

Pistepilvien tehokas yhdistäminen

Pro Gradu
Turun yliopisto
Fysiikka
2024
Teemu Saukkio
Tarkastajat:
FT Paavo Nevalainen
Prof. Jyrki Piilo

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä

TURUN YLIOPISTO

Fysiikan laitos

Saukkio, Teemu Pistepilvien tehokas yhdistäminen

Pro Gradu, 67 s., 3 liites.

Fysiikka

Maaliskuu 2024

Pistejoukko on eräs yleinen fyysistä ympäristöä kuvaava datatyyppejä, joka tuotetaan monilla erilaisilla sensoreilla ja tekniikoilla, tai niiden yhdistelmällä. Eräs yleinen ongelma on erillisten pistejoukkojen yhdistäminen. Tällöin keskeisiä ongelmia ovat yksittäisten pisteiden pariuttaminen sillä oletuksella, että kullakin pisteellä on tietty identiteetti samalla kun pisteillä on myös tietty lähes vakio sijaintitarkkuus ja vaihteleva keskinäinen tiheys. On huomattava, että oletus pisteiden identiteetistä eri pistejoukoissa vaatii käytetyiltä tekniikoilta sen, että suuri osa pisteistä, ns. kiintopisteistä, edustaa jotain tunnistettua piirrettä maailmasta.

Tässä tutkimuksessa tarkastellaan pistejoukkojen pariutusongelmaa. Käytetyt pistejoukot ovat tutkimuksen piirissä tuotettu synteettisesti tai ovat peräisin metsäalueelta saadusta mittausaineistosta. Ratkaistujen pariutusten avulla voidaan löytää ongelmaan vaikuttavia muuttujia tai luoda laajempia pistejoukkoja yhdistämällä osittain päällekkäisiä joukkoja toisiinsa. Tutkimuksen aikana tarkasteltiin koordinaatiston valinnan vaikutusta, kun pariin oletetaan olevan lähin kohdejoukon piste, sekä tapoja suodattaa tulosjoukosta selkeästi vääriä tuloksia. Tarkasteltiin myös pariutusongelman kuvantamiseen soveltuvia menetelmiä ja niiden mahdollisuuksia pariutusongelmien vaikeuden ja tuloksen oikeellisuuden arviointiin.

Tutkimuksen tuloksena havaittiin helpohkojen pariutusongelmien olevan tunnistettavissa, jonka jälkeen ongelman ratkaisuun voidaan hyödyntää laskennallisesti kevyitä menetelmiä hyvän lopputuloksen aikaansaamiseksi. Näissä tapauksissa pariutuksen tarkkuus on käytännössä riittävä, ja laskenta erittäin nopea tutkimuksessa käsitellyin menetelmin. Esitetyt menetelmät eivät ratkaise vaikeampia ongelmia, mutta toisaalta nämä voidaan tunnistaa, ja vaikeiden ongelmien tapauksessa on mahdollista joko rajoittaa pisteiden tiheyttä valitsemalla vain varminpia ja selkeimpiä kiintopisteitä, tai käyttämällä parhaillaan kehittyviä joustavan sovittamisen menetelmiä, jotka ovat hivenen kalliimpia laskennallisesti.

Omana kappaleenaan esitetään tekoälymenetelmien soveltamista pistepilvien pariuttamisessa. Tekoälyn soveltamista varten muodostetaan käytettävissä oleva data koneoppimistarkoituksiin soveltuvaan muotoon, joka mahdollistaa menetelmien hyödyntämisen pariutustehtävässä. Lisäksi testataan ongelmaan sopivien koneoppimismallien suorituskykyä tutkimuksessa käytettävissä olevalla datalla.

Asiasanat: pistejoukot, pistejoukkojen yhdistäminen, lähin naapuri, vektorikenttä, tekoäly, koneoppiminen

Sisällys

Johdanto	1
1 Pistepilvet	4
1.1 Pistepilvien syntyminen	5
1.2 Vektorikenttä	9
1.3 Pisteparit	10
1.4 Pariutustehtävän kompleksisuus	14
2 Menetelmät	16
2.1 Lähin karteesinen naapuri	18
2.2 Pituussuodatus	20
2.3 Kaksisuuntainen suodatus	22
2.4 Suodatus paikallisella yhdenmukaisuudella	22
2.5 Tuloksen parantaminen iteroimalla	23
2.6 Risteysvertaus	25
2.7 Vektorifunktion arvojoukko	30
2.8 Siirtohaku	31
3 Tulokset synteettisellä datalla	35
3.1 Karteesinen haku	35
3.2 Siirtohaku	38
3.3 Hakumenetelmien vertailu	40
4 Tulokset LIDAR- datalla	42
5 Lopputulokset hakumenetelmistä	45
6 Koneoppimismenetelmien soveltaminen	49
6.1 Datarivien muodostaminen pistepilvistä	49

6.2	Hyperparametrit	51
6.3	Validointi	54
6.4	Pariutustehtävä	56
7	Pariutusaikojen tarkastelu	61
8	Johtopäätökset	63

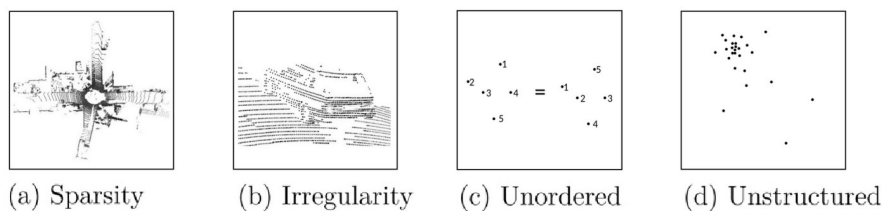
Johdanto

Pistepilvi on diskreetti kokoelma datapisteitä. Se voidaan kuvata muun muassa \mathbb{R}^2 tai \mathbb{R}^3 avaruuden koordinaatistossa. Pistepilvillä voidaan kuvata sen avaruudessa olevia muotoja jolloin pisteet kuvaavat muodon pintaa tai rajoja, tai pisteet voivat edustaa pelkkää sijaintia.

Pistepilviä käytetään esimerkiksi robotiikassa, tähtitieteessä ja konenäössä. Robotiikassa voitaisiin haluta verrata robotin sijaintisensorin ja tarkemman mittauksen välistä eroa määräämällä molempien menetelmien luomaan tietoon pistemäiset kohteet karttojen tärkeimpiin kohtiin ja etsimällä jokin kohdistus. Tähtitieteessä eräitä sovelluskohteita ovat tilanteet, joissa teleskooppien ottama kuva tähtitaivaasta halutaan yhdistää muihin kuviin samalta alueelta, siten luoden suurempia kuvia tähtitaivaasta. Konenäössä kahdessa eri sijainnissa olevien kameroiden näkemien kuvien yhdistämisellä saadaan laskettua syvyystieto, kun verrataan miten samat pisteet esiintyvät eri paikoissa suhteessa kameroihin ja siten voidaan määrittää näiden pisteiden etäisyys kameroista.

Pistepilviä ilmenee myös tietokonegrafikassa ja suunnittelussa. Tietokonepelin elementtejä varten voidaan luoda skannaus fyysisestä esineestä, jossa syntyneet mitauspisteet näyttävät tietyn pisteen esineen pinnalla tietokoneella olevassa \mathbb{R}^3 avaruudessa. Skannauksen tuloksena saaduista pisteistä voidaan yhdistää kokonaisia pintoja, jolloin saadaan aikaan haluttu esine digitaalisessa muodossa.

Skannauksessa voidaan käyttää esimerkiksi fotogrammetriaa tai valotutkaa (LIDAR light detection and ranging). Fotogrammetriassa kohteesta otetaan kaksiulotteisia kuvia jonka jälkeen kuvista etsitään tietyt pisteet, jolloin saadaan muodostettua eri kuvakulmista pisteiden sijainti ja niiden keskinäinen suhde toisiinsa [1].



Kuva 1. Pistepilvissä ja -joukoissa esiintyviä tyypillisiä piirteitä, joita tulee huomioida pistepilvi-ongelmia käsitellessä [2]. Kuvassa esitetään havainnollistava esimerkki harvasti jakautuneista pisteistä, epäsäännöllisesti sijoittuneista pisteistä, epäjärjestyksessä olevista pisteistä ja jakaumaltaan epätasaisesti sijoittuneista pisteistä. Kuvan käyttöön saatu lupa viitteen kirjoittajilta.

LIDAR-mittauksessa hyödynnetään laservalon valokeilan pientä kokoa ja sen valon tunnettua aallonpituutta heijastamalla lasersäteitä ympäristöön pulsseissa ja mittaamalla takaisinheijastukseen kulunut aika [3]. LIDAR-mittaus voidaan myös yhdistää paikkatietoon, jolloin puhutaan SLAM-mittauksesta (simultaneous localization and mapping), jossa LIDAR-mittaukseen yhdistetty paikkatieto parantaa mittauksen tarkkuutta [4].

Pistepilvien käsittelyssä esiintyy tyypillisesti muun muassa kuvassa 1 esiintyviä piirteitä. Pisteet ovat samassa avaruudessa, mutta pisteet ovat jakautuneet harvasti (sparsity), pisteet ovat epäsäännöllisesti sijoittuneet (irregularity), pisteet ovat epäjärjestyksessä (unordered) tai pistepilvessä ei ole rakenteisuutta (unstructured). Skannattaessa esimerkiksi rakennuksen sisätiloja, on tuloksena saatava pistepilvi rajoittunut rakennuksen sisätiloihin, jolloin pilviin voi muodostua alueita, joissa ei ole tuloksia, skannauksessa saatujen pisteiden ollessa epäsäännöllisesti sijoittuneita. Skannattavassa kohteessa voi olla heijastavia pintoja tai merkittävän kaukaisia kohteita, jotka johtavat alueisiin, joista pisteitä ei saada rekisteröityä. Kokonaispilvi on tällöin epäsäännöllinen. Skannaus on voitu tehdä eri kuvakulmista, jolloin saman kohteen piirteet ovat eri kuvakulmassa suhteessa skannauslaitteistoon, johtaen tällöin epäjärjestykseen pisteiden välillä tai skannattava kohde sisältää eri etäisyyksillä

olevia elementtejä, johtaen tiiviimpään rekisteröintiin lähellä olevissa komponenteissa, johtuen laasersäteiden hajonnasta suhteesta sen kulkemaan matkaan. Jokainen piirre voi esiintyä yhdessä ja erikseen.

Riippuen tavoitellusta lopputuloksesta, voivat nämä piirteet olla hyödyllisiä tai aiheuttaa haittaa. Esimerkiksi skannattaessa yksittäistä esinettä suuressa tilassa, voidaan kaukana mitattavasta kohteesta olevat asiat helposti erottaa mitattavasta kohteesta niiden mittauspisteiden harvuuden ansiosta. Rakennuksen sisätiloja mitattaessa taas voi mittaus epäonnistua, jos seinien yksityiskohdat katoavat, koska mittauspisteitä ei ole tarpeeksi.

Tässä tutkimuksessa keskityn tarkastelemaan yksittäisten pisteiden piirteitä ja ominaisuuksia, jättäen siten niistä muodostuvien pintojen ja muotojen tunnistuksen tutkimuksen ulkopuolelle. Tutkimuksen tavoitteena on tutkia pistepilvien piirteitä ja niiden käsittelyyn tarkoitettuja menetelmiä, jotka ovat laskennallisesti tehokkaita, sekä pyrkiä tarkastelemaan, onko pariutettujen pisteiden välillä olemassa yhteisiä piirteitä, joiden avulla voidaan saavuttaa ongelmaan geneerinen ratkaisu. Tutkimus keskittyy löytämään tapoja, jolla ongelman ratkaisun tuloksena saadaan mahdollisimman korkea määrä tuloksia mahdollisimman suurella varmuudella, priorisoimalla kuitenkin jälkimmäistä. Tutkimuksessa paremmaksi lopputulokseksi nähdään tulos, jossa löydetyt parit ovat mahdollisimman varmasti oikeita, välittämättä löytymättä jääneistä. Löytämällä mahdollisimman monta oikeaa paria voidaan tuloksia hyödyntää esimerkiksi etsimällä jokin funktio, jolla pisteiden kohdennusta voidaan parantaa. Tällaisen funktion löytämisen katsotaan olevan helpompaa tapauksessa, missä sen etsintään käytetty data sisältää mahdollisimman pienen määrän vääriä tuloksia. Funktion löytäminen on oleellinen osa ongelmaa, mutta sen etsiminen jätetään tutkimuksen ulkopuolelle.

1 Pistepilvet

Tutkimuksessa käytetty data on koordinaattimuodossa. Se siten sisältää koordinaatit kaikille sen avaruusakseleille. Käytettyihin pilviin viitataan käyttäen isoja kirjaimia A ja B , jolloin voidaan ilmoittaa tietty piste muodossa,

$$A_i = (x_i, y_i) \quad (1)$$

$$i = 1, \dots, n. \quad (2)$$

Tässä tutkimuksessa pidetään pistepilvi A kiinnitettynä, jolloin kahden pisteen välille voidaan ilmoittaa vektori, jonka origopiste on pisteessä A_i muodossa

$$V_i = B_i - A_i \quad (3)$$

$$i = 1, \dots, m \leq n, \text{ missä } m \text{ on pisteparien lukumäärä.} \quad (4)$$

Pisteiden välisistä vektoreista V käytetään termiä erotusvektori.

Tutkimuksessa käytetään synteettisesti luotua dataa, jossa luodaan ensin pistepilvi A , jolla on pistemäärä n . Luoduista pisteistä otetaan määrä m jota symboloi parametri $\omega = \frac{m}{n}$. Kun $\omega = 0$ kaikki pilvien A ja B pisteet ovat parittomia ja kun $\omega = 1$ kaikille pilven A pisteille on olemassa piste pilvessä B . Pilveen B luodaan myös parittomia pisteitä, jolloin molemmilla pilvillä on yhtä suuri pistemäärä n . Kun parametrin ω sanelema määrä pisteitä pilvestä A muutetaan pilven B pisteiksi, muutoksen laatua kontrolloidaan parametrilla ϵ joka on muutoksessa lisättävä satunnaistyyppinen melu sekä $vMax$ joka on suurin sallittu etäisyys, joka tuloksena saatavalla pisteellä B voi olla sen lähdepisteestä.

Käsitellessä kahta pistepilveä, jotka ovat täysin tai osittain samalla alueella, puhutaan tällöin pistepilvien A ja B leikkauksesta AB . Kahden pistepilven leikkaus voi olla osittainen, jolloin kokonaispilvestä selkeästi havaitaan alueita, joissa esiintyy

vain yhtä pilveä, tai täydellinen jolloin molemmat pilvet ovat jakautuneet suhteellisen tasaisesti koko alueelle. Tämän työn synteettisissä esimerkeissä alue AB kattaa molemmat pistepilvet, mutta LIDAR-mittauksissa alueen AB täsmällinen määrittely on laskennallinen ongelma, joka jätetään tämän tutkimuksen ulkopuolelle.

1.1 Pistepilvien syntyminen

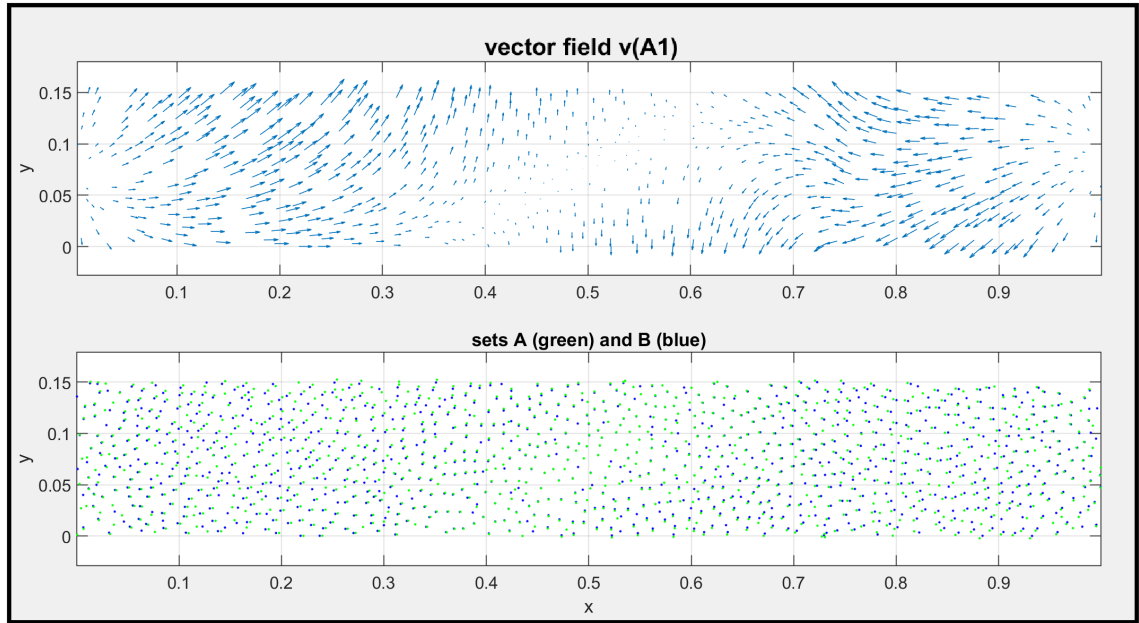
Kuten on aikaisemmin mainittu, pistepilvidataa luodaan eri yhteyksissä ja eri menetelmillä. Fotogrammetriamenetelmällä saatuja tuloksia pidetään usein epätarkkoina verrattuna LIDAR-mittauksella tuotettuun tapaan. Vaikkakin LIDAR on tarkempi, ei sekään ole täysin absoluuttinen menetelmä. LIDAR-mittauksessa voi esiintyä erinäisiä mittaustulosta heikentäviä seikkoja kuten mittaolosuhteet ja mitattavien pintojen tekstuuri. Menetelmässä esiintyy myös sen toteutustavasta riippuvia muuttujia jotka esittävät mittaustulokseen eri suuruista poikkeamaa [5].

Tutkimuksessa käytetyn synteettisen datan osalta nämä ilmiöt on pyritty ottamaan huomioon käyttäen aikaisemmin mainittuja parametreja. Tutkimuksessa käytettävän LIDAR-datan alkuperänä on SLAM-menetelmällä tehty mittaus todelliselta metsäalueelta. Tästä mittauksesta saatu data on esikäsitelty siten, että jokainen datapiste edustaa jo tunnistetun puun ydintä pisteenä kolmiulotteisessa avaruudessa. Merkittävä ero synteettiseen dataan on myös että pistepilvien A_n ja B_n koot voivat erota toisistaan. Lisäksi erotusvektorien joukon,

$$V = (b_{jk} - a_{ik}) | 1 \leq k \leq m \quad (5)$$

$$(6)$$

parien $(a_{ik}$ ja $b_{jk})$ indeksointi (ik, jk) sekä pariutuksen koko $m \leq \min(n_A, n_B)$ on tuntematon. Lisäksi joukot A ja B ovat erisuuria ja niiden yhteisalue AB on pitkänomainen ja epämääräinen.



Kuva 2. Datan generoinnin yhteydessä saatu visualisaatio pistepilvistä A ja B (alhaalla) sekä niiden väliin piirtyvät vektorit (ylhäällä). Kuvassa esiintyy kaikki oikean pariutustuloksen vektorit, joista on helppo havaita vektorikentän omaavan sille tyyppisiä piirteitä.

Käytetty synteettinen data luodaan käyttäen matlab -ohjelmaa. Algoritmista saadaan tuloksena tiedosto, joka sisältää jokaiselle pisteelle koordinaatit x ja y sekä luodun tuloksen visualisaatio jollainen esitetään kuvassa 2. Kuvassa pilvi A esitetään vihreänä pisteenä ja pilvi B sinisenä pisteenä kuvan alemmassa osassa. Pisteiden väliset vektorit esiintyvät kuvan ylemmässä osassa sinisinä nuolina. Synteettisessä datassa pilvien leikkaus on täydellinen, tarkoittaen molempien pistepilvien jakautuvan samalle alueelle tasaisesti. Synteettisesti luodusta datasta saadaan myös tieto oikeista parituksista tarkastelemalla pisteiden listausindeksiä. Pisteitä luodaan molempiin pilviin $n = 1000$ kpl mistä oikeat parit saadaan tietää parametrin ω mukaan.

Tutkimukseen luoduissa pistepilvissä parametrien ω , ϵ , $vMax$ arvo ilmoitetaan käytettävän tiedoston nimessä muodossa $PCaijk$ missä i, j ja k vastaavat edellä mainittujen parametrien taulukkoarvoja, kts. Taulukko I. Kerroin $l_0 = 0,0136$ on pilven A (ja myös pilven B) lähimpien naapuripisteiden keskimääräinen etäisyys. Metsädataesimerkissä $l_0 = 5,8$ m.

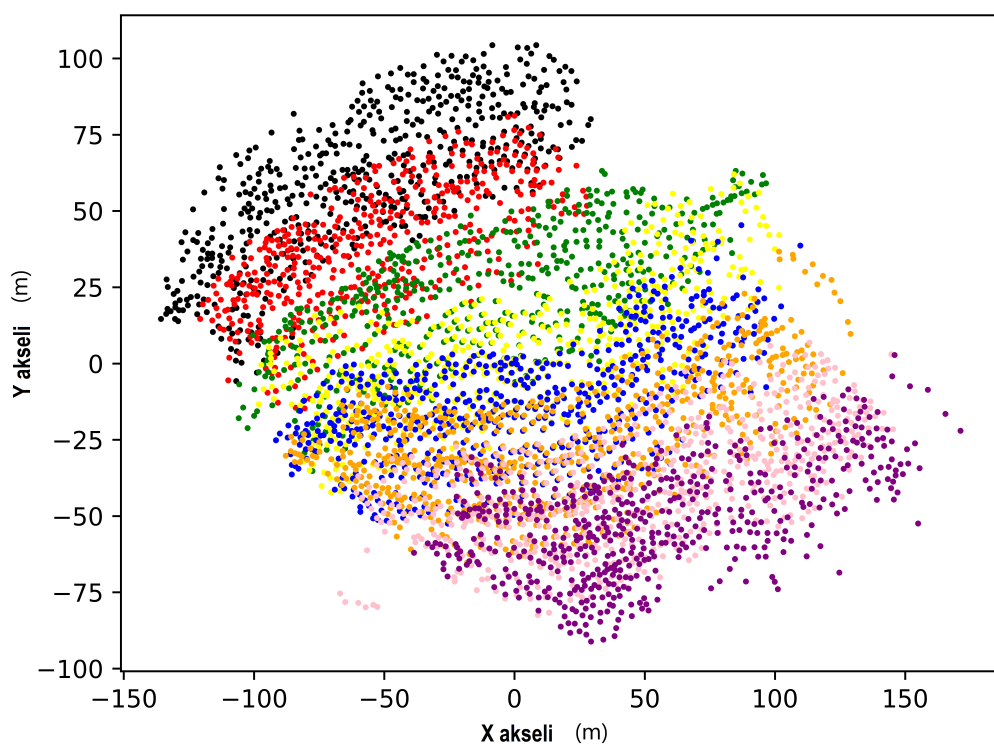
Taulukko I. Parametrien ja niille käytettyjen indeksien arvot datajoukkojen tiedostonimessä.

Parametri	Indeksi	Indeksin arvot	Parametrin arvot
ωn	i	1, 2, 3, 4, 5	980, 920, 850, 750, 600
$\frac{\epsilon}{l_0}$	j	1, 2, 3, 4, 5	0.01, 0.02, 0.04, 0.08, 0.16
$\frac{vMax}{l_0}$	k	1, 2, 3, 4	0.05, 0.1, 0.2, 0.4

Esimerkkijoukkojen indeksointi on toteutettu siten että tiedostot PCa111 ja PCb111 ovat helpoin pariutusongelma, tiedostot PCa554 ja PCb554 ovat vaikein pariutusongelma.

Johtuen tutkimuksessa käytetyn synteettisen datan akselien välisestä suhteesta (komponenttien maksimi arvot $y = 0,16, x = 1$), synteettistä dataa esittävisissä kuvissa $-y$ akseli on venytetty yksityiskohtien esittämisen helpottamiseksi. On muistettava että synteettisesti luodut pistepilvet A ja B ovat isotrooppisia ennen venytystä. Isotrooppisuus tässä yhteydessä tarkoittaa, että pisteiden ympäristö näyttää samalta kierroista ja siirroista huolimatta. Tutkimuksessa käytetyn LIDAR-datan osalta isotrooppisuus ei ole voimassa.

LIDAR-data, jota tutkimuksessa käytetään on esitetty kuvassa 3. Kuvassa esitetään useita erillisiä mittauskertoja samalla alueella erottamalla jokainen mittaus erivärisillä pisteillä. Tällöin pisteet yhdistettynä voidaan luoda koko alueen kattava pistekartta alueen puiden sijainnista. LIDAR-pisteet on annettu kolmiulotteisessa koordinaatistossa, mutta jokainen piste käsitellään tutkimuksessa siirrettynä kaksiulotteiselle pinnalle, siten jättäen niiden korkeus erot huomiotta. Alkuperäinen mittaus on toteutettu metsäkoneeseen kiinnitetyllä LIDAR ja paikannus laitteella [6]. LIDAR- ja paikkatiedon yhdistämistä kutsutaan SLAM (Simultaneous localization



Kuva 3. Tutkimuksen käyttöön saatu LIDAR-data. LIDAR-pisteet on esikäsitelty pistejoukoista pisteiksi tutkimuskäyttöön. Alkuperäiset LIDAR- pisteet sisältävät myös korkeustiedon, mutta kuvassa ne esitetään kaksiulotteiseen tasoon siirrettynä.

and mapping) tehtäväksi. SLAM on laskentatehtävä, joka esiintyy esimerkiksi robottien sijainnin määrittämisessä, SLAM-mittaus voidaan toteuttaa käyttäen lukuisia eri yhdistelmiä laitteista ja sensoreista, mittaus voidaan tehdä LIDARin lisäksi kameralla ja sijaintitiedon saamiseksi voidaan käyttää ohjaustietoa tai muuta tiedettyä informaatiota. Tutkimuksen datan keräämiseen käytetyt mittaus- ja paikannusmenetelmät ovat kuitenkin luonteeltaan sellaisia että tavalliset SLAM-ongelman piirteet eivät sisälly tutkimuksen käsittelyyn, johtuen siitä että metsädatan pisteet ovat jo prosessoidun SLAM- menetelmän lopputulos. Siksi tutkimuksessa voidaan keskittyä pelkkään pariutusongelmaan [4].

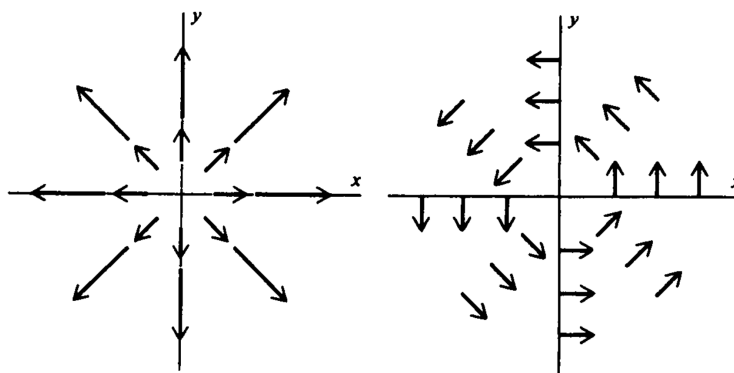
1.2 Vektorikenttä

Kuten jo aikaisemmin on todettu, voidaan kahden pisteen väliin piirtää erotusvektori. Koko pilvien leikkauksen alueella esiintyvät erotusvektorit voidaan kuvata myös vektorikenttänä joka esitettiin kuvassa 2. Vektorikenttä on matemaattinen funktio, joka määrää vektorin jokaiselle avaruuden pisteelle. Tässä oletetaan erotusvektorikentän olevan melkein kaikkialla sileän, ja joissakin paikoissa siinä on lähteitä, nieluja, spiraalimaisia sekä täydellisiä pyörteitä ja viivamaisia epäjatkuvuuskohtia.

Vektorikentässä voi olla lähteitä, nieluja ja pyörteitä. Lisäksi edellisten yhdistelmiä kuten spiraalipisteet ja viivamaisia tihentymiä [8]. Näistä käsitellään lähteet, nielut ja pyörteet. Lähdepisteestä ja pyörteestä esitetään esimerkki kuvassa 4.

Yhdistelemällä eri vektorikentän piirteitä saadaan seurauksena vektorikenttä, joka vaihtelee koko alueella pyörteen, lähteen ja nielun välillä, sekä näiden yhdistelmänä.

Tutkimuksessa käytettävän synteettisen datan tapauksessa parien välisen vektorikentän tiedetään olevan sileä funktion ja siihen lisättävän satunnaisen melun

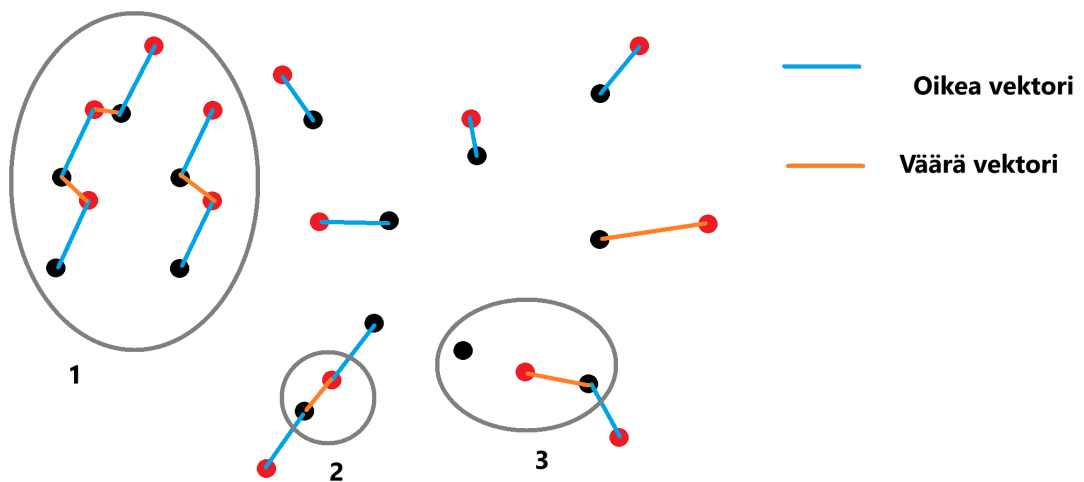


Kuva 4. Vektorikentässä esiintyvä lähdepiste vasemmalla ja pyörre oikealla. Nielu-piste vastaa negatiivista lähdettä [7].

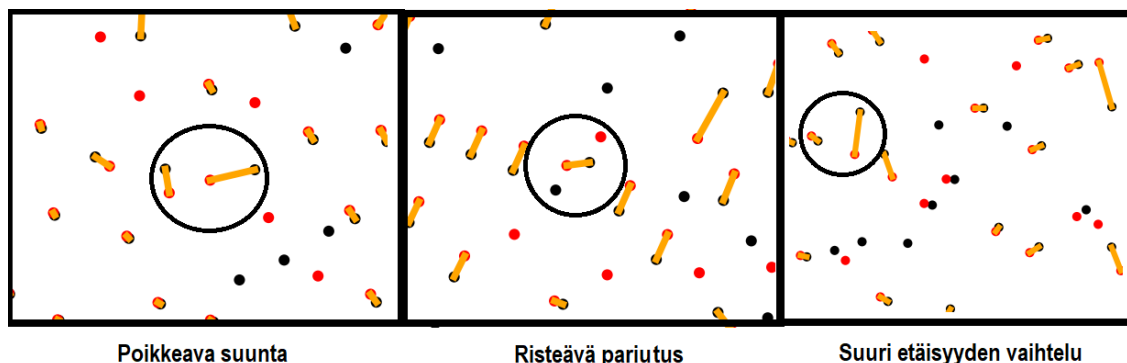
yhdistelmä. Ei voida kuitenkaan ilman erillistä todistusta tietää onko vektorikenttä yleistetyssä tapauksessa tapa esittää pisteiden väli, vai kuvaus ilmiön synnyttävästä mekanismista. Tutkimuksessa kuitenkin oletetaan, että pariutusongelmien erotusvektorit voidaan kuvata käyttäen lähes kaikkialla sileää vektorikenttäfunktioita.

1.3 Pisteparit

Tutkimuksessa tarkastellaan pääasiassa kysymystä, ovatko piste A_i ja B_j pari. Kun kaksi pistettä ovat pari, ne kuvaavat tällöin samaa paikkaa kahdesta eri näkökulmasta. Pariutuksessa tehtävää vaikeuttavana tekijänä esiintyy esimerkiksi pisteiden mahdollinen vääristyminen, jolloin oikeita pisteitä ei pystytä helposti havaitsemaan ihmisen toimesta. Eräitä tutkimuksen aikana havaittuja ilmiöitä, joiden käsittelyyn pyritään löytämään menetelmä, ovat kuvassa 5 esiintyvät ilmiöt. Kuvassa numerolla yksi merkitty on esimerkki tapauksesta, jossa pisteiden siirtymä on suuri, tämän havaittiin tutkimuksessa olevan yksi suurimmista vaikuttajista hakutulokseen. Numerolla kaksi merkityssä tilanteessa pisteitä on samassa alueessa useita ilman selviä merkkejä niiden välisistä eroista, pisteet voivat olla tällöin tiheästi sijoittuneet paikallisessa ikkunassa. Kun pilvet saattavat sisältää myös parittomia pisteitä, voi olla haastavaa päätellä onko tilanteessa kyse kahdesta parista vai yhdestä parista sekä kahdesta parittomasta pisteestä. Viimeinen tilanne kuvassa on esimerkkitapaus



Kuva 5. Havainnollistava kuva jossa esitetään eräitä tutkimuksen yhteydessä havaittuja pisteiden sijoittumiseen liittyviä ilmiöitä. Näiden ilmiöiden avulla voidaan tunnistaa pariutustehtävässä esiintyvät haasteet, joiden ohittamiseksi tutkimuksen menetelmien pirteitä ja ominaisuuksia tutkitaan. 1 pisteiden etäisyys suuri pistetiheyteen verrattaessa, pariutus tapahtuu lähempänä olevaan väärään pisteeseen joukossa pisteitä. 2 suuri paikallinen pistetiheys, pariutus tapahtuu väärään pisteeseen yhdessä parissa. 3 pisteiden siirtymä lähellä A pisteiden etäisyyksien puoliväliä, pariutus epävarma ja se voi kohdistua oikeaan tai väärään pisteeseen.



Kuva 6. Esimerkkejä tyypillisistä pariutus virheistä jotka voivat olla selkeästi havaittavissa visuaalisella tarkastelulla. Pariutuksen tuloksena saatujen vektorien suunta on epäjohdonmukainen lähiympäristön alueella, vektorit eivät siten seuraa esimerkiksi pyörteen tai lähteen suuntaa (vasemmalla). Johdonmukaisesti jakautuneiden vektorien alueella havaitaan yksittäinen poikkeava pariutus, toiset vektorit noudattavat pyörre- tai lähdepisteen suuntaa (keskellä). Pariutuksen tuloksena saadun vektorin pituus poikkeaa merkittävästi lähiympäristöstä (oikealla).

miten parittomien pisteiden samankaltainen jakauma alueella voi johtaa tilanteisiin, missä todellisuudessa pariton piste saattaa pariutua jonkin pisteen kanssa, jos sen sijainti osuu pariutuksen kannalta epäedukkaaseen paikkaan, tällöin B pisteiden etäisyys A pisteistä sijoittuu paikallisesti A pisteiden keskinäisten etäisyyksien puoliväliin, johtaen tasaisesti jakautuneeseen pistetiheyteen, jos molemmat pisteet huomioidaan.

Jotta pariutustulosta voitaisiin arvostella, tulee löytää jokin menetelmä mitata tuloksen laatua. Yksinkertaisissa tapauksissa kuten siirtymän vastatessa jäykän kappaleen siirtoa, voidaan tehtävä ratkaista yleensä vertaamalla pisteiden muodostamia kuvioita. Ihmisen tekemää havainnointia voidaan kuitenkin pitää kokonaisvaltaisesti huonona mittarina, sillä ihmisillä on taipumus löytää etsimiään kuvioita [9]. Kuvaajasta tehtyjä havaintoja voidaan kuitenkin käyttää apuna arvioidessa yksittäisiä selkeitä tapauksia, joista esitetään eräitä kuvassa 6. Poikkeava suunta ja risteävä pariutus voidaan nähdä virheenä, jos vektorikentän oletetaan olevan sileä. Tällöin on odotettavissa, että vektorien suunta esiintyy johdonmukaisena lokaalis-

sa ympäristössä. Voidaan myös olettaa, ettei vektorien pituuksissa esiinny suurta vaihtelua paikallisesti. On mahdollista luoda rajoittamattoman vaikeita tehtäviä luomalla synteettistä dataa, valitsemalla pieni parimäärä ω , suuri melu ϵ ja suuri pariutusetaisyysien amplitudi $vMax$, myös niissä tapauksissa, kun pariutukseen käytetty vektorikenttä V on suhteellisen sileä.

Pariutusta tarkastellessa saattaa esiintyä myös tilanteita, joissa pariutus voi olla oikein ja täysin ristiriidassa ympäristössä vallitsevien vektoreiden kanssa. Tämän takia pariutusongelmaa tarkastellaan pitäen tavoitteena tuloksen varmentamista tärkeimpänä. Eri pariutustuloksissa siten hyväksytään tulos, jonka pariutuksen voidaan katsoa olevan oikein korkealla varmuudella. Tutkimuksen kannalta oleellisimpia jatkotoimia katsotaan olevan tuloksena saatuun pariutukseen funktion sovittaminen ja pistepilvien välisen paikoituksen tarkentaminen. On myös oleellista ettei haku tuota hyvältä vaikuttavaa tulosta hakutehtävän ollessa erittäin huono tai mahdoton.

Pariutusongelman kannalta on oleellista, että pisteiden väliltä on poistettu suurimmat virhe-elementit, kuten jäykän kappaleen liike, sen ollessa mahdollista. Todellisissa pariutusongelmissa, jäykän kappaleen liikettä voi esiintyä, esimerkiksi asennon arvioinnin virheenä, joka voi syntyä esimerkiksi mittauslaitteiden kertyvästä virheestä (pose estimation error). Paikkavirhe voi tehdä pariutusongelman mahdottomaksi, esimerkiksi siirtämällä pisteitä toisiinsa nähden siinä määrin että parien välissä on useita muita pisteitä. Näissä tapauksissa on oleellista pyrkiä tunnistamaan miten kertyvät virhetyypit ilmenevät ja pyrkiä poistamaan virheen vaikutus, käyttäen siihen soveltuvia menetelmiä. Tutkimuksessa käytetty data on esisovitettu, siten että jäykänkappaleen liikkeen kaltaiset vaikutteet voidaan jättää tutkimuksen ulkopuolelle. Sovittamiseen soveltuva menetelmä joka sallii pitkän hakumatkan, on esimerkiksi Lego-Loam, jonka avulla data voidaan saattaa muotoon jossa sen pariuttaminen on

jo mahdollista käyttäen tutkimuksessa käytettyjä menetelmiä [10].

1.4 Pariutustehtävän kompleksisuus

Graafiteoreettisessa kirjallisuudessa [11] kompleksisuusanalyysit esitetään ongelmalle, jossa joukon koko on $2n$ ja kaikki pisteet pariutetaan (tämän työn parametrein: $\omega = 1$). Pariutuksen laatukriteeri (kustannusfunktio) on kaikkien parien välimatkojen summa. Tämän teoreettisen tehtävämuotoilun tulokset kuitenkin pätevät esillä olevaan ongelmaan.

Teoreettisen perustehtävän kompleksisuus on avoin. Kaksijakoisessa (bipartite) - ongelmassa pisteet jaetaan kahteen joukkoon (esim. punaiset ja siniset pisteet, tai kuten tämän työn tapauksessa joukot A ja B), ja pariutus tapahtuu pelkästään näiden välillä. Kompleksisuus on tällöinkin sama kuin yleisellä tapauksella. Geometrisella $1 + \epsilon$ -tarkkuudella on saatu aikaan algoritmi, joka laskee kaksijakoisen tehtävän ajassa $\mathcal{O}((n/\epsilon)^{1.5} \log^5(n/\epsilon^2))$ [12].

Tutkimuksessa esitetyssä käytännön sovelluskohteessa (metsäkoneen tunnistamien puiden sovittaminen yhteen eri ajokerroilla) on teoreettisesta perustehtävästä poiketen seuraavat lisävaatimukset:

Jako ei koske kaikkia pisteitä, vaan kummassakin joukossa A ja B on tuntematon määrä pisteitä, jotka eivät osallistu pariutukseen. Joukkojen suuruus on vain suurpiirteisesti sama $|A| \approx |B|$ ja lisäksi tehtävä ei ole täysin symmetrinen: $\omega_A = |AB|/|A| \approx \omega_B = |AB|/|B|$, missä $|B|$ on B joukon mahtavuus. Suureet ω_A ja ω_B luonnehtivat ongelmaa, ja ovat analyysin lopputulos, eivät alkusyöte. Kun parien lukumäärä laskee ($\omega \rightarrow 0$), tehtävä lähenee osajoukon valintatehtävää (etsitään osajoukko $A_1 \subset A$, joka tyydyttää tietyn loogisen testin $P(A_1)$), ja tällaisen alijou-

koetsinnän kompleksisuus on $O(2^{|A|})$.

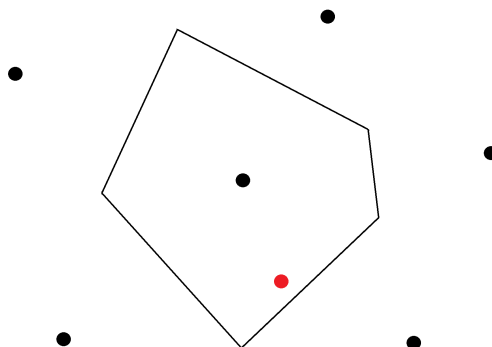
Pariutus voi kohdistua muuhun kuin lähipisteeseen, eli parin $(a, b) \in AB$ välimatka $\|b - a\|$ voi olla pidempi kuin $\|b_i - a\|$, jossa b_i on a :ta lähin piste B :ssä. Tätä kuvataan parametrilla v_{max} , joka on suurin esiintyvä parietäisyys. Kun v_{max} kasvaa, muuttuu tehtävä vaikeammaksi (eri kompleksisuusluokkaan todellisten parien siir-tyessä kauemmas kuin lähipisteet).

Parien muodostamat vektorit $v(a) = b - a$ ovat jossain määrin riippuvuudessa ympäristön vektorikentän kanssa, ja tämä helpottaa tehtävää [13], [14]. Pariutusvektorikentän koherenssi (sileys) huononee, kun pistejoukkojen A ja B pisteiden paikan mittausepävarmuus (melu) ϵ kasvaa. Tämä ϵ on eri kuin tässä kirjallisuuskatsauksessa aiemmin käytetty. Kirjallisuudessa ϵ merkitsee kustannuksen arvioinnissa sallittua suhteellista virhettä.

Yhteenvetona, ideaaliseen [12] -ongelmaan verrattuna käsitellään tässä työssä seuraavaa tapausta:

1. $\omega < 1$ eli pariutus on osittainen
2. $v_{max} > \delta$ eli pisteparin etäisyys voi olla suurempi kuin keskimääräinen naapuripisteiden etäisyys δ joukon sisällä.
3. $\epsilon \approx \delta$ eli pisteen mittaustarkkuus voi olla lähellä pisteiden välistä etäisyyttä.

Työn tarkoituksena on tutkia, miten suhteellisen yksinkertainen ja nopea pariutusalgorithmi, jonka kompleksisuus rajoittuu luokkaan $\mathcal{O}(n)$, menettää vähitellen käytettävyytensä tehtävän vaikeutuessa parametrien ω , v_{max} ja ϵ muuttuessa.



Kuva 7. Piste B (punainen) Voronoin solun alueella. Voronoin solu kattaa sen alueen, jossa solun määrittelevä A piste (kuvassa keskellä) on lähin joukon A piste.

2 Menetelmät

Pistepilvien pariutukseen voidaan käyttää useita menetelmiä, joilla on jokaisella omat ominaisuutensa ja heikkoutensa. Pareja voidaan hakea jokaisen pisteen lähiympäristöstä, jolloin puhutaan lähimmän naapurin menetelmästä. Pareja voidaan hakea myös hyödyntämällä vektorikentässä esiintyviä piirteitä, esimerkiksi jakamalla koko pistepilvi alueisiin, joista pyritään tunnistamaan paikallinen ominaismuoto, joka kuvaa tämän alueen vektorien suunnan ja suuruuden pääpiirteitä. Pääpiirteiden avulla voidaan tunnistaa ne pisteet, joiden välissä on saman kaltainen vektori, jolloin ne voidaan olettaa pariksi. Pareja voidaan etsiä myös jakamalla jokaisen kiinnitetyn pisteen A ympäristö Voronoin soluihin. Voronoin solujen avulla voidaan päätellä pilven B pisteelle tämän pisteen mahdollisuus olla alueen A pisteen pari. Voronoin solut tuotetaan rajaamalla pilven A pisteiden väli linjoilla, jotka sijoitetaan puoliväliin kaikkien A pisteiden välille. Esimerkki Voronoin solusta esitetään kuvassa 7 [6].

Pistepilvien elastisesta pariuttamisesta löytyy tutkimusta pääasiassa käyttäen kompleksisia esitysmuotoja [15] ja laskennallisesti raskaita menetelmiä [16]. Lisäksi on tutkimuksia tilanteista, joissa pyritään tunnistamaan kolmiulotteisia muotoja ja pintoja hyödyntäen pistepilviä [17].

Autonomisten koneiden yleistyessä pistepilvien ratkaiseminen johtaa tarpeeseen kasvattaa käytettyä laskentatehoa tai löytää laskennallisesti edullisempi menetelmä ongelman ratkaisemiseksi. Tutkimuksessa keskitytään pääasiassa menetelmiin, joiden laskentatehovaatimuksia voidaan pitää kevyenä. Tutkimus voidaan silloin suunnata piirteisiin ja ominaisuuksiin, joita tehtävässä ilmenee. Tällöin saatetaan löytää hyödyllistä tietoa ongelman geneerisen ratkaisun löytymistä varten.

Pistepilven alkuperästä ja koostumuksesta huolimatta sen käsittely voidaan karkeasti jakaa hakuun, suodatukseen, jälkikäsitteilyyn ja tuloksen laadun toteamiseen. Hakuun kuuluu itse algoritmi, jonka perusteella tehdään päätös, mikä piste pistepilvestä B saattaa olla pari pilven A pisteelle. Haku voidaan tehdä esimerkiksi jakamalla pisteiden ympäristöt soluihin, etsimällä lähin piste tai etsimällä pisteiden jakaumasta symmetrioita sekä samankaltaisuuksia. Hakutuloksen voidaan olettaa sisältävän virheitä, joiden poistamiseksi voidaan käyttää erilaisia suodatusmenetelmiä. Pisteitä voidaan siirtää tavoitellen joko merkkejä pariutuksen onnistumisesta, tai hakien molemmille pilville lopullista sijaintia. Jälkikäsitteilyssä voidaan myös esimerkiksi tarkastella tuloksen järjestyttä pienemmässä paikallisessa ikkunassa, jolloin voidaan arvioida vaihtoehtoisten parien sopivuutta. Laadun toteamisessa pyritään löytämään loogisia todisteita lopputuloksen järjestydestä. Pariuttaessa pistepilviä, prosessin eri vaiheita voidaan paikoitellen käyttää muiden vaiheiden tukena. Siten yllä mainittua jaottelua ei voida pitää työn suorituspolkuna sovitusta ratkaisusta.

Mikäli tutkimuksessa saataisiin tuloksena luotettava menetelmä pariutuksen varmentamiseen, voidaan tämä nähdä jo hyödyllisenä saavutuksena. Tutkimuksessa käytetyn synteettisen datan varmentaminen on triviaalia, sillä sen todellinen pari-

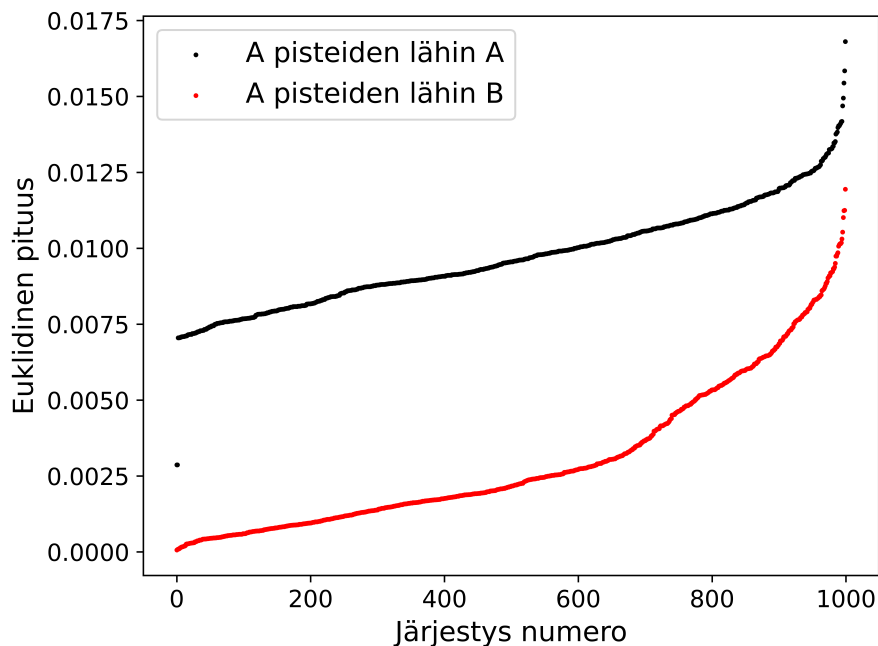
tus tiedetään ja se säilyy pilviä käsitellessä. Todellisten datajoukkojen tapauksessa pariutusta ei varmuudella voida vahvistaa. Voidaan kuitenkin pyrkiä löytämään menetelmiä synteettisellä datalla, joiden perusteella voidaan todeta pariutuksen olevan oikein määritettävissä olevalla varmuudella ja käyttää näitä menetelmiä todellisiin sovelluskohteisiin.

2.1 Lähin karteellinen naapuri

Kuten aikaisemmin mainittu, eräs tapa etsiä A pisteille naapuria, on etsiä sitä lähinnä oleva B piste. Menetelmän perusajatuksena on havaita jokin mahdollinen B piste, jonka etäisyys poikkeaa selkeästi A pisteiden välisistä etäisyyksistä. Tämä ilmiö voidaan esittää kuvana hakemalla kaikkien A pisteiden lähin A piste ja listamalla euklidiset etäisyydet, sekä hakemalla lisäksi etäisyydet A pisteistä lähimpään B pisteeseen. Järjestämällä etäisyydet pienimmästä suurempaan saadaan kuva 8, josta helposti havaitaan tietyissä tapauksissa menetelmän todennäköisesti toimivan kiitettävästi. Kuvan luontiin on käytetty synteettistä dataa tunnisteella 523, joka voidaan kääntää parimäärän, melun ja $vMax$ arvoiksi aiemmin esiteltyyn tapaan. Suuri joukko B pisteiden etäisyyksiä sijoittuu pienimpien A pisteiden etäisyyksien alapuolelle, jolloin voidaan olettaa parin löytyvän etäisyydeltä, joka on merkittävästi pienempi, kun etäisyydet muihin A pisteisiin.

Vertaamalla tiedoston 523 tuloksia tiedostoon 554, jossa parametrien arvot ovat huippuarvoissa, havaitaan mahdollisia ongelmatilanteita tämän hakumenetelmän toiminnassa. Suuremmilla parametriarvoilla ei voida enää selkeästi rajata mahdollisia etäisyyksiä B pisteiden etäisyyksien sijoittuessa lähemmän A pisteiden etäisyyksien tasoa kuten havaitaan kuvasta 9.

Hakualgoritmin toteutuksessa käydään jokaiselle lähtöpilven pisteelle läpi kaikki

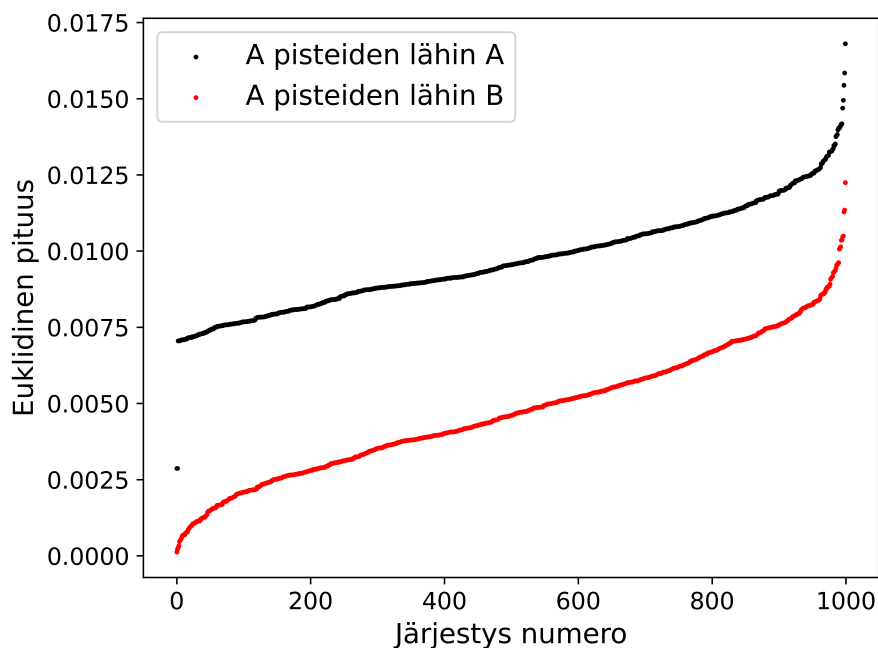


Kuva 8. Pistepilven PCa523 pisteiden etäisyys lähimpään oman pilven pisteeseen sekä pilven PCa523 pisteiden etäisyys lähimpään pilven PCb523 pisteeseen järjestyssä listassa. Pilven *A* etäisyydet merkitty mustalla ja pilven *B* etäisyydet merkitty punaisella.

vertailupilven pisteet, joista valitaan se piste, johon euklidinen etäisyys on lyhin. Euklidinen etäisyys saadaan laskettua kaavalla,

$$l_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (7)$$

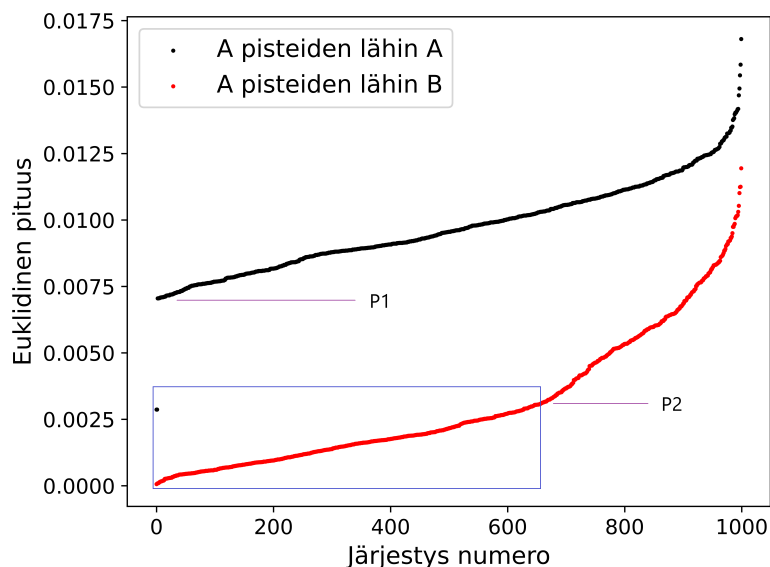
Lähimmän karteetisen naapurin haku perustuu oletukseen pilven *A* pisteiden välisten etäisyyksien tasaisesta ja yhtenäisestä jakaumasta, muodostaen siten jatkuvan käyrän kuvaajassa jonkin minimi- ja maksimietäisyyden välillä. Menetelmä muistuttaa lähestymistapaa, jossa pyritään tunnistamaan molemmista pilvistä muotoja ja piirteitä, tässä menetelmässä piirteenä voidaan katsoa olevan tasaisesti jakautunut etäisyys [18].



Kuva 9. Pistepilven PCa554 pisteiden etäisyys lähimpään oman pilven pisteeseen sekä pilven PCa554 pisteiden etäisyys lähimpään pilven PCb554 pisteeseen järjestyssä listassa. Pilven *A* etäisyyden merkitty mustalla ja pilven *B* etäisyydet merkitty punaisella.

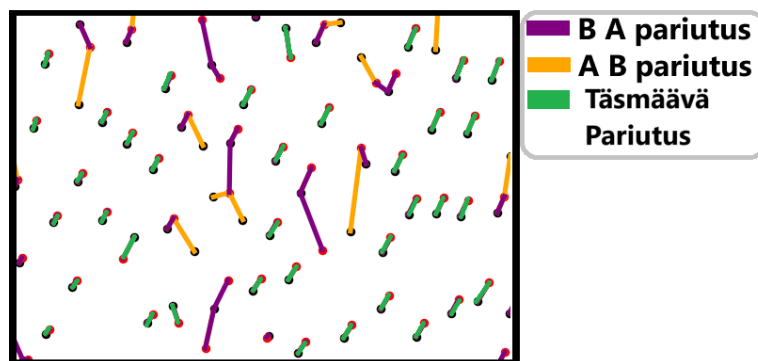
2.2 Pituussuodatus

Eräs sovellettava suodatus on poistaa pariutustulokset, joiden erotusvektorin pituus ylittää raja-arvon. Raja-arvon määrittäminen voidaan toteuttaa tarkastelemalla naapureiden välisten vektorien pituuksia kuvaajassa. Kuvaajasta etsitään pilvien *A* ja *B* välisen pariutuksen pituuksista se osuus, jossa pituudet jakautuvat tasaisesti, lyhin etäisyys pilven *A* pariin välillä on myös huomioitava. Mikäli pituudet jakautuvat tasaisesti alueella, jonka maksimiarvona on $p_2 = \frac{P_1}{2}$ voidaan P_2 valita rajaksi, missä P_1 on *A* pilven lyhin etäisyys osuudelta jossa pituusjakauma on tasainen. Rajan määrittäminen eräissä tapauksissa esitetään kuvassa 10. Kuvassa havaittu kuvaajan alue on korostettu sinisellä neliöllä sekä P_1 ja P_2 arvot korostettu lilalla viivalla. Piste P_2 soveltuu rajaksi sillä se on etäisyys, jossa piste pilvestä *B* on todennäköisemmin lähempänä pistettä pilvestä *A* kuin toista pistettä pilvestä *B*, sillä molempien pilvien etäisyysjakauma on tällöin samankaltainen.



Kuva 10. Esitys pituus suodatus rajan löytämisestä kuvaajasta. Kuvaajassa esitetään $p1$ -arvo, joka vastaa lyhintä huomioon otettavaa pilven A pisteiden keskinäistä etäisyyttä, sekä $p2$ joka vastaa suuruudeltaan puolta $p1$ -arvosta, jonka rajaamat etäisyydet ovat liilalla merkityn laatikon sisällä.

Synteettisen datan parametrien ollessa suurempia, muuttuu rajaaminen asteittain hankalammaksi, kuvaajaan muodostuvan punaisen käyrän muuttaessa muotoaan. Datan luontiin käytetyt parametrit eivät vaikuta pilven A luomiseen, jolloin sen kuvaajan muoto ei merkittävästi muutu datajoukkojen välillä. Tätä ominaisuutta käytettiin hyväksi tutkimuksessa määrittämällä sopiva raja parametrien suhteen edukkaimmilla tapauksilla, joissa sen määrittäminen on helpompaa. Näin määritettyä rajaa käytetään kaikissa tapauksissa. Raja voidaan myös nostaa korkeammaksi, jolloin hakutuloksen voidaan odottaa sisältävän enemmän virheellisiä tuloksia. Tämä voi olla oikea ratkaisu, mikäli voidaan käyttää myös muita suodatus metodeja, jolloin pituusrajaaminen toimii suodatuksen ensimmäisenä vaiheena, poistaen selkeästi väärät tulokset.



Kuva 11. Havainnollistava esitys eri pariutuksista saatavien vektoreiden eroista. Suodattamaton pariutus pilvestä A pilveen B (keltaisella) ja pilvestä B pilveen A (liilalla) sekä pariutuksen ollessa sama molemmissa tilanteissa (vihreällä).

2.3 Kaksisuuntainen suodatus

Suodatus voidaan toteuttaa myös hyödyntäen pariien välillä olevaa symmetriaa. Tilanteessa, missä pariien välinen etäisyys on lyhin mahdollinen etäisyys kahden pisteen välillä, silloin mikäli pisteet a_1 ja b_4 ovat pari, on etäisyys l_{14} lyhyempi kuin pisteiden a_1 ja b_i välillä oleva etäisyys l_{1i} , missä $i \in 0, 1, 2, 3, 5, 6, \dots, n$. Tällöin oletetaan että $l_{14} \leq l_{4j}$ joka on etäisyys pisteiden B_4 ja A_j välillä kun $j \in 0, 2, 3, 4, 5, 6, \dots, n$. Lähimmän naapurin pisteestä A pisteeseen B oletetaan olevan siten sama pariutus, kuin lähin naapuri pisteestä B pisteeseen A . Mikäli ehto ei toteudu, suodatetaan pariutustulos pois, sillä sen katsotaan olevan epävarma tai virheellinen. Eräs esimerkki ilmiöstä on esitetty kuvassa 11, jossa nähdään pariutuksia, joissa oletus lähimmästä naapurista toteutuu pisteillä, joilla havaitaan ainoastaan liila vektori. Virheellisesti pariutuneilla pisteillä havaitaan liiloja sekä keltaisia vektoreja.

2.4 Suodatus paikallisella yhdenmukaisuudella

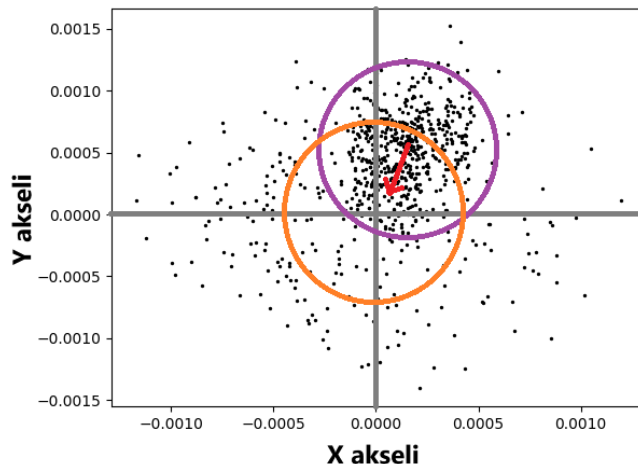
Hakualgoritmin antamaa tulosta voidaan suodattaa myös tarkastelemalla sen loogisuutta jakamalla koko alue pienempiin osiin, sekä vertaamalla löydettyjen vektorien johdonmukaisuutta. Oletetaan koko vektorikentän muodostuvan useista lähteistä, nieluista ja pyörteistä, jolloin koko kentän tarkastelu tulisi tehdä ottamalla huo-

mioon kaikki vektoriin vaikuttavat elementit. Näitä elementtejä ei kuitenkaan läh-
tökohtaisesti tiedetä, jolloin tarkastelu koko alueella ei siten onnistu. Pienemmillä
alueilla puolestaan havaitaan todennäköisemmin alueita, joissa kaikkien vektorien
voidaan katsoa noudattavat yhtä lähdeettä, nielua tai pyörrettä, ollen siten saman-
kaltaisia suunnaltaan ja suuruudeltaan. Tarkasteluun tarvittavan alueen koko riip-
puu huomattavasti pariutuksessa muodostuneen vektorikentän kompleksisuudesta.
Voi olla myös tarpeellista tehdä useampi vertailu, joissa määritetään alueet, jotka
kattavat alkuperäisten alueiden rajat. Tällainen vertailu on toteutettavissa esim.
pistepilven A delaunay -kolmioinnin kytkentämatriisin ominaismuotojen avulla [19].
Menetelmä kasvattaa kuitenkin ongelman kompleksisuutta, jonka takia se on jätetty
tämän tutkimuksen ulkopuolelle [6].

2.5 Tuloksen parantaminen iteroimalla

Suodatuksen lisäksi voidaan hakutulosta varmentaa testaamalla sen stabiiliutta.
Eräs tällainen menetelmä on toistaa hakualgoritmi useita kertoja, pyrkien muok-
kaamaan pilvien välistä sopivuutta löydettyjen erotusvektorien perusteella. Ensim-
mäisen hakutuloksen erotusvektoreista voidaan määrittää sopiva siirtovektori, jonka
avulla jokaista pilven B pistettä siirretään. Toistamalla menetelmää useita kertoja
siirtyy erotusvektorien jakauma tasaisemmaksi kaikkiin suuntiin suhteessa pilven A
pisteisiin. Erityisesti tilanteissa, joissa vektorikenttä muodostuu jäykän kappaleen
liikkeistä tai lähellä sitä olevista liikkeistä, tulee menetelmä löytämään lähes täydellisen
sovituksen.

Tutkimuksen materiaalina käytetyn synteettisen datan luomiseen käytetty vekto-
rifunktio ei kuitenkaan ole luonteeltaan jäykän kappaleen liike. Kuvassa 12 esitetään
iteroinnin tuloksena saavutettu siirtymä erotusvektorien jakaumassa.



Kuva 12. Tavoiteltu vaikutus tehtäessä pilven B pisteiden siirto iteroimalla erotusvektorien pituuden avulla. Tiheän pituusarvoja esittävän joukon (ympyröity liilalla) odotettu siirto (ympyröity keltaisella).

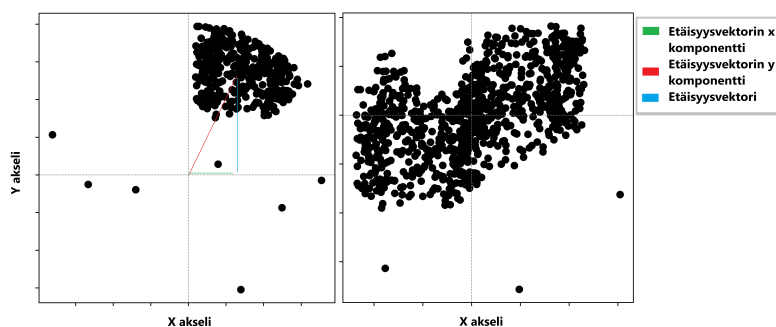
Siirtoon käytetty vektori voidaan määrittää laskemalla tuloksen erotusvektorien komponenttien keskiarvo, joka tulee skaalata sopivalla vakiolla. Menetelmä on herkkä ylisovittumiselle (engl. overfitting) sekä alisovittumiselle (engl. underfitting), jonka takia käytetyn skaalausvakion valinta on oleellista. Yli- ja alisovittuminen on seurausta lokaaleista maksimeista ja minimeistä, joita pituuksien keskiarvossa voi esiintyä.

Tutkimuksessa testatussa menetelmässä laskettiin vektorin komponentteja v_x ja v_y hyödyntämällä siirto kaavalla,

$$B' = (B_x - v_x n, B_y - v_y n). \quad (8)$$

Missä vakio n on skaalaukseen käytetty vakio ja B' siirretty pistepilvi B .

Menetelmän käyttäessä erotusvektorien pituuden keskiarvoa mittana paremmuudesta, voidaan lokaalit minimi- ja maksimi-ilmiöt ohittaa suorittamalla varmuus askelia minimiarvon löytyessä. Täten löydetty minimi edustaa siirtoa, jossa kaikkien



Kuva 13. Lähtötilanne vasemmalla ja lopputilanne oikealla. Iteroinnin vaikutus, erotusvektorien keskiarvon (punaisella) komponenttien (vihreä x , sininen y) avulla, erotusvektorien pituuden jakaumaan.

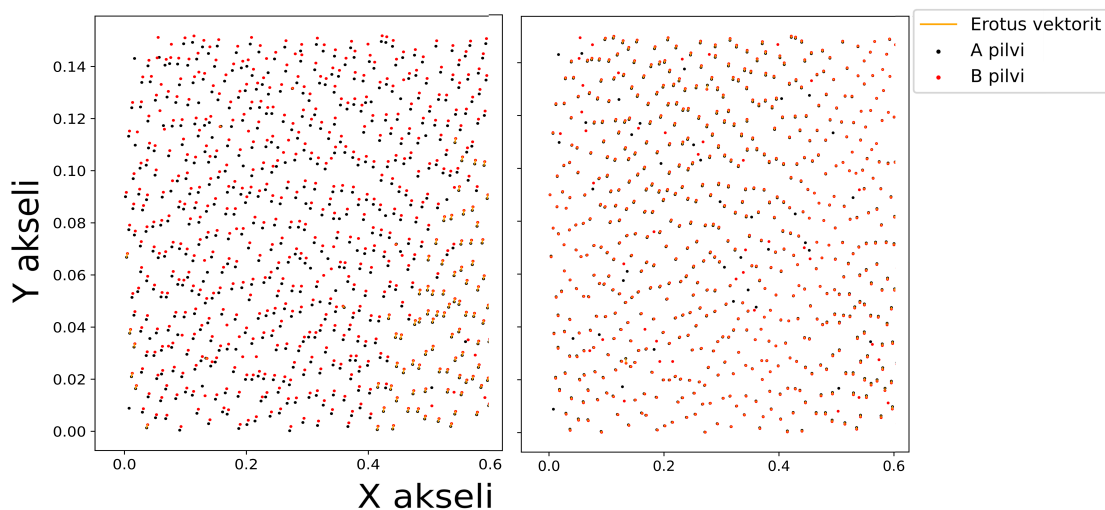
pariutustulosten erotusvektorien keskiarvo saavuttaa pienimmän arvonsa. Toisin sanoen niiden pituusjakauma on tasaisimmillaan pilven A pisteiden suhteen.

Kuvassa 13 vasemmalla on esitetty alkuperäisen pariutustuloksen erotusvektorien keskiarvo punaisella ja sen komponentit x vihreällä ja y sinisellä.

Parhaan iteraation jälkeen saadaan tuloksena oikeanpuoleinen tasaisempi jakauma erotusvektoreille. Jakauman tasoittamiseen osallistuu myös iteraation vaikutuksena löytyvät uudet pariutustulokset, jotka voidaan helposti havaita kuvasta 14, jossa esitetään kuvan 13 pariutustulos ennen iterointia vasemmalla ja iteroinnin jälkeen oikealla.

2.6 Risteysvertaus

Vektorikentän mainittuja oleellisia ominaisuuksia olivat lähteet, nielut ja pyörteet. On siten perusteltua olettaa, että jokaiselle vektorikentän synnyttämälle vektorille on olemassa jokin näistä. On myös perusteltua olettaa, että jokaiselle lähteelle, nielulle ja pyörteelle kuuluu useampia vektoreita koko pistepilven alueelta. Siten nähdään, että eräs menetelmä tuloksen löytämiseksi tai varmistamiseksi on havaita nämä lähdepisteet ja niihin sidoksissa olevat erotusvektorit.

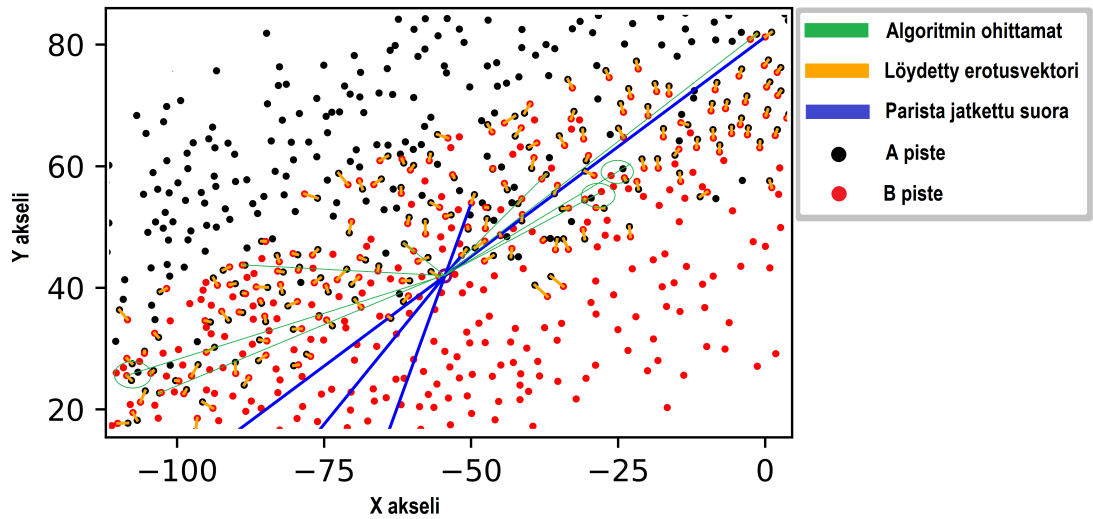


Kuva 14. Pisteet A ja B ennen iterointia vasemmalla ja iteroinnin jälkeen oikealla. Iteraatio lyhentää pariutettujen pisteiden välimatkaa merkittävästi. Kuvassa esitetään otos esimerkkinä toimivasta pariutusongelmasta menetelmän havainnollistamiseksi.

Jos etsitään kaikkien mahdollisten pistekombinaatioiden kaikkien muiden kombinaatioiden kanssa löytyviä leikkauspisteitä, puhutaan laskennallisesti erittäin raskaasta tehtävästä. Laskennallisesti edukkaampi lähestymistapa on pyrkiä tunnistamaan pisteet perustuen hakualgoritmin tulokseen ja mahdollisesti pyrkiä parantamaan kokonaistulosta ja tarkkuutta perustuen löydettyihin lähdepisteisiin. Koska vektorit sisältävät todennäköisesti myös melua, on etsintä toteutettava tämä huomioiden.

Tätä menetelmää kokeiltiin LIDAR-pisteiden hakuun menetelmän validoimiseksi ja sen havaittiin olevan käyttökelpoinen erityisesti LIDAR-datassa, sillä sen voidaan ajatella saavan alkunsa pisteestä, joka vastaa mittauslaitteen sen hetkistä todellista sijaintia.

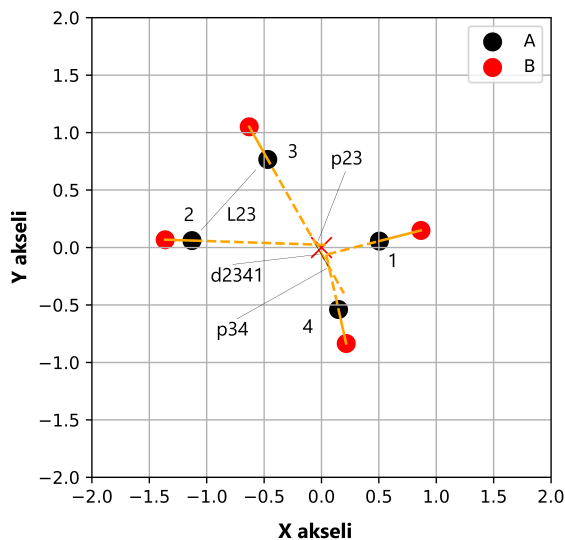
Kuvassa 15 on eräs kokeillun algoritmin tuottamista tuloksista, jossa sinisellä on esitetty pariutuksen tuloksena saadun vektorin osoittavan liilalla ympyröityyn



Kuva 15. Pariutustuloksesta jatkettujen suorien (merkitty sinisellä) risteysalue (ympyröity liilalla) ja sen avulla havaittuja ilmiöitä pariutustehtävässä. Tulokseen (vihreällä) merkittyjen havaintojen perusteella nähdään käytetyn algoritmin sisältävän virheitä, joiden johdosta mahdollisia alueeseen sopivia pareja on ohitettu.

risteysalueeseen, pariutustuloksien osoittaman alueen löytämiseksi pariutuksen erotusvektoria jatketaan alueen rajojen reunalle käyttäen B pistettä aloituspisteenä josta jatkettu vektori kulkee läpi A pisteen pilven ääri rajoille saakka. Risteysalueet määrittävät jatkettujen vektorien leikkauspisteistä. Vihreällä on esitetty hakualgoritmin ja vertausalgoritmin virheitä. Vertausalgoritmi on jättänyt joitain löydettyjä pareja jatkamatta ja joitain ei pariutettuja pisteitä voitaisiin tämän alueen näkökulmasta myös määrittää pariksi, tuottamalla niille jatkettut vektorit joiden odotetaan kulkevan risteys alueen kautta.

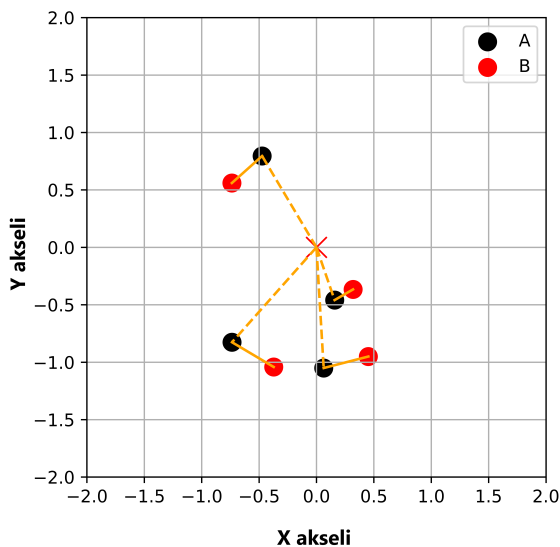
Mikäli pystytään havaitsemaan selkeitä paikallisia pisteitä, joiden ympäristössä olevat pariutukset noudattavat selkeää johdonmukaisuutta, voidaan tätä käyttää avuksi mahdollisen melutermien ϵ arvioimiseksi, pariutustuloksen varmentamiseksi tai pariutustuloksen paikalliseen silittämiseen. Melutermien arvioimiseksi tulee löytää samalla alueella esiintyvien leikkauksien piste, jonka ympäristö voidaan rajata



Kuva 16. Kaikki lokaalit pisteet samasta lähteestä. Pisteet pilvestä A mustalla, pisteet pilvestä B punaisella. Kuvasta voidaan havaita pilven B pisteiden siirtyvän kauemmas pilven A pisteistä pitkin suoria, jotka kulkevat niiden leikkauspisteistä (merkitty punaisella rastilla) pilven A pisteiden lävitse.

kohtuullisen pienisäteisellä ympyrällä. Kuvassa 16 esitetään eräs mahdollinen lokaali johdonmukainen piste, joka esitetään kuvassa punaisella rastilla merkittynä. Sen ympäristössä olevien pisteiden (A_a ja B_a) pariutus, joka on merkitty oranssilla viivalla, voidaan jatkaa kohti origoa katkoviivalla, jolloin pystytään havaitsemaan pariutustuloksien leikkauspisteitä p_{34} ja p_{23} origon ympäristössä, A pisteiden välinen pisin etäisyys L_{23} ja leikkauspisteiden välinen etäisyys d_{2341} joita voidaan käyttää avuksi melutermin arviointiin.

Samanlainen tarkastelu voidaan tehdä etsimällä pisteille lokaalia pyörähdyspistettä, kuvassa 17. Pyörähdyspisteen löytyessä voidaan edetä esimerkiksi ratkaisemalla jokaisen pariutustuloksen tuottamiseen tarvittava kulma ja pyöräytyksien keskipisteiden välinen heitto. Löytämällä kulmat ja keskipisteet, joiden keskenään verrattu vaihtelu on yhdenmukainen, voidaan hyväksyä kaikille yhteinen kulma ja origo, jonka jälkeen pisteet voidaan pyöräyttää takaisin saavuttaen tila, jossa erotusvektot



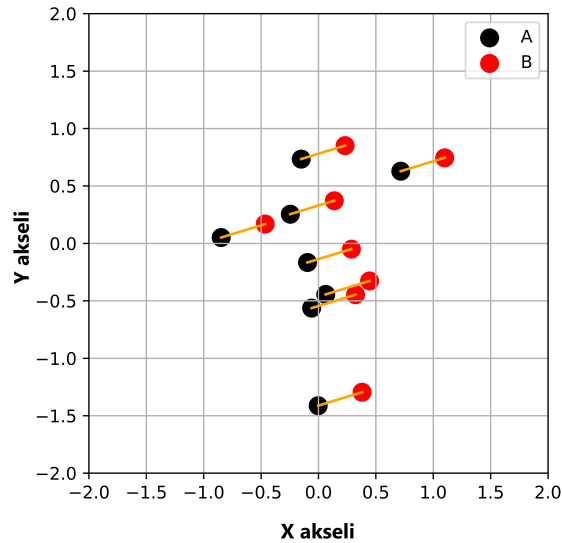
Kuva 17. Kaikki lokaalit pisteet jäykän kappaleen pyörähdyksessä. Pisteet pilvestä A mustalla, pisteet pilvestä B punaisella. Kuvassa kaikkien pilven B pisteiden voidaan havaita pyörähtävän saman pisteen (merkitty punaisella rastilla) ympärillä.

rien maksimi pituus on lähellä vaihtelun maksimia.

Yhdenmukaiseksi siirroksi voidaan katsoa pariutukset, joiden erotusvektorit ovat yhdensuuntaiset eli niiden lähdepiste tai pyörähdyspiste on pisteiden väliseen etäisyyteen verrattuna kaukana, esimerkki esitetty kuvassa 18. Tällöin pisteiden pariutus voidaan tasoittaa lokaalisti löytämällä kaikille pisteille sopiva siirto, joka minimoi pariutusvektorien summan, saavuttaen tällöin lokaalisti sileämpi tulos.

Toistamalla lokaaleja tasoituksia toisiaan leikkaavilla alueilla, voidaan lopulta saavuttaa kokonaistuloksena sileämpi pariutus koko alueella.

Menetelmän voidaan odottaa omaavan hyödyllisiä piirteitä, mutta sen laskennallinen raskaus on kuitenkin merkittävä myös pariutustuloksesta aloitettuna ja tästä syystä menetelmän syvempi tarkastelu jätetään tutkimuksen ulkopuolelle.

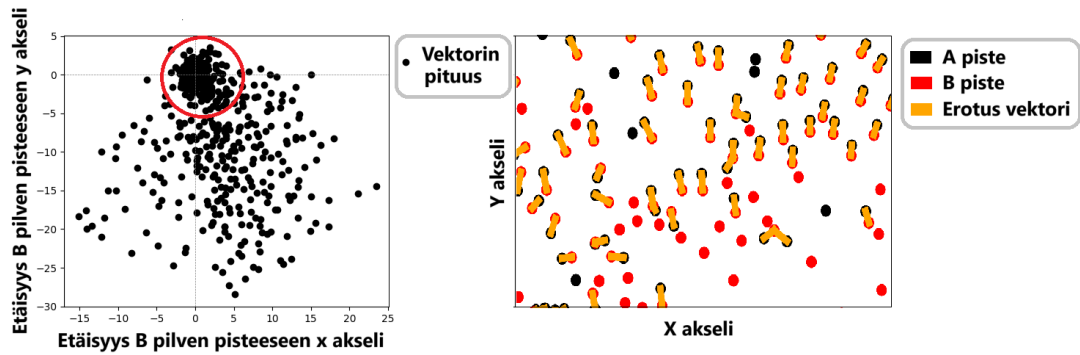


Kuva 18. Kaikki lokaalit pisteet jäykän kappaleen siirrossa. Pisteet pilvestä A mustalla, pisteet pilvestä B punaisella. Kuvassa pilven B pisteiden voidaan havaita siirtyvän saman suuntaisesti pois päin pilven A pisteistä.

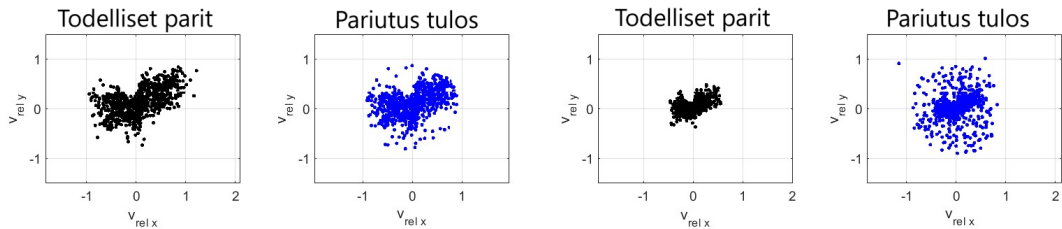
2.7 Vektorifunktion arvojoukko

Kun pisteiden siirtymisessä on olemassa vektorikenttä, on vektorikentällä olemassa myös arvojoukko. Kuvaamalla pariutustuloksen vektorien B pisteet kaikkien pilven A pisteiden ollessa kiinnitettynä origoon, nähdään miten mahdolliset erotusvektorin arvot hajaantuvat origon ympärille. Kuvassa 19 nähdään erään tuloksen arvojoukko, jonka tiheimmät alueet voidaan yhdistää tietyn tyyppisiin erotusvektoreihin pariutus tuloksessa. Kuvassa punaisella ympyröity alue vasemmassa osassa kuvaa voidaan osoittaa saavan alkunsa oikeanpuoleisen osan kaltaisista lyhyistä pariutustuloksista.

Arvojoukon tarkastelussa, tärkeimmät seikat ovat sen keskiarvo, hajontatiheys ja hajonta-alueen muoto. Kuvan 20 kaksi oikeanpuoleistakuvaparia paljastavat, miten tasaisesti origon ympärillä esiintyvä pariutustulos, jonka keskiarvo ≈ 0 selkeästi poikkeaa sitä vastaavasta oikeasta pariutuksesta. Arvioimalla arvojoukon tiheyksiä kuitenkin nähdään yhteys tiheämpien alueiden ja oikeiden tuloksien välillä. Ar-



Kuva 19. Pariutustuloksen arvojoukko (vasemmalla) ja sen tiheimmän alueen lähde (oikealla) havainnollistettuna. Vasemmalla esitetystä arvojoukosta on helposti havaittavissa B pisteitä esittävien mustalla merkittyjen pisteiden suurempi tiheys punaisella ympyröidyllä alueella. Oikealla esitetään joukko pariutettuja pisteitä, joiden B pisteiden tuloksena saadaan tiheä osa vasemmasta kuvaajasta.



Kuva 20. Kuvassa lähinaapurien keskimääräinen etäisyys. Todellinen pariutus mustalla ja pariutustulos sinisellä. Datajoukot 324 vasemmalla ja 513 oikealla. Kuvan avulla voidaan havainnollistaa arvojoukossa esiintyviä asioita, joita on huomioitava arvojoukon avulla tuloksen oikeellisuutta arvioidessa. Kuvassa v_{rel} on erotusvektorin pituus jaettuna keskimääräisellä pisteiden välisellä etäisyydellä.

vojoukkoon vaikuttaa myös mahdolliset ulkoiset ilmiöt, kuten LIDAR mittauksen aikana vallitsevat tuuliolosuhteet ja muut tekijät. Arvojoukon käyttäminen apuvälineenä on kuitenkin perusteltavaa, sillä sen voidaan odottaa kutistuvan origoon päin, kun vektorifunktion vaikutuksia poistetaan, jättäen jäljelle vain datassa esiintyvän melun arvojoukon.

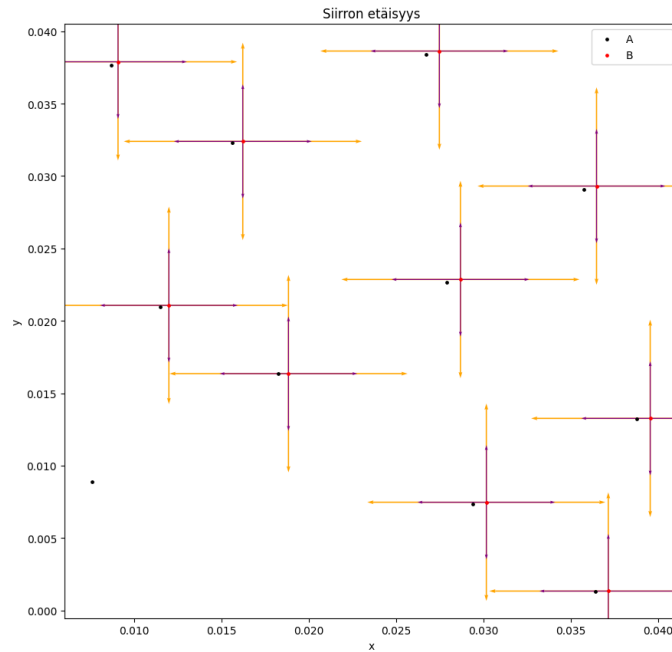
2.8 Siirtohaku

Lähimmän karteesisen naapurin vertailukohdaksi esitetään siirtohaku, jonka aikakompleksisuus on luokkaa $\mathcal{O}(p \cdot n)$ missä n on joukon koko ja p permutaatioiden

määrä. Tutkimuksessa käytetty siirtohaku hyödyntää suodattamatonta karteesista hakua sekä pistepilven B siirtoa, joten se voidaan myös nähdä eräänä karteesisen haun suodatusmenetelmänä tai iterointimenetelmänä. Tutkimuksessa käytetty siirtohaku valitsee tuloksensa perustuen laskettuun esiintymismäärään, joka poikkeaa karteesisen haun valintaperiaatteesta, jonka takia sen katsottiin olevan sopiva verrannaismenetelmä. Menetelmän pariutus toteutetaan taltioimalla jokaiselle pilven A pisteelle sen lähin karteesinen naapuri jokaisella siirtopermutaatiolla. Kun jokainen permutaatio on suoritettu, lasketaan jokaisen pilven A pisteen pariutuksista esiintymismäärä ja pariiksi valitaan eniten esiintynyt pilven B piste.

Siirtohaku toteutetaan kolmella parametrilla, jotka ovat maksimietäisyys, askelmäärä ja kulmamäärä. Maksimietäisyys-parametri vastaa esimerkiksi arvoa P_2 pituussuodatukselta. Se voidaan määrittää etsimällä lyhin A pisteiden välinen etäisyys P_1 , joka jakamalla saadaan käytetty maksimietäisyys $P_2 = \frac{P_1}{2}$. Askelmäärä parametrisoi siirrossa toteutettavien askelten suuruuden välillä $0, P_2$. Askelmäärän arvolla 1 haku etsii A pisteiden lähimmät naapurit, B pisteiden alkuperäisestä sijainnista sekä positiivisen P_2 ja negatiivisen P_2 siirron jälkeisestä sijainnista. Askelmäärän arvolla 2 siirtopisteenä on lisäksi alkuperäisen sijainnin ja P_2 etäisyyden puoliväli $\frac{P_2}{2}$. Parametrein askelmäärä 2 ja kulmamäärä 2 toteutuvat siirtovektorit esitetään kuvassa 21.

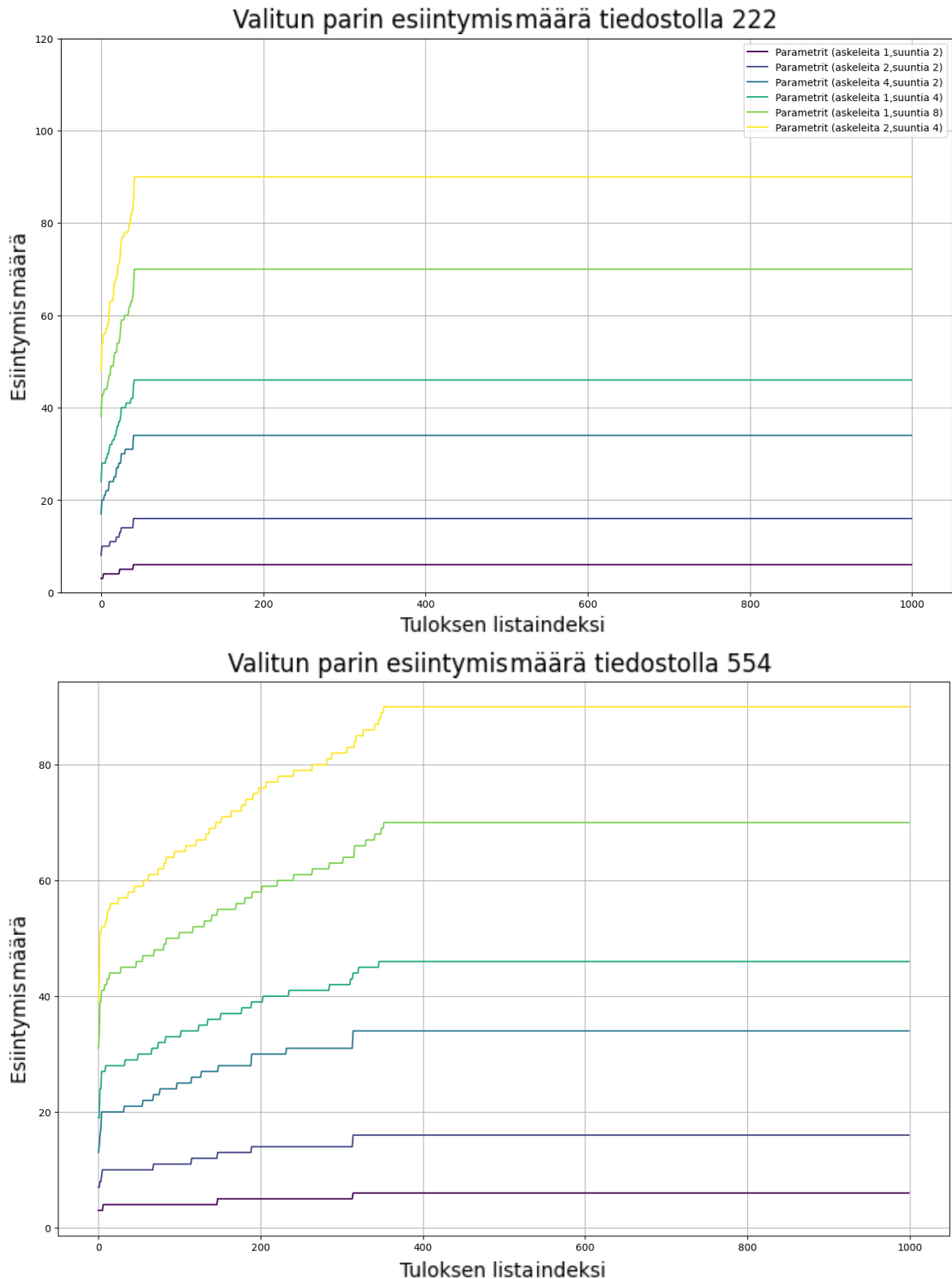
Kulmamäärää nostamalla toteutetaan siirrot tiheämmällä kulmajaolla kattaen näin useampia permutaatioita. Käyttäen tiedostoja 222 ja 554 voitiin arvioida menetelmän stabiiliutta parametrien suhteen. Tarkastelemalla pariutustuloksen valitun pilven B pisteen esiintymismääriä voidaan tuloksia tarkastella järjestettynä listana, joka esitetään kuvassa 22. Kuvasta nähdään, kuinka esiintymismäärä pysyy melko stabiilina eri parametreilla suoritetuissa siirtohaussa. Edukkaamman tiedoston 222 tapauksessa eri parametrien esiintymismäärän muutos on lähinnä nimellinen



Kuva 21. Siirtohaun kahdeksan siirtovektoria, kun askelmäärä on suuruudeltaan kaksi ja kulmamäärä suuruudeltaan kaksi. Oranssilla esitetään siirtoon käytetty vektori p_2 etäisyydellä alkuperäisestä, sekä lilalla pisteen alkuperäisen sijainnin ja p_2 etäisyyden puoliväli.

ja epäedukkaamman tiedoston 554 tapauksessa suurimman esiintymismäärän osuus eroaa noin 50 pariutuksella pienimmän ja suurimman permutaatiomäärän välillä. Menetelmä suoriutuu siten erittäin stabiilisti hakutehtävästä kaikilla parametreillä, eikä niiden odoteta tuovan merkittävää muutosta hakutuloksen tarkkuuteen tai herkkyyteen.

Siirtoon käytettävän maksimietäisyys parametrin suuruus valittiin vastaamaan A pisteiden välisten etäisyyksien keskiarvon puolikasta, tutkimuksessa tehtyjen kokeiden perusteella. Menetelmää voidaan käyttää myös muilla tavoin määriteltäviä maksimietäisyyttä kuten aikaisemmin mainittua P_2 arvoa. Tutkimuksen kannalta oleellisiin tuloksiin kuitenkin päädyttiin käyttämällä keskiarvoon perustuvaa siirtoa.



Kuva 22. Siirtohaun pariutustulokseen valittujen pilven B pisteiden esiintymismäärät eri parametreilla. Esitettynä järjestetyssä listassa molemmilla tiedostoilla. Kuvan perusteella havaitaan miten edukkaampien tiedostojen tapauksessa siirron askelmäärä ei juurikaan vaikuta esiintymismääriin. Vaikeammalla tiedostolla askelmäärän vaikutus on havaittavissa, jääden silti matalaksi suhteutettuna sen korotuksesta seuraavaan laskennallisuus vaatimusten kasvuun.

3 Tulokset synteettisellä datalla

Hakumenetelmän tulosta tarkastellaan synteettisen datan tapauksella hyödyntäen tiedossa olevaa oikeaa pariutusta. Oikean pariutuksen avulla voidaan tulokselle laskea herkkyys ja tarkkuus. Tarkkuus voidaan määrittää laskukaavalla,

$$\frac{\text{oikeita hakutuloksessa}}{\text{hakutuloksen koko}} = \text{tarkkuus}. \quad (9)$$

Sekä herkkyys laskukaavalla,

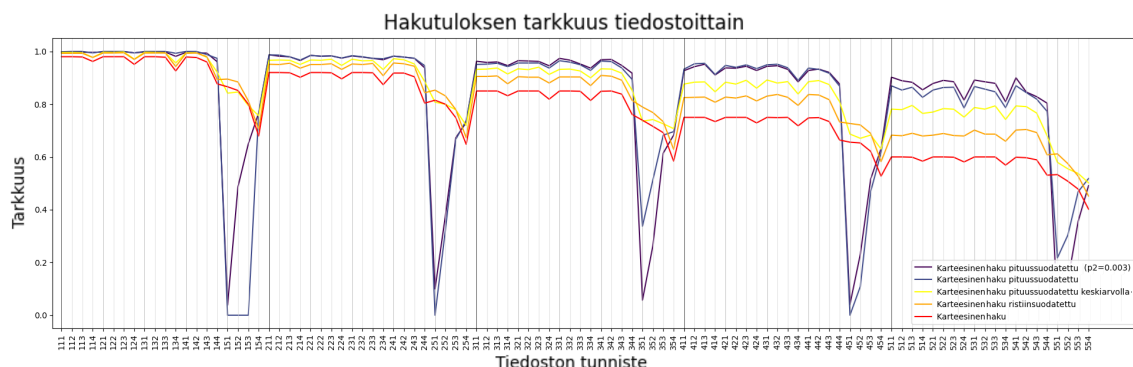
$$\frac{\text{hakutuloksen koko}}{\text{todelliset oikeat}} = \text{herkkyys}. \quad (10)$$

Näin voidaan tarkastella hakumenetelmän suoriutumista kaikilla synteettisen datan parametrien permutaatioilla. Tarkastelu tehdään käyttäen karteesisia hakua suodattamattomana ja laskennallisesti kevyimmillä suodatusmenetelmillä: pituus suodatus ja kaksisuuntaisella suodatuksella. Vertailukohtana käytetään siirtohakua.

Synteettisen datan kolmen parametrin takia tuloksien vertailun kuvaaminen osoittautui haastavaksi. Kuvaamiseen osoittautui kuitenkin parhaimmaksi kuvaaja, jossa x akselilla esitetään tiedoston tunnista muodossa ijk ja kuvaajan alueella parametrien arvon vaihtelua kuvataan mustalla viivalla i arvon vaihtumista, tumman harmaalla kuvataan j arvon vaihtumista ja vaalean harmaalla k arvon vaihtumista. Kuvaajan y akselilla on tarkasteltavan suureen suuruus.

3.1 Karteesinen haku

Hakumenetelmien arviointi aloitettiin suorittamalla karteesinen haku kaikilla tiedostoilla. Suodattamattoman haun lisäksi tuloksiin lisättiin myös pituus suodatettu karteesinen haku ja kaksisuuntaisesti suodatettu haku. Näillä valinnoilla voidaan helposti tarkastella menetelmien toimivuutta koko datajoukon yli.



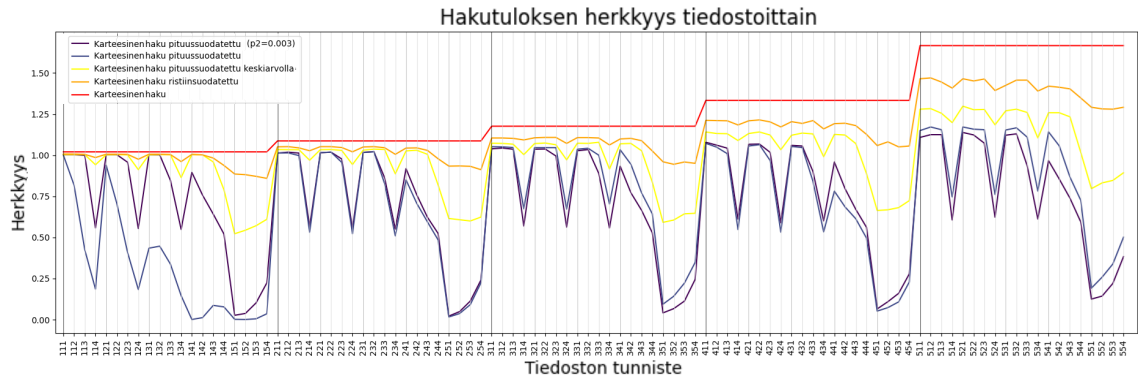
Kuva 23. Hakumenetelmien tarkkuus, koko datajoukon yli mitattuna. Kuvassa esitetään suodattamaton karteeminen haku, ristiin suodatettu haku sekä eri pituusraja valinnoilla pituus suodatus. Parametrien muutosaskeleet esitetään mustilla viivoilla (parametri i), tumman harmailla viivoilla (parametri j) ja vaalean harmailla viivoilla (parametri k).

Kuvassa 23 esitetään mainittujen hakumenetelmien tarkkuus kuvaajassa koko datajoukolle.

Kuvaajassa esitetään pituussuodatus eri raja valinnoilla. Rajoina käytetään koko joukon yli vakiona pysyvää arvoa 0.003, jonka valinta tehtiin hakemalla arvo, joka toimii parhaiten kaikilla tiedostoilla. Toinen rajavalinta perustuu menetelmät osiossa mainittuun lyhimmän A pisteiden väliseen etäisyyteen, joka vaihtelee tiedostoittain. Kolmantena rajavalintana käytettiin A pisteiden välisten etäisyyksien keskiarvon puoliväliä.

Tarkastelemalla huomataan rajavalintojen vaikutuksia tuloksiin. Käyttäen tiedostoittain vaihtuvaa rajaa tai vakiona pysyvää rajaa saavutetaan paras tarkkuus i parametrin askeleiden viimeistä neljännestä lukuun ottamatta. Kuvassa 24 esitetään tämän hakutuloksen herkkyys kaikilla tiedostoilla.

Näiden kuvaajien perusteella nähdään kaikkien menetelmien reagoivan lineaarisesti i parametrin muutoksiin. Suodatusta hyödyntävien menetelmien reagointi on loivemmin lineaarinen kuten voidaan odottaa verrattaessa suodattamattomaan mene-



Kuva 24. Hakumenetelmien herkkyys, koko datajoukon yli mitattuna. Kuvassa esitetään suodattamaton karteesinen haku, ristiin suodatettu haku sekä eri pituusraja valinnoilla pituus suodatus. Parametrien muutosaskeleet esitetään mustilla viivoilla (parametri i), tumman harmailla viivoilla (parametri j) ja vaalean harmailla viivoilla (parametri k).

telmään, joka olemassa olevasta parimäärästä riippuen johtaa väistämättä vääriinkin pariutuksiin.

Lyhimpään etäisyyteen perustuvien rajausmenetelmien tulos sisältää mielenkiintoisia piirteitä, jotka sopivat pituusjakaumaan tehtyyn tarkasteluun. Nämä menetelmät toimivat selkeästi paremmin ja tasaisemmin tiedostoilla, joissa etäisyys parien välillä on odotettavasti melko pieni. Tämä ilmiö on selkeästi havaittavissa herkkyys kuvaajan perusteella. Josta nähdään $vMax$ ja ϵ arvojen kasvun vaikuttavan voimakkaasti menetelmän herkkyyteen. Näiden kasvaessa on useampi pariutettava piste lähempänä A pisteiden voronoin solujen rajaa, jolloin menetelmä on kykenevätön havaitsemaan näitä pariutuksia joita se pitää epävarmoina. Toisaalta tämä johtaa myös lähempien pisteiden oikeellisuuden kasvuun.

Keskiarvoperusteisessa rajassa tätä ilmiötä korjataan sillä silloin menetelmän raja kattaa suuremman osan näistä ulommista pariutuksista ja siten sen tulos on merkittävästi tasaisempi koko hakualueella, verrattaessa muihin rajoihin. Tämä rajausmenetelmä kuitenkin sisällyttää myös enemmän vääriä tuloksia kautta koko ha-

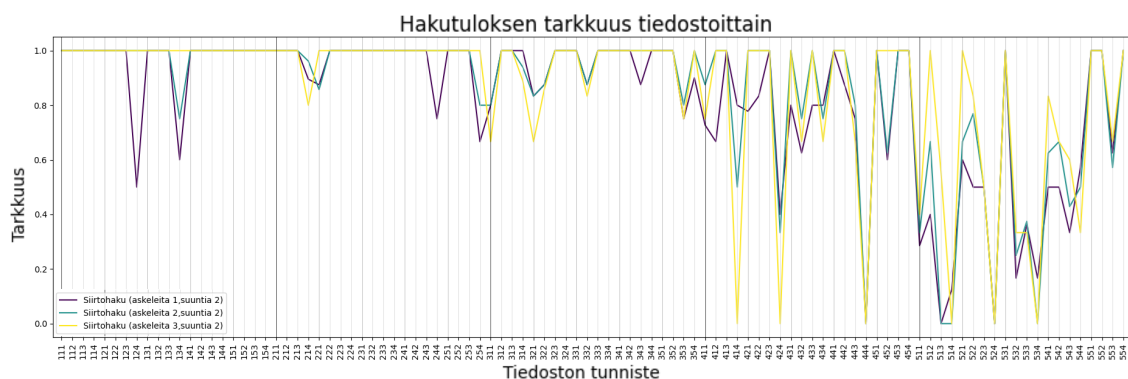
kualueen.

Suodattamattoman haun tuloksen nähdään odotetusti muuttuvan pariutusmäärän suhteessa. Sen perusteella pystytään kuitenkin toteamaan keskiarvoon perustuvan rajan ja ristiin suodatuksen toimivuus sillä molemmat menetelmät tuovat hakutulosta parempaan suuntaan pysyen melko tasaisena koko datajoukon yli. Ristiin suodatuksen nähdään kuitenkin herkkyys kuvaajan perusteella pysyvän jonkin verran tasaisempina koko datajoukon yli. Se kuitenkin sisältää myös enemmän vääriä tuloksia kuten herkkyys ja tarkkuus kuvaajan perusteella nähdään. Sen tuloksien perusteella voitaisiin kuitenkin odottaa, että käyttämällä ensin keskiarvoperusteista pituussuodatusta ja tämän jälkeen ristiinsuodatusta, voidaan tuloksen odottaa paranevan.

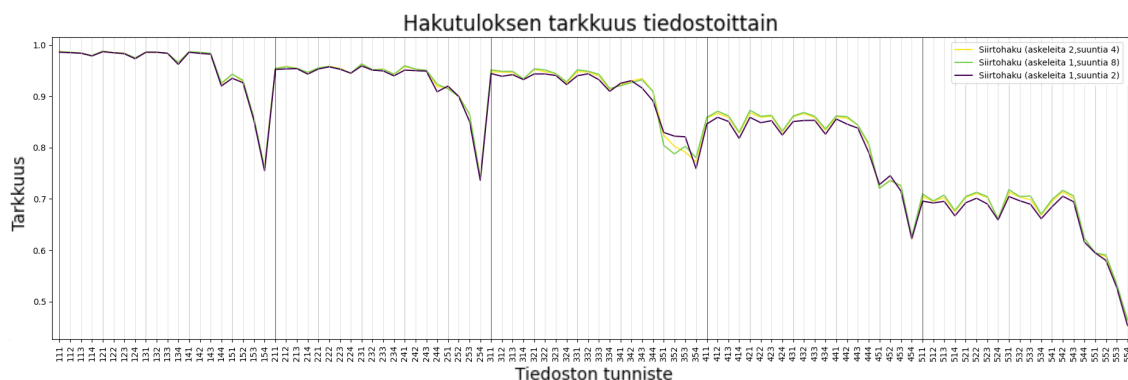
3.2 Siirtohaku

Vertailuun käytetään siirtohakua. Sen suorituskyky riippuu askel- ja kulmaparametrien valinnan lisäksi siirtoon käytettävän askeleen suuruudesta. Kuvassa 25 esitetään siirtohaun tarkkuus kaikkien tiedostojen yli kun siirron pituutena käytetään A pisteiden välisten etäisyyksien keskiarvoa kokonaisuudessaan. Tällä siirtovalinnalla menetelmä käyttäytyy erittäin epätasaisesti ja sen mitatun herkkyyden perusteella sen suorituskyky on huono. Tehdyillä kokeilla havaittiin, että kulmaparametrin muuttamisella ei ole johdonmukaista merkittävää vaikutusta, joten laskennallisen kompleksisuuden takia sen vaihtelua ei testattu kuvaajan kokeessa. Kuten kuvaajasta nähdään tällä rajavalinnalla menetelmän tarkkuus ei johdonmukaisesti muutu eri askelparametreillä.

Valitsemalla siirtoon käytetty pituus lyhimmän A pisteiden välisen etäisyyden perusteella muuttuu menetelmän suorituskyky merkittävästi tasaisemmaksi. Kuvassa

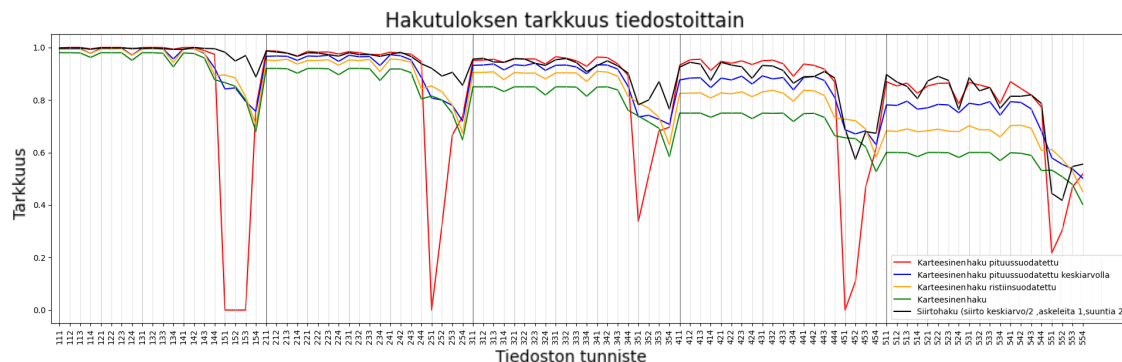


Kuva 25. Siirtohaun tarkkuus kaikkien tiedostojen yli kun siirtoetäisyys vastaa A pisteiden välisen etäisyyden keskiarvoa. Parametrien muutosaskeleet esitetään mustilla viivoilla (parametri i), tumman harmailla viivoilla (parametri j) ja vaalean harmailla viivoilla (parametri k).



Kuva 26. Siirtohaun tarkkuus käyttäen siirron pituutena A pisteiden välistä lyhintä etäisyyttä, ja vaihtamalla siirto askel ja suunta parametreja. Parametrien muutosaskeleet esitetään mustilla viivoilla (parametri i), tumman harmailla viivoilla (parametri j) ja vaalean harmailla viivoilla (parametri k).

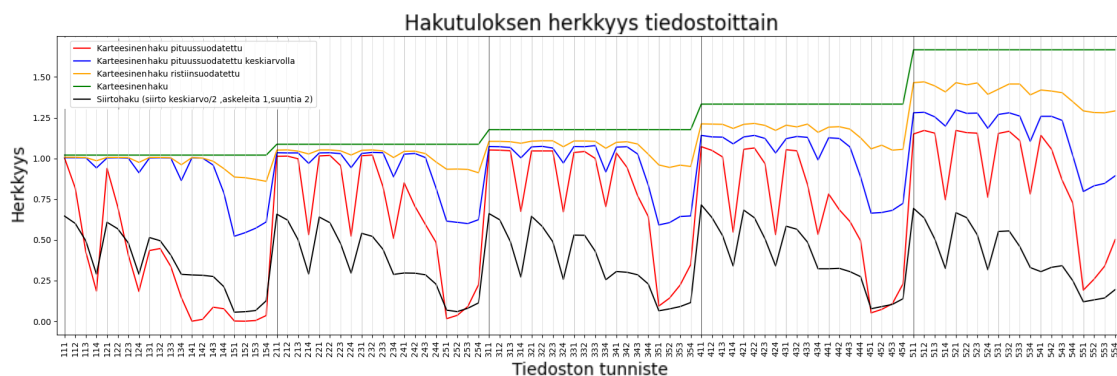
sa 26 esitetään siirtohaun tarkkuus mainitulla siirron pituudella ja eräillä parametrivalinnoilla. Kuvasta havaitaan, etteivät eri parametrivalinnat tuota merkittävää eroa hakutulosten välillä. Jonka takia voidaan katsoa valitun siirto etäisyyden olevan merkittävin valinta. Kuten karteesisen haun kokeissa nähtiin puolet A pisteiden välisestä keskiarvosta edustaa parhaiten etäisyyttä, jonka sisällä olevat pisteet ovat todennäköisimmin pari. Soveltamalla tätä etäisyyttä siirtohaakuun voidaan sen nähdä olevan riittävän erilainen ja vertailukelpoinen karteesiseen hakuun. Menetelmä saavuttaa tällöin tutkimuksen kannalta merkittäviä tuloksia.



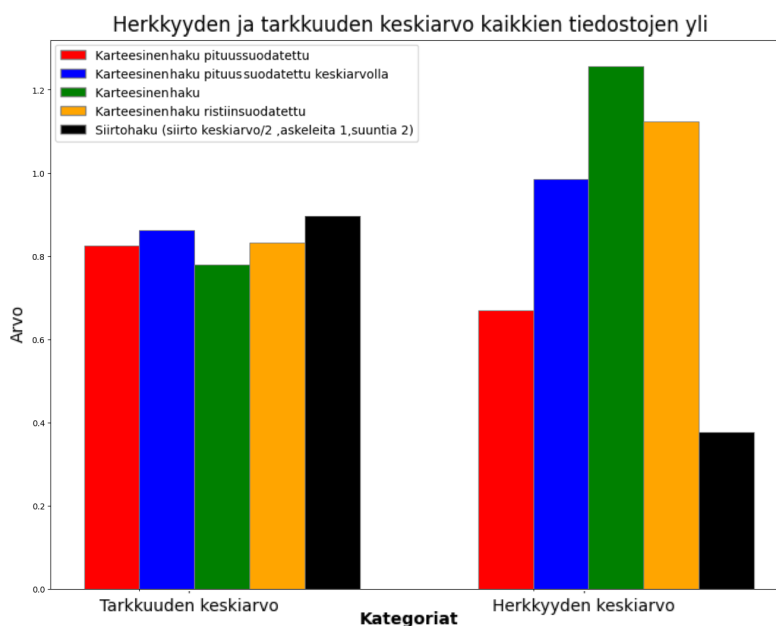
Kuva 27. Siirtohaun ja karteestisten hakujen tarkkuus kaikkien tiedostojen yli. Parametrien muutosaskeleet esitetään mustilla viivoilla (parametri i), tumman harmailla viivoilla (parametri j) ja vaalean harmailla viivoilla (parametri k).

3.3 Hakumenetelmien vertailu

Molempien hakumenetelmien tarkkuus esitetään kuvassa 27. Kuvan perusteella voidaan todeta, että keskiarvoon perustuvalla siirrolla toteutettu siirtohaku yhdistää keskiarvo perusteisen ja lyhimmän pituuden mukaisen pituus suodatuksen hyvät puolet. Siirtohaun tarkkuus pysyy suhteellisen korkealla myös jokaisen i askeleen neljänneksellä. Kuvasta 28 kuitenkin nähdään miten siirtohaun hakutulos on kooltaan merkittävästi matalampi sillä sen herkkyys on matalin kaikista hakumenetelmistä. Osoittaen sen pariutustuloksen olevan kooltaan merkittävästi pienempi kuin muilla menetelmillä. Kuvan 29 perusteella voidaan todeta että kaikki hakumenetelmät soveltuvat selkeästi tietyille osa alueille tiedostojen parametrisointeja joka johtaa tarkkuuden keskiarvojen tasaisuuteen kaikilla hakumenetelmillä. Kuvan herkkyyden keskiarvot vastaavat myös arvoja, joita herkkyys kuvaajan perusteella voidaan menetelmille odottaa.



Kuva 28. Siirtohaun ja karteesisten hakujen herkkyys kaikkien tiedostojen yli. Parametrien muutosaskeleet esitetään mustilla viivoilla (parametri i), tumman harmailla viivoilla (parametri j) ja vaalena harmailla viivoilla (parametri k).



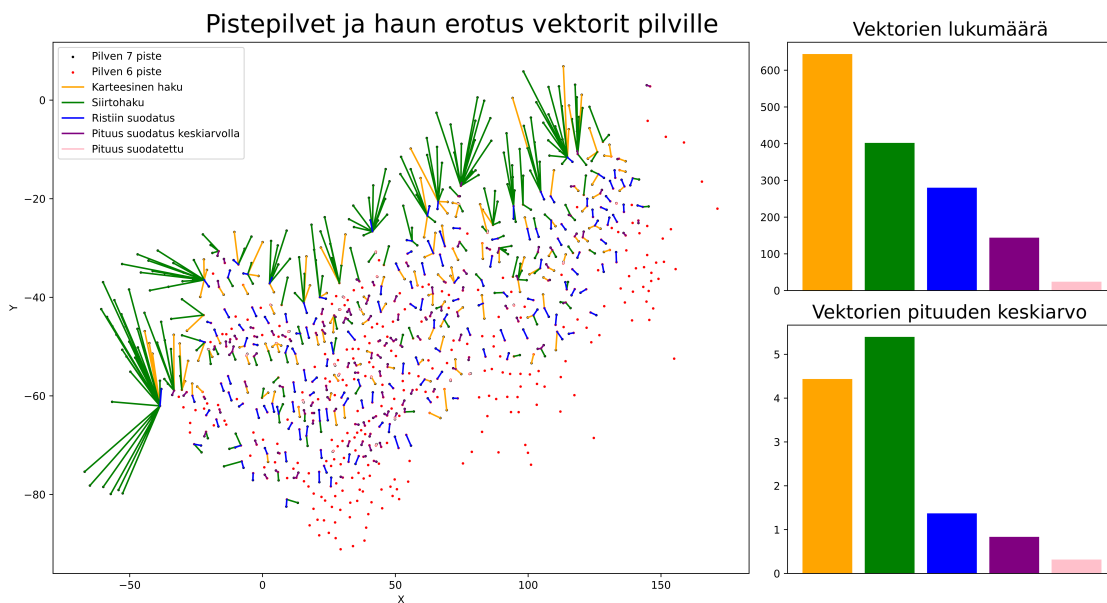
Kuva 29. Tarkkuuden ja herkkyden keskiarvo kaikilla hakumenetelmillä yli kaikkien tiedostojen. Pylväsdiagrammista nähdään tarkkuuden keskiarvojen olevan verrattain tasainen kaikilla hakumenetelmillä. Herkkyden osalta nähdään pienimmät herkkyden keskiarvot karteesiselle pituussuodatetulle haulle sekä siirtohauille. Korkeimmat herkkyysarvot saavutetaan jäljelle jääville kolmelle hakumenetelmille.

4 Tulokset LIDAR- datalla

Tutkimusta varten toimitettiin jo esikäsitelty datajoukko, joka sisältää pistekoordinaatit oletetuille puille. Menetelmien testaaminen on haastavampaa verrattuna synteettiseen dataan, sillä oikea pariutus ei ole tiedossa. Hakumenetelmien tulosta on siis arvioitava käyttäen muita mahdollisia menetelmiä. Alkuperäisen LIDAR datan pisteet sisältävät myös korkeuskomponentin, mutta niitä tarkastellaan tässä yhteydessä muuttamalla ne kaksiulotteiselle tasolle. Tulosten tarkastelun kuvaaminen synteettistä dataa muistuttavilla menetelmillä ei ole mahdollista LIDAR- datan tapauksessa. Tämän takia tuloksia tarkastellaan mielenkiintoisiksi havaittujen pariutustulosten esityksin, joiden tuloksen laatua tarkastellaan arvojoukon perusteella. Pariutustulosta esittäessä on huomioitava, että useiden päällekkäisten vektorien esittäminen kuvaajassa on haasteellista. Kuvaajat esitetään kuitenkin erojen tuomiseksi esille, niiden ollessa tutkimuksen kannalta merkityksellisimpia. Tämän johdosta tapauksissa joissa nähdään vain yhden värinen vektori, voidaan niitä pitää tutkimuksen kannalta epäoleellisina mikäli ne vastaavat tyyliltään lähiympäristössä esiintyviä muita pariutuksien vektoreita. Lisäksi voidaan tarkastella pariutusvektorien keskiarvoa, ja miten se vastaa kummankin pilven pisteiden paikoitustarkkuutta joka on noin 0,5m [6].

LIDAR- datan tapauksessa tiedostojen nimeämiskäytäntö myös poikkeaa synteettisestä. LIDAR- pistepilvet on jaettu kahdeksaan osittain päällekkäin sijoittuvaan erilliseen pilveen. Nämä pilvet voidaan testata pareittain jolloin pareja ovat $(8,4), (4,5), (5,3), (3,2), (2,1), (1,7)$ ja $(7,6)$.

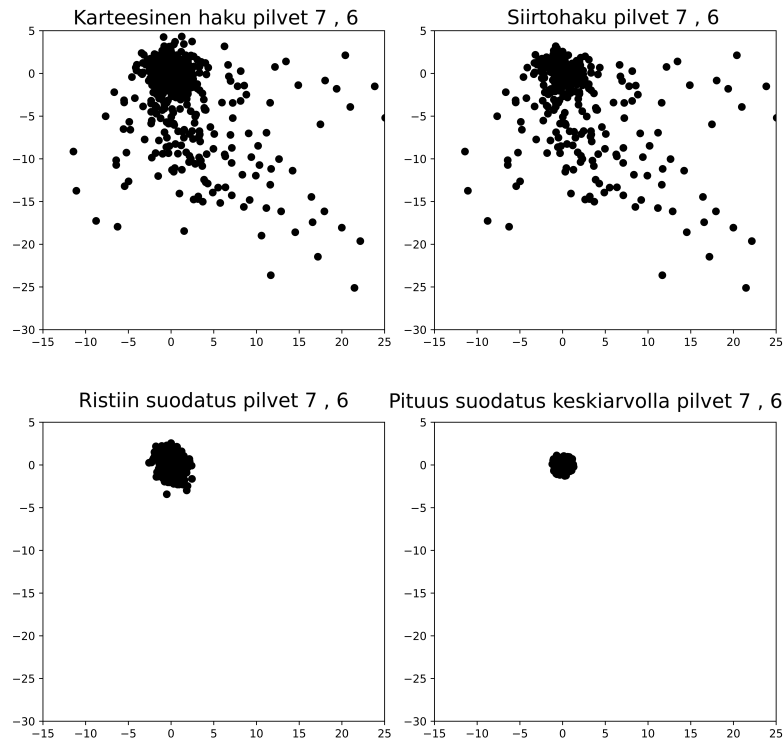
Pilviparille $(7,6)$ kuvassa 30 esitetään pariutustulos, pariutustuloksien määrä ja vektorien pituuksien keskiarvo tälle parille. Tämän parin pariutus, vaikuttaa tuloksien perusteella sisältävän pariutuksia, joiden tunnistaminen tarkastelemalla voitaisiin olettaa haastavaksi. Tarkastelussa ristiinsuodatuksella pariutuneet pisteet vai-



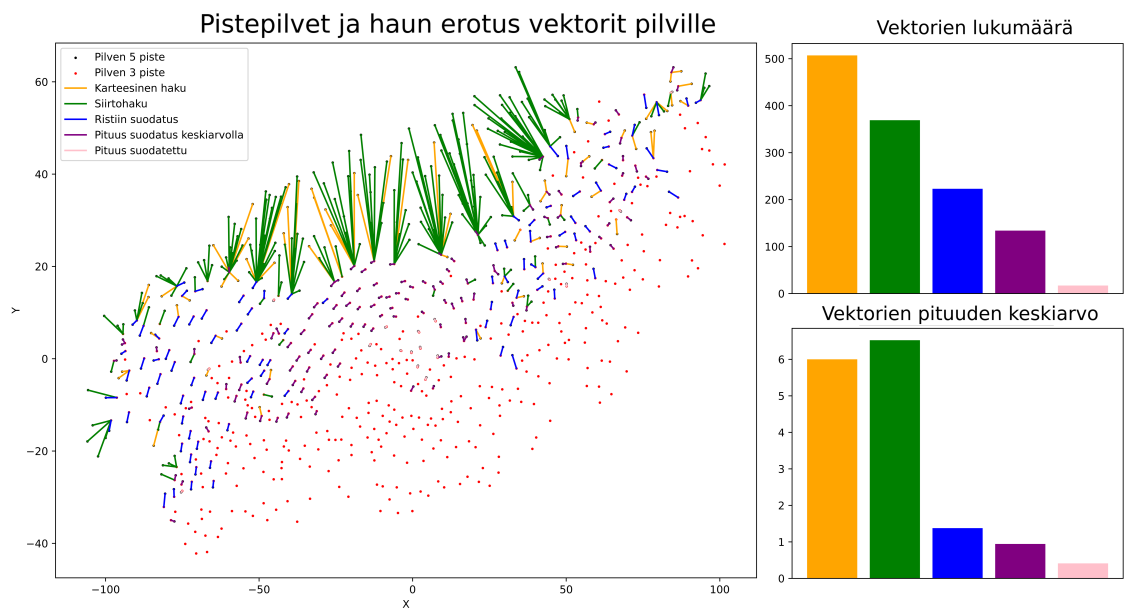
Kuva 30. Pariutustulos pistepilvi parille (7,6) eri hakumenetelmillä. Kuvassa myös hakumenetelmien tulospäämäärät ja vektorien pituuden keskiarvo pylväsdiagrammina pariutus kuvan oikealla puolella.

kuttavat olevan jakautuneet poikkeuksellisen tasaisesti pilvien päällekkäin olevalle alueelle. Kuvasta havaitaan ristiinsuodatuksen ja keskiarvo perusteisen pituus suodatuksen antavan tarkastelemalla odotettavissa olevan kaltaisia hakutuloksia. Kaikista hakumenetelmistä pituus suodatus lyhimmän etäisyyden mukaan tuottaa lähinnä paikoitus tarkkuutta olevan keskiarvon. Sen tulospäämäärä on kuitenkin verrattain pieni jonka johdosta sen oletetaan sisältävän vain pieni osuus todellisista pareista. Kuvasta 31 voidaan tarkastella hakumenetelmien arvojoukkoa, jonka perusteella ristiinsuodatetun tuloksen voidaan olettaa olevan melko oikea, arvojoukon perusteella.

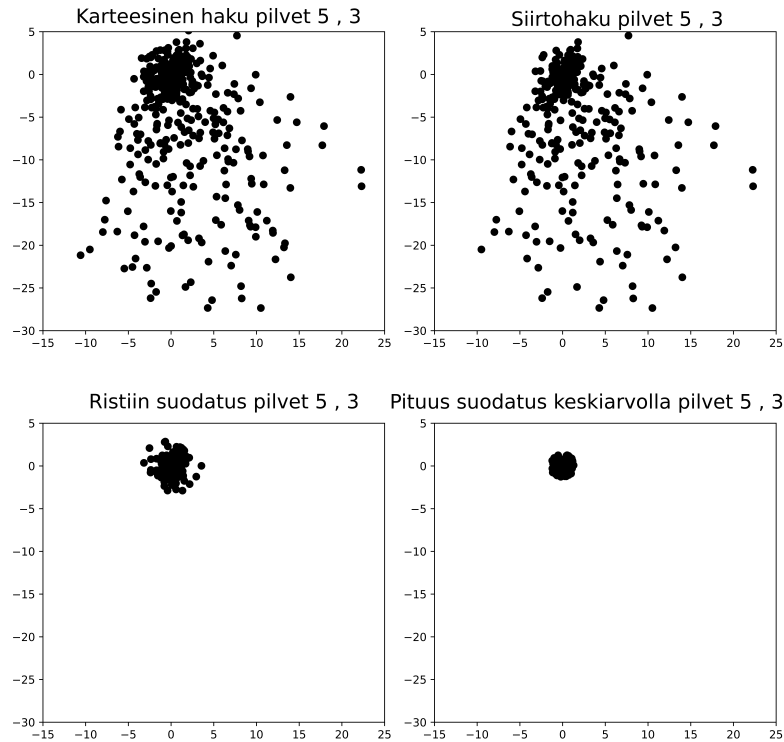
Pilviparille (5,3) saatu pariutustulos esitetään kuvassa 32 mistä nähdään pilvissä esiintyvä selkeä pyörre keskialueen paikkeilla. Tässäkin pariutuksessa voidaan pituuskeskiarvon ja arvojoukkokuvaajan joka esitetään kuvassa 33 perusteella pitää ristiinsuodatusta parhaiten suoriutuvana menetelmänä. Kuitenkin tarkastelemalla pyörteen välitöntä lähialuetta havaitaan keskiarvoperusteisen pituus suodatuksen löytäneen tarkastelemalla arvioituna varmin oikeita pariutuksia. Lisäksi



Kuva 31. Arvojoukko pistepilviparille (7,6) eri hakumenetelmillä. Pienimmän hakutuloksen antanut menetelmä on jätetty pois kuvasta sen ollessa vertailuun kelpaamaton muiden menetelmien yhteydessä.



Kuva 32. Pariutustulos pistepilviparille (5,3) eri hakumenetelmillä. Kuvassa myös hakumenetelmien tulasmäärät ja vektorien pituuden keskiarvo pylväskuvina.



Kuva 33. Arvojoukko pistepilviparille $(5,3)$ eri hakumenetelmillä. Pienimmän hakutuloksen antanut menetelmä on jätetty pois kuvasta sen ollessa vertailuun kelpaamaton muiden menetelmien yhteydessä.

ristiinsuodatuksen tuloksena, on pariutustulos kuvaajan oikeassa yläsektorissa pariutustuloksia jotka eivät tarkastelun perusteella vaikuta oikeilta. Arvojoukko kuvaajan perusteella pariutukset kuitenkin vaikuttavat oikeilta.

Kaikissa kokeiluissa tapauksissa siirtohaun ja karteesisen haun tuloksissa esiintyi merkittävä määrä selkeästi vääriä pituudeltaan pitkiä tuloksia. Tämän takia molemmat menetelmät soveltuvat sellaisenaan huonosti tämän kaltaiselle datalle.

5 Lopputulokset hakumenetelmistä

Käyttämällä karteesista hakua lyhimmän pisteiden A välisen etäisyyden mukaan suodatettuna saavutettiin huomattavan korkea tarkkuus synteettisen datan kokeilla,

joissa etäisyydet pariutusten välillä ovat pieniä. Tulos on suoraa seurausta pisteiden melko tasaisesta jakaumasta koko pistepilven alueella. Tätä oletusta tukee herkkyyden muutokset etäisyyksien kasvaessa. Suodatusta näin soveltamalla sen voidaan katsoa poimivan etäisysoletuksen perusteella parit, joiden B pisteet ovat selkeästi lähempänä niiden A pisteitä, kun tasaisesti jakautuneet A pisteet. Tämän seurauksena menetelmän kyky saavuttaa täydellinen pariutus on huono pariutusetäisyyksien kasvaessa. Menetelmä soveltuu kuitenkin tilanteisiin, joissa kaikkien parien löytäminen ei ole edellytys, vaan tavoitellaan mahdollisimman oikeaa pariutus tulosta.

Havainnot synteettisellä datalla tehdyin kokein tukevat myös havaintoja LIDAR-datalla. Tällöin oletetaan todellisten pariutusten olevan pituudeltaan lähellä mitaustarkkuutta.

Käyttämällä karteesista hakua pituus suodattaen keskiarvon puolikkaalla saadaan tulos, jonka nähtiin toimivan paremmin myös pidemmällä pariutusetäisyyksillä. Menetelmän havaittiin kuitenkin olevan herkkä poimimaan myös vääriä pareja pariutusmäärän pienentyessä (parametrin i kasvaessa). Menetelmä soveltuu kuitenkin tehtyjen kokeiden perusteella useisiin käytännön sovelluskohteisiin ja suoriutuu merkittävästi paremmin, kun lyhimmän A etäisyyden suodatus.

LIDAR- datalla tehdyissä havainnoissa saadut tulokset vaikuttavat myös olevan johdonmukaisia ja muodostuvan vektorikentän voidaan sanoa olevan pääosin melko sileä. Tässä yhteydessä sileydellä tarkoitetaan tulosta, jossa vektorien suuruus ja suunta on johdonmukainen verrattaessa välittömään lähiympäristöön. Sileydessä havaittiin kuitenkin poikkeamia pariutuksissa, joissa ei esiintynyt selkeitä pyönteitä kuten pilvien (7,6) pariutuksessa.

Ristiinsuodatuksen osalta havaittiin sen soveltuvan käytettäväksi lisämenetelmänä epävarmuuksien selvittämiseen alueilla, joissa välittömässä läheisyydessä saattaa olla useampia vaihtoehtoja. Pelkän ristiinsuodatuksen avulla saavutettiin kuitenkin vain kohtalainen pariutus sillä se sisällyttää pariutustulokseen myös suuren määrän väärää tuloksia.

LIDAR- datassa tehtyjen kokeiden perusteella ristiinsuodatuksella saadut tulokset ovat melko lähellä paikoitustarkkuuden arvoa pituudeltaan. Kuitenkin tarkasteltaessa pariutustuloksien välitöntä lähialuetta nähdään, ettei pariutustuloksesta muodostunut vektorikenttä ole suuntansa puolesta johdonmukainen. On myös mahdollista, ettei kaksiulotteiseksi muutetussa esityksessä voida olettaa parienvälisen vektorikentän olevan johdonmukainen. Tätä ajatusta osaltaan tukee menetelmän arvojoukko, jonka muoto ja tiheys vaikuttavat erinomaiselta ja sopivan oikean tuloksen kriteereihin.

Suodattamattoman karteesisen haun ja vertailukohtana käytetyn siirtohaun osalta havaittiin menetelmien suoriutuvan melko tasaisesti kaikilla tiedostoilla. Karteesisen haun osalta tämä on seurausta parimäärän ja hakutuloksen herkkyyden lineaarisesta suhteesta. Siirtohaulla vaikuttaa olevan ominaisuuksia, jonka avulla sitä voitaisiin soveltaa suodatusmenetelmänä karteesisessä haussa.

LIDAR- datan osalta molempien menetelmien tulos osoittautui odotettua huonommaksi. Molemmat menetelmät kykenivät pääasiassa havaitsemaan vain joitain tuloksia pilvien päällekkäiseltä alueelta. Tällöin, vaikka siirtohakuun sovellettaisiin pituussuodatusta on sen odotettu tulos edelleen heikko LIDAR- datan kaltaisissa pariutustehtävissä. Karteesisen haun odotettiin havaitsevan suuria määriä selkeästi liian pitkiä pariutuksia, mutta kuten siirtohaulla, myös karteesisessä haussa ei

hakutulos sisältänyt päällekkäisellä alueella olevia tuloksia, joiden voitaisiin todeta olevan selkeästi oikeita.

6 Koneoppimismenetelmien soveltaminen

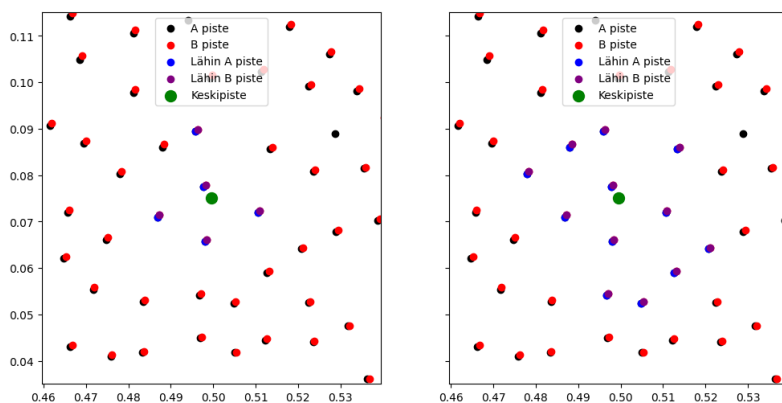
Tulokset eri hakumenetelmillä osoittavat pistejoukon pariutukseen tarvittavan tiedon todennäköisesti sisältyvän pistepilvidataan. Hakumenetelmien tulosten perusteella voidaan ajatella pariutuksen vaativan yhdistelmän eri haku- ja suodatusmenetelmiä.

Tämän tutkimuksen osalta näiden tarvittavien yhdistelmien löytäminen jätetään tutkimuksen ulkopuolelle. Kuitenkin näillä samoilla piirteillä voidaan vetää johtopäätös koneoppimismallin mahdolliselle soveltavuudelle pariutusongelman tarkastelussa. Koneoppimismallin soveltaminen ongelmaan vaatii huolellista tarkastelua käytettävissä olevan datan muotoilusta koneoppimismallille sopivaan muotoon.

Tutkimuksessa käytetään kahta moniulotteisten yhteyksien havaitsemiseen soveltuvaa koneoppimismallia: satunnaismetsä (engl. random forest) ja neuroverkko (engl. neural network). Tutkimuksessa tehtyjen tulosten toistamisen kannalta, oleellista tietoa on myös käytetyt kirjastot, mallien hyperparametrit ja arkkitehtuuri. Satunnaismetsän kokeissa käytettiin scikit-learn kirjastoa [20]. Neuroverkon kokeissa käytettiin tensorflow kirjastoa [21].

6.1 Datarivien muodostaminen pistepilvistä

Hakumenetelmien kanssa tehdyillä kokeilla pystyttiin havaitsemaan viitteitä pariutustehtävään tarvittavista tiedoista. Kahden pisteen pariutus vaikuttaa riippuvan pariutettavien pisteiden etäisyydestä ja niiden suhteessa välittömään lähiympäristöön. Koneoppimis mallin näkökulmasta tämän tiedon muodostaminen datariviksi, joka voidaan käsitellä mallin toimesta, onnistuu muodostamalla A piste keskeisiä koordinaatti vektoreita. Näiden vektorien sisältäessä N lähintä pistettä molemmista pilvistä A ja B sekä sijoittamalla opetuksessa käytetyn, näytekohtaisen lokaa-



Kuva 34. Havainnekuva lähiympäristön dataan sisällytettävistä pisteistä tarkasteltavalle A pilven pisteelle. Kuvan vasemmassa osassa esitetään viisi lähintä pistettä molemmista pilvistä ja kuvan oikeassa osassa esitetään 12 lähintä pistettä molemmista pilvistä.

lin koordinaatiston origo tarkasteltuun pisteeseen, saadaan datarivit generalisoitua koneoppimismallille sopivaksi. Tämän muotoilun ansiosta voidaan myös estää koneoppimismallia huomioimasta pisteiden geografista sijaintia. Kuvassa 34 esitetään havainnekuva erään pistepilven pisteen lähiympäristöstä. Kuvassa vasemmalla esitetään pisteet, joita koneoppimis malli käyttää, jos dataan sisällytetään 5 lähintä pistettä ja oikealla pisteet joita sisällytetään käytettäessä 12 lähintä pistettä. Kuvassa vihreällä esitetään piste $a \in A$, jonka lähiympäristöä tarkastellaan, punaisella esitetään B pilven pisteet, lilalla dataan sisällytettävät B pilven pisteet, mustalla esitetään A pilven pisteet ja sinisellä esitetään dataan sisällytettävät A pilven pisteet.

Tällä tavoin muotoiltuna saadaan data koneoppimiselle sopivaan muotoon

$$B_1, B_2, \dots, B_N, A_1, A_2, \dots, A_N, \text{luokka} \quad (11)$$

Missä termit B_n ja A_n sisältävät pisteen x ja y koordinaatit. Datarivin viimeisenä on luokka, joka on arvoltaan 0 tai jokin kokonaisluku väliltä $1 \rightarrow N$ missä $\text{luokka} = 0$ osoittaa ettei mikään mukana olevista B pisteistä ole tarkasteltavan A pisteen pari. Muut arvot osoittavat mikä mukaan valikoiduista B pisteistä kuuluu nykyisen tarkastelupisteen pariin.

Tällä tavoin saadaan tutkimuksessa mukana olevasta synteettisestä datasta luotua kattava määrä koulutusdataa, joka ei riipu pisteen sijainnista suhteessa pilven alueeseen, ja on lisäksi koneoppimismallin koulutukseen soveltuvaa. On myös tärkeää huomioida koulutukseen käytettyjen datarivien tasajakoisuus. Paremman generalisoinnin ja tasajakoisuuden luomiseksi datarivejä muodostaessa molempien pilvien pisteiden järjestys tulee sekoittaa satunnaisesti. Sekoituksen tulee tapahtua itsenäisesti, jolloin datarivin esiintymisjärjestys ei enää vastaa etäisyyttä tarkastelupisteeseen. Tällä järjestelyllä saadaan maksimoitua datan käyttöaste ja huomioitua koneoppimismallin mahdollisuus generalisoida järjestyksen perusteella, joka on epätoivottua.

Tutkimuksessa käytettävissä olevaa dataa on kolme erillistä datasettiä: kolme parametrinen synteettinen data, joka identifioitiin tunnisteella *ijk* josta käytetään termiä suuri datasetti, kaksi parametrinen synteettinen data, josta käytetään termiä pieni datasetti, joka on luotu samalla tavalla kuin kolmiparametrinen. Tässä tapauksessa pois jäävä parametri on parimäärä, joka pysyy vakiona arvolla 700 ja LIDAR data. Näillä dataseiteillä on mahdollista tehdä tutkimus eri tavoin, näistä tullaan esittämään tutkimuksen kannalta oleelliset.

6.2 Hyperparametrit

Hakumenetelmätutkimuksen jäljiltä, ei ole täysin selvää mikä on oikea määrä lähiympäristön pisteitä. Molemmissa malleissa on myös suuri määrä parametreja, joilla vaikutetaan merkittävästi mallien suorituskykyyn. Mallien parametrivalintojen hakuun käytettiin kokeilevaa menetelmää, jossa parametrien haku on automatisoitu sen koodissa. Satunnaismetsän osalta käytettiin scikit-learn kirjaston GridSearchCV toimintoa parametreilla jotka on esitetty taulukossa II.

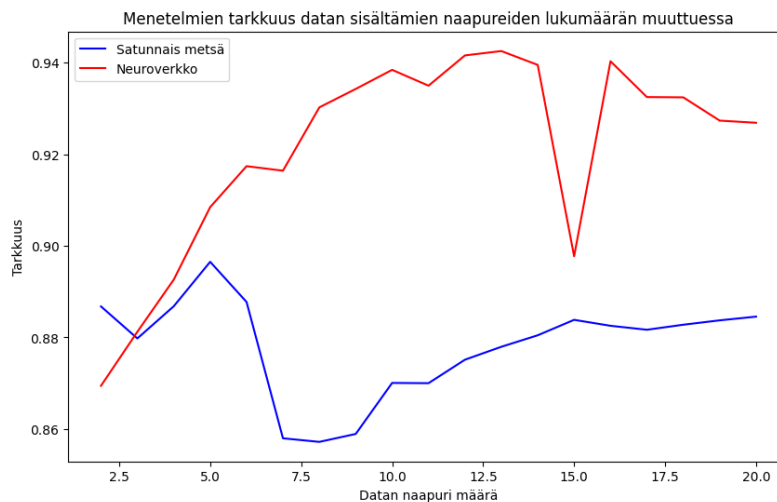
Taulukko II. Satunnaismetsän parametrihaussa käytetyt parametrivaihtoehdot. Taulukossa esitetään parametrin nimi, kuten se esiintyy käytettäessä. Lisäksi esitetään jokaisen parametrin vaihtoehtoiset arvot.

Parametri	Mahdolliset arvot
n estimators	100, 200, 500
max depth	4, 6, 8, 10
min samples split	2, 5, 10
min samples leaf	1, 2, 4

Taulukon parametreista n estimators kontrolloi satunnaismetsämallin päätöspuiden lukumäärää. Parametrillä max depth kontrolloi päätöspuiden mahdollisten jakaumien lukumäärää. Parametri min samples split on pienin määrä näytteitä, jotka vaaditaan sisäisen solmun jakamiseen. Suuremmat arvot estävät algoritmia oppimasta liian yksityiskohtaisia malleja, mikä vähentää ylisovittamista, liian suuret arvot voivat kuitenkin johtaa alisovittamiseen. Min samples leaf parametri ilmaisee pienimmän määrän näytteitä, jotka vaaditaan olemaan lehtisolmussa, päätöksentekoprosessin pienimmässä yksikössä. Jakopistettä missä tahansa syvyydessä harkitaan vain, jos se jättää ainakin tämän verran koulutusnäytteitä kumpaankin vasempaan ja oikeaan haaraan.

Neuroverkon hyperparametrien hakuun käytettiin tensorflow kirjastoon sisältyvää Hyperband algoritmia [22], jolla testattiin kahden neuronikerroksen neuronimäärää välillä 32 ja 512 askelkoolla 32. Lisäksi neuroniverkon tärkeä parametri on sen oppimismnopeus jonka vaihtoehtoiset arvot olivat 0.01, 0.001, ja 0.0001.

Tämän tuloksena saadaan molemmille malleilla tarkkuus sekä parhaat parametrit, sen hetkiselälle N arvolle. Suorittamalla parametrikäytännöllä usealla N arvolla saadaan kuvaaja 35, jossa esitetään molempien menetelmien luokittelutarkkuus, jokai-



Kuva 35. Tutkimuksen koneoppimismallien (satunnaismetsä ja neuroverkko) mitattuennustus tarkkuus N arvon muuttuessa käytettäessä parhaita parametrejä jokaiselle N arvolle. Mittauksessa käytössä suuri datasetti.

sen kierroksen parhailla parametreillä. Tämän tarkastelun tuloksena voidaan havaita satunnaismetsän parhaan arvon olevan $N = 5$ ja neuroverkolle $N = 12$. Tulos on samankaltainen suurella ja pienellä datasetillä, sisältäen vain pienikokoisia vaihteluja, joiden voidaan olettaa olevan peräisin satunnaisuudesta. Osaltaan tuloksen samankaltaisuuteen vaikuttaa datarivien tasajakoisuus, jonka johdosta luokan 0 mahdollisten rivien määrä rajoittaa muiden luokkien maksimisuuruutta.

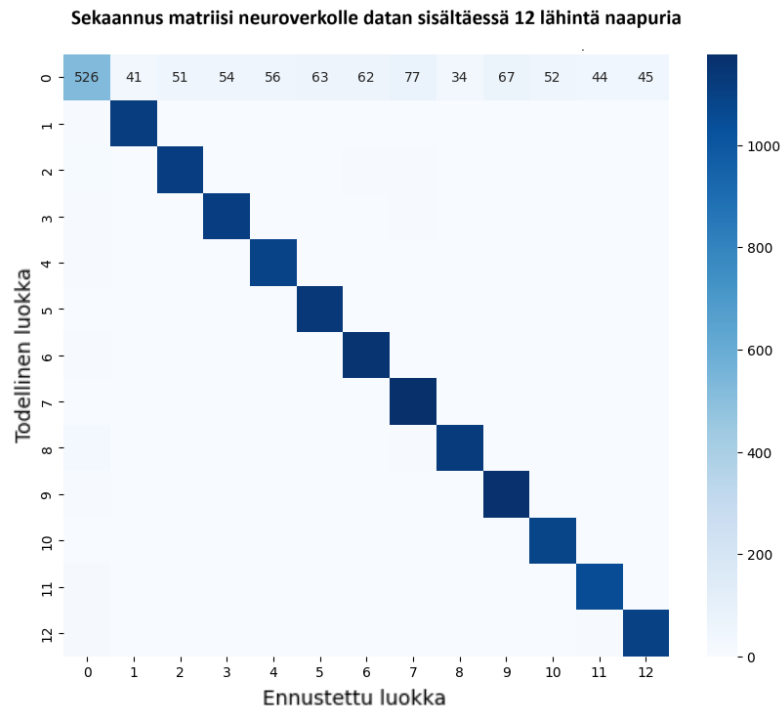
Molemmilla menetelmillä voitiin havaita onnistunut generalisaatio, jonka johdosta voidaan todeta datarivien muodostus menetelmille sopivaksi ja siten sen sisältävän mallien kannalta oleelliset piirteet. Molempien mallien osalta tarvittavan parimäärän löydettyä voidaan suorittaa mallien muiden hyperparametrien haku, joka toteutettiin erillisillä parametrimatriiseilla ja niihin tarkoitettulla hakumenetelmällä. Näiden parametrikokien jälkeen saatiin molemmille malleille hyvin toimivat parametrivalinnat.

Taulukko III. Parametrihaun tuloksena saatu tarkkuus.

Malli	Suuridatasetti	Pienidatasetti
Neuroverkko	0.93	0.94
Satunnaismetsä	0.89	0.89

6.3 Validointi

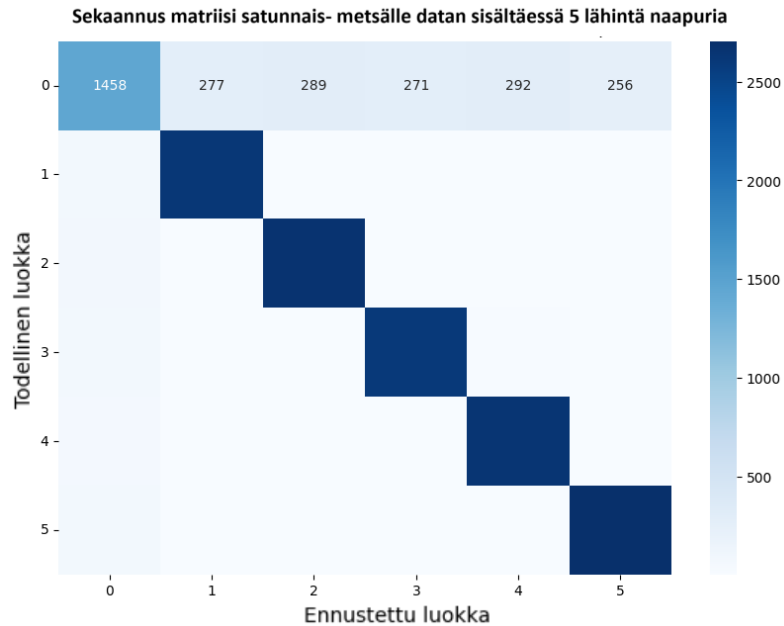
Koneoppimismallien ollessa moniluokkaisia luokittelijoita (engl. multi class classifier) voidaan mallin parametrivalinnat ja suorituskky varmentaa käyttämällä aiemmin kohtaamatonta dataa. Taulukossa III esitetään tarkkuustulokset molemmille malleille käytettäessä joko suurta tai pientä datasettiä. Suorituskkyarvojen saamiseksi validoinnissa käytettiin vain toista datasettiä jakamalla se yleisen käytännön mukaisesti koulutus- ja testaus osiin suhteilla 0.8 ja 0.2. Taulukosta havaitaan molempien datasettien osalta mallien suoriutuvan tarkkuusarvoon perustuen tasaisesti ja korkealla varmuudella. Tarkkuusarvon lisäksi on syytä tarkastella myös mallien varmuutta eri valintojen osalta. Tähän tarkasteluun sopii näiden mallien tapauksessa sekaannusmatriisi (engl. confusion matrix) jonka avulla voidaan visualisoida mallien kykyä erottaa eri luokat toisistaan. Luokkien virheluokittelujen summa osoittaa mallin kokonaisvirheluokittelun pariutuksessa. Koska molempien käytössä olevien datasettien osalta mallien suorituskky pysyy muuttumattomana, käytetään mallien koulutukseen pienempää datasettiä. Tällöin on mahdollista käyttää koulutettuja malleja isomman datasetin pariutukseen jolloin tulosta voidaan vertailla hakumenetelmien kanssa vertailukelpoisella tavalla. Kuvassa 36 esitetään pienemmällä datajoukolla koulutetun neuroverkon sekaannusmatriisi. Sen perusteella voidaan havaita mallin luokitteluvarmuuden olevan korkea ja siten osoittavan mallin toimivan pariutustehtävässä. Datarivejä tälle mallille muodostettiin luokkatasauksen jälkeen yhteensä 15236 kappaletta joista 1172 riviä kuului yhdelle luokalle. Sisällytetyn naapurimäärän ollessa viisi muodostuu mallille luodut datarivit siten kuudesta eri



Kuva 36. Neuroverkon sekaannus matriisi kun koulutus datana on pienempi data-setti. Kuvassa nähdään luokittelu varmuus kaikille käytössä olleille luokille. Kuvan perusteella mallin luokittelu varmuus on korkea.

luokasta joista yksi on luokka 0.

Kuvassa 37 esitetään satunnaismetsän sekaannusmatriisi pienemmällä datasetillä saadun koulutuksen tuloksena. Datarivejä tälle mallille oli yhteensä 17058 kappaletta, joista yhdelle luokalle kuului 2843 kappaletta. Luokitteluvarmuus on korkealla tasolla myös tällä mallilla. Mallille saatiin arvolla $N = 12$ yhteensä 13 eri luokkaa. Matriiseista voidaan todeta virheellisten luokittelujen tasaisen jakauman perusteella datarivien muodostuksen onnistuneen. Lisäksi voidaan varmentaa molempien mallien antavan hyväksyttävissä olevan tarkkuuden ja luokitteluvarmuuden. On kuitenkin huomioitava, että jokaisen luokan virheellisen luokittelun summa heijastaa tilannetta, joka on vertailukelpoinen pariutusongelman kannalta. Tämän johdosta mallien korkeista suorituskykyarvoista huolimatta on mahdollista, etteivät mallit onnistu toteuttamaan pariutusongelman ratkaisua merkittäväällä tarkkuudella. Voidaan myös havaita mallien virheellisten luokittelujen muodostuvan yksin omaan



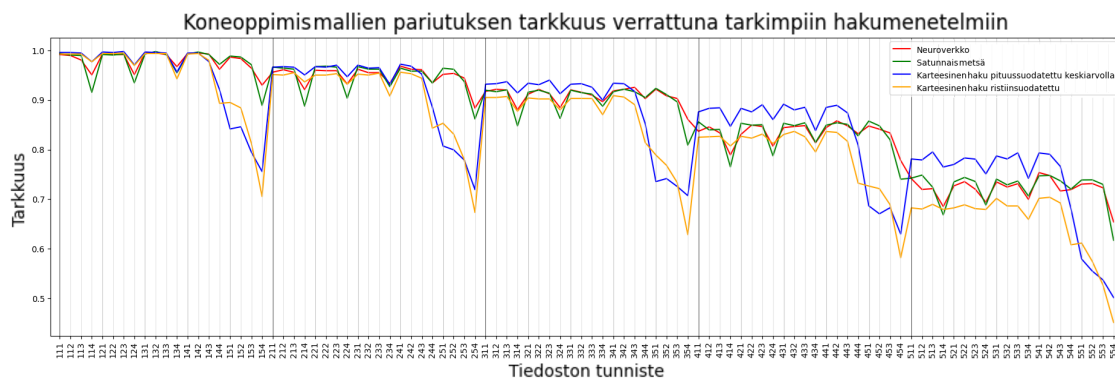
Kuva 37. Satunnaismetsän sekaannusmatriisi, kun koulutusdatana on pienempi datasetti. Kuvassa nähdään luokitteluvarmuus kaikille käytössä olleille luokille. Kuvan perusteella mallin luokitteluvarmuus on korkea.

luokkaan 0 väärin luokitelluista tapauksista.

6.4 Pariutustehtävä

Molempien mallien validoinnin yhteydessä saadut mallit tallennettiin käytettäväksi tutkimuksen seuraavaan vaiheeseen. Satunnaismetsälle lopulliset parametrit joita käytettiin tutkimuksen seuraavissa vaiheissa olivat päätöspuiden lukumäärä 500, päätöspuiden syvyys 10 ja pienin määrä näytteitä jotka vaaditaan sisäisen solmun jakamiseen 5. Lisäksi satunnaismetsän parametri max features, jonka arvo asetettiin ilman hakua määrittää käytettävien muuttujien määrää risteyksissä, tämän parametrin arvona toimii muuttujien neliöjuuri.

Neuroverkkomallin lopullinen arkkitektuuri muodostuu kahdesta peräkkäisestä tiheästä kerroksesta joissa kummassakin on 96 neuronina. Molempien kerrosten ak-

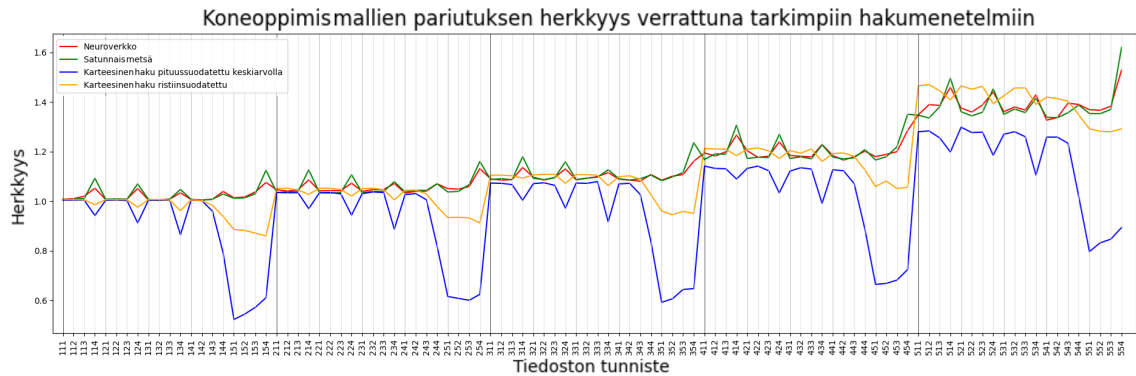


Kuva 38. Kuvassa esitetään pariutustuloksen tarkkuus kaikilla suuren datasetin tiedostoilla. Tulosten vertailuksi esitetään myös kahden hakumenetelmän tarkkuudet. Vertailu menetelminä pituus suodatus keskiarvorajalla (sininen), ristiin verrattu suodatus (oranssi). Tekoälymallien tulos esitetään punaisella (neuroverkko) ja vihreällä (satunnaismetsä).

tivointifunktiona toimii relu. Neuroverkon ulostulokerros sisältää $N = 12$ arvon mukaisesti 13 neuronina, käytetty aktivointifunktio ulostulokerroksessa on softmax. Neuroverkon optimoija oli Adam.

Tutkimuksessa käsiteltyihin pariutuksiin vertaaminen voidaan toteuttaa suorittamalla pienemmällä datasetillä koulutettujen mallien avulla ennusteet jokaiselle A pisteelle suuremmassa datasetissä. Nämä ennusteet vastaavat pariutusta. Toteuttamalla pariutus tällä tavalla ja käsitellen pariutustulosta samoin kun muilla menetelmillä, saadaan malleja hyödyntävien pariutusten tarkkuus ja herkkyys yli kaikkien suuren datasetin tiedostojen esitettyä. Kuvassa 38 esitetään pienemmällä datasetillä koulutettujen mallien pariutustulos, vertailukohtana toimivat ristiin suodatettu ja keskiarvoperusteinen pituussuodatus haku.

Tuloksista voidaan havaita miten validaatio osuudessa pohdittu luokkien virheluokittelun yhdistäminen selkeästi vaikuttaa, sillä kumpikaan malleihin perustavista hauista ei tuottanut merkittävästi parempaa pariutustulosta. Johdannaisena samasta syystä niiden hakutulos on kuitenkin merkittävästi tasaisempi yli jokaisen parametrin i askeleen. Tämän takia niiden tarkkuuden voidaan osoittavan tämän



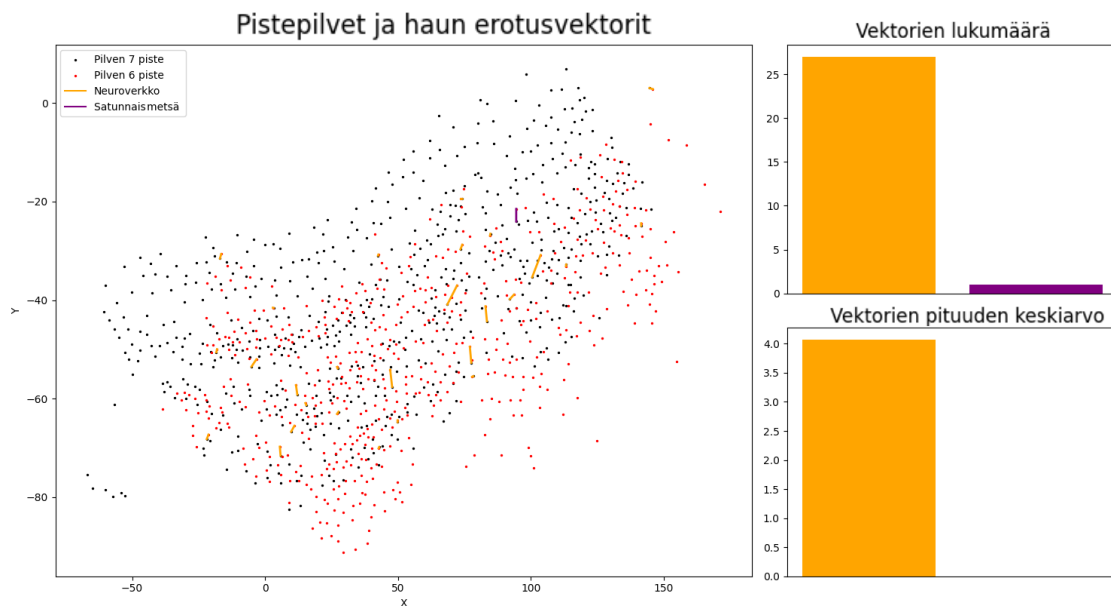
Kuva 39. Kuvassa malleja hyödyntävien hakumenetelmien ja vertailuun käytettävien hakumenetelmien herkkyys yli kaikkien suuremman datasetin tiedostojen. Vertailu menetelminä pituus suodatus keskiarvorajalla (sininen), ristiin verrattu suodatus (oranssi). Tekoälymallien tulos esitetään punaisella (neuroverkko) ja vihreällä (satunnaismetsä).

lähestymistavan olevan geneerisesti tarkempi pariutusmenetelmä.

Malleihin perustuvan pariutuksen tasaisuutta tukee myös kuvassa 39 esitetty hakumenetelmien herkkyys jonka nähdään pysyvän merkittävästi tasaisempuna koneoppimista hyödyntävillä menetelmillä.

Menetelmillä tehtiin myös kokeita käyttäen pienemmällä datasetillä koulutettuja malleja LIDAR- pilvien pariutukseen. Näistä tuloksista esitetään nyt pilvien 7 ja 6 välinen pariutus, jota tarkasteltiin myös muilla hakumenetelmillä. Kuvassa 40 esitetään tuloksena saatu pariutus ja sen vektoreihin liittyvät arvot vektorien lukumäärälle ja niiden pituuden keskiarvolle. Haussa satunnaismetsä onnistui pariuttamaan vain yhden pisteparin. Lisäksi neuroverkon pariutustulos ei tarkastelun perusteella ole johdonmukainen. LIDAR- pariutuksia tarkastellessa, tulee huomioida myös, että samojen pisteiden välisiä pariutuksia ei ole välttämättä mahdollista havaita pariutusvektorien päällekkäisyyden takia. Pidetään kuitenkin tutkimuksen kannalta oleellisempina havaita pariutukset, joissa vektorit selkeästi poikkeavat toisistaan. Tämän takia tarkastelu voidaan toteuttaa kuten aikaisemmin.

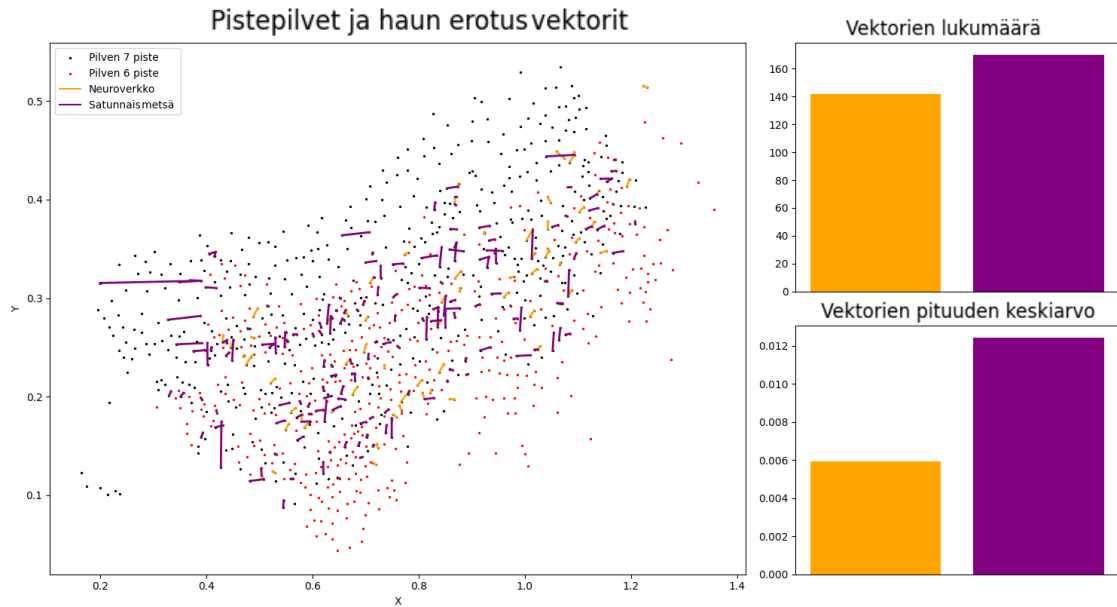
On kuitenkin huomioitava, että tämän haun yhteydessä pisteiden geografinen



Kuva 40. Malleja käyttävä pariutustulos LIDAR- datan pilville 7 ja 6. Kuvassa esitetään pisteet pilvestä 7 mustalla, pisteet pilvestä 6 punaisella, neuroverkon tuottamat vektorit oranssilla ja satunnaismetsän tuottamat lilalla. Lisäksi esitetään erotusvektoreiden lukumäärät ja niiden pituuden keskiarvo.

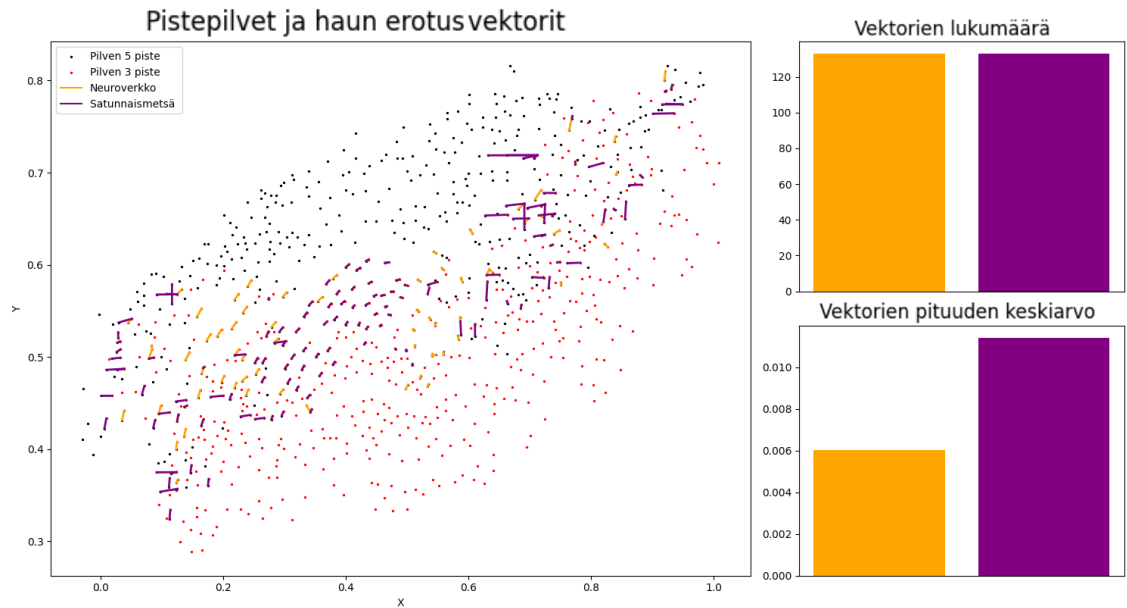
mitta poikkeaa koulutusdatana käytettyjen synteettisten pilvien mitoista merkittävästi. Kuten tiedetään, on syötetyn datan skaalauksen pysyminen vakiona oleellista koneälyn toimivuuden kannalta. Skaalaamalla pisteiden koordinaattiarvoja 200 kertaa pienemmäksi, saatiin molemmilta menetelmiltä merkittävästi parempi tulos. Kuvassa 41 esitetään skaalauksen jälkeen saatu pariutustulos pilvien 7 ja 6 välisessä pariutuksessa. Näiden pilvien pariutus on LIDAR- datasetissä yksi epäedukkaimmista sillä sen tarkastelu silmämääräisesti näyttää, että oikean pariutuksen määrittely ei ole helppo. Voidaan kuitenkin todeta molempien menetelmien saaneen myös järkevän oloisia tuloksia. Havaitaan kuitenkin satunnaismetsän taipumus poimia selkeästi pituuden mukaan väriä pariutuksia.

Kuvassa 42 esitetään eräs LIDAR- datasetin helpommin tarkasteltavista pariutuksista pilvien 5 ja 3 välillä. Tämän pilviparin pariutuksesta voidaan todeta neuroverkon pariutustuloksen olevan järkevän ja oikean näköinen. Myös satunnaismetsän



Kuva 41. Malleja käyttävä pariutustulos LIDAR- datan pilville 7 ja 6, niiden koordinaattien skaalauksen jälkeen tehdyllä pariutuksella. Kuvassa esitetään pisteet pilvestä 7 mustalla, pisteet pilvestä 6 punaisella, neuroverkon tuottamat vektorit oranssilla ja satunnaismetsän tuottamat liilalla. Lisäksi esitetään erotusvektoreiden lukumäärät ja niiden pituuden keskiarvo.

pariutustulos näyttää pääosin järkevältä, joitain selkeästi virheellisiä pariutuksia lukuun ottamatta. Aikaisemmissa kokeissa mainittu paikoitustarkkuus, joka oli suuruudeltaan $0,5m$ skaalautuu arvoon $0,0025m$ pistepilvien koordinaatistoa skaalatesa. Tämä huomioiden molemmissa tapauksissa neuroverkon pariutustulos vaikuttaa paikoitustarkkuuden puitteissa järkevältä. Tehtyjen kokeiden perusteella voidaan todeta neuroverkon soveltuvan parhaiten pistepilvien pariutustehtävään. Tehtyjen kokeiden perusteella on oletettava synteettisen ja LIDAR- datan välillä olevan jokin vaihtelun piirre, jota ei ole onnistuttu synteettisen datan generoinnin yhteydessä siihen sisällyttämään. Neuroverkon pariutustulos on kuitenkin merkittävän järkevä, josta syystä voidaan ajatella sen suoriutumiskyvyn paranevan merkittävästi, jos sen koulutukseen käytetyssä datassa voidaan edustaa paremmin LIDAR- datan kaltaisia pistepilviä. On myös huomioitava, että tutkimuksessa käytetty koordinaatiston skaalaus ei todennäköisesti vastaa täydelliseen muutokseen tarvittavaa suhdetta. Tällä



Kuva 42. Malleja käyttävä pariutustulos LIDAR- datan pilville 5 ja 3, niiden koordinaattien skaalauksen jälkeen tehdyllä pariutuksella. Kuvassa esitetään pisteet pilvestä 5 mustalla, pisteet pilvestä 3 punaisella, neuroverkon tuottamat vektorit oranssilla ja satunnaismetsän tuottamat liilalla. Lisäksi esitetään erotusvektoreiden lukumäärät ja niiden pituuden keskiarvo.

skaala- suhteella saadut tulokset kuitenkin osoittavat menetelmän toimivuuden, joten sen suuruuden tarkempi arviointi jätetään tämän tutkimuksen ulkopuolelle.

7 Pariutusaikojen tarkastelu

Taulukossa IV esitetään kokeissa käytettyjen menetelmien pariutusaikojen keskiarvo. Arvot lasketaan pariuttamalla jokainen suuren datasetin pariutus (1000 pistettä pilvessä A ja B) ja mittaamalla aika joka menetelmällä kuluu pariutukseen. Mittaukseen ei huomioida tiedostojen lataamista. Näistä arvoista lasketaan keskiarvo joka on esitetty taulukossa. Keskiarvon katsotaan heijastavan menetelmien todellista suorituskykyä hyvin sillä kaikilla menetelmillä mitatun ajan tiedostokohtainen vaihtelu oli erittäin pientä.

Mittausten vertailun mahdollistamiseksi on merkityksellistä tietää mittaukseen

Taulukko IV. Menetelmien yhden tiedoston pariutukseen käyttämän ajan keskiarvo mitattuna yli koko suuren synteettisen datajoukon. Pariutusajat mitataan pistetiedoston latauksen jälkeen ja lopetus aika otetaan pariutus tuloksen valmistuttua. Keskiarvo lasketaan jokaisen yhden tiedoston 1000 pisteen pariutukseen kuluneesta ajasta.

Menetelmä	Haun parametri	Pariutus aika [s]
Karteesinen haku	-	0.5026
Pituus suodatus	Vakio P_2	0.5029
Pituus suodatus	Tiedostokohtainen P_2	1.0056
Pituus suodatus	Keskiarvo	1.0063
Ristiin suodatus	-	1.0125
Siirtohaku	Keskiarvo / 2, askel=1, kulmia=2	32.7722
Neuroverkko	$N = 12$	58.6169
Satunnaismetsä	$N = 5$	32.7005

käytetyn tietokoneen konfiguraatio. Kokeissa käytetyn tietokoneen prosessori oli Intel core i7 10875H, käytössä oli 32Gt välimuistia ja kokeet suoritettiin Windows 11 käyttöjärjestelmässä. Koodit ajettiin Visual Studio Code kääntäjällä.

Taulukon perusteella huomataan heti miten suodatusmenetelmien käyttö kaksinkertaistaa pariutusajan. Lisäksi pariutusajan ero eri suodatusmenetelmien välillä on merkityksettömän pieni, jonka johdosta voidaan todeta että kulloiseenkin pariutustehtävään voidaan valita siihen parhaiten soveltuva suodatus, ilman että se merkittävästi vaikuttaa pariutusaikaan.

Siirtohaku ja molemmat koneoppimista hyödyntävät haut suoriutuvat pariutuksesta merkittävästi pidemmässä ajassa. On kuitenkin todennäköistä että näiden pariutusaikojä voidaan pienentää optimoimmalla käytettyjä algoritmeja, eivätkä esitettyt ajat siten täysin kuvaa menetelmien todellisia pariutusaikojä. Tämän tutki-

muksen kannalta mitatut ajat riittävät kuvaamaan menetelmien pääpiirteisiä eroja. Tulokset ovat myös erittäin mielenkiintoisia sillä ne osoittavat satunnaismetsän ja siirtohaun suoriutuvan tehtävästä lähes samassa ajassa. Huomioiden satunnaismetsän tarkkuuden ja herkkyyden tulokset jotka aikaisemmassa kappaleessa käsiteltiin voidaan siten todeta koneoppimis menetelmistä satunnaismetsän soveltuvan paremmin synteettistä dataa muistuttavaan pistepilvien pariutukseen.

8 Johtopäätökset

Tutkimuksessa kokeillut menetelmät osoittautuivat hyviksi työkaluiksi käsiteltäessä helpohkoja pariutusongelmia. Tilanteissa joissa pariutusongelman voidaan todeta liian vaikeaksi testatuille menetelmille, voidaan siirtyä hyödyntämään laskennallisesti raskaampia menetelmiä kuten ominaismuotoja [6]. Pariutusongelmaan sovellettävien menetelmien valinnassa isossa osassa on A pisteiden sekä A ja B pisteiden välisten etäisyyksien jakauma. Näitä jakaumia tarkastelemalla on mahdollista saada käsitys ongelmaan soveltuvien menetelmien vaatimuksista. Todellisten sovelluskohdeiden osalta voidaan myös todeta arvojoukko kuvaajan olevan hyödyllinen työkalu tuloksen järkevyyden arviointiin. Erityisesti tilanteissa, joissa ongelmasta tiedetään esimerkiksi arvio melun tai muun piirteen suuruudesta.

Koneoppimissovelluksien kanssa tehdyin kokein havaittiin neuroverkkomallin soveltuvan pistepilvipariutukseen, sillä sen tuottama LIDAR- datan pariutus muistuttaa eniten oikeaksi oletettavaa pariutusta. Pidetään todennäköisenä sen pariutustarkkuuden merkittävää kasvua, jos koulutusdatassa voidaan käyttää oikeanlaisia dataa. Dataan tarvittaviin muutoksiin voidaan saada johtolankoja analysoimalla tarkemmin sekaannusmatriisin osoittamien väärin luokiteltujen tapausten sisältöä. On myös oleellista tehdä lisätutkimusta LIDAR- datassa esiintyvistä piirteistä jonka puitteissa voidaan tuottaa koulutusdataa, jonka avulla mallin uskotaan saavut-

tavan merkittävä tarkkuus. Väärin luokiteltujen tapauksien huomattiin olevan yksinomaan virheellisesti 0 luokkaan jakautuneita pariutuksia, näiden perusteella voisi olla mahdollista päätellä ero parillisen ja parittoman pisteen väliltä näissä rajatapauksissa. Tämä tarkastelu jää kuitenkin tämän tutkimuksen ulkopuolelle.

Lopputuloksen kannalta on oleellista huomioida, että jokainen hakumenetelmä, joita tutkimuksessa kokeiltiin kuten myös tutkimuksen ulkopuolelle jääneet menetelmät, omaavat kukin omat rajoitteensa ongelman piirteistä ja vaikeudesta riippuen. Pariutusongelmaa käsiteltäessä on pyrittävä analysoimaan ja pohtimaan ongelman vaikeutta, voidaanko ongelman ratkaisua pitää mahdollisena ja voidaanko sitä helpottaa esikäsittelemällä. Pariutusongelma voi olla myös mahdoton ratkaista, jos pisteiden väliset yhtäläisyydet katoavat suuriin etäisyyksiin tai isoon meluun. Tutkimuksen ulkopuolelle jäi myös mahdottoman ongelman havaitseminen sen laaja-alaisuuden takia.

Tutkimus kattaa kaikki helpohkojen pariutusongelmien kannalta oleelliset seikat. Ongelman laajempi tarkastelu voisi kattaa arvojoukon kuvan ja vektorikenttäfunktioiden yhteyksien tarkastelun ja jo yllä mainitun mahdottoman ongelman rajaamisen. Nämä seikat toisivat osaltaan vastauksia tutkimuksessa avoimiksi jääneisiin kysymyksiin ja syventäisivät itsessään käsitystä ongelman luonteesta ja käyttäytymisestä eri tilanteissa.

Viitteet

- [1] Craig, N., Aldenderfer, M. & Moyes, H. (2006), "Multivariate Visualization and Analysis of Photomapped Artifact Scatters"(PDF), *Journal of Archaeological Science*, vol. 33, no. 11, pp. 1617–1627, Bibcode:2006JArSc..33.1617C, doi:10.1016/j.jas.2006.02.018, archived from the original (PDF) on 4 October 2007
- [2] Shahriar Akramullah. *Digital Video Concepts, Methods, and Metrics: Quality, Compression, Performance, and Power Trade-off Analysis*. Apress, Berkeley, CA, 2014. 2
- [3] Raj, T.; Hashim, F.H.; Huddin, A.B.; Ibrahim, M.F.; Hussain, A. A Survey on LiDAR Scanning Mechanisms. *Electronics* 2020, 9, 741. <https://doi.org/10.3390/electronics9050741>
- [4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," in *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99-110, June 2006.
- [5] Chen, Jorge & Mora, O. & Clarke, Keith. (2018). ASSESSING THE ACCURACY AND PRECISION OF IMPERFECT POINT CLOUDS FOR 3D INDOOR MAPPING AND MODELING. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. IV-4/W6. 3-10. 10.5194/isprs-annals-IV-4-W6-3-2018.
- [6] Nevalainen, Paavo & L., Qingqing & Melkas, Timo & Riekkilä, K. & Westerlund, Tomi & Heikkonen, Jukka. (2020). Navigation and Mapping in Forest Environment Using Sparse Point Clouds. *Remote Sensing*. 12. 10.3390/rs12244088.
- [7] Schey, H. M. (1997). *Div, Grad, Curl, and All That: An Informal Text on Vector Calculus* (4th ed.). W. W. Norton & Company. Third edition.
- [8] Skala, Vaclav & Smolik, Michal. A New Approach to Vector Field Interpolation, Classification and Robust Critical Points Detection Using Radial Basis Functions. *Computer Science On-line Conference* (2018).
- [9] Ohio State University. "This is your brain detecting patterns: It is different from other kinds of learning, study shows." *ScienceDaily*. ScienceDaily, 31 May 2018. <www.sciencedaily.com/releases/2018/05/180531114642.htm>.
- [10] T. Shan and B. Englot, "LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 2018, pp. 4758-4765, doi: 10.1109/IROS.2018.8594299.
- [11] TOPP. The open problems project. <https://topp.openproblem.net/p6>. Accessed: 2024-02-26.

- [12] Kasturi R. Varadarajan and Pankaj K. Agarwal. Approximation algorithms for bipartite and non-bipartite matching in the plane. In Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '99, page 805–814, USA, 1999. Society for Industrial and Applied Mathematics.
- [13] Gang Wang, Zhicheng Wang, Yufei Chen, Xianhui Liu, Yingchun Ren, and Lei Peng. Learning coherent vector fields for robust point matching under manifold regularization. *Neurocomputing*, 216:393–401, 2016.
- [14] Yikun Bai, Huy Tran, Steven B. Damelin, and Soheil Kolouri. Partial Transport for Point-Cloud Registration. arXiv e-prints, page arXiv:2309.15787, September 2023.
- [15] Fan, Taosha & Wang, Hanlin & Rubenstein, Michael & Murphey, Todd. (2020). CPL-SLAM: Efficient and Certifiably Correct Planar Graph-Based SLAM Using the Complex Number Representation.
- [16] Y. Deng, A. Rangarajan, S. Eisenschenk and B. C. Vemuri, "A Riemannian Framework for Matching Point Clouds Represented by the Schrödinger Distance Transform," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 3756-3761.
- [17] Huang, Xiaoshui, Guofeng Mei, Jian Zhang and Rana Abbas. "A comprehensive survey on point cloud registration." ArXiv abs/2103.02690 (2021).
- [18] Pomerleau, François & Colas, Francis & Siegwart, Roland. A Review of Point Cloud Registration Algorithms for Mobile Robotics. *Foundations and Trends® in Robotics*. 4. 1-104. 10.1561/23000000035. (2015).
- [19] Benjamin Ricaud, Pierre Borgnat, Nicolas Tremblay, Paulo Gonçalves, Pierre Vandergheynst, Fourier could be a data scientist: From graph Fourier transform to signal processing on graphs, *Comptes Rendus Physique*, Volume 20, Issue 5, 2019, Pages 474-488, ISSN 1631-0705,
- [20] Scikit-learn: Machine Learning in Python, Pedregosa et al., *JMLR* 12, pp. 2825-2830, 2011.
- [21] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

- [22] Li, Lisha, and Kevin Jamieson. "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization." *Journal of Machine Learning Research* 18 (2018): 1-52