



**UNIVERSITY
OF OULU**

TIETO- JA SÄHKÖTEKNIIKAN TIEDEKUNTA

**Emil Dark
Essi Passoja
Ella Juvonen**

**ETÄLÄSNÄOLOROBOTIN SUUN JA PÄÄN
LIIKKEET VIDEOANALYYSIN PERUSTEELLA**

Kandidaatintyö
Tietotekniikan tutkinto-ohjelma
Elokuu 2021

Dark E., Passoja E., Juvonen E. (2021) Etäläsnäolorobotin suun ja pään liikkeet videoanalyysin perusteella. Oulun yliopisto, Tietotekniikan tutkinto-ohjelma, kandidaatintyö, 45 s.

TIIVISTELMÄ

Etäläsnäolon ja etätapaamisen hyödyntämisen kasvaessa tarve erilaisille etäläsnäolotekniikoille kasvaa. Yksi tapa luoda etäläsnäolosta läsnäolevampaa on käyttää etäläsnäolorobottia, joka mallintaa muualla sijaitsevan ihmisen liikkeitä.

Tämä kandidaatintyö tehtiin osana isompaa ryhmää, jonka tavoitteena on tutkia ja kehittää uusia ominaisuuksia robottipään etäläsnäolon ominaisuuksien parantamiseksi käyttämällä robottien käyttöjärjestelmää (ROS) ja InMoov-robotti-alustaa. Tämä tutkielma keskittyy kuvailemaan metodeja ihmisen suun ja pään liikkeiden seuraamiseen puheen aikana ja mallintamaan liikkeet etäläsnäolorobottiin luomalla näin mahdollisimman luonnollisen kanssakäymisen robotin ja ihmisen välille etätapaamisen aikana.

Työssä toteutetun etäläsnäoloratkaisun ihmisenkaltaisuuden onnistumista pohdittiin kvantitatiivisen tutkimuksen avulla. Tutkimuksessa tehtiin kysely, jonka avulla selvitettiin ihmisten tuntemuksia robottia katsellessa verrattuna oikeaan ihmiseen. Kyselyssä haluttiin selvittää myös robottipään aiheuttamia mahdollisia Outo laakso -ilmiön aiheuttamia tuntemuksia ihmisissä eli mahdollisia epämukavuuden tunteita, jotka robotti aiheuttaa ulkonäöllään ja liikkeillään. Kyselyyn vastanneiden mukaan pään liike muistutti hieman kuvattun ihmisen liikettä, mutta suun liike ei onnistunut vastaamaan ihmisen suun liikkeitä kovinkaan hyvin suurimman osan kyselyyn vastanneiden mielestä. Lisäksi suurin osa kyselyyn vastanneista koki robottipään häiritsevänä kommunikaation kannalta.

AVAINSANAT

Etäläsnäolo, InMoov, Liikkeen seuranta, OpenCV, Videon prosessointi

Dark E., Passoja E., Juvonen E. (2021) The Mouth and Head Movements of a Telepresence Robot Based on Video Analysis. University of Oulu, Degree Programme in Computer Science and Engineering, 45 p.

ABSTRACT

As the amount of telepresence and telepresence meetings are increasing, there is a growing demand for different and more developed telepresence technologies. One way to make telepresence more present is to use a telepresence robot, which mirrors the movements of the human located elsewhere. This bachelor's thesis was made as a part of a bigger group studying and developing new features to enhance telepresence capabilities of a robot head, using Robot Operating System (ROS) and InMoov -robot platform. The thesis focuses on describing methods to capture user mouth and head movement while speaking and duplicating it on the telepresence robot creating more natural interaction between a robot and a human.

The level of success of the project was considered using a quantitative research method. A questionnaire was used to determine the feelings and emotions the participants encountered while watching videos of the robot and human next to each other. One of the questionnaire's aims was also to find out, whether the participants experience the discomfort related to Uncanny valley -phenomenon, that the robot causes with its looks and movements. According to most of the participants in the questionnaire, the movement of the robot head was rather accurate compared to the human movement. However, the mouth movement did not work too well. In addition, most of the respondents felt the robot head was at least fairly disturbing.

KEYWORDS

Telepresence, Telepresence robot, InMoov, Movement tracking, OpenCV, Video processing

SISÄLLYSLUETTELO

TIIVISTELMÄ	
ABSTRACT	
SISÄLLYSLUETTELO	
ALKULAUSE	
LYHENTEIDEN JA MERKKIEN SELITYKSET	
1. JOHDANTO	8
2. ETÄLÄSNÄOLO	9
2.1. Etäläsnäoloteknologiat	9
2.1.1. Videokonferenssipuhelut ja videopuhelut	9
2.1.2. Virtuaalitodellisuus	10
2.1.3. Telerobotiikka	12
2.1.4. Kasvojen mittaamisteknologiat	13
2.2. Ihmisen ja robotin välinen kanssakäyminen	15
2.2.1. Outo laakso -ilmiö	15
2.2.2. Kasvojen ja pään liikkeiden merkitys	16
2.3. Konenäkö ja koneoppiminen kasvojen tunnistuksessa	17
2.3.1. Kasvojen tunnistus koneoppimisen avulla	17
2.3.2. Kasvojen lokalisointi ja piirteiden tunnistus	18
2.3.3. Kasvojen kiintopistekoordinaatisto	19
2.3.4. Asento ja liikkeet koordinaatistoissa	20
2.3.5. Koordinaatisto	21
2.4. Robotin ohjaaminen	22
2.4.1. ROS	22
2.4.2. InMoov	23
3. JÄRJESTELMÄN TOTEUTUS	24
3.1. Videon prosessointi	25
3.1.1. Kameran kalibrointi	26
3.1.2. Videopuhelusovellusten integrointi	26
3.2. Datan käsittely	27
3.3. Datamalli	28
3.3.1. Mittaaminen ja normalisointi	28
3.4. Robotti	30
3.4.1. Servot ja U2D2	30
3.4.2. Servojen liikeavaruus	30
3.4.3. ROS 2 action client	32
3.4.4. Robotin kehittäminen ja säätö projektin aikana	33
3.5. Robotin liikuttaminen	35
4. TESTAUS	36
4.1. Testijärjestelyt	36
4.1.1. Testikysymykset	36
4.2. Testitulokset	37
4.3. Testitulosten pohdinta	38
5. POHDINTA	39

6. YHTEENVETO.....	41
7. AJANKÄYTTÖ	42
8. LÄHDELUETTELO.....	43

ALKULAUSE

Tämä kandidaatintyö on tehty Oulun yliopiston tieto- ja sähkötekniikan tiedekunnassa osana sulautettujen ohjelmistojen projekti -kurssia. Kurssiprojekti tehtiin ryhmätyönä ja projektin vastuut jaettiin ryhmän jäsenten kesken.

Oulussa 16. elokuuta 2021

Emil Dark
Essi Passoja
Ella Juvonen

LYHENTEIDEN JA MERKKIEN SELITYKSET

<i>AI</i>	Tekoäly (Artificial Intelligence)
<i>API</i>	Ohjelmointirajapinta (Application programming interface)
<i>CG</i>	Tietokonegrafiikka (Computer graphics)
<i>DLT</i>	Suora lineaarimuunnos (Direct Linear Transform)
<i>IP</i>	Internet-protokolla (Internet Protocol)
<i>MPU</i>	Monipisteprosessointiyksikkö (Multipoint Processing Unit)
<i>PnP</i>	Perspektiivi-n-piste (Perspective-n-Point)
<i>ROI</i>	Robot Operating Interface
<i>ROS</i>	Robot Operating System
<i>TCP</i>	Siirto kontrolliprotokolla (Transmission Control Protocol)
<i>UDP</i>	Käyttäjän datagrammi -protokolla (User Datagram Protocol)
<i>VR</i>	Virtuaalitodellisuus (Virtual Reality)

1. JOHDANTO

Maailmanlaajuisen pandemian myötä ihmiset ympäri maailmaa ovat joutuneet rajoittamaan liikkumistaan ja välttämään toistensa seuraa. Koronaviruksen leviämisen rajoittamiseksi yrityksille ja oppilaitoksille annettujen suositusten¹ myötä etäläsnäoloteknologian käyttämisestä on tullut aiempaa laajempi ilmiö. Tilastokeskuksen työvoimatutkimuksen mukaan säännöllinen etätöskentely kaksinkertaistui vuoden 2020 aikana [1] ja etäopiskelu on tullut tutuksi pienillekin koululaisille.

Etäläsnäolossa on turvavälien pitämisen lisäksi muitakin hyviä puolia. Kotona pysyminen tuo väljyyttä päivän aikatauluun, koska aikaa ei kulu matkustamiseen ja työ- tai opiskelupäivän taukoja voi hyödyntää esimerkiksi lemmikin ulkoiluttamiseen tai päivittäisten askareiden hoitamiseen. Työmatkan on todettu vaikuttavan negatiivisesti yksilön onnellisuuteen, sillä se aiheuttaa ylimääräistä stressiä ja kuluja, sekä vähentää aikaa perheen ja harrastusten parissa [2][3]. Lisäksi työmatkaamisen vähentyessä liikenteen aiheuttamat saasteet pienenevät, jolloin ympäristö hyötyy myös etäläsnäolosta.

CoCoDigi-tutkimusryhmän pitkittäistutkimuksessa [4] suomalaisten siirtymisestä etätööhön Covid-19 -kriisin aikana vastanneista 86 prosenttia oli lokakuussa 2020 tyytyväisiä etätööhön. Ihmisellä on kuitenkin luontainen tarve sosiaaliseen kanssakäymiseen, jonka määrä vaihtelee yksilöittäin. Sosiaalista eristäytymistä tunti marraskuussa 60 prosenttia CoCoDigin tutkimukseen vastanneista ja 72 prosenttia työntekijöistä koki olevansa erossa työkavereistaan. Tarve saada etätöskentelystä läsnäolevampaa innoittaa kehittämään tätä tukevia teknologioita, kuten telerobotiikkaa.

Telerobotiikalla pyritään luomaan aidomman läsnäolon tunne robotin avulla, joka jäljittelee etäosallistujan puhetta, ilmeitä ja liikkeitä. Näin etäosallistuja voidaan tuoda fyysisesti paikalle, toisin kuin esimerkiksi videopuhelun avulla. Vaikka etäläsnäolorobotit ovat olleet osa yhteiskuntaa jo miltei vuosikymmenen ajan, ovat maailmanlaajuiset mullistukset muokanneet työtä ja suunnanneet asenteita uudelleen kohti etäosallistumista. Teknologia voisi vihdoinkin olla tarpeeksi kypsää laajentuakseen kapeasta käyttäjäkunnastaan kohti valtavirtaa. Tässä opinnäytetyössä pyritään viemään etäläsnäoloa pidemmälle telerobotiikan avulla siirtämällä henkilön suun ja pään liikkeet videokuvan välityksellä robotin leuan ja pään liikkeeksi.

¹<https://valtioneuvosto.fi/-/10616/hallitus-teki-periaatepaatokset-maskisuosituksesta-ja-etatyosta>

2. ETÄLÄSNÄOLO

Etäläsnäololla viitataan joukkoon teknologioita, joiden avulla henkilö voi tuntea olevansa läsnä tai antaa läsnäolon vaikutelman muille, vaikka henkilön fyysinen keho sijaitsisikin toisessa paikassa [5]. Etäläsnäolo edellyttää tavallisesti, että käyttäjien aisteille tarjotaan sellaisia ärsykeitä, jotka saavat henkilön tuntemaan itsensä olevansa muualla. Lisäksi käyttäjille voidaan antaa mahdollisuus vaikuttaa etäsijaintiin esimerkiksi robotiikan avulla. Tässä tapauksessa käyttäjän sijainti, liikkeet, toiminnot, ääni ja muut ominaisuudet voidaan havaita, välittää ja kopioida etäpaikalle etäläsnäolovaikutuksen aikaansaamiseksi, jolloin tieto voi kulkea molempiin suuntiin käyttäjän ja etäsijainnin välillä.

Tässä työssä etäläsnäolotapaamisella tarkoitetaan tapaamista, jossa on mukana etäosallistuj(i)a, sekä yksi tai useampi paikallaolija. Etäläsnäolotapaamisista pyritään tekemään mahdollisimman läsnäolevia, jotta osallistujien kokemuksesta saataisiin tehtyä yhtä luontevia ja kanssakäymisen kannalta tehokkaita kuin läsnäolotapaamisista. Ihmisen kasvojen liikkeet eli ilmeet ovat useiden eri lihasten yhtäaikaista liikettä [6], eli vakuuttavan ja luontevan robotin aikaansaamiseksi pitäisi kyetä kopioimaan henkilön pienimmätkin liikkeet, jotta robotti voi peilata tarpeeksi vakuuttavasti ihmisen ilmeitä ja puhetta.

2.1. Etäläsnäoloteknologiat

Etäläsnäoloteknologioita on useita erilaisia, joista vanhimpana on videokonferenssipuhelu, jossa yhdistetään puhelussa ääni ja video. Muita potentiaalisia etäläsnäoloteknologioita ovat virtuaalitodellisuus-, telerobotiikka- ja hologrammiteknologiat.

2.1.1. Videokonferenssipuhelut ja videopuhelut

Yksi yleisimpiä etäläsnäolon hyödyntämistapoja ovat videoneuvottelupuhelut. Videoelementin lisääminen tuo puheluun uuden ulottuvuuden, kun osapuolet pystyvät myös näkemään toisensa. Teknologian kehityksen myötä mobiiliverkon ja -laitteiden käyttö on laajentanut videoneuvotteluiden pitomahdollisuuksia kokoushuoneen ulkopuolisiksi etätapaamisiksi kaikkien osallistujien käyttäessä henkilökohtaisia laitteitaan haluamassaan sijainnissa mobiiliverkon sisällä.

Ensimmäiset videopuhelun kehittämissyritykset aloitettiin virallisesti Yhdysvalloissa 1920-luvulla, ja 1930-luvulla videokonferensseja testattiin Saksassa [7]. Videopuhelut olivat pitkään todella kalliita ja niiden yleistymisen vei aikaa. 1990-luvulla tehtiin ensimmäinen IP-pohjainen (Internet-protokolla) videokonferenssijärjestelmä ja saataville tuli myös tehokkaampia videon kompressoititeknologioita, joiden avulla pystyttiin rakentamaan PC-perustainen videokonferenssipuhelu ja jolloin kehitettiin ensimmäinen videopuheluohjelma CU-SeeMe [8]. Videopuheluiden yleistymisen yksityis- ja yrityskäytössä alkoi kuitenkin vasta 2000-luvun alkupuolella, kun teknologian kehittyessä kaistanleveyttä saatiin kasvatettua suuremmaksi ja viivettä

pienennettyä. Ensimmäisiä hyvin tunnettuja videopuhelusovelluksia olivat Skype (2003) ja FaceTime (2010) [9].

Videokonferenssien tekniset toimintaperiaatteet ovat sama kaikilla laitteilla. Kamerasta ja mikrofonista saatu analoginen signaali muutetaan digitaaliseksi signaalin pakkaamiseen ja purkamiseen perustuvalla koodekki -menetelmällä. Pakkaamisen avulla signaali on nopeampi lähettää puhelinyhteydellä. Jos kyseessä on monipistevideokonferenssi (eng. Multipoint videoconferencing), jossa on useampi käyttäjä konferenssissa yhtäaikaaisesti, täytyy toteuttaa myös MPU eli monipisteprosessointiyksikkö (eng. Multipoint Processing Unit) tai vaihtoehtoisesti käyttää ryhmälähetys -vaihtoehtoa, jossa jokainen käyttäjä linkitetään kaikkiin vastaanottajiin, eli lähetys lähetetään kaikille vastaanottajille erikseen [9].

Videokonferenssien haasteena pidetään puheenvuorojen jakamista, joka saattaa aiheuttaa käyttäjille kiusallisia tilanteita. Erityisesti vuoroosuuntaisessa yhteydessä, eli yhteydessä, jossa vain yksi käyttäjä voi lähettää ääntä kerrallaan, pidetään keskustelua ongelmallisena, koska puheenvuoron saaminen voi olla vaikeaa kun linja varataan aina nopeimmalle. Kaksisuuntaisessa yhteydessä kaikki käyttäjät voivat puhua yhtä aikaa, mutta käyttäjät kokevat hankalaksi viiveen aiheuttamat tilanteet, jolloin saatetaan alkaa vahingossa puhumaan toisten käyttäjien puheen päälle ennen kuin toisen puhe ehditään vastaanottaa. Normaalisti tämä pystytään välttämään fyysisessä kommunikoinnissa antamalla kehonkielellä vihje toiselle osapuolelle puheenvuoron haluamisen merkiksi, tavallisesti kättä nostamalla tai henkeä ottamalla juuri ennen puhumisen aloittamista [10].

Toinen haaste konferenssipuheluissa on videon laatuun vaikuttava kuvataajuus. Verkkoliikenteen kehittyessä kaistanleveyttä on saatu kasvatettua sen verran, että videokonferenssissa saavutetaan television kuvataajuus 30 fps (eng. frames per second), eli vastaanotetaan 30 kuvaa sekunnissa. Videokonferensseissa joudutaan usein kuitenkin uhraamaan videon laatu, jotta ääni saadaan varmasti kuuluville, koska ilman sitä konferenssi menettää merkityksensä [9].

2.1.2. Virtuaalitodellisuus

Virtuaalitodellisuudella eli VR:llä (eng. Virtual Reality) viitataan ihmisen ja tietokoneen vuorovaikutuksen väliseen rajapintaan, jossa käyttäjät voivat liikkua tietokoneen generoimassa ympäristössä [11]. VR:ää käytetään paljon ihmisten kouluttamiseen eri tilanteita varten vaarantamatta heidän tai muiden turvallisuutta opetustilanteessa. Kuva 1 esittää NASAn omistamassa virtuaalitodellisuuslaboratoriossa henkilöstön koulutusta. Koulutuksen tarkoituksena on tuottaa osallistujille massan ja inertian luoma kinesteettinen tunne raskasta esinettä siirrettäessä.



Kuva 1. NASAn virtuaalitodellisuuslaboratorio (Virtual Reality Laboratory, VRL) harjoituskäytössä.²

VR:ää käytettäessä etäosallistujat voivat luoda itselleen avatarin eli itseään enemmän tai vähemmän kuvaavan hahmon, joka tuodaan virtuaaliseen tilaan. VRChat³ on yksi sovellus, jossa osallistujat voivat socialisoitua virtuaalitodellisuudessa tai pelata muiden pelaajien kanssa. Virtuaalitodellisuutta voitaisiin käyttää samalla tavoin myös opiskelu- ja työkäytössä, jolloin etäkokouksia voitaisiin järjestää kolmiulotteisena.

Virtuaalitodellisuudella on pitkälti samoja ominaisuuksia kuin telerobotiikalla, sillä siinä ihmisen eleet ja ilmeet, sekä puhe mallinnetaan toisaalle ohjelmiston avulla. VR:ssä ei kuitenkaan tarvita robottia, vaan riittää, että tietokoneella luodaan oma todellisuus johon henkilöt sijoitetaan. Robotin ja VR:n kehittäminen eroavat sillä, etteivät robotin fyysiset ominaisuudet tuo rajoitteita kehitysohjelmaan, mutta virtuaalitodellisuuden rakentaminen ja ylläpitäminen vaatii usein enemmän prosessointikapasiteettia tietokoneelta, sekä suurempaa suorituskykyä ja leveämpää kaistanleveyttä palvelimelta telerobotiikkaan verrattuna.

Kasvojen liikkeiden siirtämiseksi virtuaalitodellisuuteen tulee kasvojen liikkeitä mitata. Nykyaikaiset VR-lasit sisältävät teknologian kasvojen mittaamiseen⁴ ja ne tarjoavat hyvän alustan myös telerobotiikalle mittausten tekemiseen. Kameran ollessa jatkuvasti tietyssä sijainnissa suhteessa kasvoihin, pystytään saavuttamaan korkeampi tarkkuus kuin muilla teknologioilla ja pystytään ohittamaan pään liikkuttelun aiheuttamat haasteet mitattaessa.

²Kuvan omistaja: NASA, https://www.nasa.gov/centers/johnson/engineering/robotics_simulation/virtual_reality/index.html

³<https://store.steampowered.com/app/438100/VRChat/>

⁴<https://www.vive.com/us/accessory/facial-tracker/>

2.1.3. Telerobotiikka

Telerobotiikkaa käytetään tilanteissa, joissa ihmisen meno tarvittavalle paikalle on vaarallista tai robotin avustuksella toiminnasta saadaan tarkempaa kuin ihmisen tekemänä. Telerobotiikkaa hyödyntäviä tilanteita ovat esimerkiksi syvänmeren tai avaruuden tutkiminen, robotti-avusteiset leikkaukset ja ydinvoimaan liittyvät operaatiot. Lisäksi ihmisten työn helpottamiseksi on kehitetty esimerkiksi varastoihin ja rakennustyömaille erilaisia telerobotteja [12][13].

Lääketieteessä käytetään robotteja esimerkiksi leikkauksissa, jolloin kirurgi voi suorittaa leikkauksen toisesta lokaatiosta. Yksi tällainen laite on Da Vinci SS (Da Vinci Surgical Systems), joka on kehittyneimpiä leikkaukseen käytettäviä telerobotteja. Sitä käytettäessä voidaan nähdä tarkemmin ja suoraan operoitavaan paikkaan kameroiden avulla, eli myös kehon sisälle. Kuvassa 2 kirurgi istuu Da Vinci -laitteen ohjausyksikössä ja itse laite esitetään kuvassa oikealla.⁵



Kuva 2. Kirurgi istuu Da Vinci SS -laitteen ohjausyksikössä ja itse laite esitetään kuvassa oikealla.⁵

Kirurgisissa telerobotiikka -laitteissa stabiilin ja nopean yhteyden saavuttaminen on ollut suurin haaste [14]. Tilanteet joissa kirurgille saapuva kuva tulee viiveellä tai kirurgin tekemät liikkeet tapahtuvat eri tavalla kuin niiden on tarkoitus, useimmiten johtavat pidempiin leikkaus- ja toipumisaikoihin. Manoharan et al. ehdottavat kuitenkin tutkimuksessaan itsekorjautuvaa ja automaatiota hyödyntävää leikkauksilaitteistoa, joilla suurimpia kirurgista telerobotiikkaa vaivaavista ongelmista voitaisiin pienentää [14].

Meressä tehtäviä toimenpiteitä ja tutkimusta voidaan myös tehdä erilaisilla teleroboteilla. Yksi esimerkki on Wevolverin kehittämä robotti, joka voi sukeltaa jopa 4 kilometrin syvyyteen ohjaajan ollessa turvallisesti veden pinnalla. Wevolverin

⁵Kuvan omistaja Gustave Roussy, Vimeo: <https://vimeo.com/163681073>, CC 2.0

kehittämä robotti on niin pieni, että sillä voidaan tutkia ja korjata esimerkiksi meren pohjalla olevia putkia tai telekommunikaatiojohdoissa ilmeneviä ongelmia [15].

2.1.4. Kasvojen mittaamisteknologiat

Ensimmäisiä huomattavia teknologisia hyppyjä kasvojen ilmeiden siirtämisessä objekteihin teki Rick Lazzarini 1990-luvulla. Hän oli tuolloin kuuluisa animatroniikan ja nukkejen hallitsemisasiantuntija (eng. puppetry), joka työskenteli erikoistehosteteollisuudessa. Lazzarinilla oli tuolloin ajatus siirtää tietämyksensä, kekseliäisyytensä ja kokemuksensa animatronisista etäjärjestelmistä käytäntöön [16].

Lazzarini loi ensimmäisen version Waldo-animatroniikkajärjestelmästä, jonka avulla yksi henkilö pystyy hallitsemaan kaikkia nukan monimutkaisia liikkeitä ollessaan silti poissa näkyvistä. Kuvassa 3 esiteltävää Waldo-järjestelmä kypärää ohjataan suun liikkeillä käyttämällä leukaan kiinnitettyjä mekaanisia ja sähkömagneettisia antureita.



Kuva 3. Waldo -järjestelmän kypärä. ⁶

Lazzarini jatkoi Waldo-järjestelmän kehittämistä ja saavutti hyvän reaktioajan ja liikkeen toistokyvyn, joita hän demonstroi myöhemmin julkaistussa videossa, jossa hän käyttää teknologiaa animatronisen jääkarhun päässä 'Polar' -brändin maskotissa 'Pedroso':ssa. ⁷

Lazzarinin inspiroimana muutkin erikoistehosteteollisuuden yhtiöt näkivät kuinka paljon eloa nukkeen kyettiin tuomaan siirtämällä ihmisen ilmeet nukkerobottiin. Samaan aikaan tietokoneiden rasterointitekniikan ja laskentakyvyn parantuessa realistiset visuaaliset tehosteet olivat enemmän saavutettavissa kuin koskaan.

2000-luvun vaihteessa Weta digital -yhtiö Peter Jacksonin johdolla lähti luomaan täysin tietokonegrafiikkapohjaista (CG) hahmoa Lord of the Rings -elokuvaan. Klonkkua esitti Andy Serkis, jonka työnä oli tuoda hahmo eloon esityksen aikana. Serkisin kasvoille merkityt pisteet toimivat referenssinä animoijille Klonkun

⁶Kuvan luonut The Character Shop Inc., <http://www.character-shop.com/waldo.html>

⁷<https://www.youtube.com/watch?v=bFW2azvVEdI>, 18.6.2021

ilmeiden luomiseksi. Kuvassa 4 havainnollistetaan miten näyttelijän kasvoista voidaan jälkikäteen mallintaa animaatiohahmo.



Kuva 4. Vasemmalla Andy Serkis esiintymässä ja oikealla esityksestä luotu Klonkku-hahmo.⁸

Weta jatkoi yhtiönä menestystään Lord of the Rings -trilogian jälkeen jatkamalla teknologiansa kehittämistä, kun Weta aloitti työskentelyn James Cameronin Avatar-elokuvan tuottamiseksi. Tietokoneen prosessointivoima oli kehittynyt ja mahdollisti reaaliaikaisen kasvojen seurannan. Weta toi mukanaan projektiin myös kehittämänsä kasvojen kartoitus- ja mitoituusalgoritmit, joiden avulla ohjelmat kykenivät kartoittamaan tietokoneella generoidun hahmon piirteet näyttelijän kasvojen liikkeen mukaisesti [17], kuten Kuvassa 5 on esitetty.



Kuva 5. Kasvojen reaaliaikainen kartoitus Avatar-elokuvan kuvauksissa.⁹

Viimeisen 14 vuoden aikana mobiililaitteiden yleistyminen ja niiden komponenttien kehittyminen on johtanut samojen komponenttien käyttämiseen myös nykypäivän kasvojenmittausteknologiassa. VR-lasien olemassaolon on mahdollistanut puhelinten näyttöjen ja gyroskooppien kehittyminen nykyiseen tasoonsa, ja HTC Vive:en saa ostettua jopa suun seurantalaitteen, jonka toimiminen perustuu puhelimen kameraan ja konenäköön. Erilaisia kasvojen seurantalaitteita on esitetty Kuvassa 6.

⁸Kuvan omistaa Weta workshop, <https://gotoricks.blogspot.com/2011/08/virtual-world-of-andy-serkis.html>

⁹Kuva omistaa Media Magik Entertainment, Youtube: https://www.youtube.com/watch?v=P2_vB7zx5Q

Näiden uusien mittausmenetelmien tultua kaikkien ulottuville ja konenäköalgoritmien kehittyttyä osittain sovellusten kasvosuodattimien ansiosta voidaan kasvojen kiintopisteet merkitä kasvoille virtuaalisesti konenäön avulla sen sijaan, että ne merkittäisiin maalaamalla.



Kuva 6. Vasemmalla HTC Vive suun seurantalaitteella. Oikealla käytetään hyväksi Iphone X:ää ja ARKit-alustaa.¹⁰

2.2. Ihmisen ja robotin välinen kanssakäyminen

Ihmisen kaltaisten robottien kehittäminen on jatkuvaa, mutta kehitystyössä ei ole olemassa yhteistä linjausta siitä, kuinka paljon robottien toivotaan olevan täysin ihmisen näköisiä. Outo laakso -ilmiö kuvaa ihmisten mieltymystä robottia kohtaan kanssakäymisen aikana suhteessa siihen, kuinka paljon robotti muistuttaa ihmistä.

2.2.1. Outo laakso -ilmiö

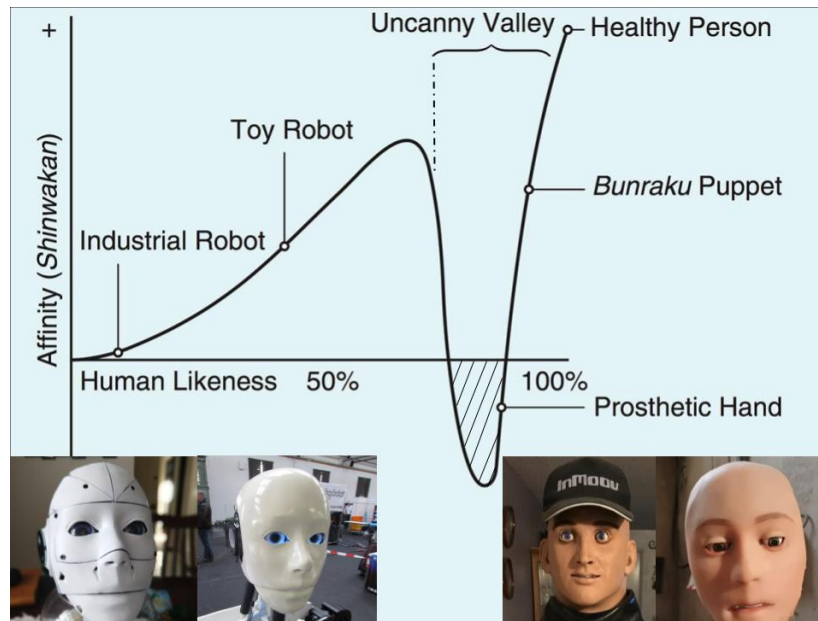
Jotta läsnäolevan ihmisen ja robotin välinen kanssakäyminen olisi luontevaa, tulee robotin olla vastaanottajalle, eli läsnäolevalle ihmiselle, mahdollisimman mieluisa katsoa ja kuunnella. Japanilainen robotiikan kehittäjä ja professori Masahiro Mori kirjoitti esseen jossa hän hypotesoi kuinka ihmisten reaktioita robotteja tai muita keinotekoisia ihmisten kasvoja kohtaan voidaan kuvata Outo laakso -ilmiön (engl. Uncanny valley) avulla [18][19].

Outo laakso -ilmiö tarkoittaa negatiivista reaktiota keinotekoisiihin ihmiskasvoihin, kun kasvot alkavat muistuttaa liikaa ihmiskasvoja. Kuvassa 7 esitetään robotin ihmisenkaltaisuuden määrän vaikutusta katsojan reaktioon. Kun humanoidirobotin ihmisenkaltaisuutta lisätään, alkaa se aluksi näyttää tutummalta ja miellyttävämmältä, mutta realismin kasvaessa riittävästi siirrytään outo laaksoon, eli koettu tutuus laskee jyrkästi. Samalla syntyy negatiivisia tuntemuksia ja outouden tunnetta. Esimerkkejä tällaisesta ovat esimerkiksi vahanuket, robotit tai animaatiofilmien ihmishahmot. Outo laakso -ilmiön vaikutuksista päästäisiin kuitenkin eroon siinä vaiheessa, kun robotti näyttää täysin samalta kuin elävä ihminen.

¹⁰Kuvan luonut Emil Dark. Kuvien omistajat vasemmalta oikealle:

6.1) Frooxius, Youtube kanava, <https://www.youtube.com/watch?v=ukKHVYwzZRA>.

6.2) Ryan Mayeda, Unreal engine blog: <https://www.unrealengine.com/en-US/blog/new-live-link-face-ios-app-now-available-for-real-time-facial-capture-with-unreal-engine>.



Kuva 7. Outo laakso -ilmiön vaikutus robotin lähestyessä ihmisen näköisyyttä ¹¹

2.2.2. Kasvojen ja pään liikkeiden merkitys

Pään ja kasvojen liikkeet vaikuttavat siihen, miten keskustelun toinen osapuoli vastaanottaa välitetyn viestin [6][20]. Eri ilmeillä ja pään liikkeillä voidaan osoittaa tunteita ja vaikuttaa siihen, onko tarkoituksena esimerkiksi osoittaa vihaisuutta vai onko viestin tarkoituksena olla hauska.

Ihmisten välisestä kommunikoinnista suuri osa on sanatonta, eli ilmeillä ja muulla kehonkielellä tehtävää kommunikointia [21]. Jotta viesti välittyy halutulla tavalla, ei siis riitä että pelkästään ääni saadaan kuuluville, vaan tarvitaan myös muita elementtejä. Kasvojen eri lihaksilla on eri vaikutuksia siihen, miten ele otetaan vastaan. Ilmeet ovat yhdistelmä eri elementtejä ja näiden elementtien ristiriidoilla voidaan hämmentää viestin vastaanottajaa. Esimerkiksi suupielet alaspäin käännettynä viittavat surulliseen henkilöön, mutta jos silmät ovat auki yhtä aikaa, kuten iloisella ihmisellä, on viesti hyvin ristiriitainen.

Ihmisen kasvoilla voidaan nähdä yli 10 000 eri ilmettä [22], joiden toteuttamiseksi tarvitaan paljon eri lihaksia. Pelkästään suun lihaksistossa merkityksellisimpiä ovat alahuulen alavetäjälihak (musculus depressor inferioris labii), suupielen alavetäjälihak (musculus depressor anguli oris), suupielen kohottajalihas (musculus levator anguli oris), hymylihas (musculus risorius) ja leukalihas (musculus mentalis).

¹¹Kuvan luonut Emil Dark. Kuvasta löytyvät robotit vasemmalta oikealle:

7.1) JC Brenes-Torres, <https://onecuriousrobot.blogspot.com/2018/11/humanoid-robot-head-based-on-inmoov.html>

7.2) rentaprinta, https://synthiam.com/uploads/user/1268392C58F0B8FAA776BFC4850EAEB7/IMG_0169-635813948838764348.jpg

7.3) merne, https://synthiam.com/uploads/user/2152039D2FE8A784D2B3E9E0909C9488/IMG_0020-636137991394125621.jpg

7.4) Mechatrons, Youtube-kanava, <https://www.youtube.com/watch?v=tJIr6Okr554>

Kaikilla näillä lihaksilla on oleellisia tehtäviä ilmeisiin suun osalta. Esimerkiksi alahuulen alavetäjälihas nimensä mukaisesti vetää alahuulia sekä alas että sivuille ja hymylihas supistuessaan vetää suupielen sivulle.

2.3. Konenäkö ja koneoppiminen kasvojen tunnistuksessa

Hyödynnettäessä telerobottiikkaa etäläsnaoloteknologiana tarvitaan erilaisia teknologioita ja ratkaisuja, joiden avulla robotti kykenee aistimaan mitä oikeassa maailmassa tapahtuu.

Konenäkö prosessoi raakadatan käskyiksi koneopetetun mallin avulla jolla se tunnistaa henkilön videosta. Tällä tavalla saadaan eroteltua videosta ensin kaikki hyödyllinen data. Sen jälkeen kasvojen eri kohdat asetellaan koordinaatistoon, jolloin voidaan laskea sopivat käskyt robotin ohjaamiseksi.

2.3.1. Kasvojen tunnistus koneoppimisen avulla

Etäläsnaolevan henkilön kasvojen liikkeiden kopioimiseksi tarvitaan dataa mallinnettavan henkilön kasvoista. Tietokone voidaan opettaa tunnistamaan henkilön kasvot esimerkiksi videolta koneoppimisen avulla. Koneoppimista pidetään yhtenä tekoälyn osa-alueena, joka on lisännyt suosiotaan massadatan ja prosessointitehon saatavuuden lisääntyessä.

Koneoppimisessa käytetään kehittyviä tietokonealgoritmeja kategorisoimaan dataa tunnistamalla siitä malleja ja kaavoja [23]. Kun videokuvaa kaapataan näytöltä, voidaan kirjoittaa ohjelmistokoodi kasvojen tunnistamiseksi esimerkiksi Python-kirjastojen avulla. Jos kaappaus sen sijaan tapahtuu kameroiden kautta, kutsutaan prosessia konenäköksi.

Koneoppiminen voidaan jakaa oppimistavan mukaan kahteen kategoriaan: ohjattuun ja ohjaamattomaan oppimiseen. Niiden ero on ulostulon ennustettavuudessa. Ryvästyksessä (clustering), joka on ohjaamatonta oppimista, datasta pyritään löytämään ennalta tuntemattomia kategoriota, joihin se jaetaan. Ohjattuun oppimiseen kuuluvassa luokittelussa sen sijaan on ennalta määritellyt kategoriat, joihin data jaetaan. Regressiossa ulostulona on diskreettien luokkien sijaan jatkuvia reaalilukuja.

Koneoppimisen työkaluihin kuuluu regressiopuu [24] eli päätöspuu, jonka ennusteet ovat jatkuvia reaalilukuja. Binääripuu on tietorakenne, joka koostuu solmuista ja suunnatuista kaarista, jotka osoittavat solmusta sen lapsisolmuun. Jokaisella solmulla voi olla enintään kaksi lapsisolmua, ja lapsettomia solmuja kutsutaan lehdiksi. Päätöspuu on binäärinen hakupuu, jonka jokaiseen sisäsolmuun liittyy testi ja jokaiseen lehteen ennuste. Sisäsolmun testi määrää, kumpaan lapsihaaraan siirrytään. Regressiopuiden avulla etsitään jatkuva-arvoista ennustetta joko jatkuvien tai epäjatkuvien tekijöiden suhteen.

Regressiopuu rakennetaan aloittaen juurisolmusta. Data jaetaan kahteen joukkoon etsimällä ennustamiseen käytetyn ominaisuuden raja-arvo, jolle datapisteiden neliövirheiden summa on pienin. Vasen lapsisolmu edustaa pienempiarvoisia datapisteitä ja oikea suurempiarvoisia. Näin syntyvien osajoukkojen jakamista jatketaan, kunnes täyttyy toinen seuraavista ehdoista: joukon kaikkien pisteiden arvo

on sama, tai joukon koko on pienempi kuin ennalta määrätty minimiarvo. Tällöin kyseisestä solmusta tulee lehti. Kun sisäsolmuja ei ole enää jäljellä, puu on valmis. Jos ennustamiseen käytettyjä ominaisuuksia on useampia kuin yksi, neliövirheiden summa lasketaan jokaiselle erikseen, ja juureksi valitaan se testi, joka tuotti pienimmän virheen.

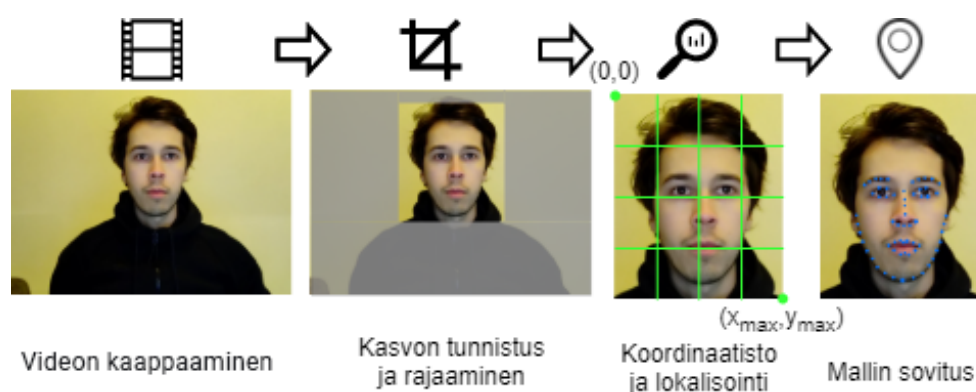
2.3.2. Kasvojen lokalisointi ja piirteiden tunnistus

Kasvot voidaan tunnistaa hakemalla kasvojen kiintopisteet. Niiden etäisyyksien vaihteluita mitataan, ja saatujen mittaustulosten avulla pystytään mallintamaan robotilla ihmisen liikkeitä ja ilmeitä. Kasvojen kiintopisteiden tunnistaminen voidaan kategorisoida muodonennustusongelman osajoukoksi. Muodonennustaja-algoritmi yrittää lokalisoida tärkeimpiä kiinnostuksen kohteita kasvoilla ja tunnistaa kasvojen tärkeät rakenteet käyttäen gradienttimenetelmää (eng. gradient descent) [25].

Kasvojen kiintopisteiden tunnistamiseen on useita erilaisia algoritmeja, mutta kaikkia menetelmiä yhdistää motivaatio yrittää paikantaa ja merkitä yleisimmät kiinnostuksen kohteet kasvoilla, kuten suu, silmät, kulmakarvat, nenä ja leuka [26][27]. Nopea ja tehokas tapa paikantaa kiintopisteet on käyttää regressiopuuta ratkaistaessa gradienttimenetelmää.

Kazemi V. & Sullivan J. [27] kuvaavat mallin, jonka kouluttamiseen tarkoitettu harjoitusdata sisältää kuvia käsin merkityistä kasvojen kiintopisteistä. Tämän jälkeen mallia jatkokoulutetaan kasvojen kiintopisteiden sijaan datalla, jossa merkitään (x,y)-koordinaateilla alueet, jotka ympäröivät jokaisen halutun kasvojen rakenteen gradienttitehostamismenetelmällä. Mallin opetusvaiheessa prioreina käytetään todennäköistä etäisyyttä syötteen pikseliparin välillä.

Kuvan 8 kasvojen kiintopisteiden tunnistaminen on siis kaksivaiheinen prosessi, jossa pyritään toteuttamaan mahdollisimman tarkka kasvojen lokalisointi kuvasta, minkä jälkeen halutaan tunnistaa kasvojen tärkeimmät rakenteet. Kuvan rajaaminen pelkkiin kasvoihin pienentää pinta-alaa, josta kasvojen kiintopisteitä haetaan.



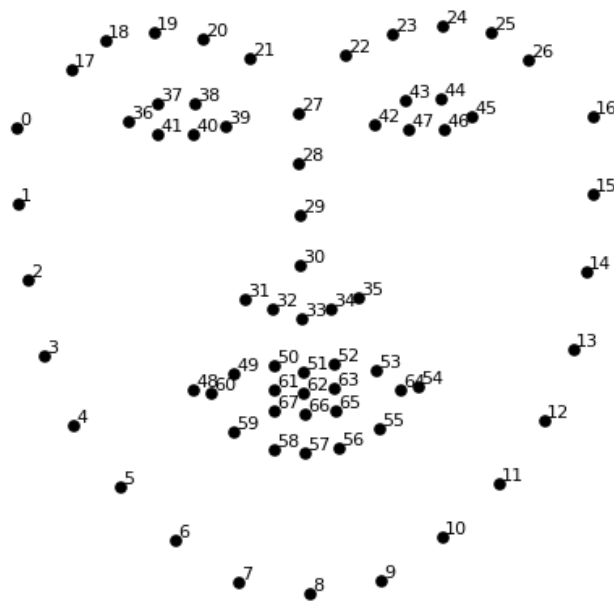
Kuva 8. Konenäköprosessi¹²

¹²Kuvan luonut Emil Dark. Kuvalle asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

Tämän harjoitusdatan perusteella joukko regressiopuita koulutetaan arvioimaan kasvojen kiintopisteiden sijainnit suoraan pikseleiden voimakkuuksista, jolloin kasvojen rakenteiden tunnistusvaihe pystytään ohittamaan kokonaan. Lopputuloksena saadaan kasvojen kiintopistetunnistin, jota voidaan käyttää kasvojen kiintopisteiden havaitsemiseen reaaliajassa korkealla tarkkuudella.

2.3.3. Kasvojen kiintopistekoordinaatisto

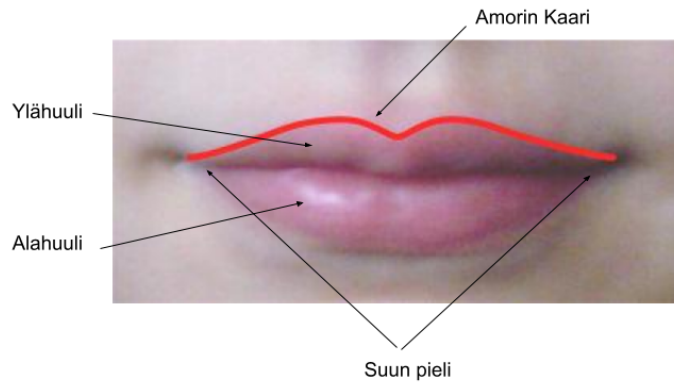
Ihmisen kasvoille voidaan asettaa kiintopisteitä Kuvan 9 mukaisesti. Kasvot voidaan kuvata koordinaattipistematriisissa asettamalla kasvoille 68 kiintopistettä. Kasvojen liikkussa koordinaattipisteiden välimatkat muuttuvat, jolloin niiden avulla voidaan kuvata kasvoilla esiintyviä ilmeitä matemaattisesti.



Kuva 9. Kasvojen kiintopistematriisi¹³

Suhteuttamalla mitattavat koordinaatit toisiinsa luodaan abstraktio koordinaatistosta. Laskemalla suhteutettujen koordinaattien etäisyys toisistaan saadaan yksinkertainen liikevektori, jonka arvo muuttuu kasvon eri osien liikkussa suhteessa toisiinsa.

¹³This image was created by Brandon Amos of CMU who works on OpenFace



Kuva 10. Suun anatomia¹⁴

Kuvan 10 suun anatomian mukaisesti suun liikehdintää eri ilmeissä voidaan laskea myös yksityiskohtaisemmin, mittaamalla huulten etäisyyden lisäksi esimerkiksi suupieltä sijaintia suhteessa suun neutraaliin asentoon. Tällöin voidaan havainnoida jo tarkemmin puhujan tunnetiloja, kuten surua tai iloa, jolloin suupielet laskeutuvat tai nousevat suhteessa neutraaliin asentoon. Virnistäessä toinen suupieli on korkeammalla kuin toinen ja ala- ja ylähuuli ovat lähempänä toisiaan.

2.3.4. Asento ja liikkeet koordinaatioissa

Konenäössä kohteen asento viittaa sen suhteelliseen orientaatioon ja sijaintiin kameraan nähden. Asentoa voi muuttaa siirtämällä kohdetta kameraan nähden tai kameraa kohteeseen nähden.

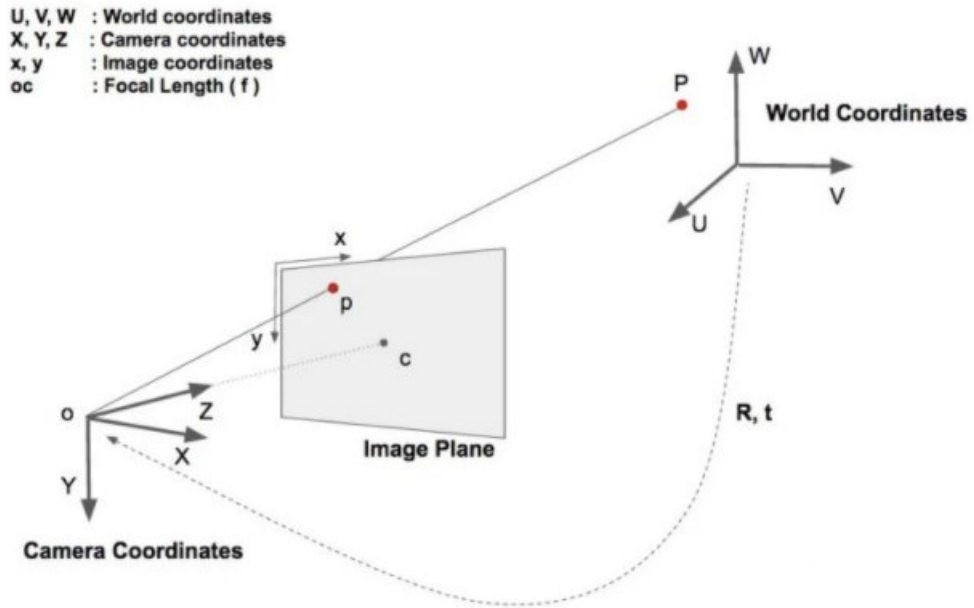
Kolmiulotteinen jäykkä objekti voi ilmentää vain kahdenlaisia liikkeitä kameraan nähden, translaatiota ja rotaatiota. Translaatio esittää kameran siirtymistä sen alkuperäisestä koordinaatista (x, y, z) uuteen sijaintiin (x', y', z') . Translaatiota kuvataan vektorilla t , joka on yhtäsuuri kuin alkuperäisen pisteen ja siirretyn pisteen erotus (1).

$$t = (X' - X, Y' - Y, Z' - Z) \quad (1)$$

Rotaatioliikkeen tapahtuessa kamera pyörii akselin X, Y tai Z ympärillä. Liikettä kuvaillaan eulerin kulmien avulla (yaw, pitch, roll), jolla on translaation tavoin kolme vapausastetta.

Kolmiulotteisen kohteen sijainnin arviointia 2D-projektion (esim. kuvan tai videon) avulla luonnehditaan konenäköammattikirjallisuudessa "Perspective-n-Point"-ongelmana [28]. PnP -ongelmassa halutaan mitata kahden pisteen $[p, P]$ välistä etäisyyttä. Pisteet esitetään Kuvassa 11.

¹⁴Kuvan luonut Emil Dark. Kuvalla asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)



Kuva 11. Kohteen asennon arviointi 2D-projektiosta¹⁵

2.3.5. Koordinaatisto

Asennon arviointi kameran kautta lasketaan käyttämällä arvoja kolmesta eri koordinaatistosta, kuten kuvassa 11 esitetään. Koordinaatistot ovat 'World'-, eli Maailma -koordinaatisto, kameran koordinaatisto ja kuvan koordinaatisto.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R \begin{bmatrix} U \\ V \\ W \end{bmatrix} + t \Rightarrow \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [R|t] \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \quad (2)$$

Otetaan laajennettu muoto kaavasta 2, missä rotaatiomatriisi ja translaatiovektori muuttuvat.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \quad (3)$$

Yhtälö (3) pystytään ratkaisemaan käyttämällä algebran suoraa lineaarimuunnosta (eng. Direct Linear Transformation, DLT). DLT ei kuitenkaan ole optimaalisin ratkaisu johtuen tavasta, jolla DLT ratkaisee matriisin, koska siitä ei löydy pakottavaa funktiota, joka pitää ratkaisun rotaatiomatriisin oikean kokoisena (3x3). Tavan kykenemättömyys minimoida uudelleenheijastusvirhettä tekee siitä vähemmän

¹⁵Kuvan luonut ProgrammerSought, <https://www.programmersought.com/article/52424662140/>

optimaalisen laskea kohteen asento. Menetelmää voidaan parantaa ottamalla huomioon, etteivät kohteen 3D-pisteet linjaudu täydellisesti kuvakoordinaatiston pisteiden kanssa. Tämä havainnollistetaan Kuvassa 11 vektorin $[p, P]$ avulla. Ottamalla 3D- ja 2D-pisteiden välisen vektorin pituuden neliöjuurien summa, pystytään laskemaan uudelleenheijastusvirheen määrä ja ottamaan se huomioon sovituksessa. [29] [30]

Useita nopeampia ja tarkempia versioita on julkaistu Gaon (2003) jälkeen, mukaan lukien Lambda Twist PnP, joka on tällä hetkellä yksi nopeimmista ratkaisuista [31]. Lambda Twist PnP tarjoaa 50-kertaisen lisäyksen nopeuteen ja samalla vähentää numeerista epäonnistumista 400-kertaisesti verrattuna Gaon PnP -menetelmään. Lambda Twist on saatavana avoimen lähdekoodin muodossa OpenMVG:stä¹⁶ ja on hyvä ohjelma jatkokehitystä mietittäessä.

2.4. Robotin ohjaaminen

Koordinaatistosta saatua dataa käytetään InMoov-robotin ohjaamiseen 'Robot Operating System':in, eli ROS:in avulla.

2.4.1. ROS

ROS on kokoelma eri työkaluja ja kirjastoja, jonka tarkoituksena on helpottaa ja yksinkertaistaa robotille kirjoitettavaa koodia [32]. ROS:ia käytetään enimmäkseen koulutuksellisiin ja esitystarkoituksiin, sillä siitä puuttuu muutamia tärkeitä ominaisuuksia [33][34].

Teollisuuden tarpeiden vastaamiseksi kehitettiin ROS 2, joka täyttää teollisuuden vaatimukset turvallisuudelle ja oikea-aikaiselle toiminnalle. ROS tukee pelkästään TCP -protokollaa, kun ROS 2 saa joustavuutta datanjakopalvelun liikenteestä. ROS 2:lla on käytössään UDP -protokolla, jolla toiminnasta saadaan luotettavampaa [35]. Lisäksi ROS 2 tarjoaa paljon ominaisuuksia, joita ei ole lainkaan alkuperäisessä ROS:ssa.

ROS (1&2) kirjastojen kehitys on viime aikoina keskittynyt tukemaan C++:aa ja Pythonia. Roscpp on C++ -asiakaskirjasto ROSille, joka on laajimmin käytössä oleva ROS kirjasto ja jonka etuna on sen korkea suorituskyky. Rospypy on ROS:lle tehty Python-asiakaskirjasto, jonka avulla ROS:lle saadaan objektorientoitunut kieli käyttöön. Se on kuitenkin suunniteltu lyhentämään ohjelmiston suunnittelu-aikaa, eikä sillä ole yhtä korkea suorituskyky kuin roscpp:llä. ROS 2:lla on eri nimiset asiakaskirjastot, rclcpp (C++) ja rclpy (Python). Näiden lisäksi tarjolla on käyttäjien tekemiä kirjastoja, jotka on jaettu julkisesti muidenkin käyttöön.

ROS 2:n asiakaskirjastoissa etuliite rcl tulee sanoista "ROS Client Library", jota käytetään kaikkien asiakaskirjastojen pohjana välittämättä ohjelmointikielestä. Näin saadaan nimettyä kaikki kirjastot johdonmukaisesti ja jokainen kirjasto käyttäytyy samalla tavoin. Tätä ei tapahdu ROS:ssa, koska siinä kaikki kirjastot tehdään alusta asti, eikä rcl:n päälle. ROS:ssa alusta asti kirjastojen tekeminen mahdollistaa

¹⁶<https://github.com/openMVG/openMVG>

kuitenkin vain yhden ohjelmointikielen käyttämisen, jolloin välttyään esimerkiksi koodin kokoamiselta (eng. compiling). Tämän vuoksi virheiden korjaaminen täytyy tehdä erikseen kaikkiin kirjastoihin ja kirjastoissa on paljon korjauksia, jotka on tehty vain kyseiseen kirjastoon. ROS 2:n asiakaskirjastojen vaatiessa korjausta voidaan usein tehdä korjaus pelkästään rcl:ään, jolloin säästetään aikaa ja vaivaa.

2.4.2. InMoov

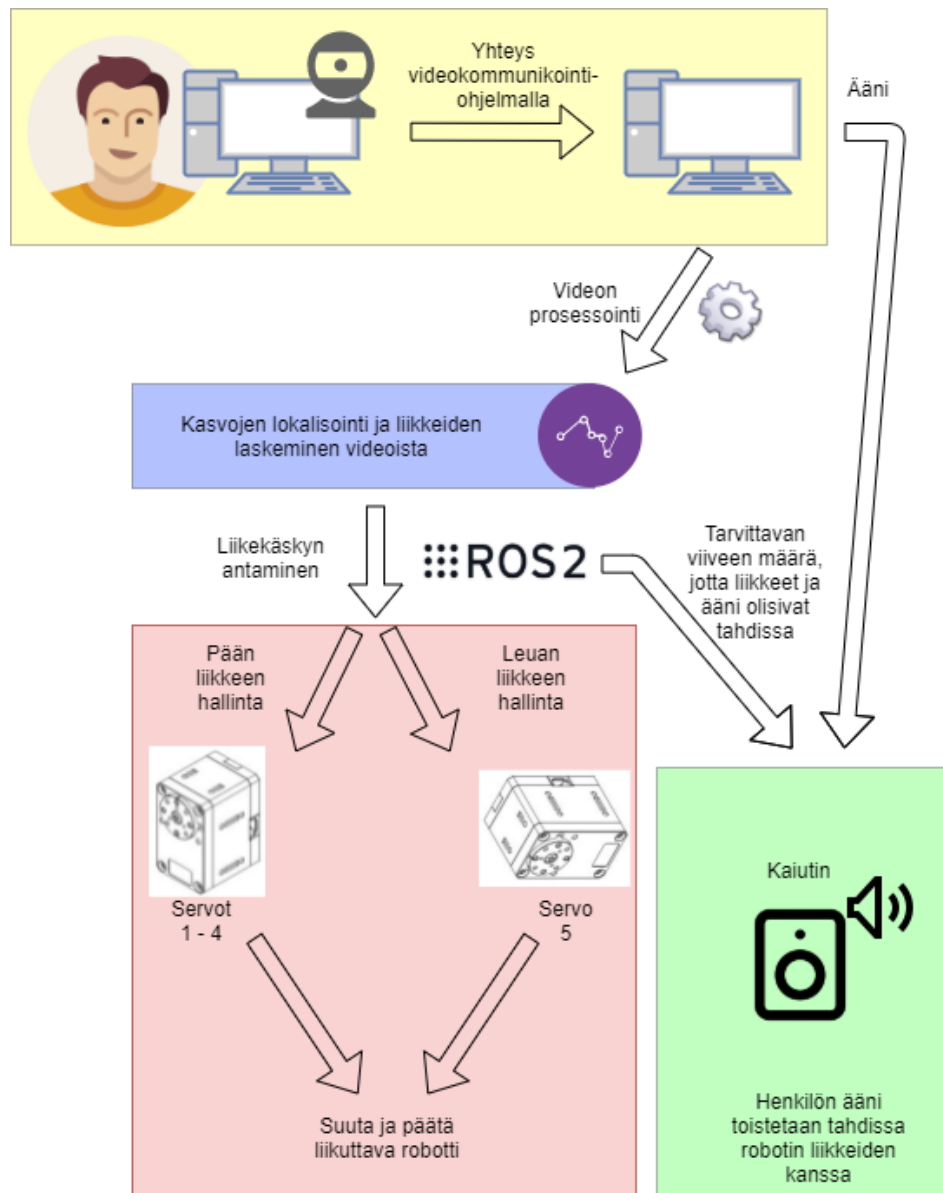
InMoov¹⁷ on ensimmäinen avoimeen lähdekoodiin perustuva täysikokoinen 3D-tulostettava robotti. Se sai alkunsa ranskalaisen kuvanveistäjän Gaël Langevin taideprojektina syyskuussa 2011. Koska InMoovin osien valmistamiseen riittää 12cm*12cm*12cm -tilavuuksinen 3D-tulostin, sen kehittäminen on mahdollista laajalle joukolle tutkijoista harrastelijoihin. InMoov tarjoaa mahdollisuuksia kehittää robottia koko kehon laajuudessa tai vain tiettyä kehon osaa, kuten päätä tai kättä. Avoimen lähdekoodin vuoksi robotin jatkuva kehittäminen on helppoa, kun jokainen kehittäjä voi tuoda robottiin oman kädenjälkensä.

¹⁷<http://inmoov.fr/project>

3. JÄRJESTELMÄN TOTEUTUS

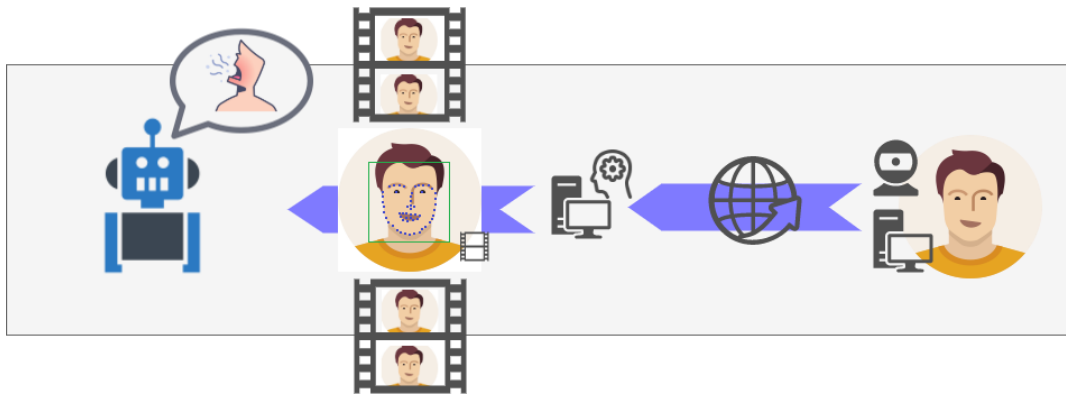
Tämän työn tarkoituksena on pyrkiä siirtämään videolta ääni ja suun liikkeet robotille niin, että puhe ja suu käyvät yhtäaikaaisesti.

Robotin etäohjauksen toteuttamiseksi kuvan videon perusteella vaaditaan ohjelmisto, joka kykenee ensin keräämään tarvittavan videomateriaalin, käsittelemään datan ja ohjaamaan robottia kerätyn datan mukaisesti. Robotin ohjaaminen tapahtuu tietokoneella ROS-alustan avulla ja saman tietokoneen ylijäänyttä prosessointivoimaa käytetään videon prosessointiin. Tällöin myös tarvittavien ohjelmien asentaminen on minimoitu ja kaikki teknistä osaamista vaativat tapahtumat hoidetaan yhdellä koneella laskien osaamisrajaa, joka vaaditaan etäläsnäolevalta henkilöltä. Prosessin eri vaiheet on esitetty Kuvassa 12.



Kuva 12. Prosessin vaiheet¹⁸

Henkilön etäläsnaolo tapahtuu käyttämällä hyväksi olemassaolevaa videokommunikointiohjelmaa (esim. Teams, Zoom, Signal, jne.), jolla hän ottaa yhteyden ROS-tietokoneeseen. Kuvassa 13 videokommunikointiohjelman tarjoama ikkuna henkilöstä antaa sovellukselle mahdollisuuden kaapata sen. Saatu data syötetään kehitettyyn kasvojen liikkeentunnistusohjelmaan, joka muuttaa videossa tunnistetut liikkeet ROS:in käyttämiksi käskyiksi. Nämä käskyt siirtyvät tietokoneesta U2D2-servokontrolleriin, minkä seurauksena saadaan servojen avulla liikutettua robotin päätä etäläsnaolevan henkilön mukaisesti.



Kuva 13. Datan kaappaaminen ja muokkaus robotin ohjaamista varten.¹⁹

Liikkeentunnistus toteutetaan käyttämällä openCV ja dlib -kirjastoja, jotka tarjoavat valmiit komponentit reaaliaikaiseen kasvojen kiintopisteiden tunnistukseen ja prosessilaskentakiihdytykseen, joita ohjelma vaatii toimiakseen mahdollisimman alhaisella viiveellä. Kirjaston käyttämä kasvojen kiintopistesijaintikartta antaa kirjastolle pisteet ja suhteellisen arvion, jossa pisteiden pitäisi olla suhteessa toisiinsa. Kirjaston käyttämä regressiopuu sovittaa tämän mallin algoritmin avulla mahdollisimman tarkasti kuvassa olevan henkilön kasvoille. Mallin sovittamat pisteet antavat pisteiden välille kvantitatiivisen etäisyyden, joka pystytään mittaamaan ja syöttämään algoritmille, joka kääntää etäisyyden ROS-käskyiksi.

3.1. Videon prosessointi

Konenäköä varten tarvittava data kerätään mss:n²⁰ ja pillow:n²¹ avulla kaappaamalla video näytöltä. Tällä tavoin videon kerääminen ei vaadi tukea tietyiltä sovelluksilta.

¹⁸Kuvan luonut Emil Dark. Kuvalle asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

¹⁹Kuvan luonut Emil Dark. Kuvalle asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

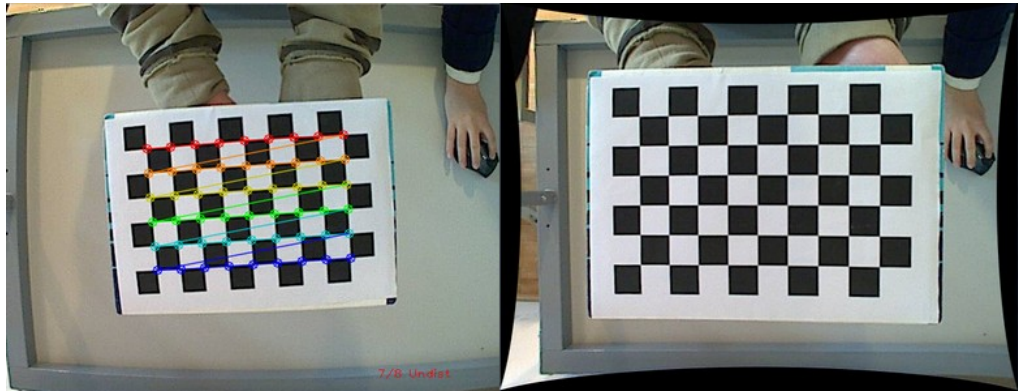
²⁰<https://pypi.org/project/mss/>

²¹<https://pypi.org/project/Pillow/>

3.1.1. Kameran kalibrointi

Tietokoneen web-kameraa käytettäessä on huomioitava kameran linssi, joka useimmiten laajentaa kameran näkökenttää suuremmaksi. Linssin näkökentän kasvaessa suuremmaksi se menettää kykynsä säilyttää linjojen suoruuden, objektien pituudet ja kulmat oikeassa suuruudessa. Ilmiö on havaittavissa kuvassa, kun reunojen koordinaattipisteet ovat kauempana toisistaan kuin pisteet kuvan keskellä, antaen kuvalle venyttävän ja taittavan efektin reunoissa. Mittaustulosten perustuessa kuvan koordinaatistoon on siitä saatavien tulosten tarkkuus täysin sidottu koordinaatiston kykyyn kuvata oikeaa maailmaa mahdollisimman tarkasti.

Kameroita kalibroitaessa käytetään ennalta tunnettua objektia, kuten kalibrointilevyä Kuvassa 14, ja verrataan sen piirteitä kuvan esittämään objektiin. Vertailun aikana huomattu ero kyetään mittaamaan, ja sitä voidaan käyttää hyväksi, kun luodaan sovitusmatriisi. Sovitusmatriisin tehtävä on vääristää kuvan objekteja sen verran, että aiemmin mainittujen piirteiden mittaustulokset ovat samansuuruisia kuin ne olisivat oikeassa maailmassa [36].



Kuva 14. Oikealla kuva ilman vääristystä ja vasemmalla vääristyksen jälkeen.²²

3.1.2. Videopuhelusovellusten integrointi

Projektin lähtökohtana oli tarjota loppukäyttäjälle mahdollisimman helppo tapa käyttää robottia etäläsnäoloon ja keskittää tekninen osaamisen henkilölle, joka on vastuussa robotin toiminnasta. Tämän saavuttamiseksi voidaan käyttää joko loppukäyttäjälle tuttua tai oppimiskäyrältään loivaa ohjelmaa. Covid-19 pandemiasta johtuva yleistynyt etätyöskentely on pakottanut suuren määrän ihmisiä opettelemaan videopuhelusovellusten, kuten Teams²³ ja Zoom²⁴, käyttämistä ja asentamista koneelleen. Sovellukset luovat hyvän alustan etäläsnäolon tarjoamiseksi massoille. Koska videopuhelusovellukset ovat jo hyvin kypsiä palveluja, ne tarjoavat paljon ominaisuuksia myös huonon yhteyden tai käyttäjien laitteistoerojen ylitsepääsemiseksi.

²² Kuvan luonut OpenCV tiimi. https://docs.opencv.org/3.4/d4/d94/tutorial_camera_calibration.html

²³ <https://www.microsoft.com/fi-fi/microsoft-teams/group-chat-software>

²⁴ <https://zoom.us/>

Videopuhelusovelluksesta voidaan kerätä dataa käyttämällä ohjelmointirajapintaa (API) videon ja äänen saamiseksi. Läpikäydyistä ohjelmista kaksi eniten tukea tarjoavaa alustaa olivat Teams ja Zoom, yritysympäristöissä suosittuja ohjelmia ovat esimerkiksi myös Slack²⁵ ja Webex²⁶.

Teamsin ja Zoomin hallitessa suurinta osaa kuluttajamarkkinoista²⁷, valittiin ne hyviksi vaihtoehtoiksi jatkokehitystä varten sopivuutensa kannalta projektin alustana. Dokumentaation perusteella Zoom tarjoaa parhaan tuen ja tarjoaa eniten ominaisuuksia kolmansille osapuolille, jotka haluavat käyttää heidän tuotettaan ohjelmansa 'back end' -ratkaisuna. Tämä ei kuitenkaan ole mahdollista ilman kalliita lisenssejä tai oman palvelimen ylläpitoa. Tästä johtuen ratkaisun käyttöönotto on kallista tai teknisesti haastavaa henkilölle, joka ei halua ylläpitää omaa palvelintaan.

Kaikkien aiemmin mainittujen yhtiöiden sovellukset tarjoavat kuitenkin ilmaisen version ohjelmastaan. Sovellusten maksulliset suunnitelmat ja palvelimien ylläpito voidaan täten kiertää käyttämällä ilmaista versiota ja sen sijaan, että käytetään sovelluksen API:a, data kerätään sovelluksen ikkunasta näyttöä kaappaamalla. Tämä ratkaisu myös tekee kehitetystä ratkaisusta riippumattoman yhdestä tietystä sovelluksesta. Tällöin henkilön, joka haluaa kommunikoida robotin kautta, ei tarvitse kuin soittaa koneelle, johon robotti on kytketty kiinni. Kaikki mittaaminen tapahtuu toisessa päässä, jolloin robotin käytön tekninen osaaminen on täysin abstraktoitu loppukäyttäjältä.

Näytöstä kaapattu kuva käsitellään ja muutetaan virtuaalikameraksi, jonka tietokone voi tunnistaa, mikä tekee tästäkin osiosta modulaarisen.

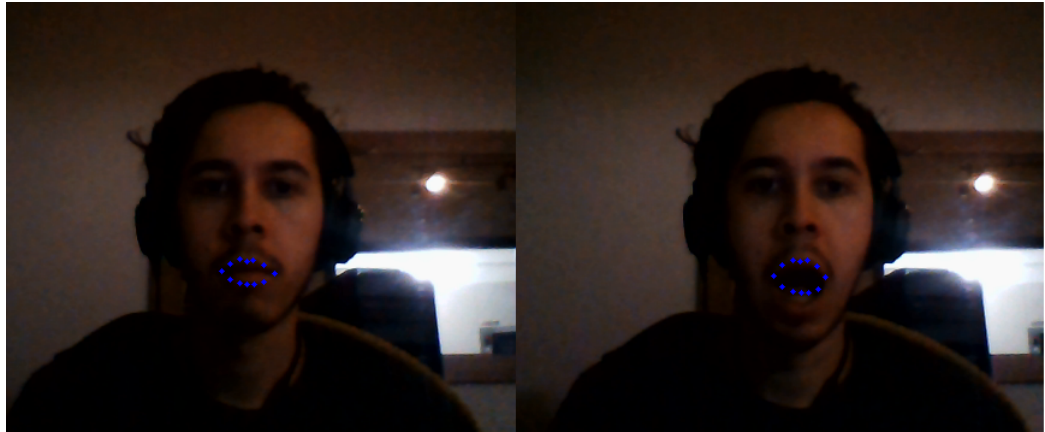
3.2. Datan käsittely

Kasvojen tunnistamiseen käytetään avointa lähdekoodikirjastoa dlib:iä joka on toteutettu C++ -ohjelmointikielellä. Dlib sopii useisiin eri käyttötarkoituksiin, kuten robotiikkaan, sulautettuihin järjestelmiin ja puhelimiin [37]. Dlibin lisäksi käytössä on tässä työssä OpenCV, eli 'Open Source Computer Vision Library' tai avoimen lähdekoodin konenäkökirjasto. OpenCV sisältää suuren määrän konenäköön ja koneoppimiseen liittyviä algoritmeja [38]. OpenCV:n avulla voidaan esimerkiksi tunnistaa videolta kasvot, eri esineitä tai objekteja ja tunnistaa joitain ihmisen toimintoja. Suun liikkeen kartoitus havainnollistetaan Kuvassa 15.

²⁵<https://slack.com/intl/en-fi/>

²⁶<https://www.webex.com/>

²⁷<https://www.emailtooltester.com/en/blog/video-conferencing-market-share/>



Kuva 15. Suun liikkeen kartoitus suun ollessa kiinni ja auki.²⁸

Samaa tekniikkaa, mutta enemmän hienosäätöä tarjoava ratkaisu on käyttää OpenTrack- ja AITrack-ohjelmia yhdessä. AITrack käyttää OpenCV-kirjaston moduuleja ja tarkempia kasvojenmittausalgoritmeja. AITrack toimii itsenäisenä ohjelmana ja lähettää mitatun datan OpenTrack-ohjelmalle UDP-paketteina. OpenTrack tarjoaa sovituskäyriä ja rajoittimet, jotka estävät liian suuria liikkeitä hajottamasta robottia. Ohjelmat on suunniteltu VR-pelejä varten.

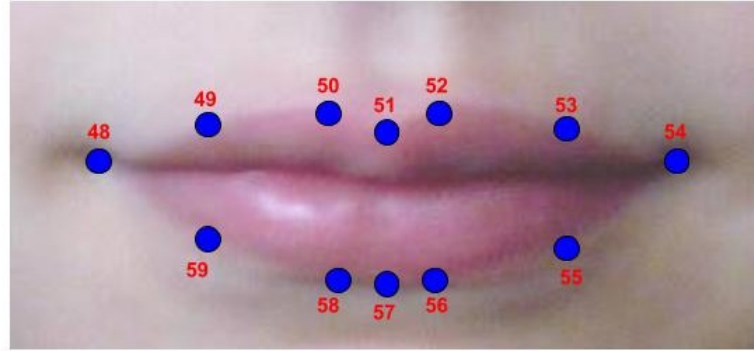
3.3. Datamalli

Robotin liikuttamiseen vaadittavan datan käsittelyyn tarvitaan erilaisia datamalleja, jotta voidaan suodattaa tarvittava data kaikesta raakadatasta. Suun ja pään liikkeille tarvitaan omat datamallinsa, joiden avulla voidaan liikutella suuta ja koko päätä halutulla tavalla, eli mahdollisimman tarkasti ihmisen liikkeiden mukaan.

3.3.1. Mittaaminen ja normalisointi

Aiemmin kuvailtu kasvojentunnistusmenetelmä mahdollistaa kasvojen kiintopisteiden sijaintien mittauksen koordinaattimatriisissa. Koordinaattien etäisyys lasketaan käyttämällä kolmea pistettä alahuulen alarajasta ja vastaavasti kolmea pistettä amorinkaaren huipusta. Huulten pisteet on esitetty Kuvassa 16. Pisteiden ollessa anatomian huippuja ne antavat suurimman erotusarvon ja liikkuvat ensimmäisinä ihmisen puhuessa.

²⁸Kuvan luonut Emil Dark. Kuvalle asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)



Kuva 16. Huulten kiintopisteet.²⁹

Koordinaattipistematriisissa ylähuulta vastaavat pisteet 48-53 ja alahuulta vastaavat pisteet 54-60. Puhetta voidaan mallintaa yksinkertaisesti liikuttamalla suuta ylös ja alas, eli laskemalla ylä- ja alahuulen etäisyys toisistaan koordinaattipisteiden avulla eri ajanhetkillä. Huulten etäisyyden kasvaessa arvo kasvaa ja minimiarvo saavutetaan suun ollessa kiinni.

Jokainen kolmesta pisteestä saa parin vastakkaiselta huipulta. (Tässä tapauksessa pisteet 50:58, 51:57 ja 52:56 ovat pareja.) Kun tiedetään pisteiden $P1 = (x_1, y_1)$ ja $P2 = (x_2, y_2)$ koordinaatit, voidaan laskea niiden välinen etäisyys, joka on myös niiden yhdysjanan pituus. Tämä voidaan vektorialgebrallisesti ilmaista Kaavalla (4).

$$\text{etäisyys} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4)$$

Saatujen arvojen jakauman normalisointiin voidaan käyttää tavallista aritmeettista keskiarvoa (5), koska vierekkäiset pisteet ovat kasvojen tunnistuksen sovittaman mallin takia aina toistensa vieressä.

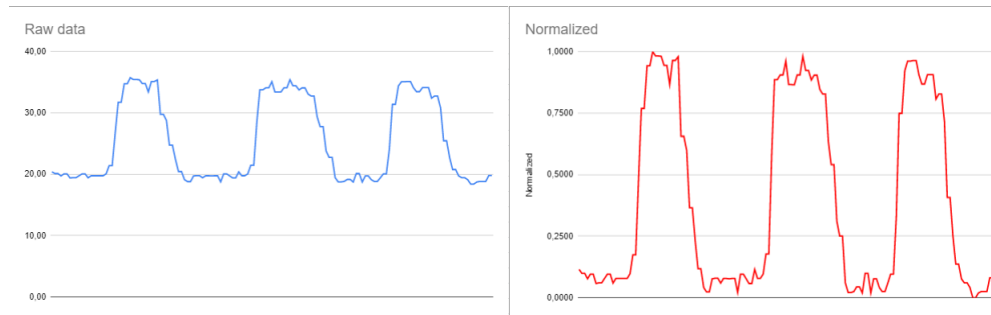
$$\bar{x} = \frac{1}{n} \sum_{n=1}^n x_n \quad (5)$$

Dataa normalisoidessa yritetään säätää mitattujen arvojen tulokset eri mittakaavoista yhteiselle asteikolle. Tällöin robottia voidaan liikuttaa hallitusti raudan rajoitusten mukaisesti. Kaavan (6) Min-Max -normalisoinnin avulla voidaan säätää arvot välille $[0, 1]$ [39]. Normalisoimalla jo prosessointivaiheessa vältetään ylimääräisen raakadatan siirtämiseltä edestakaisin.

$$\hat{x}_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (6)$$

Datan normalisoinnin jälkeen datasta piirretty viivakaavio näyttää muodoltaan samalta kuin raakadata, mutta arvot ovat nollan ja yhden välillä. Kuvassa 17 on esitetty vasemmalla raakadata ja oikealla normalisoitu data.

²⁹Kuvan luonut Emil Dark. Kuvalla asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)



Kuva 17. Huulten välinen etäisyys käsittelemättömänä datana ja normalisoituna datana.³⁰

Kullekin servolle sovelias arvoväli saadaan mittaamalla robotin leuan ja pään liikkeen ääriarvot ja suhteuttamalla normalisoitu arvoväli servon mittakaavaan Kaavalla (7).

$$x_{servo} = \frac{max_{servo} - min_{servo}}{max_1 - min_1} \times (x_1 - min_1) \quad (7)$$

3.4. Robotti

Työssä käytettävällä InMoov -robottipäällä on suu, joka liikkuu vain ylös ja alas, eli suu voi olla täysin kiinni tai sitä voidaan aukea halutun verran äärirajoiden sisällä. Toisin sanoen robottipäällä on tällä hetkellä käytössään vain alahuulen alasvetäjälihäs. Kasvojen tunnistukseen käytettävä ohjelma kykenee tunnistamaan kasvojen eri piirteitä ja liikkeitä paljon laajemmin, joten robotin fyysisiä ominaisuuksia parantamalla robottia voitaisiin ohjata huomattavasti realistisemmin.

3.4.1. Servot ja U2D2

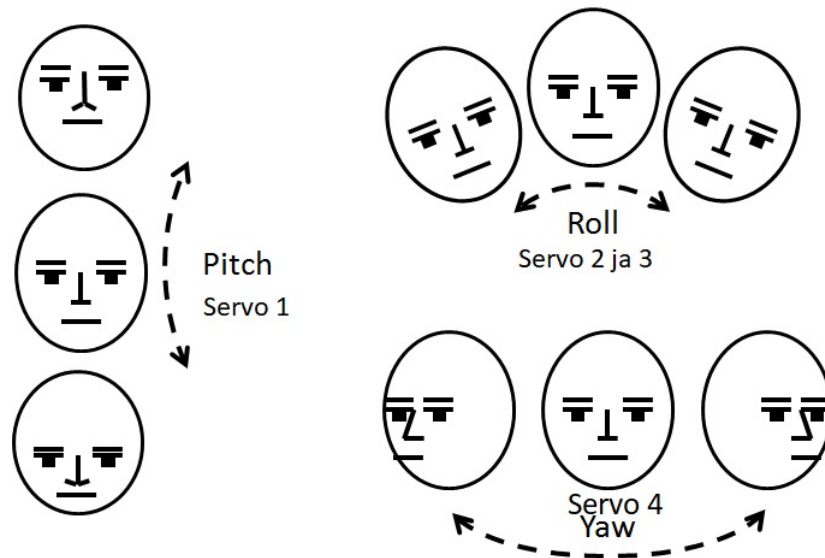
Leuan liikuttamiseen käytetty servo on Dynamixel XL-320 ja pään liikuttamiseen käytetään neljää XL-430-servoa. [40]. Niitä ohjataan U2D2-adapterin avulla, joka näkyy tietokoneelle sarjaporttilaitteena. Baud-nopeutena käytetään 57 600.

3.4.2. Servojen liikeavaruus

Robotin pään ja leuan liikkeiden tuottamiseksi käytetään viittä servoa, neljää liikuttamaan päätä ja yhtä liikuttamaan leukaa. Pään kääntäminen Kuvan 18 mukaisesti horisontaalisen akselin ympäri (pitch) ja pään kallistaminen (roll) toteutetaan servoilla 1-3, jotka yhdessä luovat Stewart-alustan, johon pää on kiinnitetty. Servo 1 liikuttaa päätä vaaka-akselin ympäri, luoden pään nykytoteutetun liikkeen ylös ja alas. Servot 2 ja 3 robotin kaulassa mahdollistavat pään kallistamisen vasemmalta oikealle. Neljännen servon liike tapahtuu alustan suuntaisesti, ja se kiertää päätä pystyakselin

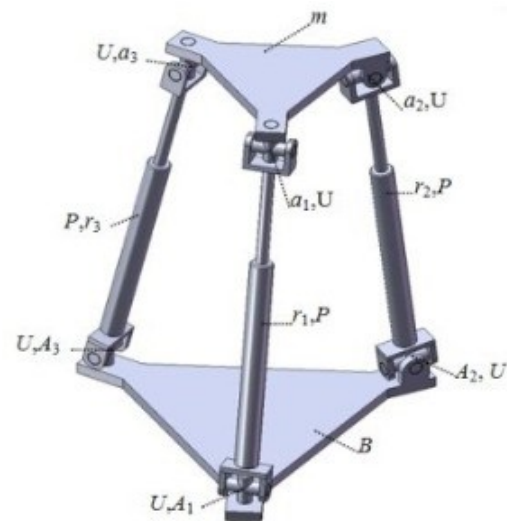
³⁰Kuvan luonut Emil Dark. Kuvalla asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

ympäri vasemmalta oikealle (yaw). Leuan servo liikuttaa alaleukaa pystysuorassa suunnassa.



Kuva 18. Pään rotaatioliikkeet ja niitä kontrolloivat servot.³¹

Robotin pään liikeavaruus perustuu Stewart-alustan mekaniikkaan, jonka avulla pään kallistaminen ja kääntäminen vertikaalisen akselin ympärillä pystytään toteuttamaan kolmella servolla. Stewart -alusta esitetään Kuvassa 19.



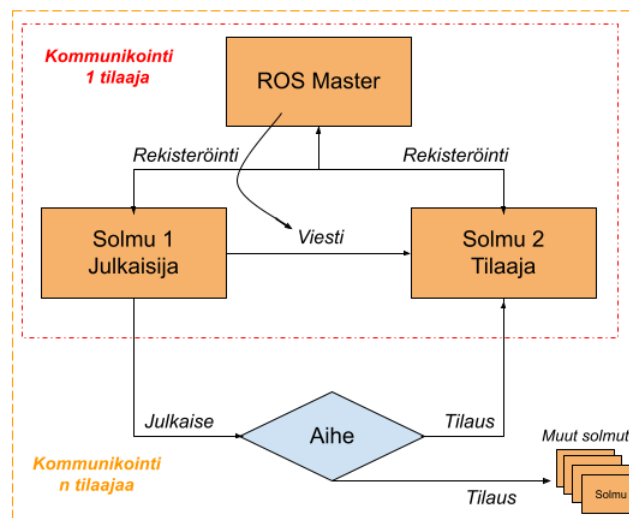
Kuva 19. Vasemmalla robotti ja oikealla kaulan mekanismia kuvaava 'Stewart platform' -malli.³²

³¹Kuvan luonut Emil Dark. Kuvalle asetettu lisenssi CCBY4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

³²Kuvan luonut Emil Dark. Kuvalle asetettu lisenssi CCBY4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

3.4.3. ROS 2 action client

Modulaarisen ROS 2:n rakenne voidaan esittää Kuvan 20 graafina, joka yksinkertaisimmillaan koostuu *solmuista* (node) ja *aiheista* (topic) [41]. Jokaisella solmulla on oma rajattu tarkoituksensa, kuten robotin renkaiden ohjaaminen. Solmut ovat prosesseja, jotka kommunikoivat toistensa kanssa ja suorittavat tehtäviä. Aiheet, joilla on nimensä mukaisesti tietyt aihealueensa, toimivat väylänä solmujen välillä kulkeville viesteille. *Julkaisijana* toimiva solmu lähettää aiheen kautta viestin *tilaajasolmulle*. Aiheella voi olla useita julkaisijoita ja vastaanottajia ja solmu voi toimia julkaisijana ja tilaajana usealle aiheelle.

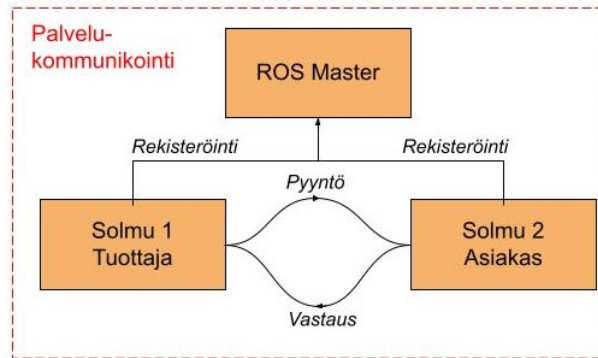


Kuva 20. ROS 2:n kommunikointitavat.³³

Yksinkertaisen, dataa yhteen suuntaan liikuttavan julkaisija/tilaaja -rakenteen lisäksi solmut voivat kommunikoida *palveluiden* avulla. Palvelukommunikointi -malli esitetään Kuvassa 21. Siinä palvelun *tuottajasolmu* lähettää dataa vain silloin, kun *asiakassolmu* lähettää sille pyynnön. Palvelu sopiikin yksittäisten tehtävien suorittamiseen, ja jatkuvaa datavirtaa vaativiin tehtäviin, kuten renkaan liikkeen ohjaamiseen, kannattaa käyttää aihetta tai *toimintoa* (action).

Toiminto (Service) on kommunikointityyppi, joka on tarkoitettu pitkäkestoisten tehtävien suorittamiseen. Sen kolme osaa ovat *tavoite* (goal), *tulos* (result) ja *palaute* (feedback). Toiminto käyttää palvelun tavoin asiakas/palvelija -mallia. Asiakassolmu lähettää palvelijasolmulle tavoitteen, jonka palvelija kuittaa vastaanotetuksi. Palvelijasolmu lähettää takaisin jatkuvaa palautetta sekä lopulta tuloksen. Toiminto, jota ei ole suoritettu loppuun, voidaan myös keskeyttää.

³³Kuvan luonut Emil Dark. Kuvalla asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)



Kuva 21. ROS 2:n palvelukommunikointi. ³⁴

3.4.4. Robotin kehittäminen ja säätö projektin aikana

Kesken projektia robotin pään liikkeitä kontrolloivien servojen kiinnityslevy hajosi ja vaati uuden valmistamisen. Tämä antoi mahdollisuuden jatkokehittää robottia eteenpäin ja korjata robotista löytyneitä ongelmia.

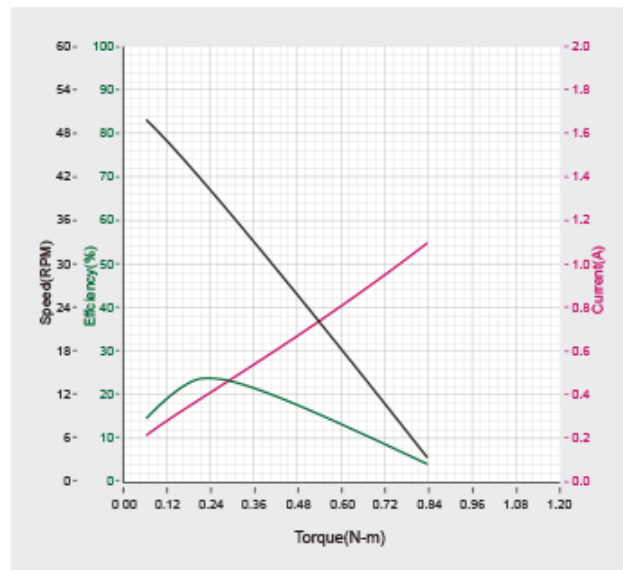
Servot olivat kiinnitettyinä puolijäykkään muovilevyyn, josta oli poistettu ylimääräistä materiaalia lisäjouston saavuttamiseksi, mikä kohdisti rasituksen pienemmälle pinta-alalle. Muovi jousti ylös ja alas kaulamäntien laajentuessa ja supistuessa aiheuttaen ajan kuluessa muovin rasisuuden kolmeen eri kohtaan. Tämä teki levystä käyttökelvottoman.

Vanha kiinnityslevy vaihdettiin 8 millimetriä paksuun akryylilevyyn, joka kestää servojen tuottaman vääntömomentin helposti ilman vääntymistä tai uudelleen halkeamista. Samalla pystyttiin suunnittelemaan kiinnityslevyyn servojen asentojen säätämiseen uusi mekanismi, jonka avulla servojen asentoa levyllä ja niiden kulmaa suhteessa kaulaan kyetään hienosäätämään mäntien kitkan ja juuttumisen vähentämiseksi. Kuvan 22 mukaisesti paras nopeus/vääntö -suhde saavutetaan, kun servo liikkuu noin 40 kierrosta minuutissa (tehokkuus 24%).

Mäntien rasva vaihdettiin paremmin muoville sopivaan ja ohuempaan, jottei männän sisälle syntyisi ilmapilaria. Boylen lain kuvaileman vastavoiman välttämiseksi männän päätyyn tehtiin reikä, jotta ilma pääsisi karkaamaan ja paine männän sisällä olisi sama kuin ulkopuolella. Nämä tulokset näkyvät Kuvassa 23.

³⁴Kuvan luonut Emil Dark. Kuvalle asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

Performance Graph

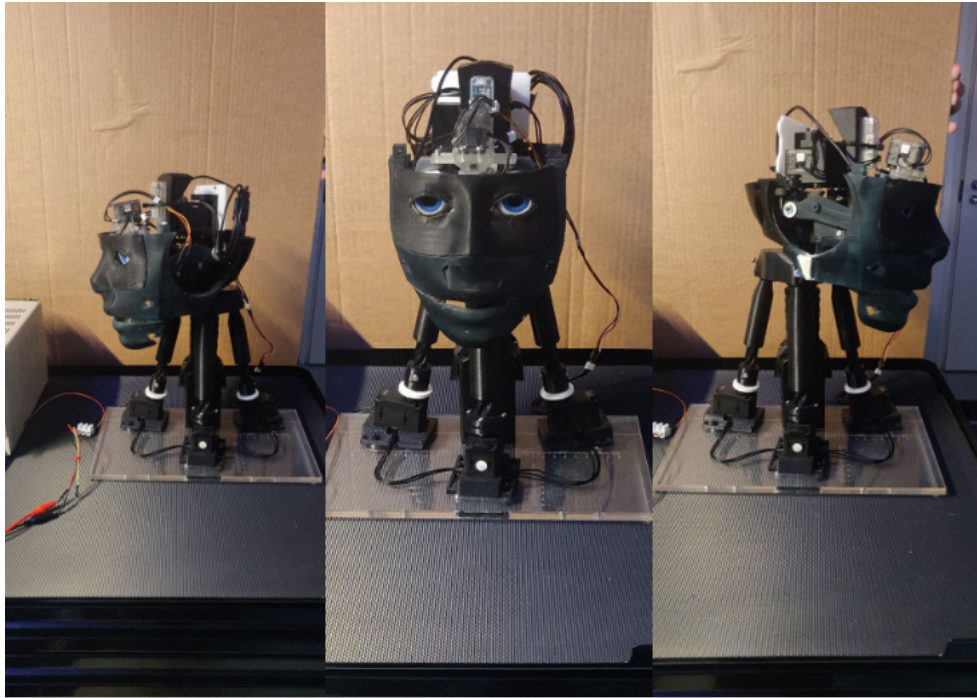


Kuva 22. Servon voimakaavio.

Puheominaisuutta kehitettäessä robotille 3D-tulostettiin kaiutinteline, mutta robotin näkömekanismi oli suurempi kuin InMoov-dokumentaatiossa³⁵ kuvailtu. Kaiuttimen integrointi robotin sisälle täten ei ollut mahdollista ilman näkömekanismin uudelleen suunnittelua tai pienemmän kuin dokumentaatiossa mainitun kaiuttimen hankkimista. Molemmat näistä vaihtoehdoista olivat tämän projektin laajuuden ulkopuolella. Päädyttiin sen sijaan kytkemään kaiuttimet tietokoneeseen ja asettamaan ne robottipään viereen.

³⁴Kuvan luonut Dynamixel, <https://emanual.robotis.com/docs/en/dxl/x/x1430-w250/>

³⁵<http://inmoov.fr/eye-mechanism/>



Kuva 23. Päivitetty robotti

3.5. Robotin liikuttaminen

Robotin ohjaamiseen käytettiin OUSPG:n ROS 2:lle toteutettavaa SOP-robottia, jolle toteutettiin ROS 2 -paketti leuan liikuttamista varten. Paketti koostui kahdesta solmusta, joista yksi toimi julkaisijana ja toinen tilaajana ja asiakkaana. Julkaisijasolmu tuotti videokuvan perusteella puhujan ylä- ja alahuulen etäisyyteen pohjautuvaa dataa. Toinen solmu tilasi aihetta, ja lähetti asiakkaana pyyntöjä leukaa liikuttavalle toiminnolle.

Kurssin osallistujien käytössä oli yksi InMoov-projektiin perustuva robottipää. Koska robotin käytössä täytyi vuorotella muiden ryhmien kanssa, suurin osa kehitystyöstä tehtiin RVIZ-simulaattorin avulla. Simuloitu robotti toimi servolle annettavien sijaintitavoitteiden absoluuttisia arvoja lukuunottamatta kuten fyysinen robotti, mutta oli saatavilla olevilla laitteilla käytettynä erittäin hidas. Simulointi kuitenkin auttoi varmistamaan, että paketin rakenne oli oikea ja toiminto-asiakas-suhde toimi.

Fyysisen robotin servoihin oltiin yhteydessä U2D2-adapterin avulla. Virtaa servot saivat erillisestä jännitelähteestä. Leukaa liikuttavan Dynamixel XL-320:n liikeradalle etsittiin sopivat ääriarvot ohjaamalla sitä Dynamixel Wizard 2.0:n avulla, ja data normalisoitiin niiden perusteella.

³⁵Kuvan ottanut Emil Dark, Kuvalle asetettu lisenssi CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.fi>)

4. TESTAUS

Projektin onnistumista testattiin kyselyn avulla, jonka tarkoituksena oli selvittää kyselyyn osallistujien mielipiteitä robotin toiminnasta ja ihmisenkaltaisuudesta sekä liikkeiden että ulkonäön osalta. Testikysymyksillä haluttiin siis selvittää myös mahdollisia epämukavuuden tunteita liittyen outo laakso -ilmiöön.

4.1. Testijärjestelyt

Testit toteutettiin nauhoittamalla videota samanaikaisesti robotista ja ihmisestä. Videot editoitiin yhteen niin, että ihmisen ja robotin liike ovat mahdollisimman yhtäaikaiset, jotta kyselyyn vastaajat arvostelivat puhtaasti robotin liikkeen aitoutta eivätkä viivettä. Testikysymyksiä varten oli kaksi eri videota, joista toinen on kategoriata 1 ja toinen kategoriata 2 varten. Kysymyksessä 5 viitattiin molempiin videoihin ja kyselyyn osallistujien yleiseen mielipiteeseen robottien kautta kommunikointiin. Kyselyyn vastaaminen oli täysin anonymia ja kyselyyn käytettiin Google Forms - verkkosovellusta.

4.1.1. Testikysymykset

Kategoria 1. Suun liikkeen arviointi:

1. Arvostele robotin suun liikkeen aitoutta verrattuna ihmisen suun liikkeeseen asteikolla 1-5.	
1	Robotin suun liike ei muistuta lainkaan ihmisen suun liikettä.
2	
3	Robotin suun liike muistuttaa jokseenkin ihmisen suun liikettä.
4	
5	Robotin suun liike muistuttaa täysin ihmisen suun liikettä.
2. Kuinka luontevalta robotin puheen seuraaminen sinusta tuntuu asteikolla 1-5?	
1	Robotin puhetta on vaikea seurata.
2	
3	Robotin puhetta on kohtuullisen vaivatonta seurata.
4	
5	Robotin puhetta on helppo seurata.

Kategoria 2. Pään liikkeen arviointi:

3. Arvostele robotin pään liikettä verrattuna ihmisen pään liikkeeseen asteikolla 1-5.	
1	Robotin pään liike ei muistuta lainkaan ihmisen pään liikettä.
2	
3	Robotin pään liike muistuttaa jokseenkin ihmisen pään liikettä.
4	
5	Robotin pään liike muistuttaa täysin ihmisen pään liikettä.

4. Miten pään liike vaikuttaa mielestäsi keskustelun laatuun asteikolla 1-5?	
1	Erittäin häiritsevästi.
2	
3	Ei vaikutusta.
4	
5	Parantaa keskustelun laatua huomattavasti.

Outo laakso:

5. Tuntuisiko vuorovaikutus tällaisen robotin kanssa häiritsevältä?	
1	Kyllä, erittäin.
2	
3	Kyllä, hieman.
4	
5	Ei tuntuisi häiritsevältä.

4.2. Testitulokset**Yhteenveto**

Osallistujien lukumäärä	32
Keskimääräinen arvosana	2,7
Suun liike (keskiarvo)	2,2
Pään liike (keskiarvo)	3,0
Outo laakso	2,8

Kysymyskohtaiset tulokset

Kysymys	Tulos (keskiarvo)	Mediaani
1.	2,5	3
2.	1,9	2
3.	3,2	3
4.	2,9	3
5.	2,8	3

4.3. Testitulosten pohdinta

Kyselyyn vastanneiden mukaan pään liike koettiin paremmin ihmisen liikkeen kaltaiseksi kuin suun liike. Heistä yli puolet kallistuivat siihen, etteivät kumpikaan, suun tai pään liike, paranna keskustelun laatua. Viimeisessä kysymyksessä nähdään myös, että suurempi osa kokee robotin kanssa kommunikoinnin häiritseväksi. Kyselyn tulokset olivat hieman odotettua paremmat, sillä robottia ei saatu toistamaan kaikkia sille lähetettyjä liikekäskyjä.

Robottia ei saatu vietyä sille tasolle mikä oli alkuperäisenä tavoitteena, mutta kyselyn tulokset osoittivat, että ainakin robotin pään liikkeet muistuttivat jonkin verran jo ihmisen pään liikettä.

5. POHDINTA

Projektin tavoitteena oli kehittää ja toteuttaa uusi tapa ihmisten väliseen vuorovaikutukseen etäläsnäoloteknologian avulla ja laskea käyttäjältä vaadittavan teknisen osaamisen tarvetta teknologian etujen hyödyntämiseksi.

Etäläsnäolorobottien tuoma etu eri käyttötarkoituksissa on ilmeinen, mutta niitä käytetään vielä hyvin spesifisiin ongelmiin, eivätkä ne ole saavuttaneet yleistä käyttäjäryhmää. Projektin tarkoituksena oli yrittää rikkoa näitä rajoja tarjoamalla hyvin modulaarinen ratkaisu, jonka eri osia voitaisiin hyödyntää tulevien kehittäjien toiveiden mukaisesti. Projekti voidaan jakaa neljään eri osa-alueeseen: datan siirtoon, liikkeiden mittaamiseen, robotin liikuttamiseen ja kommunikointiin/vuorovaikutukseen.

Tutkitusta kirjallisuudesta ei löytynyt vastaavaa tapaa soveltaa kuluttajille laajasti tunnettua teknologiaa muihin etäläsnäoloteknologioihin. Teknologian adoptiota helpottaa se, että kuluttajan näkökulmasta sen käyttämiseen riittää videopuheluserveyshyönteiden ottaminen robotia ohjaavaan tietokoneeseen. Projektia suunniteltaessa tämä koettiin hyväksi tavaksi varmistaa, että kuluttajalta vaadittavat tekniset toimenpiteet eivät nosta teknologian käyttöönoton kynnyksiä liian korkealle.

Liikkeiden tunnistus ja mittaaminen oli projektin onnistunein osa. Tutkitun kirjallisuuden perusteella tämä oli ensimmäinen kerta, kun liikkeiden mittaaminen on tehty OpenCV:n ja Inmoov -robotin kanssa. Mittaussovellus ohittaa käytetyn robotin rajoitteet reilusti ja antaa loistavan pohjan tuleville kehittäjille. Sovellus kykenee mittaamaan käyttäjän liikkeitä 30 kertaa sekunnissa. Kasvojen koordinaatistopisteitä on 68 kappaletta, joten dataa kerätään enemmän kuin roboti pystyy toistamaan, kuten esimerkiksi suun kohdalla.

Robotin liikuttaminen oli projektin suurin haaste. Tarkoituksena oli saavuttaa saman tasoinen liikkuvuus kuin kirjallisuudesta löytyneissä esimerkeissä, mutta tähän ei päästy. Nopeiden liikkeiden toteutus kävi haastavaksi servomoottoreissa syntyneiden vääntövoimien takia. Erityisesti pään kääntäminen pystyakselin ympärillä vasemmalta oikealle (yaw) tuotti hankaluuksia, sillä sen aikana servomoottori ylikuormittui herkästi, jolloin se asettui suojattuun tilaan. Aiheutettu vikaviesti joudutaan käymään poistamassa manuaalisesti Dynamixel Wizard -ohjelmassa. Ongelma pyrittiin korjaamaan suunnittelemalla uusi pohjalevy, johon servot asennetaan parantaen kulmaa suhteessa päähän, vähentäen syntyvää kitkaa mekanismissa. Uusi pohjalevy vähensi kitkaa, mutta sitä esiintyi edelleen liikaa.

Projektin alussa käytetty simulaattori toimi vain koodin toimivuuden tarkistamiseksi, mutta sen alhaisen virkistystaajuuden takia sillä ei kyetty kehittämään mitään muuta toimintaa eteenpäin. Jälkikäteen katsottuna simulaattorin käyttö tuotti enemmän päänvaivaa ja hukattuja tunteja kuin hyötyä projektin näkökulmasta.

Taustakirjallisuutta tutkiessa huomattiin, että ihmisen ilmeisiin vaikuttaa suun alueella useampi suurempi lihas. Robotin suosaa voitaisiin kehittää jakamalla suu esimerkiksi kolmeen tai useampaan osaan, jolloin suupieliä saataisiin käännettyä, jotta ilmeitä saataisiin hieman enemmän esille. Odotetusti robotin suun liike ei muistuttanut testikyselyssä vastaajien mielestä kovin selvästi ihmisen liikkeitä, koska robotille lähetetyt datapaketit eivät kaikki toistuneet robotissa. Tämän lisäksi robotin suun liikkeen ollessa vain vertikaalisessa suunnassa, on ihmisen hienovaraisen eleiden

vastaanottaminen täysin mahdotonta robotin kautta, eikä ihmismäisyyttä siksi ole juurikaan havaittavissa robotin liikkeissä.

Robottia voitaisiin kehittää ihmismäisemmäksi lisäämällä sille esimerkiksi otsan ja poskien lihasten liikuttelu sen lisäksi, että suu jaettaisiin pienempiin osiin. Tällöin robotille mahdollistetaan enemmän ilmeitä ja tuodaan robottiin laajemmin etäläsnäolevan henkilön luonnetta ja kasvojen eleitä. Robotin fyysisiä ominaisuuksia lisäämällä tarvittaisiin tarkempaa ja laajempaa kasvojen mittausta ja tunnistusta, mutta esimerkiksi käytössä olleella kiintopistemenetelmällä voitaisiin jo tunnistaa suun liikkeet ja käyttää saatua dataa laajemmin laskemalla myös muita suun liikkeitä kuin huulten vertikaalista etäisyyttä.

Keskustelua yhteistyömahdollisuuksista muiden robottia kehittävien ryhmien kanssa käytiin, mutta huomattiin, että robotin kehittäminen olisi pitänyt aloittaa alusta alkaen yhdessä, jotta yhteistyö olisi onnistunut. Yhteistyökeskusteluita käytäessä todettiin, että data kerätään joka työssä eri tavoin, jolloin projekteissa olisi tapahtunut valtavia askeleita taaksepäin, jotta yhteistyö olisi onnistunut. Silloinkin datan keruu ja käsittely olisi ollut mahdollisesti melko ongelmallista, jotta kaikki oltaisiin saatu toimintaan yhtäaikaisesti, sillä tässäkin työssä pään ja suun liikuttelua ei saatu toimimaan yhtäaikaisesti.

Telerobotiikkaa kehitettäessä tulee miettiä, mihin suuntaan halutaan lähteä robottien kehittämisessä. Jos robotista halutaan kehittää ihmisen näköinen, on lopputuloksen todella oltava luonnollisen näköinen, jotta outolaakso -ilmiöltä välttyttäisiin. Toinen mahdollisuus on siirtyä tarpeeksi kauas ihmisen kasvoista ja tehdä hahmo, joka pystyy edelleen kuvaamaan etäläsnäolevan henkilön tunnetiloja välttäen kuitenkin outo laakso -ilmiön.

Etäläsnäolorobotin käyttö toisi etätyöskentelyyn lisää ulottuvuuksia, jos henkilö voisi ohjata kehonsa avulla koko ihmiskehon omaavaa robottia toisessa sijainnissa. Tällaisen robotin avulla henkilö voisi avustaa työtehtävissä jopa eri maassa, tai esimerkiksi opastaa ja jakaa tietotaitoa eri lokaatioissa. Tällöin robottien ansiosta välttyttäisiin turhalta matkustelulta, kun tehtäviä voisi suorittaa eri lokaatioista. Yksi jo olemassa oleva esimerkki on tässäkin paperissa aiemmin mainittu lääketieteessä käytetty Da Vinci -robotti, joskin sitä ei ole tehty ihmisten väliseen kommunikointiin, vaan pelkästään toimenpiteen suorittamiseen.

6. YHTEENVETO

Projektissa kehitettiin modulaarista ratkaisua etäläsnäolon saavuttamiseksi videopuhelusovelluksen kautta InMoov-robottiin, missä etäläsnäolevan henkilön liikkeet peilataan robottiin luoden etäläsnäolon tunteen henkilölle ja luonnollisen kanssakäymisen robotin kanssa kommunikoiville.

Projektin osat onnistuttiin tekemään irtonaisiksi, jolloin ne toimivat käytännössä millä tahansa videopuhelusovelluksella. Videosta kyettiin onnistuneesti tunnistamaan henkilön kasvot ja mittaamaan suun liike ja pään eulerin kulmat.

Sekä robotin pään ja leuan liikkeille luotiin ROS-solmut, jotka normalisoivat mittaukset servokohtaisesti ROS2 julkaisija/tilaaja-mallin mukaisesti.

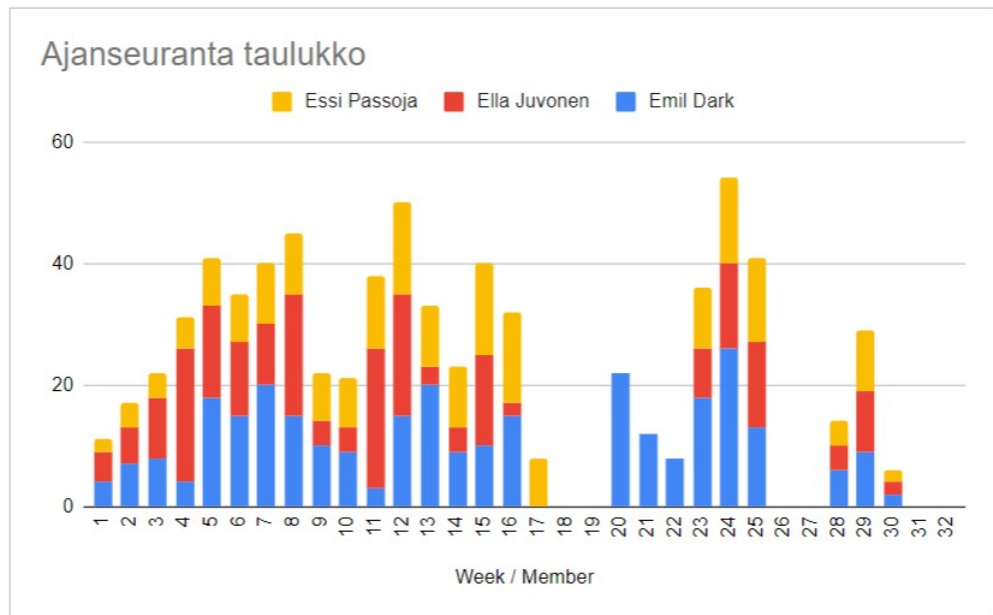
Ihmisen suun ja pään liikkeet onnistuttiin muuttamaan robotin liikkeiksi, mutta käytetyn robotin fyysiset rajoitteet estivät sulavan liikkumisen ja viive vahingoitti kanssakäymisen laatua.

Robotin uskottavuutta testattiin kyselyn avulla, jossa osallistujille näytettiin videoita robotista. Osallistujien tehtävänä oli arvostella robotin suoriutumista arvosteluasteikolla 1-5. Osallistujien mielestä robotissa on vielä parantamisen varaa, eikä se varsinkaan suun liikkeiden osalta ole vielä luonnollinen. Pään liikkeet saivat hieman paremman arvostelun.

Kasvonliiketunnistusmalli antaa hyvän pohjan jatkokehitykselle. Enemmän työtä vaaditaan robotin reaaliaikaisen liikkumisen parantamiseksi. Toimivaa toteutusta voi luontevasti hyödyntää tilanteissa, missä henkilö ei kykene itse olemaan paikan päällä ja haluaa antaa paikan päällä olevalle henkilölle tavan kanssakäydä hänen kanssaan muuten kuin ruudun kautta.

7. AJANKÄYTTÖ

Työtehtävät jaettiin tasaisesti kaikkien kesken.



8. LÄHDELUETTELO

- [1] Leskinen T. (2020) Säännöllisesti kotona työskenteleminen on kaksinkertaistunut. Tieto ja Trendit, Tilastokeskus URL: www.stat.fi/tietotrendit/blogit/2020/saannollisesti-kotona-tyoskenteleminen-on-kaksinkertaistunut/.
- [2] Stutzer A. & Frey B.S. (2008) Stress that doesn't pay: The commuting paradox. *Scandinavian Journal of Economics* 110, ss. 339–366.
- [3] Oguz S. (2014) Commuting and personal well-being, 2014. Office for National Statistics .
- [4] Blomqvist K., Sivunen A., Vartiainen M., Olsson T., Ropponen A., Henttonen K. & Van Zoonen W. (2020) Etätyö suomessa koronaviruspandemian aikana. CoCoDigi URL: <https://cocodigiresearch.files.wordpress.com/2020/12/etacc88tyocc88-suomessa-koronaviruspandemian-aikaan-2020.pdf>.
- [5] Sheridan T.B. (1992) Telerobotics, automation, and human supervisory control. MIT press.
- [6] Auvinen P. (2018) Ilmeet kasvoillamme: Ihmiskasvoista, ilmeistä ja taiteesta .
- [7] Business Matters (2015) The history of Video Conferencing. URL: <https://bmmagazine.co.uk/tech/history-video-conferencing/>.
- [8] editorial staff of vcs.su (2019) A Brief History of Video Conferencing: From the Beginning to Full Commercial Use. Habr. URL: <https://habr.com/en/post/465459/>.
- [9] Senft T.M. Videoconferencing communications. Encyclopedia Britannica. URL: <https://www.britannica.com/technology/videoconferencing>.
- [10] Marmpena M., Lim A. & Dahl T.S. (2018) How does the robot feel? perception of valence and arousal in emotional body language. *Paladyn, Journal of Behavioral Robotics* 9, ss. 168–182. URL: <https://doi.org/10.1515/pjbr-2018-0012>.
- [11] Zheng J.M., Chan K.W. & Gibson I. (1998) Virtual reality. *IEEE Potentials* 17, ss. 20–23.
- [12] Richert D., Macnab C.J.B. & Pieper J.K. (2012) Adaptive haptic control for telerobotics transitioning between free, soft, and hard environments. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 42, ss. 558–570.
- [13] Sheridan T. (1989) Telerobotics. *Automatica* 25, ss. 487–507. URL: <https://www.sciencedirect.com/science/article/pii/0005109889900939>.

- [14] Manoharan S. & Ponraj N. (2019) Precision improvement and delay reduction in surgical telerobotics. *Journal of Artificial Intelligence* 1, ss. 28–36.
- [15] Nancekievill M. (2020) Robotics for Harsh Environments: Subsea. Wevolver. URL: <https://www.wevolver.com/article/robotics-for-harsh-environments-subsea>.
- [16] Baker T. & Baker T. (2020) The History of Motion Capture Within The Entertainment Industry. väitöskirja. URL: <http://urn.fi/URN:NBN:fi:amk-202005057375>.
- [17] Weta, Facets. URL: <https://www.wetafx.co.nz/research-and-tech/technology/facets/>.
- [18] Mori M., MacDorman K. & Kageki N. (2012) The uncanny valley [from the field]. *IEEE Robotics Automation Magazine* 19, ss. 98–100.
- [19] Brenton H., Gillies M., Ballin D. & Chatting D. (2005) The uncanny valley: does it exist? Teoksessa: Proceedings of conference of human computer interaction, workshop on human animated character interaction, Citeseer.
- [20] Tiede M., Mooshammer C. & Goldstein L. (2019) Noggin nodding: Head movement correlates with increased effort in accelerating speech production tasks. *Frontiers in Psychology* 10, s. 2459. URL: <https://www.frontiersin.org/article/10.3389/fpsyg.2019.02459>.
- [21] Livingstone S. & Palmer C. (2016) Head movements encode emotions during speech and song. *Emotion (Washington, D.C.)* 16, ss. 365–380.
- [22] Ekman P. (2003) Emotions revealed. Times Books.
- [23] Alpaydin E. (2014) Introduction to machine learning. Adaptive computation and machine learning, MIT Press, Cambridge, Massachusetts, third edition p., 616 s. Includes index.
- [24] Tutz G. (2012) Regression for categorical data. Cambridge series on statistical and probabilistic mathematics, Cambridge University Press, Cambridge ; New York, 574 s. Description based upon print version of record.
- [25] Jin X. & Tan X. (2017) Face alignment in-the-wild: A survey. *Computer Vision and Image Understanding* 162, ss. 1–22. URL: <https://www.sciencedirect.com/science/article/pii/S1077314217301455>.
- [26] Çeliktutan O., Ulukaya S. & Sankur B. (2013) A comparative study of face landmarking techniques. *EURASIP Journal on Image and Video Processing* 2013, ss. 1–27.
- [27] Kazemi V. & Sullivan J. (2014) One millisecond face alignment with an ensemble of regression trees. 2014 IEEE Conference on Computer Vision and Pattern Recognition , ss. 1867–1874.

- [28] Fabbri R., Kimia B. & Giblin P. (2012) Camera pose estimation using first-order curve differential geometry. ss. 231–244.
- [29] Gao X., Hou X., Tang J. & Cheng H. (2003) Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, ss. 930–943.
- [30] Hesch J.A. & Roumeliotis S.I. (2011) A direct least-squares (dls) method for pnp. *Teoksessa: 2011 International Conference on Computer Vision*, ss. 383–390.
- [31] Persson M. & Nordberg K. (2018) Lambda twist: An accurate fast robust perspective three point (p3p) solver. *Teoksessa: Proceedings of the European Conference on Computer Vision (ECCV)*.
- [32] Open Robotics About ROS. URL: <https://www.ros.org/about-ros/>.
- [33] Haukijärvi M. (2020) Robot operating system 2 (ros 2) is here - why should you care? *Atostek* URL: <https://atostek.com/en/robot-operating-system-2-ros-2-is-here-why-should-you-care/>.
- [34] Kutluca H. (2020) Robot operating system 2 (ros 2) architecture. *Medium* URL: <https://medium.com/software-architecture-foundations/robot-operating-system-2-ros-2-architecture-731ef1867776>.
- [35] Open Robotics ROS 2 Documentation. URL: <http://docs.ros.org/en/foxy/>.
- [36] OpenCV team Camera calibration With OpenCV. URL: https://docs.opencv.org/3.4/d4/d94/tutorial_camera_calibration.html.
- [37] Dlib C++ Library. URL: <http://dlib.net/>.
- [38] OpenCV team About OpenCV. URL: <https://opencv.org/about/>.
- [39] Grus J. (2015) *Data Science from Scratch: First Principles with Python*. O’Reilly Media, Inc., 1st p., 121-140 s.
- [40] Dynamixel, Documentation of dynamixel xl430-w250-t servo. URL: <https://emanual.robotis.com/docs/en/dxl/x/xl430-w250/>.
- [41] Morgan Quigley B.G. & Smart W. (2015) *Programming Robots with ROS*. O’Reilly Media, Inc., 1st p., 110-145 s.