videokuvaalta

Eletunnistusvideokuvaalta koostuu seuraavista vaiheista. Kuvaa esikäsitellään, siitä poistetaan esim häiriötä tai Kinectin tapauksessa taustaa. Kuvasta valitaan piirteet joiden mukaan sitä luokitellaan. Piirrevalintaa voisi tarkoittaa vaikka sitä että jos haluamme luokitella ihmisiä valitsemme pituuden tai painon tai jonkin täysin muun tekijän kuten laulutaidon. Piirteitä voidaan valita useita jolloin meillä on moniulotteinen avaruus, jossa pisteitä käsitellään. Tällainen on hankala piirtää 2D -tasoon ja se tekee laskennasta usein raskasta mutta on välttämätön jotta voidaan erottaa useampia eri luokkia. Jostain tilanteissa piirteiden määrää voidaan vähentää sopivalla lineaarikuvauksella.

Videokuvan tapauksessa suoritetaan seuraavaksi ajallinen jako. Videokuva jaetaan siis pätkiin joiden sisällä ei ole liikettä. Tyypilliset videopätkät tiivistetään yksittäiseen kuvaan, hävittäen aikaulottovuus. Videokuvan ajallinen jako voidaan suorittaa etukäteen vertailemalla esimerkiksi vertailemalla peräkkäisten framejen samankaltaisuutta. Esimer-

kiksi videokuvassa jossa ihminen ensin istuu ja sitten nousee seisomaan voidaan erottaa ainakin kolme erillistä jaksoa: istuminen, nouseminen ja seisominen. Jako voi tapahtua myös tunnistuksen aikana.

Tämän jälkeen pyritään löytämään jokin yhteys piirteiden ja luokan välillä. Ongelmamme on luokitusongelma; tiedämme etukäteen mitä luokkia on olemassa. Opetusvaiheessa tiedämme luokan voimme johtaa jonkin funktion näytteen ja luokan välille. Tutkimus riippuu valitsemistamme piirteistä, joita voisivat olla esimerkiksi: onko kuvassa tapahtunut vaakasuuntaista liikettä? Onko kuvissa havaittavissa paljon pystysuoria viivoja? Tässä voidaan käyttää erilaisia luokittelualgoritmeja lähin naapuri esimerkiksi.

Luotuaamme funktion ja näytteen välille voimme siirtyä opetusdataan. Opetusdatassa tiedämme oikeat vastaukset, mutta suoritamme laskennat kuin emme tietäisi. Opetusdataa voidaan käyttää arvoimaan luokittimen kykyä arvioida muuta dataa kuin sitä millä se on luotu.

Ihminen on erittäin taitava eletunnistuksessa. Ihminen oppii uuden eleen yhdestä näytteestä. Ihminen tunnistaa liikkeen tai eleen vaikka kuva oli sumea eikä yksittäisestä framesta erottaisi hahmoa. Eletunnistuksessa ollaan kiinnostuneita ihmisen kyvystä tunnistaa eleitä ja pyritään matkimaan ihmisen eletunnistus näköä. Kuvista ei esimerkiksi pyritään erottamaan hahmoa vaan ihmisen havaitsemiskyvyn tavoin tutkitaan pelkkää liikettä.

## 3 ChaLearn Gesture Challenge -kilpailu

### 3.1 Kilpailun esittely

ChaLearn Gesture Challenge kilpailun tarkoituksena oli lisätä kiinnostusta eletunnistukseen syvyyskameroilla. Kilpailussa jaettiin 50 000 Kinect-sensorilla kuvattua videonäytettä. Videonäytteet sisälsivät esimerkiksi viittomia tai poliisin käsimerkkejä. Kilpailijoiden tarkoitus oli kehittää eletunnistusmenetelmä jonka avulla eleet opitaan yhdestä opetusnäytteestä. Eleitä oli jaettu kategorioihin käyttötilanteen mukaan esimerkiksi poliisin käsimerkit olivat yksi kategoria. Eleet joilla kilpailutoita testattiin olivat eri eleitä kuin opetusdatassa mutta samoista kategorioista.

Videoklipeillä esiintyi aina yksi ihminen kerrallaan suorittamassa tiettyä elesarjaa. Kuva rajattiin yläruumiiseen ja eleet tehtiin pääasiallisesti käsillä. Liikkeet lopetettiin ja

aloitettiin aina samasta lepoasennosta. Videonäytteet sisälsivät syvyyskamerakuvan sekä värikuvan mutta ei ranganseurausta. Haasteita toivat vaihtelevat taustat ja valaistukset videoilla.

Kilpailijoille jaettiin kolme datajoukko: kehitysdata, validointidata ja lopullinen arviointidata. Dataa oli luokiteltu kategorioihin sen mukaan oliko kyseessä esimerkiksi viittoma, luonnollinen ele vai tanssiasento. Opetusdatassa näytteille annettiin laabelit kuvaamaan oikeaa ratkaisua. Validointi ja arviointi datassa vain osalla oli oikeat laabelit. Kilpailun loppupuolella paljastettiin lopullinen testidata jonka avulla tuloksia arvioitiin

The Recognition of Human Movement Using Temporal Templates 00910878.pdf

## 3.2 Katsaus kilpailutöihin

Kilpailijoiden metodeja selvitettiin lyhyellä kyselyllä, johon vastasi 20 ryhmää 22 parhaan ryhmän joukosta. Lähes kaikki tiimit tekivät jonkinlaista kuvien esikäsittelyä. Kuvista poistettiin häiriötä, asiaanliittymättömiä kohteita tai tausta. Huomioitavaa on kuitenkin, että jotkin hyvin menestyneistä ryhmistä eivät tehneet minkäänlaista esikäsittelyä kuville.

Suurin osa osallistujista käytti HOG/HOF - piirteitä, SIFT/STIP piirteitä, kulmien tai nurkkien tunnistusta tai kehitti omia tälle datalle soveltuvia ominaisuuksia. Suurin osa käytti pelkkää syvyyskuvaa, osa molempia, sekä väri että syvyyskuvaa. Toisen sijan voittaja käytti pelkkää RGB-videokuvaa. Tyypillisesti kilpailijat käyttivät ominaisuuksien valintaa, tiivistystä tai kuvausta toiseen lineaariavaruuteen.

Videokuva jaetaan ajallisesti osioihin perustuen videokuvan samankaltaisuuteen. Tunnistuksessa voidaan käyttää erilaisia graafisia malleja kuten Hidden Markov Model ja Conditional Random Fields. Itse luokittelu tapahtui KNN tai muilla yksinkertaisilla keinoilla. Kilpailun järjestäjien odottamaa metodia Transfer learning käytettiin, tosin kukaan menestyneistä kilpailijoista ei käyttänyt sitä.

Kilpailutöissä huomattavaa oli että ne käyttivät hyvin erilaista lähestymistapaa. Ryhmä xiaozhuwudi käytti työssään MHI-tekniikkaa eli tutkittiin viimeaikaisen liikkeen määrää videokuvassa eri ajan hetkinä.

Ryhmä zunga valitsi piirteiksi horisontaalisen ja vertikaalisen liikkeen sekä kuvan ulkoasun. Tämä jälkeen yritetään löytää regressio eli funktio joka kuvaa piirteet oikeaan luokkaan. Oikea funktio haetaan minimoimalla virheen neliötä.

Ryhmä immortals lähti liikkeelle ajatuksesta että liikkeet koostuvat sarjaista yksittäisiä eleitä. Immortals käytti valmista databasea useista liikkeista ja koosti liikkeistä malleja joita se käytti tunnistuksessa.

Ensimmäinen ryhmä näytti keskittyvän piirrevalintaan, toinen ryhmä regressioon ja kolmas ryhmä videokuvan tunnistuspuoleen.

## 4 Loppuluku



# Lähteet

Sirkka Hirsjärvi, Pirkko Remes ja Paula Sajavaara. *Tutki ja kirjoita*. Tammi, Hämeenlinna, 2009. ISBN 978-951-31-4836-2. 15. uudistettu painos.

IPDC Forum. About IP Datacasting - Overview, 2004. URL <http://www.ipdc-forum.org/about/index.html>. IPDC Forumin WWW-sivu. Viitattu 18.2.2004.

Salli Kankaanpää, Elina Heikkilä, Riitta Korhonen, Sari Maamies ja Aino Piehl, toimittajat. *Kielitoimiston oikeinkirjoitusopas*. Kotimaisten kielten tutkimuskeskus, 2010. 8. painos.

Ilkka Kauranen, Mikko A. Mustakallio ja Virpi Palmgren. *Tutkimusraportin kirjoittamisen opas opinnäytetyön tekijöille*. Teknillinen korkeakoulu, Espoo, 2006. ISBN 951-22-8359-X (nid.). Lisäpainokset: 2. korj. p. 2007.

Simo K. Kivelä. LaTeX-kurssi. Saatavissa <http://matta.hut.fi/matta2/latex/index.html>. Viitattu 28.1.2011.

Frank Mittelbach, Michel Goossens ja Johannes Braams. *The LaTeX companion*. Addison-Wesley, Boston, toinen painos, 2004. ISBN 0-201-36299-6 (nid.). Lisäpainokset: Repr. 2006.

Tobias Oetiker. Pitkänpuoleinen johdanto LaTeX2e:n käyttöön. Saatavissa <ftp://www.ctan.org/ctan/info/lshort/finnish/lyhyt2e.pdf>. Viitattu 28.1.2011.

Mikael Puolakka. *BiBTeX-tyylin tklt käyttöohje*. Helsingin yliopiston tietojenkäsittelytieteen laitos, 2002.

Teemu Teekkari. Diplomityöni. Diplomityö, Aalto-yliopiston perustieteiden korkeakoulu, Espoo, 2010. Saatavissa <http://www.cis.hut.fi/teemu/katkotaan/rivia/jotta/helpompi/katkaista/sopivasta/kohtaa/tamakin/litania/poikki/rtfm.html>. Viitattu 25.1.2011.

## A Esimerkkiliite

Jos työhön kuuluu suurikokoisia (yli puoli sivua) kuvia, taulukoita tai karttoja tms., jotka eivät kokonsa puolesta sovi tekstin joukkoon, ne laitetaan liitteisiin. Liitteet numeroidaan. Jokaiseen liitteeseen tulee viitata tekstissä, eikä liitteisiin ole tarkoitus laittaa “mitä tahansa”, vaan vain työlle oikeasti tarpeellista materiaalia. Liitteisiin voidaan sijoittaa esim. malli kyselylomakkeesta, jolla tutkimushaastattelu toteutettiin, pohjapiirustuksia, taulukoita, kaavioita, kuvia tms.

**TIK.kand suositus: Vältä liitteitä.** Jos iso kuva, mieti onko sen koko pienettävissä (täytyy olla tulkittavissa) normaalin tekstin yhteyteen. Joskus liitteeksi lisätään matemaattisen kaavan tarkempi johtaminen, haastattelurunko, kyselypohja, ylimääräisiä kuvia, lyhyitä ohjelmakodeja tai datatiedostoja.

Työtä varten mahdollisesti tehtyjä ohjelmakodeja ei tyypillisesti lisätä tänne, ellei siihen ole joku erityinen syy. (Kukaan ei ala kirjoittaa tai tarkistamaan koko koodia paperilta vaan pyytää sitä sinulta, jos on kiinnostunut.)