Ihmisen havaitsemiseen käytetyt tekoälytekniikat autonomisissa koneissa

Turun yliopisto
Tietotekniikan laitos
TkK-tutkielma
Tietotekniikka
Kesäkuu 2024
Lauri Levälehto

TURUN YLIOPISTO
Tietotekniikan laitos

Lauri Levälehto: Ihmisen havaitsemiseen käytetyt tekoälytekniikat autonomisissa koneissa

TkK-tutkielma, 26 s.
Tietotekniikka
Kesäkuu 2024

Tässä kirjallisuuskatsauksessa käydään läpi tekoälytekniikoita, joita voidaan käyttää autonomisissa koneissa ihmisten havaitsemiseen. Havainnointiin voidaan käyttää eri sensoritekniikoita, joista tässä ovat mukana näkyvän valon kamerat, sisältäen stereokamerat, sekä lämpökamerat ja lidar. Eri datatyypeillä on omat vahvuutensa ympäristön ja ihmisen hahmottamiseen, ja vaativat omanlaisensa käsittelyn. Pääosin tunnistuksessa käytetään kuitenkin konvoluutioneuroverkkoja, ja tekoälyyn toteutetaan jollain tasolla muisti.
Lisäksi katsauksessa pohditaan sensoritekniikoiden yhdistämisen mahdollisuutta samaan järjestelmään, millä voidaan mahdollisesti parantaa toimintavarmuutta. Riippuen käyttökohteesta, eri sensoriyhdistelmät voivat olla järkeviä. Todettiin olevan tapoja luoda kamerakuvan pohjalta 3D-avaruus, jolloin siitä saadaan yhteensopiva 3D-dataa tuottavien sensoritekniikoiden kanssa. Lopuksi katsauksessa käydään lyhyesti läpi liikkeenseurantatekniikoita ja todetaan, että jo perinteisen objektinseurannan avulla neuroverkon tunnistustehtävää saadaan helpotettua. Selvisi myös, että koneessa voidaan hyödyntää toiminnantunnistusta (engl. *human action recognition*), jotta tiedetään mitä ihminen tekee ja osataan päätellä mitä koneen halutaan tehtävän.

Asiasanat: tekoäly, autonominen kone, robotti, sensoritekniikat, ihmisen havaitseminen

Sisällys

Kuvat

1 Johdanto

Autonomisten robottien ja koneiden kehittyessä ne on saatu hahmottamaan ympäristönsä paremmin ja paremmin. Perinteisiä robotteja on ollut tehtaissa jo pitkän aikaa, mutta niiden ympäristö tarvitsee pitää tarkoin kontrolloituna. Hitsausrobotti voi tehdä saman virheettömän sauman miljoona kertaa, mutta osa on joka kerta sen edessä samassa paikkaa. Älykkäät robotit pystyvät kartoittamaan ympäristöään ja havaitsemaan sen muutoksia, mikä mahdollistaa autonomisten koneiden käytön myös ulkona, muun muassa metsä- ja kaivosteollisuudessa, jossa liikutaan haastavassa maastossa.

Ehkä vielä mielenkiintoisempaa kuitenkin on, kuinka robotin älykkyys mahdollistaa ihmisen työskentelyn robotin rinnalla, samassa tilassa. Monessa työvaiheessa vaaditaan edelleen ihmistyöntekijää suorittamaan hienomotoriikkaa ja sovelluskykyä vaativia tehtäviä, sekä tarkkailemaan työn laatua. Työn jakamista ihmisten ja robotin välillä helpottaa merkittävästi, jos robotti ei ole eristettynä toisessa tilassa. Vaikka ihminen tällöin tekisi joka toisen työvaiheen, robotin apu silti puolittaisi hänen työmääränsä ja auttaisi fyysisissä tehtävissä. Ihminen pystyisi myös helposti (ihmiselle intuitiivisesti) näyttämään robotille, mitä tarvitsee tehdä, esimerkiksi mihin kohtaan kappaletta tulee leikata reikä tai hitsata sauma. Edellä kuvattu visio edellyttää roboteilta ihmisten virheetöntä tunnistamista, jotta työskentely on turvallista.

Tämä kirjallisuuskatsaus kartoittaa tekniikoita, joita havainnointiin voidaan käyttää, ja yrittää vastata, missä järjestelmässä on eniten potentiaalia. Tältä pohjalta valittiin seuraavat tutkimuskysymykset:

1. Mitä sensori- ja tekoälytekniikoita ihmisen havainnointiin yleisesti käytetään?

2. Voidaanko eri sensorien dataa yhdistää; miten? Onko tämä järkevää?

3. Miten liikkeenseurantaa voi hyödyntää havainnoinnissa?

Tietoa haettiin sekä Google Scholarista että tavallisesta Googlesta, pääosin englanniksi. Alkuperäisiin hakutermeihin lukeutui muun muassa *artificial intelligence/AI, robot, human recognition, gesture recognition, computer vision, lidar, infrared/IR, stereo camera*. Hakuja kuitenkin tarkennettiin tulosten perusteella, ja sisällytettiin hakulauseisiin termejä kuten *human motion recognition* ja *human action recognition/HAR*. Esimerkiksi HAR-termin olemassaolo selvisi löytyneen tutkielman abstraktia lukemalla, ja se osoittautui oikein päteväksi hakusanaksi. Stereokamerajärjestelmien olemassaolosta oltiin tietoisia jo ennen kirjoitusprosessia, joten niiden sisällyttäminen tutkittaviin sensoritekniikoihin oli luonnollista. Tuloksista ohitettiin hyvin spesifit, esimerkiksi tiettyyn neuroverkkotyyppiin porautuvat tutkielmat, sillä ne menivät syvemmälle yksityiskohtiin kuin tässä kirjallisuuskatsauksessa aiotaan. Tarkoituksena on kartoittaa kenttää, tekniikoita ja mahdollisuuksia laajalla tähtäimellä, ja tämä vaikutti myös lähteiden sopivuuteen. Spesifimpiä lähteitä sisällytettiin lähinnä silloin, jos ne tarjosivat ajantasaista tietoa eri tekniikoiden suorituskyvystä.

Työn sisältö on jaettu lukuihin seuraavanlaisesti: Luvussa 2 esitellään ensin ihmisen havaitsemiseen käytettyjä sensoritekniikoita ja niiden tuottamaa dataa. Sitten käydään läpi käyttötarkoituksessa yleisimmin käytetyt neuroverkkotyypit ja pohditaan niiden vajaavaisuuksia eri sensoritekniikoihin yhdistettynä. Luvussa 3 pohditaan eri sensorityyppien yhdistämisen potentiaalia ja toteutustapoja. Luvussa 4

kuvataan ihmisen liikkeiden seuraamisen ja ennustamisen mahdollisuuksia ja haasteita. Siinä peilataan, miten liikkeenseuranta parantaa aikaisemmin esiteltyjen tekniikoiden tarkkuutta ja käyttökelpoisuutta. Luvussa 5 kerätään ajatukset yhteen ja pohditaan tekniikkojen potentiaalia sekä nyt että tulevaisuudessa.

2 Paikannustekniikat

Autonomisten koneiden ympäristön hahmottamiseen on kaksi vallalla olevaa sensorikategoriaa: kamerat ja tutkat. Käymme tässä luvussa läpi molempien vahvuuksia ja heikkouksia, sekä erittelemme kategorioiden sisältämiä tekniikoita hieman tarkemmin. Lisäksi perehdymme tekoälytekniikkoihin, joita käytetään ihmisen tunnistamiseen datasta.

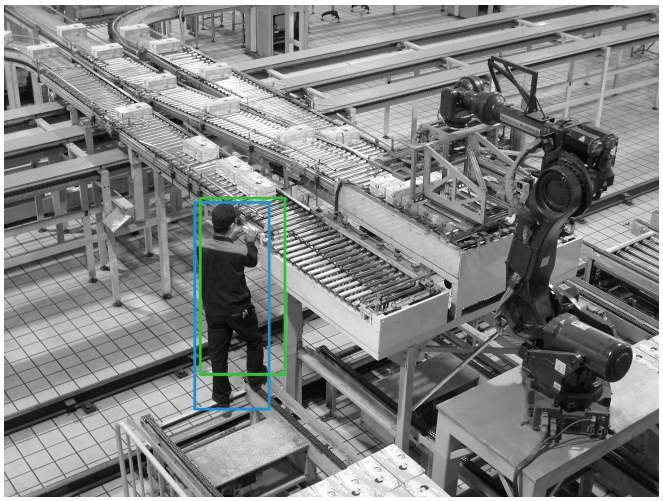
On huomattava, että autonomiset koneet voivat hyödyntää toiminnassaan monia muitakin tekniikoita, kuin mitä käsittelemme tässä. On muun muassa olemassa tavarankuljetusrobotteja, jotka seuraavat ennalta määritettyä reittiä ja havaitsevat mahdolliset esteet ultraäänianturilla [1][2]. Näinkin voi saada aikaan huomattavan ”älykkään” ja toimivan järjestelmän. Keskitymme nyt kuitenkin tekniikoihin, jotka mahdollistavat ihmisen tarkan paikantamisen ja sitä myötä robotin ja ihmisen välisen yhteystyön.

2.1 Kamerat

Yksinkertaisimmillaan pelkät näkyvän valon kamerat riittävät havainnointiin, niin kuin ihminenkin pystyy tarkasti hahmottamaan ympäristönsä pelkällä katseellaan. Etuna muihin sensoritekniikoihin nähden on kuvan tarkkuus ja värien erottaminen. Tällainen konenäkö kuitenkin edellyttää erittäin edistynyttä kuvantunnistusta, jotta kuvasta hahmotettaisiin tila ja tilasta objektit. Ihmisaivot kykenevät suorastaan hämmästyttävän tehokkaaseen kuvadatan prosessointiin ja tunnistavat sil-

mänräpäyksessä objektin geometrian, koon ja etäisyyden. Saman toteuttaminen tekoälyllä vaatii suurta laskentatehoa, ja virheettömän toiminnan varmistaminen on vaikeaa.

Toimintavarmuuden suureena tekoälykehityksessä käytetään usein keskimääräistä tunnistustarkkuutta (engl. *average precision, AP*), eli sitä kuinka iso osa kaikista kuvista tunnistetaan oikein, prosenttiarvona. On eri standardeja siihen, mikä lasketaan oikein tunnistamiseksi; esimerkiksi AP50-mittaustavalla tekoälyn rajaama alue täytyy olla yli 50-prosenttisesti oikein. Se ei siis saa jättää objektista liikaa merkitsemättä, mutta ei myöskään merkitä liikaa tyhjää objektin ympäriltä. Kuvassa 2.1 tätä on havainnollistettu.



Kuva 2.1: Havainnollistava kuva keskimääräisen tunnistustarkkuuden (AP) määrittämisestä. [3]

Siinä on työntekijä tehtaassa, joka tekoälyn tulee tunnistaa. Sinisellä suorakulmiolla on käsin merkitty oikea alue, joka tarkkaan rajaa työntekijän. Vihreällä on kuvitteellisen tekoälyn merkitsemä alue. Tekoäly on jättänyt jalkaterät ja vasemman olkapään merkitsemättä ja on puolestaan merkinnyt liikaa oikealta ylhäältä, liukuhihnan päältä. AP50-mittaustavalla suoritus laskettaisiin kuitenkin onnistumiseksi, sillä oikean ja tekoälyn rajaaman alueen leikkaus on yli 50 % niiden unionista (engl. *intersection over union, IoU*). [4]

Yksinkertaisimmillaan järjestelmässä on yksi tavallinen näkyvän valon kamera, jonka tuottamasta videokuvasta ympäristö päätellään. Yhtä kameraa käyttävät järjestelmät etsivät kuvasta objektien reunat pelkästään kontrasti- ja värieroja havainnoimalla. Tyhjästä taustasta objektin erottaminen olisi tietenkin helppoa, mutta todellisuudessa ympäristö on monimutkainen, taustalla voi olla puita ja kasvillisuutta. Jotta ihmistä ei tarvitse erottaa kuvan keskeltä täysin tekoälylogiikan avulla, on hyödyllistä käyttää muitakin tekniikoita ihmisen tunnistamiseen.

2.1.1 Stereokamerat

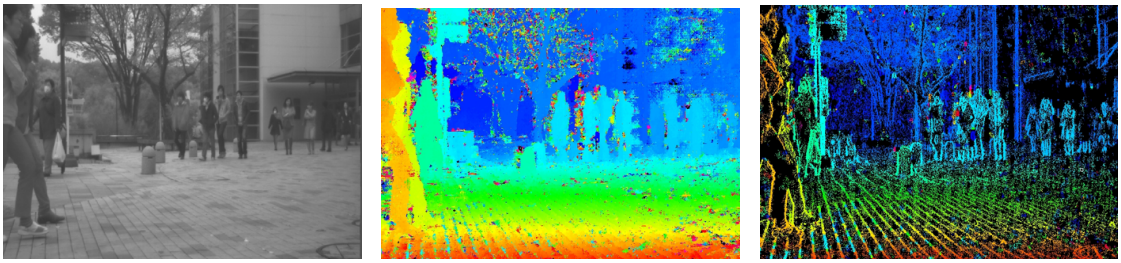
Tilan syvyysinformaatio voidaan päätellä käyttämällä kahta kameraa; samaan tapaan kuin petoeläimillä ja ihmisillä on tarkka syvyysnäkö kahden eteenpäin osoittavan silmän ansiosta. Stereonäön avulla saadaan havaittua ja paikallistettua objektit oikein silloinkin, kun tekoälyn kuvantunnistus ei toimi täydellisesti. Oleellisintahan on, että objekti erotetaan taustasta ja sen sijainti arvioidaan oikein. Se, onko väistettävä kohde ihminen vai jotain muuta, on usein toisarvoista.

Stereokameraratkaisuja on jo autoissa. Esimerkiksi Subarun vuonna 2008 esittelemä EyeSight-järjestelmä perustuu kahteen kameraan tuulilasin yläreunassa, eikä siinä käytetä normaalia tutkaa ollenkaan [5]. Myös Teslan Autopilot-järjestelmä sisältää kaksi eteenpäin osoittavaa kameraa stereonäön luomiseksi [6]. Siinä käytettiin aluksi etäisyyden mittaamiseen myös tutkaa, mutta se on jätetty malleista vaiheittain pois 2021 alkaen [7]. Vaikka stereokamerajärjestelmät käyttävät näkyvän valon kameroita, niiden datankäsittely muistuttaa paljon enemmän lidarjärjestelmää kuin kuvantunnistusta. Kahden kameran kuvasta päätellään syvyysinformaatio kuvan eri osille.

Kuvassa 2.2 näkyy ITD Labin[1] stereokamerajärjestelmän dataa. Vasemmalla (a) on tällaisen järjestelmän kameran tuottama harmaasävyinen kuva. Harmaasävyku-

[1]ITD Lab on entisen Subaru Eysightin pääinsinöörin, Keiji Saneyoshin perustama yritys.

van käsittely on värillistä helpompaa, ja bonuksena aidon harmaasävysensorin hämäränäköominaisuudet ovat myös värisensoria paremmat. Keskellä (b) näkyy kuvista prosessoitu syvyysdata eri värein. Kuva muistuttaa paljon lidarjärjestelmän tuottamaa, mutta se ei koostu pisteistä ja kuvassa näkyy siellä täällä pieniä artefakteja. Oikealla (c) on uudempi ja kehittyneempi järjestelmä, joka tunnistaa ja korostaa objektien reunat.



(a) Kameran kuva (b) Syvyysdataa (c) Reunantunnistus

Kuva 2.2: Stereokamerajärjestelmän dataa. (ITD Lab) [8]

Pelkkään ihmisen tunnistukseen ei välttämättä edes tarvita varsinaista tekoälyä, vaan esimerkiksi ITD Labin hätäjarrutusjärjestelmä tunnistaa objektien reunat perinteisellä ei-oppivalla algoritmilla. Perinteisessä toteutuksen eduiksi kerrotaan se, että joten se soveltuu heti ilman koulutusta käytettäväksi myös vieraassa ympäristössä ja toimii varmemmin tilanteessa kuin tilanteessa. Tämä voi olla käyttökelpoinen lähestymistapa myös koneisiin, joiden toiminta vaatii tekoälyä, sillä perinteisen algoritmin tuottama kuva (kuten kuva 2.2 c) voidaan antaa eteenpäin tekoälylle, jolloin sen tehtäväksi jää ihmisten tunnistaminen valmiiksi korostetusta kuvasta. [8]

Korkeamman resoluution vastapainona stereokameran tuottama syvyysinformaatio ei ole niin tarkkaa kuin lidarin. Käytetty algoritmi, kameran laatu ja objektin etäisyys vaikuttavat suuresti saavutettavaan tarkkuuteen. ITD Labin ISC-100XC-kameralle ilmoittama virhemarginaali on metrin päässä 0,2 %, 10 m päässä 2,3 %, 30 m päässä 7 % ja 60 m päässä 14 % [9]. Virhe on siis 60 m päässä jo yli 8 metriä, mutta toisaalta ihmisenkin etäisyysarviot alkavat näin kaukana heittää; 8 metriä on

kuitenkin vain kaksi autonmittaa. Lisäksi esimerkkinä käytetyn kameran tarkkuus on vain 1024x720 pikseliä, ja nykyisin tavanomaisella 4K-resoluutiolla (3840x2160) pitäisi teoriassa yltää vielä 100 m etäisyydellä edellä mainittuun 8 metrin tarkkuuteen [10]. Riippuu käyttötarkoituksesta, onko tämä tarkkuus tarpeeksi; Ihmisen kokoiselle koneelle saattaa olla, mutta kymmeniin metreihin yltävälle tehdasnosturille varmasti ei.

2.1.2 Infrapunakamerat

Infrapunakamerat muistuttavat perinteisiä näkyvän valon kameroita, mutta ne taltioivat valon sijaan lämpösäteilyä. Infrapunakameran kuvassa ihminen loistaa kirkkaana ympäristöstään ruumiinlämmön vuoksi. Kuvassa 2.3 ihminen erottuu savuverhon takaakin.

Kuva 2.3: Lämpökameran kuva ihmisestä savuverhon takana. [11]

Näytöllä kuva havainnollistetaan yleensä harmaasävyisenä, niin että lämpimimmät alueet ovat valkoisia ja kylmimmät tummanharmaita/mustia. Myös laitteelle kuvan prosessointi on hyvin samanlaista kuin normaalin harmaasävyisen näkyvän valon kuvan. Kameran pikselit mittaavat säteilyn voimakkuuden eri pisteissä ja arvot tallentuvat laitteelle digitaalisesti. Jokaisella kuvan pikselillä on siis jokin numeerinen kirkkausarvo, mutta se vain kuvaa lämpösäteilyn, eikä valon, voimakkuutta.

Lämpökamera voi myös aktiivisesti muuttaa herkkyyttään ja siten mittausaluettaan ja -tarkkuuttaan. Tämäkin vastaa näkyvän valon kameroita, jotka saavat sää-

deltyä sensorille pääsevän valon määrää aukkoa ja valotusaikaa muuttamalla. Lämpökameroiden heikkouksia on alhainen resoluutio, yleensä vain 320x240 tai 640x480 pikseliä. Näkyvän valon kameroilla tämä on moninkertainen, nykyään yleinen 4K-resoluutio on 3840x2160 pikseliä[2]. Toisaalta ihminen erottuu kuvasta lämmön ansiosta, joten kovin yksityiskohtaiselle kuvalle ei ole edes tarvetta. [12] [13]

Infrapunakamerat ovat myös kalliimpia kuin perinteiset näkyvän valon kamerat, tosin entistä edullisempia malleja on ilmestynyt viime aikoina. Autonominen kone toiminee varsin maltillisissa olosuhteissa, joten yleisesti edellytetyille kalliimpien mallien erityisominaisuuksille, kuten lämmönkestolle palokuntakäytössä, ei ole tarvetta.

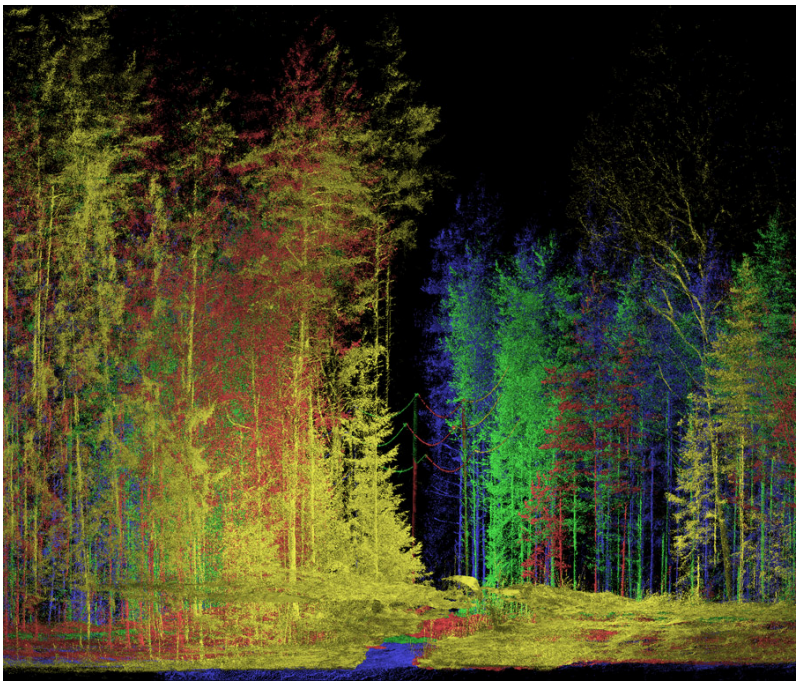
2.2 Tutkat

Missä kamerat taltioivat ympäristön säteilyä, etäisyyssensorit, kuten lidar, perustuvat kaikuluotaamiseen. Sensori lähettää ääntä tai säteilyä ja mittaa, kuinka kauan sillä kestää heijastua takaisin. Signaalin kulkuajasta saadaan laskettua kohteen etäisyys. Yksinkertaisin on ultraäänianturi (engl. *sonar*), joka käyttää mittaamiseen ääntä ihmisen kuuloalueen ulkopuolella, yli 20 kHz taajuudella. Käyttö rajoittuu lyhyen kantaman sovelluksiin, kuten autojen pysäköintisensoreihin. Tutka (engl. *radar*) käyttää radioaaltoja, yltäen paljon suurempaan kantamaan, mutta sen tarkkuus ei riitä ihmisen tunnistamiseen.

Viimeisenä lidar käyttää mittaamiseen laservaloa, mikä tekee siitä käyttötarkoitukseemme mielenkiintoisimman. Laserin saa kohdistettua tarkasti haluttuun suuntaan, joka mahdollistaa ympäristön skannaamisen muutama piste kerrallaan. Lasereita voi olla laitteessa lukuisia, mutta niitä ei saa mahdutettua kuin pikseleitä kamerakennoon. Niitä on yleensä tietty määrä päällekkäin (esimerkiksi 128), ja ne

[2]Nykyään 4K:sta puhuttaessa tarkoitetaan yleisimmin Ultra HD -resoluutiota, joka ylläkin ilmoitetaan, eli 3840x2160 pikseliä. 4K:lla voidaan tarkoittaa myös resoluutiota 4096x2160. Tarkalla resoluutiolla ei tässä yhteydessä ole merkitystä, vaan oleellista on suuruusluokka.

pyörivät, skannaten joka suuntaan ympärilleen [14]. Vaikka skannaus toimii hyvin suurella nopeudella, se häviää silti kameralle, millä on merkitystä, kun lidaria halutaan käyttää reaaliaikaisessa toiminnassa, eikä vain tilan kartoittamiseen etukäteen.



Kuva 2.4: Lidarin avulla luotu pistekartta ympäristöstä metsäkoneen käyttöön. [15]

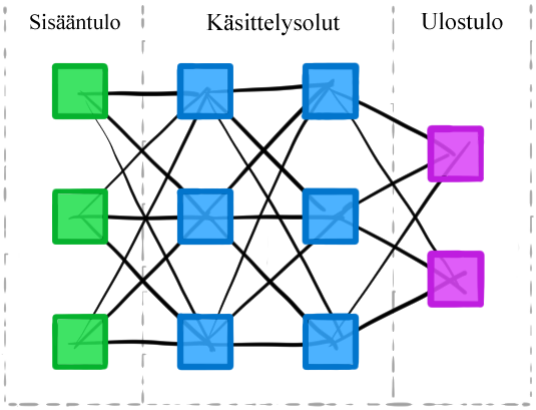
Kuvassa 2.4 näkyy lidarjärjestelmän luoma pistekartta. Kuva näyttää ihmissilmälle varsin selkeältä kaksiulotteisenakin, mutta varsinainen etu näkyvän valon kameraan verrattuna on syvyysinformaation tallentuminen. Kuvassa tätä havainnollistetaan karkeasti eri värein, mutta todellisuudessa sadan metrin etäisyydelläkin päästään jopa 2 cm mittaustarkkuuteen. [16]

Lidarin heikkouksia ovat laitteiden kalleus ja kameraa hitaampi toiminta, mikä rajoittaa niiden käyttökohteita. Etenkään hitaasti liikkuvassa ja kalliissa työkoneessa nämä eivät kuitenkaan ole merkittäviä ongelmia, ja toimintavarmuus ja tarkkuus kuvantunnistukseen verrattuna saattaa merkitä enemmän. Pimeys ei myöskään haittaa lidaria kuten näkyvän valon kameraa, mutta toisaalta vesihöyry, kuten sumu, heijastaa ja taittaa valoa, ja haittaa siten järjestelmän toimintaa. [17]

2.3 Tekoälytekniikat

Oli kuva tuotettu millä antureilla tahansa, ihmisen tunnistaminen siitä vaatii joka tapauksessa lisäprosessointia. Datan analysointiin käytettäviä tekoälytyyppejä on lukemattomia, mutta ne voidaan jakaa muutamaan pääkategoriaan. Konvoluutioneuroverkot (engl. *convolutional neural network, CNN*) ovat yleisesti käytössä kuvantunnistuksessa ja takaisinkytketyt neuroverkot (engl. *recurrent neural network, RNN*) kuvasarjoja käsiteltäessä [18].

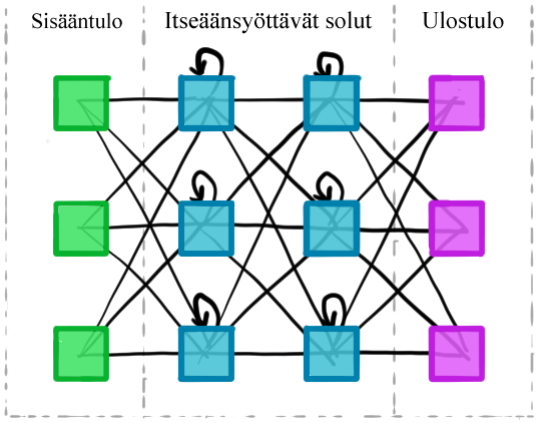
Käydään ensin läpi neuroverkoista yksinkertaisin, eteenpäin syöttävä neuroverkko (engl. *feed forward neural network, FFNN*), sillä takaisinkytketyt- ja konvoluutioneuroverkot pohjautuvat samaan ideaan. Neuroverkko koostuu solmuista, neuroneista, joita on tämän alaluvun kuvissa havainnollistettu neliöin. Kuvassa 2.5 on vasemmalla vihreällä syöte (lähtödata), keskellä sinisellä ”piilossa” olevia käsittelykerroksia ja oikealla violetilla lopputulos.



Kuva 2.5: Eteenpäin syöttävän neuroverkon rakenne

Data kulkee eteenpäin syöttävissä neuroverkoissa rinnakkaisesti eli useampaa reittiä kerroksesta toiseen. Yleensä jokainen kerroksen neuroni on yhdistetty seuraavan kerroksen neuroneihin. FFNN:n opettamiseen käytetään yleensä ohjattua oppimista, jossa neuroverkolle annetaan syötteen lisäksi myös haluttu lopputulos, ja

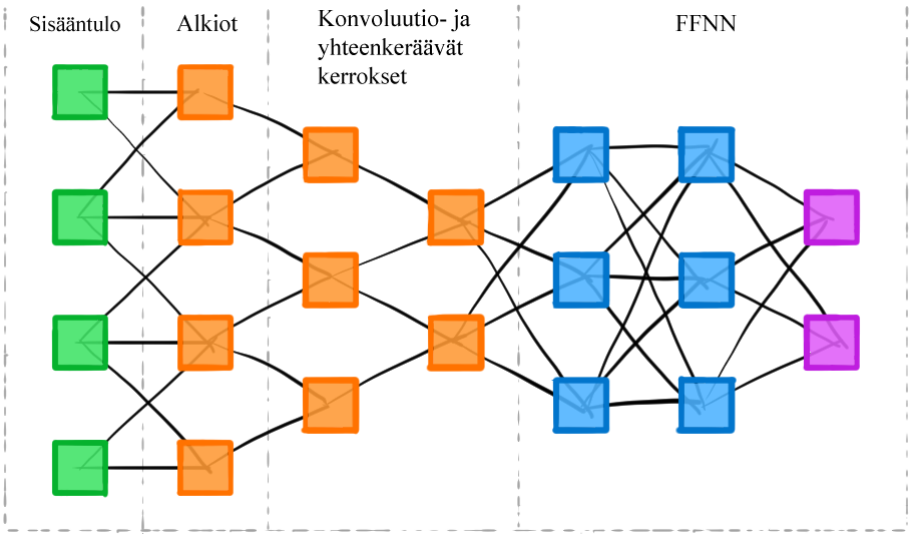
verkko aktivoi kytköksiä niin, että se pääsisi syötteestä lopputulokseen. Kerroksien syöte-seuraus-eroja seuraamalla saadaan tietoa neuroverkon toiminnasta. [19]



Kuva 2.6: Takaisinkytketyn neuroverkon rakenne

Takaisinkytketyillä neuroverkoilla on ”muisti”. Ne ovat varsin suora modifikaatio eteenpäin syöttävistä verkoista, kuten kuvasta 2.6 huomaa. Erona on, että neuronit tallentavat tuloksensa ja käyttävät sitä syötteenä seuraavassa ajossa. Ne siis saavat syötteen sekä edeltävältä kerrokselta että itseltään, ja näin historiatieto vaikuttaa tuloksiin. Historia kuitenkin häviää nopeasti, sillä verkko muistaa vain edellisen arvonsa. Vaikka tuohon edelliseen tulokseen onkin vaikuttanut sitä edeltävä arvo ja niin edelleen, arvojen on mahdollista vaihtua varsin nopeaan. Videota analysoidessa historiaa on näin vain muutaman ruudun verran, ja jos joka ruutu analysoitaisiin ja videokuva sisältää 60 kuvaa sekunnissa, historia yltää vain sekunnin sadasosia tai kymmenyksen taaksepäin. On olemassa neuroverkkoja, joissa on fyysistä muistia joka solulla. Näin historia pysyy pidempään, mutta rauta muuttuu monimutkaisemmaksi, rajoittaen puolestaan neuroverkon kokoa. [19]

Konvoluutioneuroverkkojen (kuva 2.7) kerrokset suppenevat käsittelyn edetessä, muokaten dataa vaihe vaiheelta abstraktimmaksi. Löydetyt ja abstraktoidut objektit lajitellaan sitten kategorioihin. Prosessin aikana kameran raaka pikselidata muuttuu käsitykseksi siitä, mitä kuva todennäköisimmin esittää. Tulos-soluja voi olla monta,



Kuva 2.7: Konvoluutioneuroverkon rakenne

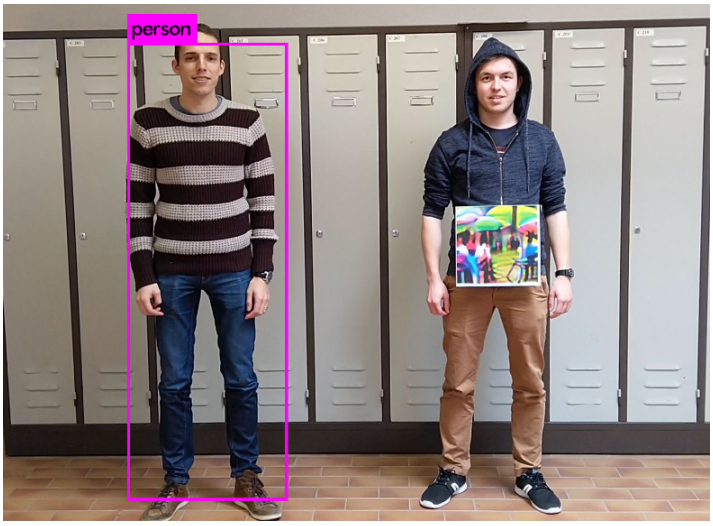
koska kuvassa voi olla vaikka ”koira kuussa”, eli siitä tunnistetaan monta ominaisuutta. Käytännössä kuva käydään läpi pala kerrallaan; tarkasteltava alue voi olla esimerkiksi 20x20 pikseliä. Prosessi aloitetaan vasemmalta ylhäältä ja joka kierroksella siirrytään pari pikseliä oikealle. Alueissa on siis paljon päällekkäisyyttä. [19]

Konvoluutioverkkojen ongelma on, että niiden toimintaa ja oppimista on käytännössä mahdotonta seurata, koska ne perustuvat niin suureen määrään muuttujia, niiden välisiä vuorovaikutuksia ja laskutoimituksia. Jos verkko toimii pääosin oikein, mutta ei selviydykään jostain hyvin spesifistä tilanteesta, tätä ei saada selville ennen kuin juuri kyseistä tilannetta on testattu tai se on tullut vastaan oikeassa käytössä. Turvallisuuteen vaikuttavissa järjestelmissä, kuten ihmisen tunnistamisessa, tällaisten puutteiden mahdollisuus on saatava hyvin pieneksi.

Konvoluutioneuroverkkojen ryhmän alle mahtuu valtava määrä erilaisia arkkitehtuureja. Muutamia yleisiä ovat Region-based Convolutional Neural Network (R-CNN), Fast R-CNN, Mask R-CNN, You Only Look Once (YOLO) - listaa voisi jatkaa loputtomiin, ja monesta on vielä useampi kehitysversio [4]. Ihmisenhavaitsemisverkon luominen on siis paljon monimutkaisempaa kuin tyypin valinta ja koon

määrittäminen. Neuroverkkoarkkitehtuureissa on vielä paljon vapautta, eivätkä parhaat käytännöt ole vielä lukkiintuneet.

Yksi havainnollistava esimerkki neuroverkkojen tekemistä virheistä näkyy kuvassa 2.8, johon liittyvässä tutkimuksessa testattiin, kuinka helppoa tekoälyä hyödyntävän valvontakamerajärjestelmän huijaaminen olisi.



Kuva 2.8: Taulu henkilön edessä huijaa kuvantunnistusta. [20]

Tarkoitukseen suunniteltu fyysinen taulu kaulassa riittää hämäämään tekoälyä, ja ihminen jää havaitsematta, vaikka järjestelmä normaalisti toimiikin tarkasti. [20] Tämän vuoksi pelkkään näkyvän valon kameraan perustuva ihmisentunnistus on niin ongelmallista. Tunnistus voi toimia täydellisesti lähes aina, mutta koneen työskennellessä ihmisen lähellä ei saisi täysin epäonnistua koskaan, ei edes kertaa miljoonassa. Kun katsoo kuvantunnistuksen tekemiä hassuja virheitä - kun ne luulevat koiraa appelsiiniksi tms. - teknologian tulevaisuus saattaa vaikuttaa toivottomalta.

On kuitenkin mielenkiintoista huomata, että ihmiset tekevät ensisilmäyksellään vastaavanlaisia virheitä. Google AI Recidency -ohjelmaan kuuluvassa tutkimuksessa *Adversarial Examples that Fool both ComputerVision and Time-Limited Humans* [21] väläytettiin koehenkilöille kuvia ja heidän tuli nappuloilla valita, kumpaan kahdesta ennakkoon annetusta kategoriasta kuva kuuluu (esim. kissa vai koira). Kuvia

kuitenkin näytettiin vain 63 millisekunnin ajan. Osaan kuvista lisättiin kohinaa sellaisiin taktisiin kohtiin, että se olisi saanut tekoälyn luokittelemaan kuvan väärin. Lisäksi näytettiin valekuvia, jotka eivät kuuluneet kumpaankaan annetusta kategoriasta, ja myös niitä käsiteltiin kohinalla. Havaittiin, että tekoälyyn tepsivä kuvankäsittely vaikutti myös ihmisten reaktionopeuteen ja virhemäärään ja valekuvien arvaustulokseen saatiin vaikutettua kohinalla. Tämä on merkityksellistä ja hieman hämmästyttävää, sillä kohinaan kompastuminen on sellaista absurdia käytöstä, jota odottaisi vain tekoälyltä, ei ihmiseltä.

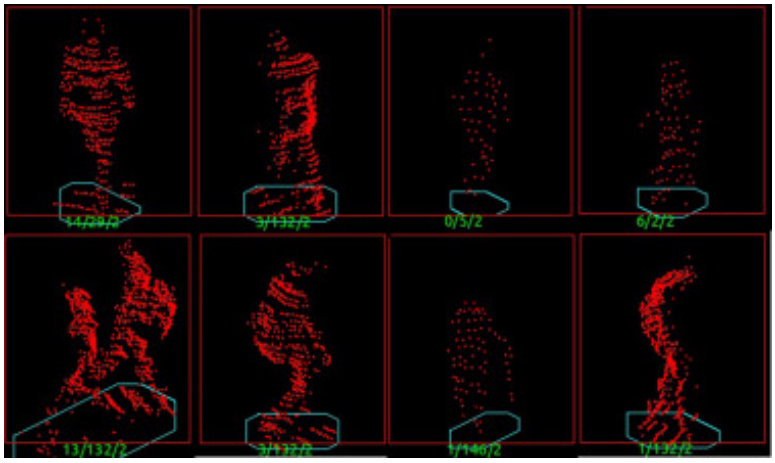
Ihmiset ovat toki kuvantunnistuksessa edelleen ylivoimaisia, mutta nämä tulokset osoittavat, että on ainakin mahdollisuus, että tekoälyn virheet karsiutuvat laitteiston nopeutuessa ja verkkojen koon kasvaessa. Toisin sanoen se, että tekoälyn päätöksentekoon pystyy vaikuttamaan pienellä kuvan hienosäädöllä, ei välttämättä tee järjestelmistä ”luonnottomia” ja pohjimmiltaan erilaisia kuin ihmisnäöstä. Ongelma saattaa johtua täysin suorituskyvyn puutteesta. Ei siis ole itsestään selvää, miten verkot saadaan paremmiksi: Tarvitseeko niiden olla vain suurempia vai hyvin erilaisia, vai molempia? Vai onko nopean kehityksen kannalta parasta lisätä eri sensoreita, joiden datasta ihminen on helpompi tunnistaa?

2.3.1 3D-datan erityispiirteet

Lidarpohjaiset järjestelmät ovat pohjimmiltaan erilaisia näkyvän valon kuvantunnistusjärjestelmiin verrattuna. Ne saavat ympäristön muodoista 3D-dataa, mutta eivät toisaalta havaitse värejä. Syvyysinformaation avulla objektien koon ja sijainnin määrittäminen on verrattain helppoa ja tarkkaa, mutta toisaalta niiden määrittämistä ihmiseksi vaikeuttaa mainittu väri-informaation puute sekä etenkin pistekartan hyvin karkea resoluutio.

Kameraan pohjautuva järjestelmä voi teoriassa määrittää hyvinkin tarkasti, mitä ihminen tekee, minne se katsoo ja minne se on todennäköisesti liikkumassa seu-

raavaksi. Lidarjärjestelmän alhainen tarkkuus tekee tästä paljon haastavampaa ja kuvaa (2.9) katsoessa tuntuu, että ihmiselläkin olisi täysi työ tunnistaa eri asennoissa olevat hahmot jalankulkijoiksi. On hyvä kuitenkin muistaa, että pistekartta on oikeasti kolmiulotteinen, eikä oheinen kuva siten täysin totuudenmukaisesti kuvasta neuroverkon käytettävissä olevaa dataa.



Kuva 2.9: Lidardataa käyttävän tekoälyjärjestelmän koulutukseen käytettyjä kuvia jalankulkijoista. [22]

2017 julkaistussa tutkimuksessa *Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle* [22] kehitettiin järjestelmää jalankulkijoiden tunnistamiseksi lidarin avulla. Rajallisesta resoluutiosta huolimatta järjestelmä kykeni muun muassa tien reunaa ja jalkojen asentoa havainnoimalla arvioimaan, onko jalankulkija astumassa tielle.

Lisäksi lidarjärjestelmätkin voivat hyödyntää tietoa ajasta ja liikkeestä. Jos objekti viime ruudussa tunnistettiin ihmiseksi, ja se on nyt vain liikkunut hieman, sen tiedetään olevan edelleen ihminen, eikä kategorisointia tarvitse käydä läpi uudelleen. Niinpä ei myöskään haittaa, vaikka yhdessä ruudussa objektista saataisiin niin vähän dataa, ettei sitä tunnisteta ihmiseksi. Kunhan vain jokin ruutu sisältää hyvää dataa riittävän lyhyen ajan sisällä, järjestelmä havaitsee ihmisen (suuremmitta viiveittä).

2.3.2 IR-kuvantunnistus

Kuten aiemmassa luvussa todettiin, infrapunakamera helpottaa tunnistusta sikäli, että ihminen loistaa kuvassa ympäristöön verrattuna. Infrapunakameran tuottama data on pohjimmiltaan videokuvaa, joten teknisesti se olisi helppo syöttää normaaliin näkyvän valon kuvia analysoivaan neuroverkkoon. Kuvat on kuitenkin värjätty täysin eri logiikalla: kirkkaat kohdat merkitsevät lämpimiä kohteita eivätkä valoa. Mikään, mitä tekoäly on oppinut kuvan kirkkauden (valon) käytöksestä objektien ympärillä, ei pädekään lämpökamerakuvan kohdalla. Tutkimuksessa *Human Detection in Thermal Imaging Using YOLO* [4] testattiin, voisiko verkkoja kuitenkin käyttää sellaisenaan, ilman uudelleenopetusta. Lopputuloksena oli, että näkyvän valon kameroilla opetettu YOLO-verkko ei tuottanut hyvää tarkkuutta lämpökameroiden kuvaa analysoidessa. Opetus lämpökamerakuvalla paransi suoritusta huomattavasti, tosin jääden silti huomattavasti näkyvän kuvan verkosta.

Ehkä odotetusti täysin ilman opetusta tai rajallisella opetuksella näkyvälle valolle tehty neuroverkko ei näytä soveltuvan lämpökamerakuvan analysointiin. Tästä olisi kuitenkin syytä tehdä lisätutkimusta. Kuinka paljon opetusta korjaus vaatii; voiko näkyvän valon verkkoa käyttää pohjana? Auttaako suorituskykyisempi laitteisto ja suurempi neuroverkko? Koska halutaan varmistua, että robotti toimii oikein kaikissa tilanteissa, olisi potentiaalisesti erittäin hyödyllistä, jos opetuksessa pystyttäisiin hyödyntämään näkyvän valon verkkoja, eikä tarvitsisi tukeutua vain suppeahkosti saatavilla olevaan lämpökamerakuvaan.

3 Useamman sensoritekniikan käyttäminen

Kuten edeltävästä luvusta käy ilmi, eri paikannustekniikoilla on hyvät ja huonot puolensa. Näkyvän valon kamerat voisivat toimia yksinään, mutta kuvankäsittelyn toimintaan ei voida ainakaan vielä luottaa riittävän suurella varmuudella. Lidar ja lämpökamerat taas eivät yksinään yllä riittävän suureen tarkkuuteen tai nopeaan toimintaan. Parasta olisi käyttää useampaa tekniikkaa yhtä aikaa, mutta miten yhdistää eri sensorien tuottama data?

3.1 Mahdolliset yhdistelmät

Paras sensoritekniikka riippuu käyttökohteesta, ja niinpä myös mahdollisia sensoriyhdistelmiä on lukuisia. Tunnistimme muutamia yhdistelmiä eri käyttötarkoituksiin, ja listaamme ne tässä, ennen kuin pohdimme datan yhdistämistä.

- Kaksi kameraa: stereokamera, mahdollistaa etäisyyden laskemisen. Etäisyyden lasku voidaan suorittaa ilman tekoälyä, jolloin objektien sijainnin hahmotus ei riipu tekoälystä.
- Näkyvän valon kamera ja lämpökamera: lämpökameran kuvasta havaitsee ihmisen.
- Yksittäinen näkyvän valon kamera ja lidar: lidar mittaa etäisyyden.

- Stereokamera ja lidar: stereokamera mittaa etäisyyden suuremmalla resoluutiolla ja nopeammin, mutta lidar tarkistaa mitat suuremmalla tarkkuudella.

Edellämainituissa järjestelmissä on nähtävissä kaksi kategoriaa: Yksissä kumpikin sensori tuottaa uniikkia dataa ja on välttämätön toiminnalle, joten niitä voisi kutsua yhdenvertaisiksi järjestelmiksi. Esimerkki tällaisesta järjestelmästä on sellainen, jossa on yksi näkyvän valon kamera ja lidar, ja jonka tekoäly ei kykene kuvasta arvioimaan objektien etäisyyttä. Tällöin molempien sensorien data on neuroverkon toiminnan kannalta välttämätöntä. Toisen kategorian järjestelmissä yksi sensori riittäisi toimintaan yksinään, ja toisen sensorin tuottama data varmentaa toimintaa tai tarkentaa tuloksia. Tällaisille yhdistelmille kuvaava nimi voisi olla hierarkkinen järjestelmä, sillä toinen sensoritekniikka ei ole välttämätön toiminnalle. Nimityksessä on tosin ongelmansa, sillä sen antaa helposti ymmärtää, että ristiriitatilanteessa primäärisensorin data painaa toisiosensorin dataa enemmän; näin ei kuitenkaan ole. Oleellista on vain, että toisiosensori tukee toimintaa. Hierarkkisessa järjestelmässä voi olla esimerkiksi näkyvän valon kamera yhdistettynä lämpökameraan, ja siinä neuroverkko kykenee tunnistamaan ihmisen näkyvän valon kameran kuvastakin, mutta lämpökameraa käytetään välttämään virheitä.

3.2 Sensoreiden datan yhdistäminen

Lämpökameran voi helposti kuvitella toimivan tavallisen kameran kanssa, sillä molemmat tuottavat kaksiulotteista kuvaa. Kunhan kamerat osoittavat samaan suuntaan, voidaan kuvia suoraan verrata. Intuitiivisinta ihmiselle lienee ajatella, että kuvat ovat päällekkäin. Neuroverkon näkökulmasta taas kuvan alueille löytyy värin informaation lisäksi lämpötila. Täytyy vain kiinnittää erityistä huomiota siihen, että kameroiden sijainti ja näkökentän laajuus (engl. *field of view*) ovat riittävän lähellä toisiaan, jotta kuvat olisivat vertailtavissa. Näitä aspekteja voidaan myös kom-

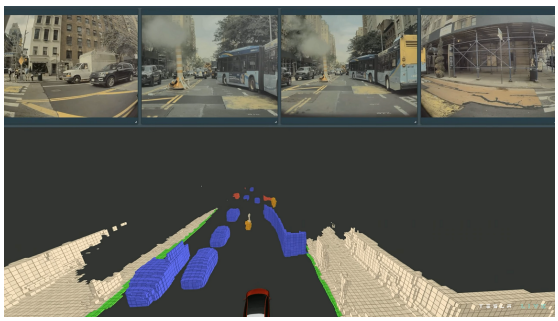
pensoida jälkiprosessoinnilla, jos virhe on tiedossa. On kuitenkin huomattava, että kameroiden sijainti saattaa heittää tuotannollisista syistä; eli ei voida olettaa, että kompensoitava virhe on aina sama, kuin mitä ollaan speksattu, vaan kamerat tarvitsee kalibroida. Tämä pätee myös useampaa näkyvän valon kameraa käytettäessä[23].

Yhtä triviaalia ei ole stereokameran ja lidarin datan yhdistäminen, koska stereokameran syvyystarkkuus ei riitä objektin muotojen hahmottamiseen ja lidar taas tallentaa vain joukon pisteitä. Molemmat tekniikat siis tuottavat 3D-dataa, mutta se on kummankin osalta niin vaillinaista, että yhdistäminen tarkasti on miltei mahdotonta. Jos stereokameran tuottamista tarkoista ääriviivoista tunnistetaan ihminen, sen etäisyys voidaan lidarin avulla toki mitata tarkemmin. Tässä kohtaa objekti on kuitenkin jo tunnistettu. Datojen yhdistäminen etukäteen tarkemmaksi 3D-maailmaksi, josta ihmisen havaitseminen olisi helpompaa, tuskin on mahdollista. Se vaatisi erityistä prosessointialgoritmia ja 3D-maailman tarkkuus rajoittuisi joka tapauksessa stereokameran tuottamiin ääriviivoihin ja lidarin tuottamiin pisteisiin. Toisin sanoen stereokameran data on tarkkaa vain siltä osin, että se perustuu näkyvän valon kuvaan; resoluutio on korkea, mutta syvyystarkkuus ei. Samoja hyötyjä siis saavutetaan, jos tekoälylle syötetään lidar-datan tueksi tavallista näkyvän valon kameran kuvaa.

2D-kuvan ja lidarin tuottaman 3D-datan yhdistäminen vaikuttaa haastavalta, mutta siihen on keinoja. Teslan on esitellyt AI Day -tapahtumissaan itseajojärjestelmäänsä, joka mallintaa kameroidensa kuvien perusteella kolmiulotteisen virtuaalisen vektoriavaruuden. Järjestelmässä kaksi kameraa osoittaa eteenpäin muodostaen stereokuvan, vaikka kameroissa onkin eri levyinen näkökenttä. Lisäksi auton ympärillä on kuusi muuta kameraa, niin että kuvaa saadaan koko 360 asteen alalta.[24]

Vektoriavaruudesta autojen liikkeiden hahmottaminen ja ennustaminen on huomattavasti tarkempaa kuin suoraan kameroiden kuvista. Kuvassa 3.1 nähdään Teslan järjestelmän dataa.

(a) Objektintunnistus vektoriavaruudessa [25]



(b) Hallussapitoverkko [26]

Kuva 3.1: Tesla Vision -järjestelmän muodostamaa 3D-dataa.

Vasemmalla on objektien tunnistusdataa (engl. *labeling*) vektoriavaruudessa. Oikealla näkyy tilan hallussapitoverkko (engl. *occupancy network*). Molemmat tekoäly on muodostanut kameroiden kuvista. Teslan järjestelmä siis muodostaa 3D-avaruuden puhtaasti toiminnan parantamiseksi, eikä lidaria ajatellen, mutta on helppoa nähdä kuinka se tarvittaessa mahdollistaisi myös lidarin lisäämisen järjestelmään.

4 Ihmisen liikkeiden huomioon ottaminen

Liikkeenseuranta (engl. *motion tracking*) on oleellinen osa ihmisiä havainnoivia järjestelmiä, sillä se helpottaa edellisissä luvuissa esiteltyjä objektin tunnistus- ja paikannustehtäviä. Käsittelemme tekniikkaa lyhyehkösti, sillä se on yleiskäyttöistä eikä vaadi suuria kustomointeja juuri tekoälykäyttöön. Objektien seuraaminen helpottaa tunnistusta ja paikannusta siksi, että sen avustaessa ympäröivää tilaa ei tarvitse joka ruudulla analysoida uudelleen tyhjästä. Yksinkertaisimmillaan se toimii siis ennen itse analysointia, rajaten sille siirtyvää dataa. Toisaalta kehittyneemmät järjestelmät voivat tallentaa liikkeitä ja niiden perusteella ennustaa, mitä ihminen on tekemässä ja minne liikkumassa.

4.1 Liikkeenseuranta

Liikkeenseuranta on monesta kuluttajakamerastakin tuttua teknologiaa. Kamera lukitsee tarkennuksen valittuun kohtaan, ja jos kohdassa oleva objekti liikkuu, kameran tarkennusalue seuraa. Autonomisissa laitteissa voi hyötyä samasta tekniikasta. Kyse on tavallaan muistin lisäämisestä tunnistusjärjestelmään, vaikka toteutuksessa käytettävä algoritmi voi olla yksinkertainenkin. Jos objekti havaitaan yhdessä ruudussa ihmiseksi, sen tiedetään seuraavassakin ruudussa olevan sitä edelleen. Kun käytetään takaisinkytkettyä neuroverkkoa (RRN), objektin tunnistaminen samaksi

kuin viime ruudussa on huomattavasti helpompaa kuin päättää tyhjältä pohjalta, onko kyseessä ihminen vai ei.

Näiden järjestelmien kohdalla puhutaan usein liikkeenseurannasta, vaikka periaatteessa kyse on objektin sijainnin seurannasta; liikkeitä ei taltioida mihinkään. Ero on siksi merkityksellinen, että objektille voidaan myös laskea varsinaiset liikeradat ja niiden avulla arvioida objektin sijainti tulevaisuudessa. Esimerkiksi aiemmin mainitussa tutkimuksessa, jossa lidaria käytettiin jalankulkijoiden tunnistamiseen [22], tallennettiin tunnistettujen ihmisten koordinaatit, ja useasta ruudusta tuotettuja koordinaatteja vertaamalla laskettiin ihmisen liikkeen nopeus ja suunta. Sen ja hahmon asennon avulla pystyttiin arvioimaan, onko jalankulkija astumassa tielle.

4.2 Käytöksen ennustaminen

Jos ajatellaan miten ihminen ennakoisi toisen henkilön liikkeitä, kyse on kuitenkin paljon enemmästä, kuin vain fyysisen liikeradan arvioinnista. Luonnollisesti katsomme, mitä toinen tekee. Kun työkoneen kuski havaitsee edessään työntekijän, ohjaa toimintaa se, saako hän luotua katsekontaktin. Eli merkittävää ei ole vain se, kuinka robotti havaitsee ihmisen, vaan myös se, onko ihminen havainnut robotin. Myös autonomisille koneille siis toisi merkittävää hyötyä ihmisen havainnoinnissa, jos ne pystyisivät arvioimaan ihmisen käytöstä. Yksi tapa arvioida, minne ihminen on liikkumassa, on juuri katseen seuraaminen. Muun muassa Teslan itseajaviin autoihin kehittämä järjestelmä pyrkii tunnistamaan, mihin ihmiset katsovat, ja sitä myötä arvioimaan, mihin he aikovat liikkua. Pyöräilijän pään liikkeet näkee selvästi, ja sitä myötä voi ennakoida, että hän on kääntymässä.

Toisaalta voidaan haluta tietää, mitä ihminen tekee, vaikka mitä työvaihetta hän on suorittamassa. Tämä voi helpottaa sen arvioinnissa, minne hän loogisesti on liikkumassa, sekä pidemmälle vietynä myös sen, kuinka robotti voi tehtävässä auttaa. Tekniikasta käytetään laajalti termiä *human action recognition* (HAR), jonka voi-

si kääntää toiminnantunnistukseksi. Toiminnantunnistus perustuu ihmisen asennon tunnistamiseen, joka ei pelkällä raa'alla sensoridatalla ole helppoa, kuten edeltävissä luvuissa käsiteltyjen haasteiden perusteella voi kuvitella. Apuna voidaan kuitenkin hyödyntää tietoa ihmiskehon liikeradoista (engl. *skeleton-based modeling*): Ihmisen asento selviää kuvasta helpommin, kun mittasuhteet, nivelien sijainnit ja liikeradat ynnä muu ovat karkeasti tiedossa. Ihmisen jalkaterä esimerkiksi osoittaa aina suurin piirtein samaan suuntaan kuin jalkakin, ei 90 astetta sivulle, ja näin voidaan sulkea pois suuri määrä virheellisiä vaihtoehtoja. Toiminnantunnistusta on tutkittu tehdaskäyttöä ajatellen, niin että robotti tunnistaa milloin ihminen on kumartunut tarkastelemaan kappaletta, kurottaa sitä kohti, näyttää sitä käsissään, tai viittoo seismerkin [27]. Toiminnantunnistusta on eri yhteyksissä tehty myös eri sensoritekniikoilla, kuten lämpökameroiden kuvasta [28].

5 Yhteenveto

Yhteenvetona voidaan todeta, että on vaikea löytää yhtä yksittäistä tekniikkaa, jolla autonominen kone havaitsisi ihmisen ja muuten pystyisi toimimaan luotettavasti. Näkyvän valon kamerat yhdistettynä tehokkaaseen tekoälyyn on yleiskäyttöisin tutkituista sensoritekniikoista, ja se soveltuu koko muuttuvan ympäristön hahmottamiseen, ei pelkästään ihmisten havainnointiin. Ongelmana on kuitenkin virheiden mahdollisuus, sillä ihmisen havaitseminen ja sijainnin määrittäminen nojaa täysin konenäköön. Lämpökameroiden, lidarin tai stereokamerajärjestelmän avulla ihmisen sijainti on helpompi todeta suoraan datasta, yksinkertaisemmalla prosessoinnilla. Toistaiseksi vaikuttaa, että eniten potentiaalia usean sensoritekniikan yhdistämisessä, jolloin näkyvän valon kameroihin perustuvan järjestelmän toimintavarmuutta saadaan tuettua. Pelkkiin näkyvän valon kameroihin perustuvat järjestelmät ovat tulevaisuudessa neuroverkkojen kehittyessä ja laskentatehon kasvaessa mahdollisia, mutta vielä toistaiseksi kehityksen aallonharjalla olevien järjestelmienkin toiminnassa voi tulla virheitä. Tekniikan jalkautuminen joka yrityksen saataville ottanee vielä kauemmin.

Yksinkertaisilla lämpökameraa tai lidaria hyödyntävillä ratkaisuilla on myös paikkansa, sillä järjestelmän ei välttämättä tarvitse havainnoida ympäristöä kovin tarkasti tai nopeasti. Lämpökamerat ja lidarit maksavat, mutta yksikköhinta muutamalle laitteelle tuotannossa ei usein ole ratkaiseva tekijä, jos järjestelmän kehitys on

halvempaa. Vaatii massavalmistusta, jotta kehityskustannuksista tulisi mitättömiä laitteiston hintaan nähden.

Nykytekniikalla aivan yksinkertaisten, magneettijuovaa seuraavien kuljetusrobottien käyttö voi silti olla käytännössä suurimmin näkyvä muutos. Ne tiputettiin yksinkertaisuutensa vuoksi tämän työn ulkopuolelle, mutta niitä on mahdollista hyödyntää monella alalla, ja pienet kustannukset mahdollistavat niiden käytön pieniinkin tehtäviin.

Lähdeluettelo

[1] MHI. ”Automatic Guided Vehicles”. (2024), url: https://www.mhi.org/fundamentals/automatic-guided-vehicles (viitattu 19. 04. 2021).

[2] M. Brownlee. ”Tesla Factory Tour with Elon Musk!” (21. elokuuta 2018), url: https://youtu.be/mr9kK0_7x08?t=404 (viitattu 19. 04. 2021).

[3] C. M. Grenier. ”Chinese Factory Worker”. (2024), url: https://www.flickr.com/photos/26087974@N05/9595016509 (viitattu 02. 06. 2024).

[4] M. Ivašić-Kos, M. Krišto ja M. Pobar, ”Human Detection in Thermal Imaging Using YOLO”, teoksessa *Proceedings of the 2019 5th International Conference on Computer and Technology Applications*, sarja ICCTA 2019, ACM, huhtikuu 2019. DOI: 10.1145/3323933.3324076.

[5] New Atlas. ”Subaru develops advanced stereoscopic vision system for cars”. (22. huhtikuuta 2010), url: https://newatlas.com/subaru-new-eyesight-stereoscopic-vision-system/14879/ (viitattu 19. 04. 2021).

[6] S. Crowe. ”Researchers back Tesla’s non-LiDAR approach to self-driving cars”, The Robot Report. (25. huhtikuuta 2019), url: https://www.therobotreport.com/researchers-back-teslas-non-lidar-approach-to-self-driving-cars/ (viitattu 19. 04. 2021).

[7] Tesla. ”Tesla Vision Update: Replacing Ultrasonic Sensors with Tesla Vision”. (2024), url: https://www.tesla.com/support/transitioning-tesla-vision (viitattu 05. 06. 2024).

[8] J. Yoshida. ”Subaru EyeSight Father Returns in Stereo Vision”, EE Times. (21. tammikuuta 2018), url: https://www.eetimes.com/subaru-eyesight-father-returns-in-stereo-vision/ (viitattu 12. 04. 2021).

[9] ITD Labs, *Intelligent Stereo Camera ISC-100VM, ISC-100XC*, lokakuu 2019. url: https://itdlab.com/wordpress/wp-content/uploads/2019/10/isc100vm_ENG.pdf (viitattu 19. 04. 2021).

[10] ITD Labs. ”Features of Intelligent Stereo Camera”. (2024), url: https://itdlab.com/wordpress/en/technology/ (viitattu 19. 04. 2021).

[11] Opticsplanet.com. ”A Look at Thermal Imaging vs. Night Vision”. (13. heinäkuuta 2016), url: https://www.opticsplanet.com/howto/how-to-thermal-imaging-vs-night-vision-devices.html (viitattu 14. 01. 2022).

[12] Opgal. ”Intro to IR (Part 3): Sensitivity, resolution and frame rate”. (huhtikuu 2018), url: https://www.opgal.com/blog/thermal-cameras/intro-to-ir-part-3-sensitivity-resolution-and-frame-rate/ (viitattu 11. 05. 2021).

[13] Flir. ”Flir infrared camera models’ specifications 2024”. (2024), url: https://www.flir.eu/browse/professional-tools/thermography-cameras/ (viitattu 18. 05. 2024).

[14] Hesai. ”Things you need to know about LiDAR: the more lasers, the better?” (21. maaliskuuta 2023), url: https://www.hesaitech.com/things-you-need-to-know-about-lidar-the-more-lasers-the-better/ (viitattu 18. 05. 2024).

[15] Pointcloud-tutkimushanke. ”Vaikutukset metsiin”. (2016), url: https://pointcloud.fi/vaikutukset-metsiin/ (viitattu 19. 04. 2021).

[16] Velodyne Lidar. "Velodyne Lidar HDL-64E S3 spec sheet". (2024), url: https://velodynelidar.com/downloads/#datasheets%5C%20first (viitattu 22. 10. 2021).

[17] S. Odeh. "Camera-Based Vision vs. LiDAR in Autonomous Vehicle Technology". (28. helmikuuta 2024), url: https://www.linkedin.com/pulse/camera-based-vision-vs-lidar-autonomous-vehicle-technology-odeh-d5lgf (viitattu 18. 05. 2024).

[18] P. Wang, W. Li, P. Ogunbona, J. Wan ja S. Escalera, "RGB-D-based human motion recognition with deep learning: A survey", *Computer Vision and Image Understanding*, vol. 171, s. 118–139, kesäkuu 2018, ISSN: 1077-3142. DOI: 10.1016/j.cviu.2018.04.007.

[19] F. van Veen. "Neural Network Zoo", The Asimov Institute. (syyskuu 2016), url: https://www.asimovinstitute.org/neural-network-zoo/ (viitattu 12. 04. 2021).

[20] S. Thys, W. V. Ranst ja T. Goedemé, "Fooling automated surveillance cameras: adversarial patches to attack person detection", 2019. DOI: 10.48550/arXiv.1904.08653.

[21] G. F. Elsayed, S. Shankar, B. Cheung et al., "Adversarial examples that Fool both Computer Vision and Time-Limited Humans", *Advances in Neural Information Processing Systems*, 2018. DOI: 10.48550/arXiv.1802.08195.

[22] H. Wang, B. Wang, B. Liu, X. Meng ja G. Yang, "Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle", *Robotics and Autonomous Systems*, vol. 88, s. 71–78, helmikuu 2017, ISSN: 0921-8890. DOI: 10.1016/j.robot.2016.11.014.

[23] Tesla. "Tesla AI Day 2021, 1:01:32, camera calibration". (20. elokuuta 2021), url: https://youtu.be/j0z4FweCy4M?t=3692 (viitattu 16. 06. 2024).

[24] Tesla. ”Tesla AI Day 2021, 0:49:04, vector space”. (20. elokuuta 2021), url: https://youtu.be/j0z4FweCy4M?t=2944 (viitattu 16. 06. 2024).

[25] Tesla. ”Tesla AI Day 2021, 1:32:53, auto labeling”. (20. elokuuta 2021), url: https://youtu.be/j0z4FweCy4M?t=5573 (viitattu 16. 06. 2024).

[26] Tesla. ”Tesla AI Day 2022, 1:15:49, occupancy network”. (1. lokakuuta 2022), url: https://youtu.be/ODSJsviD_SU?t=4549 (viitattu 16. 06. 2024).

[27] A. Roitberg, A. Perzylo, N. Somani, M. Giuliani, M. Rickert ja A. Knoll, ”Human activity recognition in the context of industrial human-robot interaction”, teoksessa *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, s. 1–10. DOI: 10.1109/APSIPA.2014.7041588.

[28] A. Akula, A. K. Shah ja R. Ghosh, ”Deep learning approach for human action recognition in infrared images”, *Cognitive Systems Research*, vol. 50, s. 146–154, elokuu 2018, ISSN: 1389-0417. DOI: 10.1016/j.cogsys.2018.04.002.