

# 3D-overflatemodellering basert på stereoskopiske bilder

**Thomas Thorsen Røren**

Master i elektronikk

Innlevert: Juli 2012

Hovedveileder: Lise Lyngsnes Randeberg, IET

Medveileder: Lukasz Paluchowski, IET

Norges teknisk-naturvitenskapelige universitet  
Institutt for elektronikk og telekommunikasjon



# Oppgavetekst

---

Oppgaven går ut å danne en prototype 3D-modell av stereoskopiske bilder for anvendelse ved dybdeestimering og visualisering av sår i hud. Det er en føring for oppgaven at man ønsker en prototypemodell som visuelt gjengir mest mulig nøyaktig det aktuelle avbildningsobjektet som skal modelleres. Oppløsningen for dybdeforskjellene i modellen bør maksimalt være 1,5 mm. Prosesseringstid er ikke relevant ved implementering av prototypemodellen på dette stadiet av prosjektet, men bør være av en rimelig karakter for testing og videre undersøkelser.





# Abstract

---

The goal of this project has been to develop a prototype 3D model based on stereoscopic images. The purpose is to model chronic ulcers for visualization and depth estimation of the wounds. The model can be used for area and volume calculations of the wounds which will give valuable diagnostic information. The stereo images were subject to geometrical calibration, image rectification and image matching to conduct the modeling. The challenges of these procedures were related to the image matching and the height resolution in the model. Image matching can be improved by preprocessing the stereo images to enhance the contrast, and the height resolution can be increased by changing the hardware used in the project. As a continuation of this project the 3D model will be combined with hyperspectral imaging to also give spectral information in the wound.

# Sammendrag

---

Prosjektet har omhandlet utvikling av en prototype 3D-modell som er basert på stereoskopiske bilder. Hensikten er å modellere kroniske hudsår for visualisering og dybdeestimering av sårene. Modelleringen vil kunne brukes til areal- og volumberegninger av sårene som gir verdifull diagnostisk informasjon. Det ble gjort geometrisk kalibrering, bilderektifikasjon og bildematching av stereobildene for å kunne foreta modelleringen. Utfordringer under disse prosessene var knyttet til presisjon av bildematchingen av homogene områder og høydeoppløsningen i modellen. Bildematchingen kan forbedres ved å forbehandle stereobildene for å øke kontrasten, og høydeoppløsningen kan økes ved å endre det eksperimentelle oppsettet. Som en videreføring av dette prosjektet skal 3D-modelleringen kombineres med hyperspektrale avbildninger for i tillegg å kunne gi spektral informasjon om såret.



# Forord

---

Denne masteroppgaven ble utført våren 2012 ved institutt for elektronikk og telekommunikasjon, Norges Teknisk-Naturvitenskapelige Universitet. Hovedveileder har vært førsteamanuensis Lise Lyngsnes Randeberg og biveileder har vært doktorgradsstudent Lukasz Paluchowski. Jeg vil rette en stor takk til Lukasz Paluchowski for uvurderlig hjelp under masteroppgaven. Han er svært kompetent på fagfeltet sitt og har kunnet bidratt med mye gjennom dette halvåret. Jeg vil også takke Lise Lyngsnes Randeberg som foreslo og gav meg muligheten å jobbe med denne masteroppgaven. I tillegg vil jeg anerkjenne Norsk Elektro Optikk AS som har bidratt med teknisk utstyr. Fagfeltene omhandlet i dette prosjektet er relativt nye områder for meg, noe som har gjort dette til et utfordrende prosjekt og har bidratt til at jeg har lært svært mye gjennom dette semesteret.

Trondheim,  
Thomas Thorsen Røren

# Innholdsfortegnelse

---

<b>1</b>	<b>INNLEDNING</b>	<b>6</b>
<b>2</b>	<b>TEORI</b>	<b>8</b>
2.1	INTRODUKSJON TIL FOTOGRAMMETRI	9
2.1.1	Skalering	9
2.1.2	Koordinatsystemer	10
2.1.3	Perspektivprojeksjon	11
2.2	KALIBRERING	12
2.2.1	Radiell og tangentiell distorsjon	13
2.2.2	Sirkeldeteksjon	15
2.3	STEREOSKOPI	16
2.3.1	Parallakse	16
2.3.2	Synsfelt i stereoskopi	17
2.4	STEREODISPARITET	18
2.4.1	Bildematchingsalgoritmer	19
2.4.2	Metriske metoder for matching	21
2.5	BILDEREKTIFIKASJON	23
2.5.1	Epipolare linjer og plan	24
2.5.2	Rektifikasjon av bildematrixene	25
2.6	MODELLERING	26
2.6.1	Koordinater i objektrommet	27
2.6.2	Høydeopløsning	27
<b>3</b>	<b>METODE</b>	<b>29</b>
3.1	EKSPERIMENTELT OPPSETT OG PROGRAMMERINGSSPRÅK	30
3.1.1	Utstyr	30
3.1.2	Akvisisjon av stereobilder	31
3.1.3	Fremgangsmåte	32
3.1.4	Programmeringsspråk	33
3.2	KALIBRERING	33
3.2.1	CalDe	33
3.2.2	CalLab	34
3.2.3	Implementasjon av kalibreringsresultatene	38
3.3	REKTIFIKASJON	40
3.3.1	Transformering av bildematrixene	40
3.3.2	Punkt-deteksjon og matching	41
3.3.3	Estimering av fundamentalmatrisen	41

3.3.4 Perspektivtransformering.....	42
3.4 MATCHING.....	43
3.4.1 Valg av metrisk matchingsmetode.....	43
3.4.2 Fremgangsmetode for matching.....	43
3.4.3 Koordinatsystem og parallakseberegning.....	45
3.5 MODELLERING .....	46
3.5.1 Interpolering av Z-verdier.....	46
3.5.2 Koordinater i objektrommet.....	47
<b>4 RESULTATER.....</b>	<b>48</b>
4.1 KALIBRERING.....	49
4.1.1 CalDe .....	49
4.1.2 CalLab .....	50
4.1.3 Implementasjon av kalibreringsresultatene.....	51
4.1.4 Verifikasjon av kalibreringsresultatene.....	54
4.2 REKTIFIKASJON .....	58
4.2.1 Punktdeteksjon og matching.....	60
4.2.2 Perspektivtransformering.....	62
4.3 MATCHING.....	63
4.3.1 Korresponderende punkter i stereobildene .....	64
4.3.2 Parallakseberegning.....	65
4.4 MODELLERING .....	67
4.4.1 Fjerning av ekstrempunkter .....	67
4.4.2 Modellen.....	67
<b>5 DISKUSJON.....</b>	<b>70</b>
5.1 YTELSE.....	70
5.2 MODELLERING AV SÅR .....	71
<b>6 KONKLUSJON .....</b>	<b>73</b>
<b>7 BIBLIOGRAFI.....</b>	<b>75</b>
<b>8 VEDLEGG.....</b>	<b>77</b>
8.1 KODEOVERSIKT .....	77
8.2 DIVERSE MODELLER.....	78

# Figurliste

---

Figur 2.1: Illustrasjon av koordinatsystemene i bilderommet og i objektrommet .....	10
Figur 2.2: Illustrasjon av linsedistorsjon .....	12
Figur 2.3: Nåleputedistorsjon rutenett (t.v.) og tønne-distorsjon rutenett (t.h.) .....	13
Figur 2.4: Illustrasjon av parallakse ved punktet A .....	17
Figur 2.5: Bilde 1 viser utsnitt av referansebildet. Dette utsnittet kalles referansevinduet og brukes til å søke gjennom bilde 2 for matching. ....	20
Figur 2.6: Stereobilder før (1) og etter (2) rektifikasjon. Etter rektifikasjon er søkeområdet kun i horisontal retning, det vil si kun over en rad .....	23
Figur 2.7: Illustrasjon av et punkt i rommet som blir projisert på to bildeplan, og den korresponderende epipolare geometrien er også illustrert.....	25
Figur 2.8: Illustrasjon av bildeplanene etter rektifisering .....	26
Figur 3.1: Eksperimentelt oppsett med kamera, linse og translasjonstrinn. Den røde sirkelen viser kameraet og linsen. ....	30
Figur 3.2: Illustrasjon av sjakkbrettmønsteret med tre punkter og deres koordinater .....	34
Figur 3.3: Delaunay triangulering .....	39
Figur 3.4: Rutenett ved matching. De grønne punktene indikerer de punktene som blir matchet. ....	44
Figur 3.5: Matrise med korrelasjonskoeffisienter ved matching av et punkt. X og Y-aksen representerer rader og kolonner respektivt, og Z-aksen representerer korrelasjonskoeffisientene.....	45
Figur 3.6: Interpolert overfalte og den faktiske overflaten i 3D-modellen .....	46
Figur 4.1: Detekterte hjørner på sjakkbrettet i CalDe.....	49
Figur 4.2: De fem lokale pikselfeilhistogrammene samt det globale pikselfeilhistogrammet markert i rødt .....	50
Figur 4.3: Sjakkbrett før kalibrering.....	52
Figur 4.4: Sjakkbrett etter kalibrering .....	52
Figur 4.5: Intensitetsforskjell mellom figur 4.3 og figur 4.4 .....	53
Figur 4.6: Konturbildet av pikselskiftet før kalibrering og etter kalibrering. Tallene til høyre er oppgitt i piksler.....	54
Figur 4.7: Lagerkuler før kalibrering. Den grønne markeringen tilsvarer den optimale formen på kulen, mens den røde tilsvarer den faktiske formen på kulen.....	55
Figur 4.8: Lagerkuler etter kalibrering. Den grønne markeringen tilsvarer den optimale formen på kulen, mens den røde tilsvarer den faktiske formen på kulen.....	56
Figur 4.9: Stereobildene av modelleiren med en basislinje på 10 cm .....	59
Figur 4.10: Anaglyfisk visning av stereobildene. 3D-effekten kan ses med ordinære 3D-briller farget rødt og cyan. ....	59

Figur 4.11: De røde sirklene markerer detekterte punkter i stereobildene. Her er kantpunktene fjernet. ....	60
Figur 4.12: Punktene som er igjen etter fjerning av punkter som ikke kan matches eller er matchet feil .....	61
Figur 4.13: Korresponderende punkter i anaglyfisk visning. De røde sirklene markerer punktene i bildet til venstre og de grønne tegnene markerer punktene i bildet til venstre.....	61
Figur 4.14: Korresponderende punkter i anaglyfisk visning etter RANSAC.....	62
Figur 4.15: Korresponderende punkter i anaglyfisk visning etter rektifikasjon samt tre utvalgte punktpar. Alle tre punktpar er rette og like lange.....	63
Figur 4.16: Feilmatching i stereobildene .....	64
Figur 4.17: Forbedring av matching ved å begrense søkeområdet.....	65
Figur 4.18: Området i bildet som har blitt matchet .....	66
Figur 4.19: Interpolerte parallakseverdiene av figur 4.18.....	66
Figur 4.20: 3D-modellen av modelleiren beregnet på bakgrunn av parallakseverdiene som vist i figur 4.19. Modellen vises fra det nordøstlige hjørnet av figur 4.18. X-, Y- og Z-aksene er gitt i centimeter. ....	68
Figur 4.21: 3D-modellen med referansevindu på $40 \times 40$ piksler. Modellen vises fra det nordøstlige hjørnet av figur 4.18 .....	69



# 1 Innledning

---

Hensikten med dette prosjektet er å prøve å danne en prototype av en topografisk 3D-modell av kroniske hudsår som er avbildet todimensjonalt. Målet er å kombinere en slik 3D-modell sammen med hyperspektral avbildning til diagnostisk bruk. Til dette formålet vil et høyoppløselig hyperspektralt kamera sammen med et 3D-avbildningssystem gi spektral informasjon om såret i tillegg til visualisering og dybdeestimering av såret. Den spektrale informasjonen vil kunne gi viktig diagnostisk data som oksygenering av vev, pigmentering samt klassifisering av såret. 3D-modellen vil kunne bidra med areal- og volumberegninger av såret. Ved å følge pasienter over tid vil areal- og volumforskjellene i såret kunne være et verdifullt diagnostisk verktøy. Et sår gror nedenfra og opp og dermed vil man kunne trekke slutninger om det skjer en forbedring i såret eller en forverring. Det er spesielt tre typer kroniske sår som det skal fokuseres på:

1. Venøse leggsår
2. Arterielle leggsår
3. Diabetiske fotsår

Venøse leggsår skyldes at veneklaffene i blodårene ikke fungerer som de skal. Dette gjør at blodet kun renner nedover i beina og medfører økt trykk i blodårene, og dette gir igjen åreknuter og fører til at væske siver ut i vevet. Konsekvensene av dette kan være hevelser og sår som ikke gror av seg selv. Arterielle leggsår er en konsekvens av utilstrekkelig blodtilførsel som gir oksygenmangel i hud og vev. Den vanligste årsaken til at dette skjer er åreforkalkning. Nevropati og arteriosklerose er to senkomplikasjoner ved diabetes som kan gi diabetiske fotsår. Amputasjon er en vanlig konsekvens av disse typer sår, og 85 % av alle amputasjoner hos pasienter med diabetes skyldtes denne type sår.

I dag måles og diagnostieres disse sårene på en svært enkel og unøyaktig måte med bruk av målebånd og visuell inspeksjon. Dette prosjektet vil dermed kunne være en ny og ikke-invasiv metode for karakterisering av kroniske sår. Systemet vil kunne tas i bruk av dermatologer, karkirurger, ortopeder og plastikkirurger.

Dette prosjektet vil altså kun fokusere på dannelsen av en prototype for topografisk 3D-modellering. Det finnes en del eksisterende teknikker for modellering i 3D. Eksempler på slike teknikker er LIDAR, interferometri og stereoskopi. I dette prosjektet har 3D-modellering basert seg på stereoskopiske bilder ved translasjon av et kamera ved å bruke et translasjonstrinn. Stereoskopi

er et begrep som ble først beskrevet i 1838 av Charles Wheatstone. Han gjorde en rekke fysiologiske og optiske eksperimenter for å prøve å forstå dybdesyn hos mennesker. Stereoskopi prøver nemlig å emulere dybdesynet til mennesker ved enten bruke to kameraer eller translere et kamera mellom tagging av to bilder. I dag blir 3D-teknologien stadig mer utbredt. Den tas i bruk av NASA, underholdningsindustrien i både spillindustrien og filmindustrien og i kartografi, for å nevne noen områder.

Utviklingen av denne prototypen har basert seg på ny kodefunksjonalitet istedenfor å forbedre eksisterende koder. For å begrense lengden på rapporten og gi mest mulig flyt i lesningen tar ikke rapporten utgangspunkt i å forklare den implementerte koden i detalj, men heller forklare den prinsipielle virkemåten og resultatene av kodene.

Hovedkapitlene i rapporten er strukturert på følgende måte:

- (2) Dette kapitlet gir en kort innføring i den teoretiske bakgrunnen for de respektive emnene aktuelt for dette prosjektet. For å ha forståelse for metodebruk og resultater er det viktig å ha en forståelse av den teorien som har blitt lagt til grunn ved prosjektgjennomføring. Kapitlet vil også bidra til en større forståelse av de problemene og utfordringer som kommer ved implementasjon av teorien.
- (3) Kapitlet gir en oversikt over metodene som har blitt brukt underveis i prosjektet. Her beskrives fremgangsmåten som har blitt benyttet for å komme fram til resultatene. Det gis også begrunnelser for hvorfor disse fremgangsmetodene har blitt benyttet framfor andre metoder.
- (4) Her blir resultatene som er oppnådd beskrevet. Kapitlet viser den endelige implementasjonen og hvordan forskjellige parametere endrer resultatene. Det forklares også eventuelle endringer av fremgangsmetoden som måtte gjøres på bakgrunn av tidligere resultater.
- (5) Her gis det en evaluering og en diskusjon av resultatene som har blitt oppnådd. Det kommer også en rekke forslag til hvordan systemet kan endres for å oppnå enda bedre resultater.
- (6) Konklusjon som oppsummerer rapporten. Resultatene sammenlignes med det målet som er satt i oppgaveteksten. Det gis også anbefalinger for hvilke områder som bør forbedres eller endres. Tilslutt gis det en kort beskrivelse av mulighetene for videre arbeid med dette prosjektet.

# 2 Teori

---

Kapitlet gir en kort innføring i det teoretiske grunnlaget for prosjektet. Først belyses fotogrammetri generelt, deretter de fotogrammetriske begrepene og aktuelle beregningsmetoder innenfor fotogrammetri. Andre del av kapitlet tar så for seg geometriske kalibrering av kamera for korreksjon av linsedistorsjon. Deretter forklares stereoskopi, stereodisparitet og bilderektifikasjon, som er sentrale begreper i dette prosjektet. Tilslutt utdypes det nærmere hvordan 3D-overflatemodellen beregnes og hvordan forskjellige parametere påvirker modelleringen.

## 2.1 Introduksjon til fotogrammetri

Fotogrammetri er et sett med teknikker som brukes til å bedømme geometriske egenskaper til objekter i fotografiske bilder (1). Mer konkret er hovedformålet til fotogrammetri å fastslå det geometriske forholdet mellom bildet og avbildningsobjektet som det eksisterte på den tiden da det ble avbildet. Når dette forholdet er fastslått kan informasjon om avbildningsobjektet innhentes ved kun å analysere bildet. Analyseringsteknikkene som brukes i fotogrammetri bygger på klassisk fotografitolkinger ved bruk av dataanalyseteknikker i tillegg til visuelle tolkninger. Et prosjekt som innebærer bruk av fotogrammetriske teknikker vil bestå av to hoveddeler:

1. Akvisisjon og forbehandling av bildene
2. Bildeprosessering for uthenting av aktuell informasjon

Den første hoveddelen innebærer å bestemme hvordan avbildning av det aktuelle objektet skal foretas, og preprosessering av disse bildene for å fremheve strukturer eller generelt sett gjøre det enklere å foreta bildebehandling senere. Den andre hoveddelen vil bestå av å avgjøre hvilken type fotogrammetrisk system, og derunder også hvilke metoder og algoritmer, som skal benyttes for uthenting av den informasjonen som er ønskelig.

### 2.1.1 Skalering

For å kunne beregne de metriske verdiene i bilder eller kart må skaleringsfaktoren kjennes. Denne faktoren relater lengdeenhetene i bildet til lengdeenhetene i virkeligheten og er gitt av forholdet mellom distansen mellom to punkter i bildet og distansen mellom de samme punktene i virkeligheten. Forholdet mellom fokallengden og høyden fra perspektivsenteret i kameraet er det samme som forholdet mellom bildedistansen mellom to punkter og den faktiske distansen mellom de samme punktene. Dette konstante forholdet skyldes formlike geometrier mellom de angitte distansene. Ved å bruke dette får vi altså følgende:

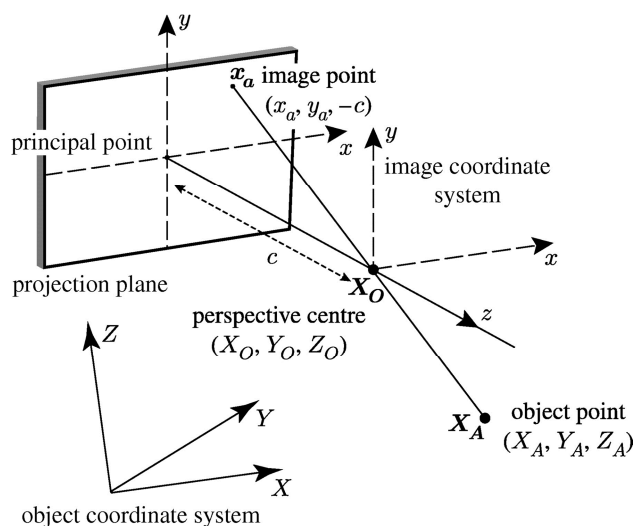
$$\text{skalering} = \frac{\text{bildedistanse}}{\text{objektdistanse}} = \frac{f}{h} \quad (2.1)$$

der  $f$  er fokallengden og  $h$  er avstanden mellom avbildningsobjektet og perspektivsenteret i kameraet.

Ideelt sett er skaleringsfaktoren uavhengig av om den måles i x-retning eller y-retning i bildet. Vertikale fotografier regnes som å ha retningsuavhengig skaleringsfaktor. Er derimot bildeplanet ikke parallelt med planet til avbildningsobjektet vil skaleringsfaktoren variere med retningen i bildet. Graden av denne variasjonen er relatert til hvor tiltet kameraet er i forhold til avbildningsobjektet. Selv om denne variasjonen ikke virker særlig stor vil et lite avvik ha stor betydning når dette beregnes med andre svært store eller små faktorer.

### 2.1.2 Koordinatsystemer

Den optiske akse til et kamera passerer gjennom perspektivsenteret og treffer bildeplanet i punkt som kalles prinsipalpunktet eller prinsipalpunktet av autokollimasjon (PPA) (1) som vist i figur 2.1.



Figur 2.1: Illustrasjon av koordinatsystemene i bilderommet og i objektrommet (2)

Prinsipalpunktet sammenfaller sjelden med bildets midtpunkt, men befinner seg i nærheten. Dette avviket mellom PPA og bildets midtpunkt bestemmes gjennom kalibreringsprosedyrer og utdypes nærmere i delkapittel 2.2. Figur 2.1 viser i tillegg et punkt i rommet,  $\mathbf{X}_A$ , som blir avbildet til et punkt i bildeplanet,  $\mathbf{x}_A$ . Punktet  $\mathbf{X}_A$  er gitt et koordinatsystem som refereres til som objektrommet og er gitt med koordinatene  $(X_A, Y_A, Z_A)$ . Dette punktet blir avbildet til et koordinatsystem som refereres til som bildeplanet eller bilderommet.

Koordinatene i bilderommet blir da gitt ved  $(x_a, y_a, -c)$ , der  $c$  er fokallengden. For å oppnå koordinater i bilderommet som er i henhold til perspektivsenteret må det

introduseres et skift i koordinatene slik at origo i bildekoordinatsystemet tilsvarer PPA-koordinatene i bildet.

En alternativ form for representasjon av koordinatene er bruk av såkalte homogene koordinater (1). Her er kun det relative forholdet mellom koordinatene av interesse. Denne typen koordinatsystem brukes ofte i projektiv geometri, siden de muliggjør punktrepresentasjon ved uendelig. Todimensjonale bildekoordinater blir representert som

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ w \end{bmatrix} \text{ der } x = \frac{x'}{w} \text{ og } y = \frac{y'}{w}. \quad (2.2)$$

### 2.1.3 Perspektivprojeksjon

For å kunne relatere koordinatene i objektrommet til koordinatene i bilderommet brukes en lineær transformasjon som kalles perspektivtransformasjon.

Hensikten med denne transformasjonen er å projisere et punkt i objektrommet til et punkt i bilderommet. Homografi er et matematisk begrep i geometri som knytter seg til projeksjon av rette linjer. De optiske strålene fra perspektivsenteret krysser to plan, bildeplanet og objektplanet. Forholdet mellom avstandene i bildeplanet og avstandene i objektplanet er gitt ved et unikt, konstant kryssforhold. Dermed kan punkter i bildeplanet beregnes på bakgrunn av punkter i objektrommet og omvendt. Alle bilder som tas av det samme objektplanet er relatert via homografi (3). Dette brukes i blant annet bilderektilifikasjon og beregning av rotasjon og translasjon av kameraet mellom to bilder.

Perspektivsenteret i figur 2.1 er gitt ved koordinatene  $(X_0, Y_0, f)$ , der  $Z_0$  tilsvarer fokallengden  $f$ . Ved å bruke dette samt antagelsen at koordinatsystemet i objektrommet og koordinatsystemet i bildeplanet sammenfaller, kan man utlede følgende basert på formlikhet i figur 2.1:

$$\frac{x}{f} = \frac{-X}{Z-f} = \frac{X}{f-Z} \Rightarrow X = \frac{x}{f}(f-Z) \quad (2.2)$$

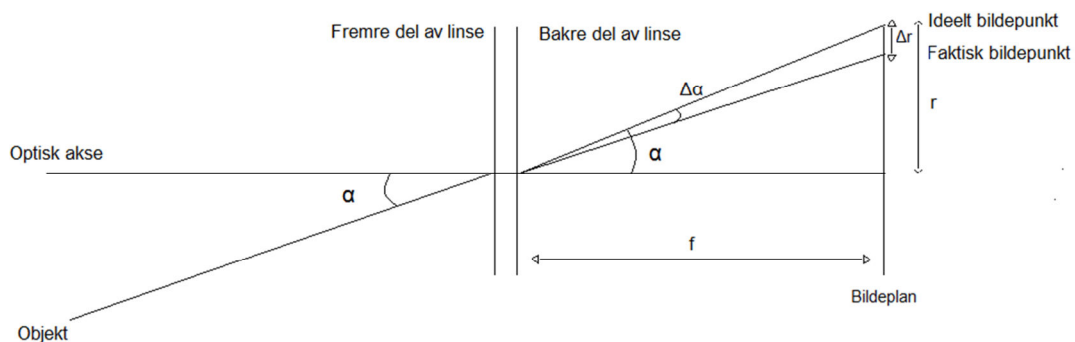
og

$$\frac{y}{f} = \frac{-Y}{Z-f} = \frac{Y}{f-Z} \Rightarrow Y = \frac{y}{f}(f-Z) \quad (2.3)$$

der  $(x, y)$  er koordinater i bildeplanet,  $(X, Y, Z)$  er koordinater i objektrommet og  $f$  er fokallengden.

## 2.2 Kalibrering

Alle bilder tatt med linser som er sfærisk formet vil ha aberrasjoner i mer eller mindre grad. Disse aberrasjonene gjør at bildet ikke representerer avbildningsobjektet perfekt i samtlige punkter i bildet, noe som konkret kan føre til feil når distansen mellom to punkter skal kalkuleres ved å skalere opp distansen mellom de tilsvarende pikslene. Det finnes fem hovedtyper aberrasjoner: Sfærisk aberrasjon, koma, astigmatisme, bildekrumming og distorsjon (1). Det er sistnevnte, linsedistorsjon, som direkte påvirker målenøyaktigheten i bildet og må derfor korrigeres i fotogrammetriske anvendelser gjennom geometrisk kalibrering. De andre aberrasjonene påvirker stort sett bildekvaliteten og ikke målenøyaktigheten. I tillegg til økning av den metriske nøyaktigheten i bildet vil også presisjonen ved bildematching øke ved geometrisk kalibrering. Figur 2.2 illustrerer at linsesystemer i praksis har et lite avvik i vinkelen mellom den optiske akse og den optiske strålen fra objektet inn til linsen og mellom vinkelen mellom den optiske akse og lysstrålen utfra linsen. Hvis linsen hadde hatt perfekt utforming uten distorsjon ville vinkelavviket vært lik null.



Figur 2.2: Illustrasjon av linsedistorsjon

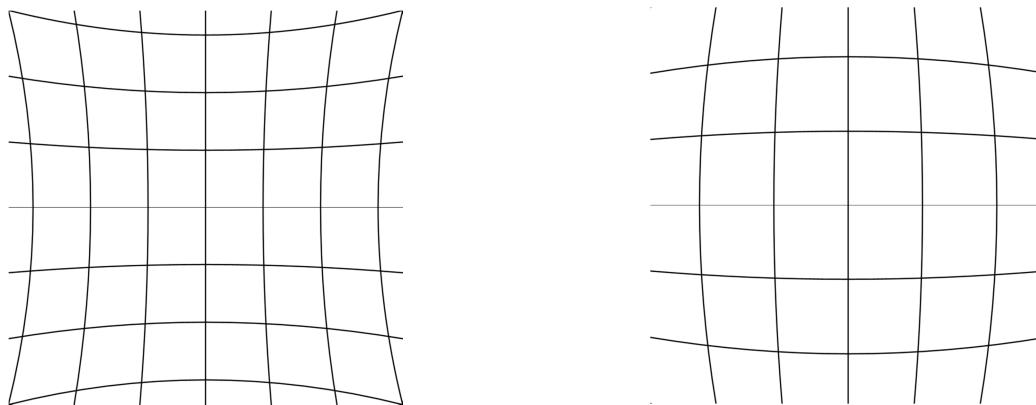
Distorsjon i bilder kan ha mange mønstre, inkludert irregulære mønstre, men stort sett beveger distorsjonen seg radielt utover i bildet med utgangspunkt i prinsippunktet. Radiell distorsjon er altså en radiell forskyvning av de korrekte

posisjonene til pikslene i bildet, og dette skyldes at vinkelen mellom den optiske aksene og lysstrålen forandrer seg radielt utover i bildet. Det vil blant annet si at distorsjonen øker jo lengre koordinatene i bildet er fra prinsippunktet (PPA). Den radielle komponenten av distorsjonen er gitt av ligning 2.4 og viser at den radielle distorsjonen endres ulineært med vinkelavviket.

$$\Delta r = r - f \tan(\alpha - \Delta \alpha) \quad (2.4)$$

### 2.2.1 Radiell og tangentiell distorsjon

Radiell distorsjonen kan vanligvis deles opp i to hovedtyper: nåleputedistorsjon og tønne-distorsjon (1). Figur 2.3 viser hvordan rutenettet blir endret til et ikke-uniformt fordelt rutenett ved nåleputedistorsjon til venstre i figuren og tønne-distorsjon til høyre i figuren. I tønne-distorsjon har den radielle distorsjonen positive verdier, det vil si at distorsjonsvektorene peker utover og vekk fra PPA i bildet. Sidekantene vil nå krumme utover og rutenett ligner dermed på en tønne. Nåleputedistorsjon har derimot negativ radiell distorsjon, og distorsjonsvektorene vil dermed peke innover mot PPA. Dette medfører blant annet at sidekantene blir krummet inn mot midten og formen til rutenettet minner om en nålepute.



Figur 2.3: Nåleputedistorsjon rutenett (t.v.) og tønne-distorsjon rutenett (t.h.) (4).

I tillegg til radiell distorsjon kan også distorsjonen være tangentiell og skyldes som regel produksjonsfeil av linsekomponentene (1). Distorsjonen kommer av at CCD-enheten i kameraet ikke ligger perfekt parallelt med bildeplanet.

Tangentiell distorsjon fører til en forskyvning av bildepunktene som beveger seg normalt på den radielle distorsjonen. For å korrigere for både tangentiell og radiell distorsjon kan de korrigerede koordinatene i bildet beregnes etter Browns modell (5) gitt av ligning 2.5 nedenfor.



$$\begin{aligned}
x_u &= x_d + (x_d - x_c)(K_1 r^2 + K_2 r^4 + \dots) + \\
&P_1(r^2 + 2(x_d - x_c)^2) + 2P_2(x_d - x_c)(y_d - y_c)(1 + P_3 r^2 + \dots)
\end{aligned}
\tag{2.5}$$

$$\begin{aligned}
y_u &= y_d + (y_d - y_c)(K_1 r^2 + K_2 r^4 + \dots) + 2P_1(x_d - x_c)(y_d - y_c) + \\
&P_2(r^2 + 2(y_d - y_c)^2)(1 + P_3 r^2 + \dots)
\end{aligned}$$

der

$(x_u, y_u)$  = ikke forvrent bildepunkt,

$(x_d, y_d)$  = forvrent bildepunkt,

$(x_c, y_c)$  = PPA i bildet,

$K_n$  = N-te radiell distorsjonskoeffisient,

$P_n$  = N-te tangentiell distorsjonskoeffisient,

$$r = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2},$$

... = uendelig serie.

Koordinatene  $x_u$  og  $y_u$  gjør det mulig å fastslå hvor de forvrente koordinatene ville befunnet seg med null distorsjon. Siden de ikke-forvrente koordinatene sjelden passer perfekt på et uniformt fordelt rutenett må det interpoleres for å kunne beregne intensitetsverdien til det uniformt fordelte rutenettet med samme størrelse som det opprinnelige bildet. For å kunne anvende ligning 2.5 må distorsjonskoeffisientene og PPA-koordinatene estimeres gjennom kalibreringsprosedyrer. For å kunne estimere distorsjonskoeffisientene og PPA er det hensiktsmessig å bruke et veldefinert mønster som kalibreringsobjekt. Dette kan for eksempel være mønsteret på et sjakkbrett. Hvis kameraet har betydelig distorsjon vil man kunne se hvordan mønsteret forandrer seg, spesielt i ytterkantene i bildet. Utfra disse endringene vil man kunne beregne de intrinsiske parameterne av kameraet, og dette er som regel første del av kalibreringen. Den intrinsiske matrisen inneholder fem parametere og består av (6):

$$\mathbf{A} = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}
\tag{2.6}$$

der  $\alpha_x = f \cdot m_x$ ,  $\alpha_y = f \cdot m_y$ , og der  $f$  er fokallengden i piksler  $m$  er

skaleringfaktorer i x- og y-retning.  $\gamma$  representerer en faktor som beskriver

skjevhetsforholdet mellom x- og y-aksen, og koordinatene  $u_0$  og  $v_0$  tilsvarer PPA i bildet. Kalibreringens virkemetode blir beskrevet i mer detalj i delkapittel 3.2.

### 2.2.2 Sirkeldeteksjon

For å verifisere kalibreringen eller estimere hvor vellykket kalibrering ble gjennomført er det hensiktsmessig å avbilde objekter med veldefinerte fysiske utforming for så å undersøke hvordan den geometriske formen eventuelt endret seg i bildet før og etter kalibreringen. Formen bør være enkel og presis slik at verifiseringen er enklest mulig å gjennomføre. Sirkler eller kuler er godt egnet til dette formålet siden de har enkel geometrisk utforming med konstant distanse fra kanten av sirkelen til sentrum. Ved å se på sirklene i et bilde med betydelig distorsjon vil man kunne se klare endringer i den geometriske formen til sirkelen. Krumningen i sirkelen vil kunne være et mål på hvor godt sirkelen i bildet representerer den geometriske formen til sirkelen som blir avbildet. Et annet mål på dette vil være å beregne kompaktheten til sirklene. Kompaktheten til en sirkel er gitt av ligning 2.7 (7). Ved å bruke like store kuler og ved å plassere kulene ved kantene og i midten av bildet vil man kunne sammenligne hvordan den geometriske utformingen forandrer seg på forskjellige steder i bildet.

$$\text{kompakthet} = \text{omkrets}^2 / \text{areal} \quad (2.7)$$

For å kunne måle avvikene i den geometriske utformingen i bildet må sirklene detekteres ved bruk av bildesegmenteringsalgoritmer. Hough-transformen er en slik algoritme som anvendes til å finne diverse geometriske former. Bildet blir transformert til et såkalt akkumulatorrom ved å bruke en funksjon som beskriver den geometriske formen man leter etter (8). Punkter som ligger på en sirkel er beskrevet av ligning 2.8.

$$(x - c_1)^2 + (y - c_2)^2 = c_3 \quad (2.8)$$

Fremgangsmåten i Hough-transformen er først å gjøre kantedetektering for så å konstruere sirkler for hver kantpixel inn i akkumulatorrommet. Deretter inkrementeres  $c_1$  og  $c_2$  i ligning 2.8 og verdiene som passer  $c_3$  best blir valgt basert på en terskel som blir satt (8). Bli denne terskelgrensen satt for lav kan det hende at sirklene ikke blir detektert og satt for høy kan det føre til falske deteksjoner.

## 2.3 Stereoskopi

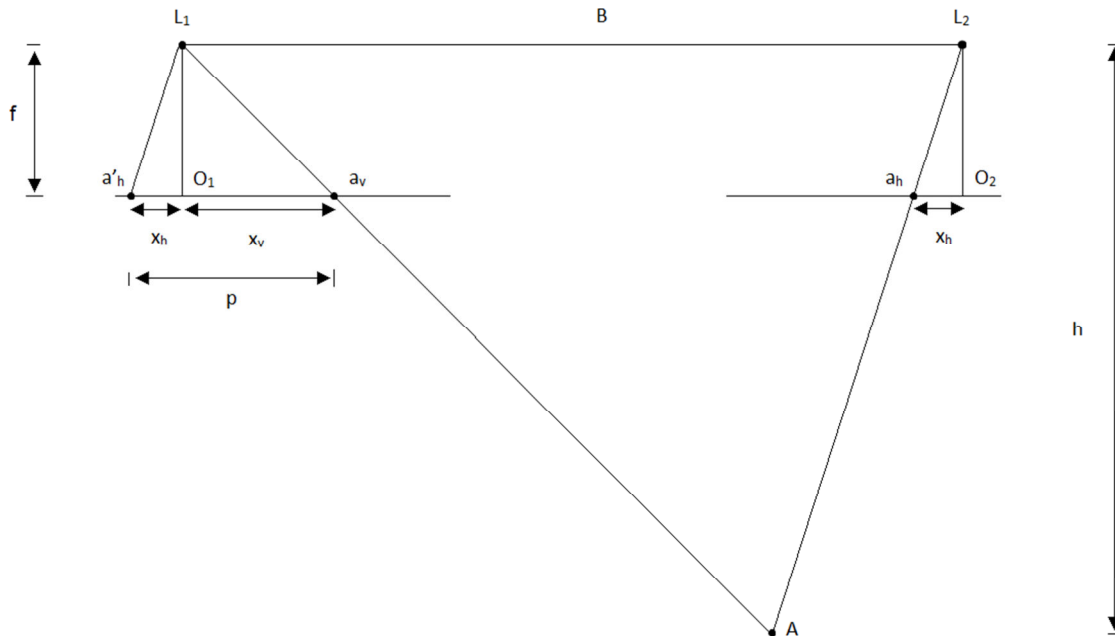
Begrepet stereoskopi beskriver en teknikk eller egenskap som relaterer seg til evnen å oppfatte dybde i rommet (1). Hos mennesker gjøres dette ved at objekter i rommet blir oppfattet fra hvert øye og dermed fra to ulike vinkler. Dette gjør hjernen i stand til å oppfatte forskjellige avstander til forskjellige objekter i rommet. Det er flere andre metoder hjernen bruker for å oppfatte dybde, men dette er primærteknikken som gjør oss i stand til å ha en tredimensjonal oppfatning av det vi ser. Hjernen er kapabel til å kombinere to ulike bilder om til et tredimensjonalt bilde i komplekse omgivelser og i tillegg gjøre denne prosesseringen i sanntid. Det bør derfor også være mulig å utarbeide en teknisk løsning som utøver den samme funksjonen.

### 2.3.1 Parallakse

For å anslå avstanden til objekter i rommet benytter blant annet hjernen seg av det faktum at objekter som er nærmere øynene er mer forflyttet i hvert øye enn objekter som er langt fra øynene. Det vil si at hvis du lukker det ene øyet og ser på et nærmere objekt for deretter å lukke det andre øyet og ser på det samme objektet, vil du se at det har oppstått en forskyvning av objektet. Hvis du derimot ser på et objekt lengre unna vil denne forskyvningen være mindre eller tilnærmet null. Et annet klassisk eksempel er hvis du beveger deg i relativ høy hastighet i bil vil objekter som lyktestolper nær veien bevege seg raskt forbi deg mens fjell for eksempel som er lengre unna vil bevege seg sakte. Dette konseptet kalles parallakse og kan beskrives som en tilsynelatende forflytting av objektet på grunn forflytting av den som observerer objektet. Parallakse kan uttrykkes kvantitativt som  $p = x_{\text{venstre}} - x_{\text{høyre}}$ , der  $x_{\text{venstre}}$  og  $x_{\text{høyre}}$  er koordinatene langs den positive x-akse til et bestemt objekt henholdsvis før og etter observanden har forflyttet seg en viss distanse. Denne distansen omtales som basislinje, og ved stereoskopi i menneskesyn tilsvarer denne basislinjen distansen mellom øynene.

Figur 2.4 illustrerer to bilder tatt ved punktene  $L_1$  og  $L_2$  av det samme punktet, punkt  $A$ . Kameraet blir altså forflyttet langs basislinjen  $B$ . Punkt  $A$  blir avbildet i punktene  $a_v$  og  $a_h$  i bildeplanet. Punkt  $A$  kan for eksempel være et elevert punkt i et ujevnt terreng som det tas bilde av, og avstanden, eller høyden, mellom kameraet og det aktuelle punktet er gitt ved  $h$ .  $O_1$  og  $O_2$  representerer prinsippunktene i bildeplanet for hvert bilde. Distansen i bildet mellom de respektive prinsippunktene og  $a_v$  og  $a_h$  er henholdsvis gitt av  $x_v$  og  $x_h$ . Disse representerer x-koordinatene i bildet siden prinsippunktene er origo i

koordinatsystemet i bildet. Fokallengden til kameraet er vist ved  $f$  i figuren. Parallaxen er, som før, gitt ved  $p = x_v - x_h$ . Merk at  $x_h$  er til venstre for origo, og dermed negativ, noe som gjør at lengden av parallaxen blir summen av distansene  $x_v$  og  $x_h$ .



Figur 2.4: Illustrasjon av parallaxe ved punktet A

Hvis det legges til grunn perfekte geometriske former, slik som i figur 2.4, kan parallaxen ved punkt A gi informasjon om avstanden mellom punktet og kameraet. Det vil si at man kan få informasjon om for eksempel hvor elevert punktet er i forhold til resten av terrenget. Ved å anerkjenne at trekantene  $(L_1, L_2, A)$  og  $(L_1, a_v, a'_h)$  er formlike har vi at,

$$\frac{B}{h} = \frac{p}{f} \rightarrow h = \frac{Bf}{p} \quad (2.9)$$

Ligning 2.9 bekrefter det som allerede er nevnt over, at avstanden til et punkt er invers proporsjonal med parallaxen. Høy parallaxe medfører at punktet er nærme, og lav parallaxe medfører at distansen til punktet er lengre unna.

### 2.3.2 Synsfelt i stereoskopi

Som det fremkommer av figur 2.4 må punkt A befinne seg på begge bildene tatt ved  $L_1$  og  $L_2$ , med andre ord må punktet være innenfor synsfeltet til kameraet på

begge steder. Dette medfører begrensinger på hvor lang basislinjen kan være og hvor nærme kameraet kan være til punktet. Fordelen med å øke basislinjen er å øke nøyaktigheten i parallaksen, spesielt når punktet er lengre, eller dypere, unna. Dette gir en økning i oppløsning i høyde- eller dybdeberegningene som gjøres på bakgrunn av parallaksen. Det kan også vises at skaleringsfaktoren, som er gitt av fokallengden og avstanden fra kameraet til objektet, påvirker oppløsningen til dybdeberegningene. Parallaksen kan skrives som summen av det parallakseskiftet i x-koordinatene som uansett oppstår grunnet forflytning av kameraet og det parallakseskiftet som er direkte relatert til dybdeforskjellene.

Den førnevnte parameteren,  $p_0$ , er konstant og er uavhengig av

dybdeforskjellene, mens den sistnevnte parameteren,  $\Delta p$ , tilsvarende parallaxsevariasjonene som er knyttet til dybdeforskjellene og er dermed en variabel. Ved å bruke dette sammen med ligning 2.9 får vi følgende:

$$p = p_0 + \Delta p = \frac{Bf}{h} + \Delta p = B \cdot \text{skalering} + \Delta p \quad (2.10)$$

der  $\Delta p \ll p_0$ .

Økning av skaleringen og basislinjen fører altså til større differanser i  $p$  og dermed større oppløsning i dybdeberegningene, siden dybdeforskjellene er invers proporsjonale med differansene i parallaksen. Hvis kameraet har fast fokallengde er det kun mulig å øke skaleringen ved å redusere avstanden mellom kameraet og objektet. Ulempen med å øke basislinjen, skaleringen eller begge for mye er at det aktuelle området med punkter der parallaksen skal måles kan havne utenfor synsfeltet når kameraet forflyttes til den andre enden av basislinjen. Synsfeltet til linsen setter dermed begrensinger for oppløsningen til dybdeberegninger i objektet.

## 2.4 Stereodisparitet

Som beskrevet tidligere kan dybden, eller høyden, i et område beregnes ved å analysere to bilder tatt ved hver sin ende av basislinjen. Disse to separate bildene må delvis sammenfalle og sammenfalle helt over det aktuelle området der dybden skal evalueres. Bildene kalles da stereobilder. For å kunne beregne

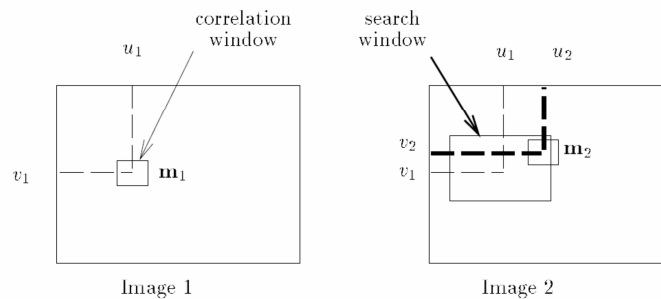
parallaksen i hvert enkelt punkt må punktene i det sammenfallende området i det ene bildet lokalisere i det andre bildet. Dette omtales som stereokorrespondanseproblemet. Som med andre bildeanalyseteknikker så er implementasjonen av en nøyaktig, effektiv matchingsalgoritme vesentlig vanskeligere enn den stereomodellen mennesker relativt lett klarer å danne. Bildematching er i seg selv en vanskelig utfordring. I et bilde med svært mange piksler, noe som alle moderne bilder innehar, vil det som regel finnes to eller flere områder med piksler som har en stor grad av ambiguitet. Dette fører til falske matchinger som da gir feil parallakse mellom stereobildene. Støy i bildene kan i tillegg føre til fullstendig utelukkning av det korrekte punktet som skulle matches. Strukturene i bildene kan også forandre seg betraktelig når synsvinkelen endres. En linje på et vist antall piksler i det første bildet kan muligens bare ha en lengde på halvparten av den opprinnelige lengden i det andre bildet. Bildematchingsalgoritmen må ta høyde alle disse scenarioene på en effektiv måte.

### 2.4.1 Bildematchingsalgoritmer

Det finnes to hovedmetoder som matchingsalgoritmene kan grupperes inn under; områdematching og karakteristisk matching (1). Den sistnevnte type matching beror på karakteristiske trekk i bildet. Dette kan være linjesegmenter, spesielle interessepunkter eller lukkede konturer som matches og er best benyttet i topografiske bilder der klare strukturer forekommer i bildet. Denne matchingen anses ikke relevant for dette prosjektet siden strukturene av de aktuelle objektene er av høyst varierende grad, og det blir dermed vanskelig å gi en generell karakteristikk av strukturene ved implementering av kode. Fokuset vil derfor være utelukkende på områdematching videre i dette prosjektet.

Områdematching bruker simpelthen pikselverdiene innen et begrenset område fra det ene bildet for å matche disse pikselverdiene til pikselverdiene i det andre bildet. Det ene bildet brukes da som referansebildet slik at det søkes i det andre bildet etter det tilsvarende punktet gitt i referansebildet. Dette gjøres ved å ta et utsnitt av referansebildet rundt det aktuelle punktet som skal matches. Dette utsnittet kalles referansevinduet og består som regel av  $N \times N$  antall piksler. Kun matching av én og én piksel istedenfor blokker av piksler ville gitt lave prosesseringstid, men dette hadde ført til mange falske matchinger siden de samme pikselverdiene opptrer på mange forskjellige steder i bildet. Fremgangsmetoden er illustrert i figur 2.5, der referansevinduet er vist i bilde 1 og søkevinduet som referansevinduet blir matchet opp mot vist i bilde 2. Det

utsnitt av søkevinduet som passer best med referansevinduet blir antatt til å være det aktuelle matchede punktet i referansevinduet. Et stort referansevindu gjør matchingen mer pålitelig og mer resistent for påvirkning av støy i bildet, men ulempen er at matchingen nå ikke fanger opp brå endringer i parallakse. Dette skyldes at dybdeforskjellen medfører endringer som nå blir midlet ut fordi vinduet egentlig er ved to forskjellige parallakser. Referansevinduet med forholdsvis stor størrelse fungerer dermed som et lavpassfilter anvendt på parallakseverdiene. I tillegg vil et stort vindu medføre høyere prosesseringstid enn et lite vindu. Bildematching går ut ifra at det er matematiske likheter mellom pikslene i referansevinduet og det bildet som skal matches. Det medfører at pikselverdiene i bildene bør normaliseres før de matches.



Figur 2.5: Bilde 1 viser utsnitt av referansebildet. Dette utsnittet kalles referansevinduet og brukes til å søke gjennom bilde 2 for matching. (9)

Bildeområdet referansevinduet skal søke i angir størrelsen på søkevinduet. Er den maksimale parallaksen kjent kan søkevinduet begrenses, og matchingen vil dermed ha lavere prosesseringstid. Den totale parallaksen er gitt av ligning 2.10, men dette krever blant annet kjennskap til den maksimale dybdevariasjonen i avbildningsobjektet. Er den maksimale parallaksen ikke kjent må det søkes på tvers av hele bildet. Antall rader det må søkes over bestemmes av i hvilken grad stereobildene er rektifisert. Ved fullstendig rektifikasjon av stereobildene kan søket begrenses til en rad i bildet. Dette vil utdypes nærmere i delkapittel 2.5. Kvaliteten på punktmatchingen avgjøres av bildenes karakteristikk. Sannsynligheten for å gjøre feil matching er invers proporsjonal med signal-støy forholdet i bildet. Områder i stereobildene med mye høyfrekvent innhold øker presisjonen til matchingen siden detaljnivået i disse områdene er høyere. Områder med lavfrekvent innhold, som for eksempel bakgrunnen til avbildningsobjekter eller veldig homogene områder, kan by på utfordringer i matchingen siden punktene i disse områdene kan lett forveksles av matchingsalgoritmen. Kvaliteten på matchingen degraderes også ved

linsedistorsjon i kameraet. Linsedistorsjon kan, som nevnt i delkapittel 2.2, korrigeres for ved geometrisk kalibrering. Det kan derfor konkluderes med at bildematchingssystemer har betydelige begrensninger. Mange piksler vil ikke kunne matches grunnet lavt frekvensinnhold eller store, bråe endringer i parallaksen. Å prøve matching av slike områder kan degradere modellen og kaste bort prosessorkraft. Det er derfor bedre og prøve å detektere disse feilmatchingene og heller tilpasse disse verdiene til naboområdet.

### 2.4.2 Metriske metoder for matching

Det finnes en del ulike metoder for å måle de matematiske likhetene i bildene på, der hver metode varierer i prosesseringstid og nøyaktighet. Disse metodene bruker forskjellige metriske mål når forskjellen mellom referansevinduet og vinduet som skal matches skal beregnes. Nedenfor er det tatt med et utvalg av disse metodene som anses som relevante for matchingen i dette prosjektet (10).

#### Sum av den absolutte differanse (SAD)

Denne metoden summerer absoluttverdien av differansen mellom pikselverdiene i referansevinduet og pikselverdiene i det vinduet som skal matches. Matematisk tilsvarer dette:

$$d_1(I_j, T) = \sum_{i=1}^n |I_{i,j} - T_i| \quad (2.11)$$

der  $I$  er  $j$ -te bildet eller vinduet som skal matches og  $T$  er templatene eller referansevinduet som det også kalles. SAD beregnes over  $n$  antall piksler i vinduene. Den laveste SAD-verdien av gir den beste estimeringen av lokasjonen til det vinduet som matcher best med referansevinduet ved bruk av denne metoden.

#### Sum av den kvadrerte differanse (SKD)

Forskjellen mellom SKD og SAD er i all hovedsak at absoluttverdien av differansen til pikselverdien blir kvadrert. SKD-matching er dermed gitt av:

$$d_2(I_j, T) = \sum_{i=1}^n |I_{i,j} - T_i|^2 \quad (2.12)$$



Maksimum absolutt differanse (MaxAD)

Her blir maksimum av absoluttverdien til differansen mellom pikselverdiene til referansevinduet og vinduet som matches summert. Dette gir dermed:

$$d_{\infty}(I_j, T) = \lim_{p \rightarrow \infty} \sum_{i=1}^n |I_{i,j} - T_i|^p = \max_i |I_{i,j} - T_i|^p \quad (2.13)$$

Normalisert krysskorrelasjon

Normalisert korrelasjon mellom vinduene i de to stereobildene tar hensyn til forskjeller i pikselverdi og variansen mellom bildene. Korrelasjonen mellom vinduene blir normalisert ved å trekke fra gjennomsnittet for så å dividere på standardavviket av vinduene. Denne beskrivelsen er gjengitt matematisk i ligning 2.14 i todimensjonal form. Normalisering av korrelasjonen er nødvendig for å unngå mismatch mellom vinduene grunnet regioner der pikslene har høyere gjennomsnittlig verdi, eller regioner med høyere varians som vil gi høyere korrelasjon enn den korrekte matchede regionen med mer normale pikselverdier (1). Disse avvikene kan skyldes at stereobildene har for eksempel blitt tatt under forskjellige lysforhold, eller skyldes at i enkelte regioner oppstår det mettede piksler som igjen skyldes refleksjoner eller lignende.

$$\gamma(u, v) = \frac{E[I \cdot T] - E[I]E[T]}{\sigma(I)\sigma(T)} = \frac{\sum_{x,y} [I(x, y) - \bar{I}_{u,v}][T(x-u, y-v) - \bar{T}]}{\sqrt{\sum_{x,y} [I(x, y) - \bar{I}_{u,v}]^2 \sum_{x,y} [T(x-u, y-v) - \bar{T}]^2}} \quad (2.14)$$

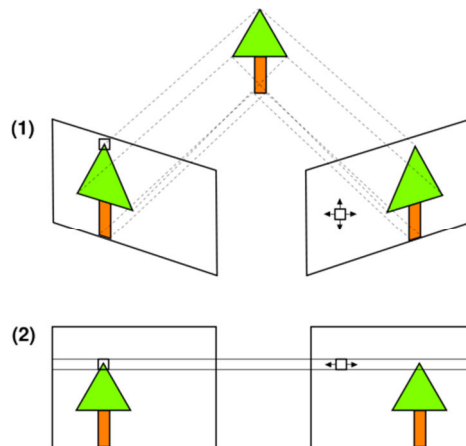
der  $\bar{T}$  er gjennomsnittsverdien til  $T$  og  $\bar{I}_{u,v}$  er gjennomsnittet av  $I(x, y)$  i den regionen templatene befinner seg i.

Som det fremkommer av ligning 2.14 er denne metriske metoden for matching vesentlig mer komplisert enn de foregående. Ved implementering har derfor denne metoden høyere prosesseringstid enn de andre. Tatt i betraktning at denne matchingen skal gjøres for opp mot millioner av piksler er prosesseringstiden for en enkelt match av stor betydning.

## 2.5 Bilderektifikasjon

Vanligvis vil ikke stereobilder være fullstendig på linje vertikalt eller være ved samme høyde ved translasjon av kameraet mellom endene av basislinjen. Dette er det praktiske tilfellet selv om translasjonen av kameraet, teoretisk sett, bare er i horisontal retning i bildet. Hvis forskyvingen av kameraet skjer ved å bruke et translasjonstrinn kan dette skyldes ufullkommen presisjon i de mekaniske bevegelsene til translasjonstrinnene. En forskyvningsfeil i translasjonen på bare noen titalls mikrometer kan resultere i pikselforskyvning i bildet. Hvor mye metrisk forskyvning som resulterer i pikselforskyvning avhenger av den metriske pikselstørrelsen til kameraet og skaleringsfaktoren. Konsekvensene av pikselforskyvning i vertikal retning i stereobildene er at søkevinduet i bildematchingen også må bevege seg i vertikale retning i bildet.

Høydeforskyvninger ved mekanisk translasjon kan føre til feil matching i stereobildene ved enkelte punkter, men dette medfører som regel ikke problemer og er underordnet problemet med vertikal forskyvning i stereobildene. Som nevnt i det foregående delkapittel kan søkeområdet ved bildematching reduseres betraktelig ved fullstendig rektifikasjon av stereobildene. Rektifikasjon er en type transformasjon som brukes til å projisere stereobilder til et fellesbildeplan. Dette medfører at eventuelle høydeforskyvninger og forskyvninger i-y-retning i bildet blir borte. Figur 2.6 illustrerer denne bilderektifikasjonen.



Figur 2.6: Stereobilder før (1) og etter (2) rektifikasjon. Etter rektifikasjon er søkeområdet kun i horisontal retning, det vil si kun over en rad i bildet (11).

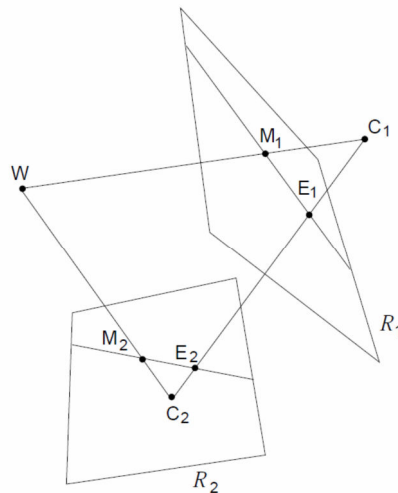
### 2.5.1 Epipolare linjer og plan

Før ble stereobilder rektifisert ved å bruke optiske teknikker, men i det siste tiåret har rektifikasjon primært blitt gjort ved såkalt planar rektifikasjon (12). Ved bruk av planar rektifikasjon blir vanligvis geometrien av den optiske projeksjonen modellert ved å bruke en enkel lineær transformering av hvert bilde. Dette medfører i realiteten at begge bildeplanene blir rotert til de sammenfaller langs den ene bildeaksen. Fordelen med denne metoden er at den er simpel å implementere, har rask prosesseringstid og bevarer bildekarakteristikken (12).

Et punkt i rommet,  $W$ , blir projisert til to forskjellige bildepunkter,  $M_1$  og  $M_2$ , som vist i figur 2.7.  $M_1$  og  $M_2$  er skjæringspunkter i sine respektive bildeplan,  $R_1$  og  $R_2$ , mellom kameraenes perspektivsenterer  $C_1$ ,  $C_2$  og  $W$ . Som nevnt i delkapittel 2.1 kalles linjen som inneholder  $C$ , og er ortogonal til planet  $R$ , den optiske akse. Prinsipalpunktet (PPA) er skjæringspunktet mellom den optiske akse og planet  $R$ . Distansen mellom  $C$  og  $R$  tilsvarer fokallengden.  $M_1$  og  $M_2$  konstituerer et konjugert punktpar. I det ene bildepunktet  $M_1$  ligger det konjugerte punktet,  $M_2$ , i det andre bildet på en linje som kalles for den epipolare<sup>1</sup> linje (13). Det finnes uendelig antall epipolare linjer som korresponderer til vilkårlige punkter i bildeplanet (1). Epipolarlinjen er projeksjonen gjennom  $C_2$  av den optiske strålen fra  $W$  til  $M_1$ . Alle de epipolare linjene i et bestemt bildeplan går gjennom et felles punkt i begge bildeplanene. Dette punktet kalles en epipol<sup>1</sup> og er i figur 2.7 vist ved betegnelsen  $E_1$  og  $E_2$ . Disse punktene tilsvarer projeksjonen av perspektivsenteret til det andre kameraet.

---

<sup>1</sup> Epipolar og epipol er ikke godkjente norske ord, men det finnes heller ikke noen gode norske ord for de engelske ordene *epipolar* eller *epipole* ved anvendelser i denne konteksten.

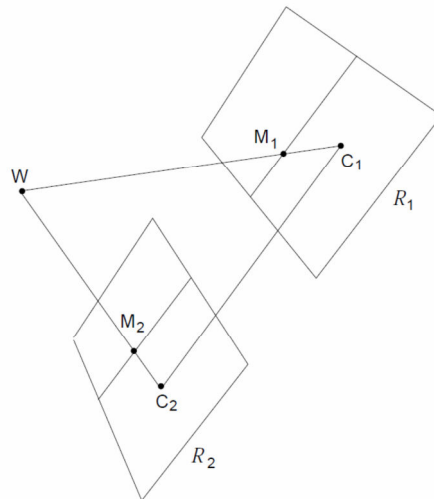


Figur 2.7: Illustrasjon av et punkt i rommet som blir projisert på to bildeplan, og den korresponderende epipolare geometrien er også illustrert. (13)

Når  $C_1$  er i det fokale planet til kameraet til høyre i figuren over er den tilsvarende epipolen ved uendelig og de epipolare linjene er en bunt av parallelle linjer i bildet til høyre. Basislinjen tilsvarer linjen mellom  $C_1$  og  $C_2$ . Når denne linjen omfatter begge fokalplanene er begge epipolene ved uendelig, det vil si at begge bildeplan er parallelle til basislinjen. De epipolare linjene danner da en bunt av parallelle linjer i begge bilder. Denne prosessen kalles rektifikasjon og er altså en transformering av stereobildene slik at de epipolare linjene er parallelle og horisontale (13).

### 2.5.2 Rektifikasjon av bildematrixene

Formålet til rektifikasjon er å rotere bildematrixene rundt sine perspektivsentre slik at fokalplanene befinner seg i samme plan. Dette gjør, som nevnt tidligere, at fokalplanene inneholder basislinjen og de epipolare linjene er parallelle. For å oppnå horisontale, epipolare linjer må basislinjen være parallell med den nye x-aksen etter rotering. Figur 2.8 vise to rektifiserte bildeplan der alle de konjugerte bildepunktene har de samme vertikale koordinatene. Kravet for dette er at begge kameraene har de samme intrinsiske parameterne etter rotasjon. Hvis stereobildene tas av det samme kameraet vil de intrinsiske parameterne, gitt av ligning 2.6, være de samme for begge bilder, og bildeplanene vil i tillegg ligge i samme plan.



Figur 2.8: Illustrasjon av bildeplanene etter rektifisering

For å representere den epipolare geometrien brukes det noe som refereres til som fundamental matrisen (3). Dette er en  $3 \times 3$  matrise som relaterer de korresponderende punktene i stereobildene. For et punkt med homogent bildekoordinat  $x$  i et av stereobildene eksisterer det et korresponderende punkt med homogent bildekoordinat  $x'$  i det andre bildet. Disse to punktene ligger på den epipolare linjen  $l'$ . Det skjer dermed en transformering slik at  $x \rightarrow l'$ , og denne transformering er en perspektivtransformasjon fra punkter til linjer. Fundamentalmatrisen  $\mathbf{F}$  er definert slik at den tilfredsstiller  $x^T \mathbf{F} x = 0$  og  $l' = \mathbf{F} x$  (3).

## 2.6 Modelling

Etter at parallaksen er blitt beregnet i de aktuelle punktene i avbildningsobjektet kan 3D-modellen startes å beregnes. Siden modellen skal representere avbildningsobjektet i objektrommet vil man måtte beregne X-, Y- og Z-koordinatene på bakgrunn av parallaksen og koordinatene i bildeplanet slik som forklart i avsnitt 2.1.3.

### 2.6.1 Koordinater i objektrommet

Etter bilderektifikasjon er begge bildeplanene, på hver sin ende av basislinjen, koplanare. Dette medfører at x-koordinatene i begge bildeplan er på samme linje. I tillegg er PPA kjent etter kalibreringsprosessen noe som muliggjør introduisering av et skift i koordinatene i bildeplanet slik at PPA utgjør origo i bildet. Koordinatene i bildeplanet sammenfaller dermed med koordinatene i objektplanet. Det vil si at x-koordinatene i hvert av de to bildeplanene til stereobildene er sammenfallende, og x-koordinatene i objektrommet er kun skiftet med lengden på basislinjen. Sammen med ligningene 2.2 og 2.3 gir dette følgende:

$$\begin{aligned}
 X_2 &= X_1 + B \Rightarrow \\
 \frac{x_2}{f}(f - Z) &= \frac{x_1}{f}(f - Z) + B \Rightarrow \\
 Z &= f - \frac{Bf}{x_2 - x_1}
 \end{aligned} \tag{2.15}$$

der  $x_2 - x_1$  tilsvarer parallaksen som definert i avsnitt 2.3.1. Z-verdiene i ligning 2.15 tilsvarer dermed høydeberegningen i ligning 2.9, men er negativ og fokallengden er trukket fra høyden i forhold til ligning 2.9. Ligningene 2.2, 2.3 og 2.15 gir dermed alle koordinatene i objektrommet basert på parallaksen, fokallengden og lengden på basislinjen.

### 2.6.2 Høydeoppløsning

Siden Z-verdiene er proporsjonale med parallakseverdiene vil den minste mulige endringen av Z-verdiene forekomme når  $\Delta p$  i ligning 2.10 også endres med den minste mulige verdien. Den minste mulige verdien av  $\Delta p$  tilsvarer et skift av en piksel mellom x-koordinatene til de to stereobildene. Oppløsningen i Z-aksen blir derfor:

$$\Delta Z = Z_1 - Z_2 = f - \frac{fB}{p_1} - \left( f - \frac{fB}{p_2} \right) = fB \left( \frac{1}{p_1} - \frac{1}{p_2} \right) \tag{2.16}$$

der  $p_1 = B \cdot \text{skalering}$  og  $p_2$  er gitt av ligning 2.10. Ligning 2.16 gir altså den endringen i Z-verdien som oppstår når parallaksen endrer seg med en piksel.

For å kunne relatere pikslene i bildene til metriske verdier må pikselstørrelsen kjennes. Dette er en faktor som er gitt av CCD-enheten i kameraet, og denne

faktoren oppgis i meter per piksel(MPP). For å kunne beregne hvor mange meter en piksel utgjør må man også ta i betraktning skaleringsfaktoren siden MPP-størrelsen kun gir de metriske målene i forhold til bildeplanet og ikke objektplanet. Dermed blir  $\Delta p = 1 \text{ [pix]} \cdot MPP \text{ [m/pix]} / skalering \text{ [ ]}$  som da tilsvarer det minste mulige skiftet i parallaksen i meter. Ligning 2.16 kan da skrives om til:

$$\begin{aligned}
 \Delta Z &= fB \left( \frac{1}{B \cdot skalering} - \frac{1}{B \cdot skalering + \Delta p} \right) \\
 &= fB \left( \frac{1}{B \cdot skalering} - \frac{1}{B \cdot skalering + MPP / skalering} \right) \\
 &= \frac{f \cdot MPP}{skalering(B \cdot skalering^2 + MPP)} = \frac{h \cdot MPP}{B \cdot skalering^2 + MPP} \\
 &\approx \frac{h^3 \cdot MPP}{B \cdot f^2} \tag{2.17}
 \end{aligned}$$

der det antas at  $MPP \ll B \cdot skalering^2$ . For å få overflatemodellen så nøyaktig som mulig må  $\Delta Z$  være så lav som mulig. Som nevnt tidligere i avsnitt 2.3.2 kan dette gjøres ved øke skaleringen og lengden på basislinjen, men utfra ligning 2.17 bør også kameraets MPP-størrelsen være så lav som mulig. I tillegg viser ligningen at lavere høyde er viktigere enn høyere fokallengde for å oppnå høyere høydeoppløsning i modellen.

# 3 Metode

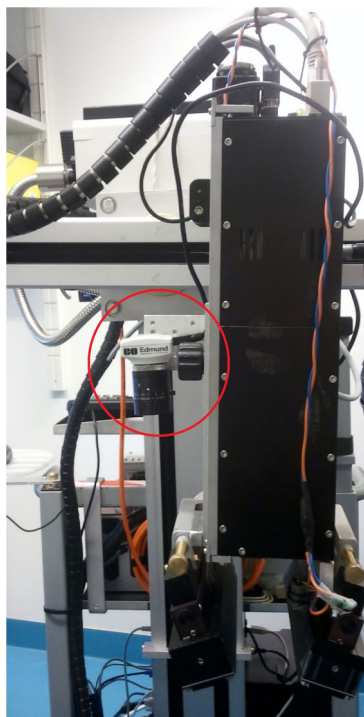
---

Her gis det en beskrivelse av metodebruken i prosjektet. Først blir utstyret og det oppsettet som har blitt benyttet ved akvisisjon av stereobildene beskrevet. Det gis en oversikt over hvordan stereobildene blir prosessert etter akvisisjonen. Deretter gjennomgås metodebruken ved geometrisk kalibrering og bilderektifikasjon. Her vil også spesifikk teori benyttet for de konkrete metodene forklart. Valg av matchingsmetode blir begrunnet og i tillegg blir fremgangsmetoden for matching beskrevet. Tilslutt forklares det hvordan modellen dannes.



## 3.1 Eksperimentelt oppsett og programmeringsspråk

Ved avbildning av objektet som skal overflatemodelleres har det blitt brukt et monokromatisk USB-kamera samt en separat linse med fast fokallengde montert på kameraet. Detaljert informasjon om kameraet og linsen er gitt i avsnitt 3.1.1. Kameraet ble indirekte festet til et translasjonstrinn, som vist i figur 3.1. Som figuren viser er kameraet egentlig festet til et hyperspektralt kamera som igjen er festet til translasjonstrinnet. Det hyperspektrale kameraet, som omtalt i introduksjon, skal sammen med overflatemodellen i tillegg gi spektral informasjon om avbildningsobjektet. Translasjonstrinnet ble brukt til å forflytte kameraet i x-retning ved akvisisjon av stereobildene. Lengden på denne forflytningen tilsvarer som kjent basislinjen. I tillegg er det mulig å forflytte translasjonstrinnet i Z-retning slik at høyden til kameraet kan justeres.



*Figur 3.1: Eksperimentelt oppsett med kamera, linse og translasjonstrinn. Den røde sirkelen viser kameraet og linsen.*

### 3.1.1 Utstyr

Kameraets og linsens tekniske spesifikasjoner er oppgitt i tabellene under. Det oppgis kun de dataene som anses som relevante for dette prosjektet.

<b>Kamera</b>	
Sensortype	EV76C560A
Sensorstørrelse (H × V) [mm]	6.79 × 5.43
Antall piksler	1280 × 1024
Pikselstørrelse (MPP) [μm]	5.3 × 5.3
Pikseldybde	10 bit
Dimensjoner [mm]	44 × 44 × 25.4

Tabell 3.1: Tekniske spesifikasjoner for monokromatisk USB-kamera (14)

<b>Linse</b>	
Fokallengde [mm]	35.00
Synsfelt, 2/3" sensor	39.3mm - 14.3°
Synsfelt, 1/2" sensor	28.5mm - 10.4°
Synsfelt, 1/3" sensor	21.4mm - 7.8°
Apertur	F1.65 - F22
Lengde [mm]	36.1

Tabell 3.2: Tekniske spesifikasjoner for linse montert på kamera (15)

### 3.1.2 Akvisisjon av stereobilder

Ved akvisisjon av stereobildene av avbildningsobjektet ble USB-kameraet koblet til en pc, og bildene ble tatt ved bruk av egen programvare. Dermed kan bildene tas uten berøring av kameraet. Dette er en stor fordel siden kontakt med kameraet trolig ville ført til en mekanisk forflyttelse av kameraet. En forflyttelse av kameraet på bare noen titalls mikrometer ved en rimelig høyde ville ført til pikselskift i bildet.

Siden dette prosjektet er kun startfasen av et større prosjekt, og prosjektoppgaven baserer seg på dannelse av en prototypemodell vil ikke faktiske hudsår bli hovedfokuset for modellering i dette prosjektet. Grunnen til dette er at det ikke forelå adekvate bilder av sår som kunne modelleres godt med oppsettet i dette prosjektet. Sårene var for grunne til å kunne bli modellert ordentlig med den maksimale høydeoppløsning som er mulig å oppnå med dette oppsettet. Som avbildningsobjekt har det i stedet blitt en spesiell type modelleire. Fordelen med modelleire er muligheten å kunne forme den geometrisk etter ønske. Modelleiren brukt i dette prosjektet er estimert til å være 2,5 × 2,5 cm i bredden og lengden, og høyden varierer over hele objekt.

Oppløsningen langs Z-aksen er blant annet gitt av høyden mellom kameraet og avbildningsobjektet og basislinjen (jf. avsnitt 2.6.2). De er de eneste parameterne som lar seg justere med dette oppsettet. Det ble tatt totalt 12 bilder av modelleiren. For hvert nytt bilde som ble tatt ble kameraet forflyttet 1 cm. Det vil si at basislinjen mellom det første bildet og de andre bildene varierer mellom 1 til 11 cm. Høyden mellom kameraet og modelleiren ved akvisisjon av stereobildene var 30 cm. Dette var i praksis den minste høyden som kunne oppnås med en maksimal basislinje på 11 cm uten at avbildningsobjektet forsvant ut av synsfeltet til linsen.

### 3.1.3 Fremgangsmåte

Etter det første bildet ble tatt av modelleiren ble kameraet forflyttet langs basislinjen i kun x-retning. Etter geometrisk kalibrering av bildene ble gjort ble bildene visuelt inspisert for eventuelle skift i y-retning i de to stereobildene. Det er særdeles viktig at skiftet er null i y-retning ved bildematching som forklart i delkapittel 2.5. Det ble funnet at stereobilder med en basislinje på 1 cm hadde en forskyvning i y-retning på en piksel. I tillegg viste det seg ved flere undersøkelser at denne forskyvning i y-retning økte proporsjonalt med lengden på basislinjen. En mekanisk forflytning i translasjonstrinnet på bare noen titalls mikrometer i y-retning ved rimelig høyde vil resultere i pikselskift i y-retning. Dette nødvendiggjør derfor bilderektifikasjon av stereobildene før pikslene blir matchet.

Fremgangsmåten i bildeprosesseringen av stereobildene blir da som følger:

1. Geometrisk kalibrering
2. Bilderektifikasjon
3. Bildematching
4. Modellering

Geometrisk kalibrering må først gjøres for å korrigere for feilrepresentasjon av avbildningsobjektet i bildene. Deretter gjøres rektifikasjon av stereobildene slik at det oppnås epipolar geometri, det vil si null skift i y-retning mellom stereobildene. Dette muliggjør at matching av piksler kan gjøres ved kun å søke over en rad i bildet, og dette er dermed neste steg. Når pikslene i det ene bildet er matchet og lokalisert kan parallaksen beregnes. Disse parallakseverdiene brukes så for å beregne Z-verdiene, som igjen blant annet brukes til å beregne X- og Y-koordinatene. Den videre strukturen av dette kapittelet samt resultatene i neste kapittel følger denne fremgangsmåten kronologisk.

### 3.1.4 Programmeringsspråk

Valg av programmeringsspråk avhenger av flere faktorer. De viktigste faktorene er som regel prosesseringstid og implementasjonstid av algoritmene. Som det følger av oppgaveteksten er prosesseringstid en underordnet faktor i denne fasen av prosjektet. I tillegg må en del forskjellige algoritmer testes og undersøkes. Rask implementasjon gjør det mulig å teste forskjellige algoritmer på et tidlig stadium og vurdere graden av funksjonalitet og kvalitet. Hvor rask en algoritme kan implementeres avhenger av hvor stor grad basisfunksjonalitet er implementert i programmeringsspråket. Denne oppgaven baserer seg på en god del bildebehandling. Matlab har egne integrerte moduler for både generell bildebehandling og stereoskopisk behandling av bilder. I tillegg er det raskt å utføre matriseoperasjoner i Matlab. Ulempen med Matlab er lang prosesseringstid, men som nevnt er ikke dette av høy betydning i denne fasen av prosjektet. All kode i dette prosjektet har derfor blitt implementert i Matlab.

## 3.2 Kalibrering

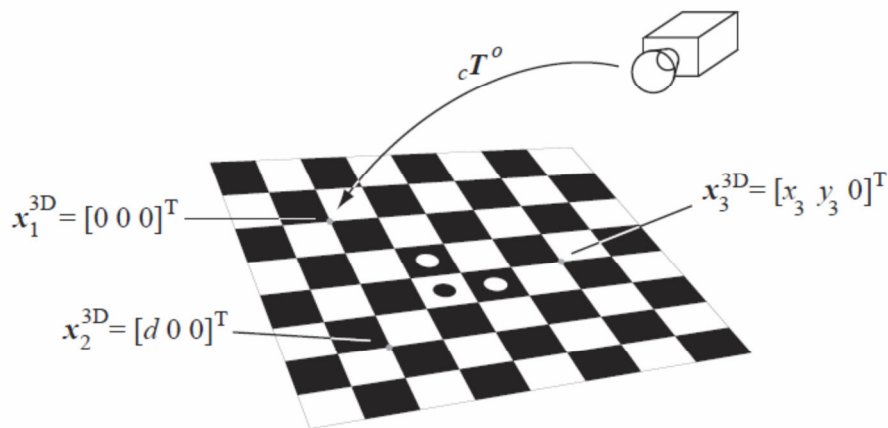
I dette prosjektet har det blitt innhentet programvare fra det tyske romfartssenteret(DLR), institutt for robotikk og mekatronikk, for kalibrering av kameraet (16). Denne programvaren består av to separate og uavhengige program; camera Calibration Detection tool (CalDe) og camera Calibration Lab (CalLab). DLR CalLab og CalDe ble begynt utviklet i 2005 og var et resultat fra kravet om å ha en uavhengig applikasjon for kamerakalibrering. Programvaren ble utviklet i IDL for å oppnå plattformuavhengighet, men resulterte også i redusert utviklingstid og høyere ytelse (16).

### 3.2.1 CalDe

Hensikten med CalDe er å detektere og måle mønsterpunktene i avbildningsobjektet som brukes til å kalibrere kamera. Det ble i begynnelsen av prosjektet vurdert å bruke et diffraksjonsmønster som avbildningsobjekt for kalibrering siden punktene i mønsteret er presist uniformt fordelt, og det er enkelt å beregne hvor de forskjellige punktene i diffraksjonsmønster skal befinne seg. Eksperimenter ble utført ved å bruke en laser som lyskilde og sende

laserlyset gjennom et diffraksjonsgitter, men denne ideen ble senere forkastet noe senere grunnet pragmatiske problemer og usikkerheten knyttet rundt deteksjonen av punktene i diffraksjonsmønsteret av kalibreringsprogramvaren. Et sjakkbrett er standard avbildningsobjekt i kalibrering og garanterer nøyaktighet og repeterbarhet i mønsteret. Dette ble derfor brukt i stedet i dette prosjektet, selv om det er mulig å bruke andre avbildningsobjekter med et gitt mønster med denne programvaren.

Ved bruk av et sjakkbrett som avbildningsobjekt vil CalDe detektere hjørnene mellom de svarte og hvite kvadratene. Et sjakkbrettmønster er pragmatisk siden det oppnås subpixel lokaliseringspresisjon i hjørnene (17). Størrelsen på mønsteret bør velges i henhold til linseaperturen og distansen mellom linsen og mønsteret, slik at avbildning blir så skarp som mulig. For å oppnå en mest mulig korrekt kalibrering bør mønsteret dekke hele bildet. Sjakkbrettmønsteret i dette prosjektet ble skrevet ut på et A4-ark, og hver kvadrant hadde en størrelse på  $5 \times 5$  mm. I sjakkbrettmønsteret er det også to hvite sirkler og en sort sirkel innenfor enkelte av kvadrantene. Disse gir CalDe informasjon om x- og y-retning samt lokasjonen til origo for sjakkbrettet som vist i figur 3.2. Prosessen med mønstergjennkjennning i CalDe er fullstendig automatisk, men krever at brukeren definerer størrelsen på kvadrantene og at startpunktet for mønsteret blir definert i programvaren.



Figur 3.2: Illustrasjon av sjakkbrettmønsteret med tre punkter og deres koordinater (17)

### 3.2.2 Callab

De punktene CalDe har detektert blir deretter overført og prosessert av Callab. Callab har her som hensikt å beregne de intrinsiske parameterne gitt i ligning

2.6. CalLab krever detekterte punkter fra minst tre separate bilder tatt fra ulike vinkler for å beregne korrekt. Brukeren har mulighet til å fjerne de punktene som er detektert feilaktig, eller som har høy pikselfeil, før de intrinsiske parameterne beregnes. Det er svært avgjørende for nøyaktigheten av kalibreringen at punktene er presist og robust detektert. Det kan argumenteres for at kalibreringsresultatet med denne programvaren ikke avhenger av nøyaktigheten av mønsteret, men bare avhenger av nøyaktigheten av mønsterdeteksjonen (17).

Den første estimeringen av de forskjellige parameterne i den intrinsiske matrisen er basert på homografi. Deretter blir denne estimeringen numerisk optimalisert i CalLab. Hvis et todimensjonalt punkt betegnes som  $\mathbf{m}=[x,y]^T$  og et tredimensjonalt punkt betegnes som  $\mathbf{M}=[X,Y,Z]^T$  kan disse vektorene utvides til homogene koordinater slik at  $\tilde{\mathbf{m}}=[x,y,1]^T$  og  $\tilde{\mathbf{M}}=[X,Y,Z,1]^T$ . Forholdet mellom et tredimensjonalt punkt i objektrommet og et todimensjonalt punkt i bilderommet er da gitt ved å gjøre perspektivprojeksjon:

$$k\tilde{\mathbf{m}} = \mathbf{A}[\mathbf{R} \quad \mathbf{t}]\tilde{\mathbf{M}} \quad (3.1)$$

der  $k$  er en vilkårlig skaleringsfaktor siden bildet er en todimensjonal skalert versjon av det som blir avbildet i objektrommet (6).  $\mathbf{A}$  er den intrinsiske matrisen gitt i ligning 2.6, og  $(\mathbf{R}, \mathbf{t})$  er ekstrinsiske parametere som består av rotasjonsmatrisen  $\mathbf{R}$  og translasjonsmatrisen  $\mathbf{t}$ . Både rotasjon og translasjon er nødvendige for å kunne relatere koordinatene i objektrommet til koordinatene i bilderommet. For å kunne produsere koordinater i bilderommet som er parallelle til koordinatene til objektrommet må det skje en rotasjon av matrisen  $\tilde{\mathbf{M}}$  siden det konverteres fra 3D til 2D. Matrisen må i tillegg gjennom translasjon med mindre de to koordinatsystemene sammenfaller, noe de sjelden gjør. Denne translasjonen introduserer et skift i koordinatene i objektrommet slik at disse koordinatene sammenfaller med perspektivsenteret til kameraet.

Videre defineres  $Z=0$  slik at avbildningsplanet befinner seg ved null i  $Z$ -koordinatene i objektrommet. Ligning 3.1 kan da skrives om til

$$k \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (3.2)$$

der  $\mathbf{r}_1$ ,  $\mathbf{r}_2$  og  $\mathbf{r}_3$  er kolonner i rotasjonsmatrisen  $\mathbf{R}$ . Siden  $Z = 0$  kan  $\tilde{\mathbf{M}}$  skrives om til  $\tilde{\mathbf{M}} = [X, Y, 1]^T$  som dermed relaterer koordinatene i objektrommet til koordinatene i bilderommet ved en homografi  $\mathbf{H}$  fra ligning 3.2 (6):

$$\lambda \tilde{\mathbf{m}} = \mathbf{H} \tilde{\mathbf{M}} \text{ der } \mathbf{H} = \mathbf{A}[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \quad (3.3)$$

Videre kan en homografi betegnes som

$$\mathbf{H}' = [\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3] \quad (3.4)$$

Fra ligning 3.3 har vi da at

$$[\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3] = \lambda \mathbf{A}[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}] \quad (3.5)$$

der  $\lambda$  er en vilkårlig skaleringsfaktor. Rotasjonsmatrisen  $\mathbf{R}$  er alltid en ortogonal matrise noe som i dette tilfellet medfører at  $\mathbf{r}_1 \cdot \mathbf{r}_2 = 0$  som da gir to følgende krav utfra ligning 3.5:

$$\mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 = 0 \quad (3.6)$$

$$\mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 \quad (3.7)$$

Ligning 3.6 og 3.7 beskriver to krav til de intrinsiske parameterne gitt en homografi. Nå som kravene er definert kan det oppnås en analytisk løsning for parameterne i den intrinsiske matrisen. Videre blir denne løsningen numerisk optimalisert basert på maksimums sannsynlighetskriteriet. Først lar vi

$$\mathbf{B} = \mathbf{A}^{-T} \mathbf{A}^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2 \beta} & \frac{v_0 \gamma - u_0 \beta}{\alpha^2 \beta} \\ -\frac{\gamma}{\alpha^2 \beta} & \frac{\gamma^2}{\alpha^2 \beta^2} + \frac{1}{\beta^2} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{v_0}{\beta^2} \\ \frac{v_0 \gamma - u_0 \beta}{\alpha^2 \beta} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{v_0}{\beta^2} & \frac{(v_0 \gamma - u_0 \beta)^2}{\alpha^2 \beta^2} + \frac{v_0^2}{\beta^2} + 1 \end{bmatrix} \quad (3.8)$$

Ved å definere  $\mathbf{B}$ -matrisen på denne måten oppnås det en viss symmetri i matrisen og dette kan defineres ved en 6D-vektor (6):

$$\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T \quad (3.9)$$

Hvis den  $n$ -te kolonnen i homografimatrisen  $\mathbf{H}$  beskrives som  $\mathbf{H}_n = [\mathbf{h}_{n1} \ \mathbf{h}_{n2} \ \mathbf{h}_{n3}]$  får vi følgende:

$$\mathbf{h}_n^T \mathbf{B} \mathbf{h}_m = \mathbf{v}_{nm}^T \mathbf{b} \quad (3.10)$$

der

$$\mathbf{v}_{nm} = [h_{n1}h_{m1}, h_{n1}h_{m2} + h_{n2}h_{m1}, h_{n2}h_{m2}, h_{n3}h_{m1} + h_{n1}h_{m3}, h_{n3}h_{m2} + h_{n2}h_{m3}, h_{n3}h_{m3}]^T$$

Med dette og ved å sette inn at  $\mathbf{B} = \mathbf{A}^{-T} \mathbf{A}^{-1}$  kan begge de to kravene gitt i ligning 3.6 og 3.7 skrives om til

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = 0. \quad (3.11)$$

Hvis  $k$  antall bilder tas av avbildningsobjektet for kalibreringen vil ligning 3.11 bestå av  $k$  separate ligningssett og vi får

$$\mathbf{V} \mathbf{b} = 0 \quad (3.12)$$

der  $\mathbf{V}$  er en  $2k \times 6$  matrise. Det er bare for  $k \geq 3$  som gir en unik løsning av  $\mathbf{b}$  og som vil kunne bestemme alle parametere i den intrinsike matrisen. Hvis mindre enn tre bilder er tilgjengelig vil man måtte sette enkelte parametere til antatte eller kjente verdier for å kunne beregne de andre parametere i den intrinsike matrisen. Gitt at  $k \geq 3$  kan  $\mathbf{b}$  løses for i ligning 3.12, og dermed kan alle de intrinsike parametere beregnes ved følgende ligninger:

$$\begin{aligned} v_0 &= (B_{12}B_{13} - B_{11}B_{23}) / (B_{11}B_{22} - B_{12}^2) \\ \lambda &= B_{33} - [B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})] / B_{11} \\ \alpha &= \sqrt{\lambda / B_{11}} \\ \beta &= \sqrt{\lambda B_{11} / (B_{11}B_{22} - B_{12}^2)} \\ \gamma &= -B_{12}\alpha^2\beta / \lambda \\ u_0 &= \gamma v_0 / \alpha - B_{13}\alpha^2 / \lambda \end{aligned} \quad (3.13)$$



Alle parameterne oppgitt ovenfor, i tillegg til de de radielle distorsjonskoeffisientene blir optimalisert og raffinert ved å minimalisere ligning 3.14 (6). Det her er antatt at det er gitt  $k$  bilder med  $l$  bildepunkter. Hvis alle disse bildepunktene inneholder hvit støy estimeres maksimumsannsynlighetskriteriet ved å minimalisere ligning 3.14, der  $\mathbf{m}_{ij}$  representerer punkt  $j$  i bildet  $i$  og  $\hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j)$  tilsvarer projeksjonen av det punktet  $\mathbf{M}_j$  i bildet  $k$ . Dermed vil den minste avstanden mellom disse to parameterne tilsvare den optimaliserte løsningen.

$$\sum_{i=1}^k \sum_{j=1}^l \|\mathbf{m}_{ij} - \hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j)\|^2 \quad (3.14)$$

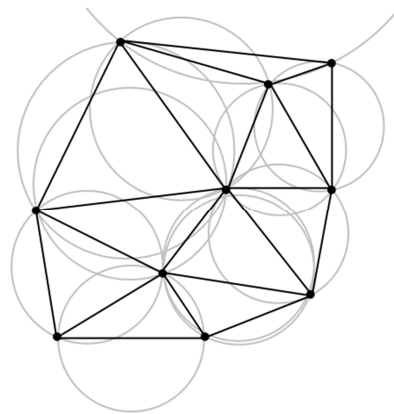
For å oppsummere er altså den estimerte fokallengden gitt ved en skalering av  $\alpha$  og  $\beta$ , koordinatene til PPA i bildet er gitt ved  $(u_0, v_0)$ , skjevhetsforholdet mellom  $y$ - og  $x$ -aksen er gitt ved  $\gamma$  i ligning 3.13 og de radielle distorsjonskoeffisientene blir estimert i ligning 3.14. De numeriske verdiene av de forskjellige parameterne, og de konkrete resultatene av kalibreringsprosessen er beskrevet i delkapittel 4.1.

### 3.2.3 Implementasjon av kalibreringsresultatene

CalLab og CalDe har da gitt alle de numeriske verdiene for å kunne beregne de nye bildekoordinatene som er uten forvrenging. Ligning 2.5 viser hvordan disse beregnes. CalLab beregner ikke de tangentielle distorsjonskoeffisientene og medfører at disse settes lik null i ligning 2.5. Det er oppgitt en kode på DLR sin hjemmeside for hvordan beregningen av disse koordinatene kan implementeres (18). Kort forklart baserer denne metoden seg på forholdet i jacobimatrisen mellom de kalibrerte koordinatene og residualfeilen i de opprinnelige koordinatene. Kalibreringen av koordinatene forløper helt til dette forholdet er under en fastsatt grense. I dette prosjektet er denne forvrengningsfeilen satt til maksimum 0,1 piksel. I tillegg blir koordinatene omgjort til vektorer i  $x$ - og  $y$ -retning når de kalibrerte koordinatene skal beregnes, og dette gjøres for å øke kjørehastigheten til koden. Etter at de kalibrerte koordinatene er kalkulert blir disse vektorene igjen omgjort til  $M \times N$  matriser med samme størrelse som det opprinnelige bildet.

Siden ligning 2.5 består av ikke-lineære ledd vil de nye kalibrerte koordinatene ikke ha konstant differanse mellom seg over en gitt avstand. Det vil si de

oppholder seg på steder innimellom og utenfor rutenettet til de opprinnelige koordinatene. Etter at de kalibrerte koordinatene har blitt estimert må derfor også hver pikselverdi i bildet beregnes på nytt på bakgrunn av de opprinnelige pikselverdiene. DLR sin kode tar bare for seg beregning av de kalibrerte koordinatene og ikke pikselforflytningen disse nye koordinatene medfører. I dette prosjektet har det blitt brukt en spesiell type interpolering for å estimere de nye pikselverdiene til de korresponderende koordinatene. Dette gjøres ved å benytte en av Matlabs innebygde funksjoner, *TriScatteredInterp*, som brukes til å gjøre interpolering ved spredte datapunkter. Denne interpoleringen benytter Delaunay triangulering som for et sett med datapunkter triangulerer slik at ingen andre punkter i settet er innenfor den omskrevne sirkelen for trianglet rundt datapunktet (19). Dette er illustrert i figur 3.3 der de sorte prikkene representerer datapunktene. På bakgrunn av de kalibrerte koordinatene, og pikselverdiene i bildet, lager *TriScatteredInterp* en matematisk overflate som gir pikselverdiene som funksjon av koordinatene i bildet. Denne funksjonen kan evalueres i samtlige punkter i bildet og blir evaluert i de opprinnelige koordinatene til bildet. Dette gjøres fordi pikselverdiene i det opprinnelige bildet egentlig er representert ved de kalibrerte koordinatene og ikke det opprinnelige koordinatsystemet. For å finne ut hva de egentlige pikselverdiene er i det opprinnelige koordinatsystemet blir altså de pikselverdiene fra det opprinnelige bildet nå representert ved de kalibrerte koordinatene interpolert.



Figur 3.3: Delaunay triangulering (20)

## 3.3 Rektifikasjon

Etter kalibrering foretas rektifikasjon av stereobildene. Siden translasjonstrinnet introduserer et lite skift i y-retning i stereobildene er det nødvendig å rektifisere bildene for å kunne matche bildepunktene på en best mulig måte. Her gjennomgås transformeringen av bildene for rektifikasjon.

### 3.3.1 Transformering av bildematrixene

Bilderektifikasjon, som nevnt i avsnitt 2.5.2, gjøres med utgangspunkt i perspektivprosjeksjon. I ligning 3.3 kan homografimatrisen skrives om til (13)

$$\mathbf{H} = \mathbf{A}[\mathbf{R} | \mathbf{t}] = \left[ \begin{array}{c|c} \mathbf{q}_1^T & q_{14} \\ \mathbf{q}_2^T & q_{24} \\ \mathbf{q}_3^T & q_{34} \end{array} \right] = [\mathbf{Q} | \tilde{\mathbf{q}}] \quad (3.15)$$

der koordinatene i bildeplanet blir da gitt ved

$$\begin{aligned} x &= \frac{\mathbf{q}_1^T \mathbf{w} + q_{14}}{\mathbf{q}_3^T \mathbf{w} + q_{34}} \\ y &= \frac{\mathbf{q}_2^T \mathbf{w} + q_{24}}{\mathbf{q}_3^T \mathbf{w} + q_{34}} \end{aligned} \quad (3.16)$$

Koordinatene  $\mathbf{c}$  i perspektivsenteret  $C$  er gitt ved

$$\mathbf{c} = -\mathbf{Q}^{-1}\tilde{\mathbf{q}} \quad (3.17)$$

Med dette kan  $\mathbf{H}$  skrives om til

$$\mathbf{H} = [\mathbf{Q} | -\mathbf{Q}\mathbf{c}] \quad (3.18)$$

Tanken bak rektifikasjon er å rotere de opprinnelige homografimatrisene til hvert bilde rundt sine respektive perspektivsenter til fokalplanene til bildene er koplanare. Dette medfører at begge fokalplanene inneholder basislinjen, og basislinjen er dermed parallell til den nye x-aksen i begge bildeplan. Siden de intrinsiske parameterne er like i begge de opprinnelige homografimatrisene vil de nye homografimatrisene etter rotasjon kun være forskjellige i perspektivsenterkoordinatene. De nye homografimatrisene kan skrives som

$$\begin{aligned} \mathbf{H}_{n1} &= \mathbf{A}[\mathbf{R} | -\mathbf{R}\mathbf{c}_1] \\ \mathbf{H}_{n2} &= \mathbf{A}[\mathbf{R} | -\mathbf{R}\mathbf{c}_2] \end{aligned} \tag{3.19}$$

De nye perspektivsentrene  $\mathbf{c}_1$  og  $\mathbf{c}_2$  er gitt av de opprinnelige perspektivsentrene gitt i ligning 3.17. Denne transformering av koordinater som følge av de nye homografimatrissene gjøres altså på bakgrunn av fundamental matrisen (jf. avsnitt 2.5.2). For å kunne estimere fundamentalmatrisen må bildeplanenes opprinnelige orientering før rektifikasjon kjennes. Homografimatrissen gir denne orienteringen og beregnes ved å sammenligne korresponderende punkter i stereobildene og hvordan disse står i forhold til hverandre.

### 3.3.2 Punktdeteksjon og matching

Deteksjon av punkter gjøres i Matlab med funksjonen *CornerDetector* og gjøres separat i begge stereobildene. Denne funksjonen ble brukt til deteksjon og lokalisering av hjørner basert på sammenligning av lokale intensiteter av pikselverdiene. For hvert detektert hjørne blir et kvadratisk område på  $21 \times 21$  piksler rundt hjørnet ignorert ved videre deteksjon.

Etter at det har blitt foretatt punktdeteksjon i begge bilder blir det hentet ut karakteristiske trekk rundt hvert hjørne, og dette gjøres med funksjonen *extractFeatures*. Dette gjøres for å kunne korrespondere de detekterte punktene i det ene bildet til de detekterte punktene i det andre bildet. Størrelsen på området rundt punktetene som skal matches har en vindusstørrelse på  $9 \times 9$  piksler.

Matchingen av de korresponderende punktene gjøres deretter med funksjonen *matchFeatures*. Metoden som brukes ved matching er normalisert korrelasjon som er beskrevet i avsnitt 2.4.2 ved ligning 2.14. Punktene i det ene bildet blir korrelert opp mot punktene i det andre bildet. Den med høyest korrelasjonsverdi anses å være to korresponderende punkter. Her har terskelgrensen blitt satt til 10 % for at to punkter skal anses som korresponderende til hverandre. Det vil si at hvis ingen punktpar ikke har korrelasjonsverdi over denne terskelgrensen blir punktet ikke matchet.

### 3.3.3 Estimering av fundamentalmatrisen

Etter at de korresponderende punktene i stereobildene har blitt matchet kan fundamentalmatrisen bli estimert. For at denne estimeringen skal bli så god som mulig er det viktig å få fjernet punkter som er feil matchet, med andre ord

punkter som ikke er korresponderende til hverandre i bildene. Avstanden mellom de korresponderende punktene skal være omtrentlig de samme for alle de matchede punktene i stereobildene (jf. ligning 2.10). De punktparene som har en urimelig høy eller lav distanse mellom seg bør derfor fjernes. For å ekskludere disse brukes blant annet en robust estimeringsteknikk kalt Random Sample Consensus (RANSAC). Dette er en iterativ metode for å estimere parametere for et datasett (21). Denne estimering blir bedre jo flere iterasjoner RANSAC-algoritmen har muligheten til å kjøre, og her er antall iterasjoner satt til 1000. Distanseterskelen for å finne feil matchede punkter er satt til 0,01, og prosentandelen for å finne maksimalt antall gyldige matchinger er satt til 99,99. De matchede punktene med avvikende distanseverdier blir fjernet og fundamentalmatrisen blir deretter estimert.

### 3.3.4 Perspektivtransformering

Nå som fundamental matriser er estimert kan den faktiske rektifikasjonen foretas. I funksjonen *estimateUncalibratedRectification* brukes fundamentalmatrisen samt de korresponderende punktparene i stereobildene til å beregne de to nye homografimatrissene. Det blir beregnet to separate homografimatrissene til hvert stereobilde. Funksjonen krever ikke intrinsiske eller ekstrinsiske parametere fra kameraet.

Deretter blir funksjonen *GeometricTransformer* benyttet til å transformere de opprinnelige stereobildene til rektifiserte bilder på bakgrunn av de nye homografimatrissene. Hver homografimatrise blir anvendt på sitt respektive bilde. Etter at stereobildene har blitt rektifisert har koordinatene til pikselverdiene endret seg og befinner seg ikke lenger i et uniformt rutenett. Dette er samme situasjon som oppstår etter kalibrering (jf. avsnitt 3.2.3) og nødvendigvis interpolering for ny beregning av pikselverdiene i et uniformt rutenett. Den interpoleringen brukt her er av typen bilineær interpolering.

## 3.4 Matching

For å kunne beregne parallaksen i punktene i bildet må altså punktene i det ene stereobildet lokaliseres i det andre stereobildet (jf. delkapittel 2.4).

Bildematchingen etter rektifikasjon av stereobildene kan gjøres vesentlig enklere siden det nå kun søkes etter det korresponderende punktet i den samme raden som det punktet som skal matches befinner seg i.

### 3.4.1 Valg av metrisk matchingsmetode

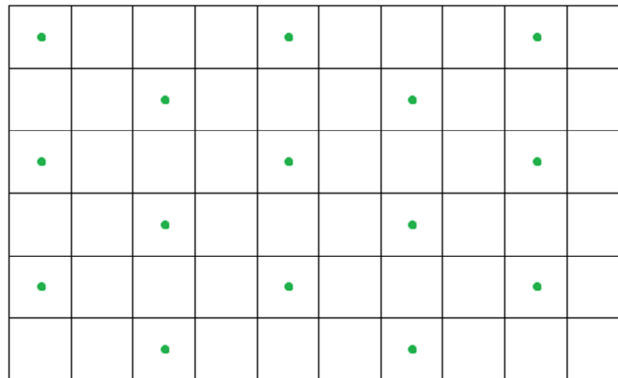
Alle de metriske matchingsmetodene i avsnitt 2.4.2 ble evaluert i stereobildene. Konklusjonen ble at SKD-, SAD- og MaxAD-metodene hadde rask prosesseringstid, men presisjonen av matching av en del punkter var for unøyaktig. Utfallet av matchingsmetoden som bruker normalisert krysskorrelasjon var derimot motsatt. Denne metoden brukte lengre prosesseringstid enn de overnevnte, men hadde forholdsvis god nøyaktighet ved matching av punktene. Selv om prosesseringstiden for korrelasjonsmetoden bare er ørlite større enn de andre metodene så vil denne differansen bli skalert opp når alle punktene i bildet skal matches. I tabell 3.1 er det vist at kameraet har en oppløsning på  $1280 \times 1024$ , og det vil totalt si i overkant av 1,3 millioner piksler som skal matches. Det er derimot særdeles viktig for modellering at matchingen av punktene i stereobildene matches perfekt. En mismatch på bare en piksel vil føre til at 3D-modellen endres med et helt nivå, det vil si høyden i modellen endres med  $\Delta Z$  (jf. avsnitt 2.6.2). I tillegg er prosesseringstid en underordnet faktor i dette prosjektet.

På dette grunnlaget har derfor normalisert krysskorrelasjon blitt brukt som matchingsmetode videre i prosjektet.

### 3.4.2 Fremgangsmetode for matching

For å redusere prosesseringstiden til matchingen slik at den blir av en rimelig karakter innføres det et nytt rutenett slik som vist i figur 3.4. Figuren viser et  $6 \times 10$  rutenett, og dette illustrer et vilkårlig utdrag fra bildet. De grønne markeringene indikerer hvilke piksler som blir matchet. Dette medfører altså at matchingen hopper over tre piksler i hver rad for hver matching som gjøres, og dette bidrar effektivt til å redusere prosesseringstiden med en tredjedel. En prosesseringstid for matching av hele bildet på 30 minutter vil nå ta 10 minutter. Ved det oppsettet som er benyttet vil tre piksler i bildet tilsvare en metrisk

størrelse på 0,14 mm. Det er derfor ikke forventet at høyden eller parallakseverdiene vil forandre seg i særlig grad over disse tre pikslene som hoppes over. Som vist i figur 3.4 er rutenettet i tillegg utformet slik at annenhver rad er skiftet langs kolonnen for å miste minst mulig informasjon. Siden avbildningsobjektet aldri opptar hele bildet er det meningsløst å matche hele bildet. Det er derfor implementert funksjonalitet som muliggjør markering av det området som skal matches. Dette vil også redusere prosesseringstiden ved matching betydelig.

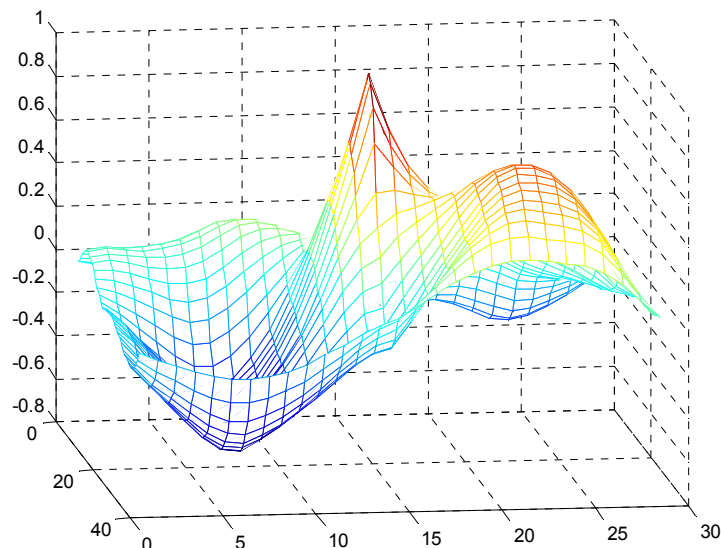


*Figur 3.4: Rutenett ved matching. De grønne punktene indikerer de punktene som blir matchet.*

For hvert punkt som matches tas det et utdrag av piksler rundt det punktet i det ene bildet. Dette kalles referansevinduet, og størrelsen på dette vinduet har blitt satt til  $15 \times 15$  piksler. Referansevinduet er alltid kvadratisk, og det aktuelle punktet som skal matches befinner seg alltid i midten av vinduet. Dette referansevinduet blir krysskorrelert med et vindu fra det andre bildet over den samme raden som det punktet som skal matches er i. Vinduet i det andre bildet refereres til som slavevinduet. Dette vinduet består av like mange rader som referansevinduet, men antall kolonner er lik det totale antallet kolonner i hele bildet. Det søkes altså kun over den midterste rad, men over alle kolonner i det bildet som punktet skal matches. Som nevnt tidligere bør ikke antall rader i disse vinduene settes til en selv om bildene er rektifisert. Dette ville ført til for mange ambiguiteter ved korrelasjon og dermed også mange feilmatchinger. Hvis størrelsen på vinduene blir for store derimot vil dette føre til en midling av parallakseverdiene. Bråe endringer i høyden hos avbildningsobjektet vil dermed ikke bli detektert(jf. avsnitt 2.4.1).

For å gjøre denne krysskorrelasjonen mellom vinduene i de to stereobildene brukes funksjonen *normxcorr2*. Denne funksjonen beregner de normaliserte krysskorrelasjonskoeffisientene i todimensjonal form gitt ved ligning 2.14.

Verdiene på disse koeffisientene varierer fra -1,0 til 1,0 og er gitt av en todimensjonal matrise. Et eksempel på hvordan denne korrelasjonsmatrisen kan se ut er vist i figur 3.5. Figuren viser kun en liten del av kolonnene det søkes over. Korrelasjonsmatrisen vil alltid bestå av dobbelt så mange rader som finnes i enten referansevinduet eller slavevinduet (siden begge vinduene innehar like mange rader). Det søkes dermed etter den maksimale positive verdien i den midterste raden siden bildene er rektifiserte og korrelasjonsverdiene av interesse vil derfor alltid være i denne raden. Legg merke til i figur 3.5 at maksimal verdien befinner seg i rad 15 (langs x-aksen), og den maksimale korrelasjonsverdien vil som regel befinne seg i den raden som har samme verdi som antall rader i referansevinduet eller slavevinduet. Denne beregningen av krysskorrelasjonsmatrisen mellom vinduene gjøres for hvert punkt som skal matches.



*Figur 3.5: Matrise med korrelasjonskoeffisienter ved matching av et punkt. X og Y-aksen representerer rader og kolonner respektivt, og Z-aksen representer korrelasjonskoeffisientene.*

### 3.4.3 Koordinatsystem og parallakseberegning

For å beregne parallaksen mellom de matchede punktene må koordinatsystemet i stereobildene settes slik at PPA utgjør origo. Etter rektifikasjon av stereobildene har PPA endret seg i det ene bildet. For å gjenfinne dette punktet blir PPA-punktet i det første bildet matchet og lokalisert i det andre bildet. Origo i koordinatsystem i det andre bildet utgjør dermed dette matchede punktet.



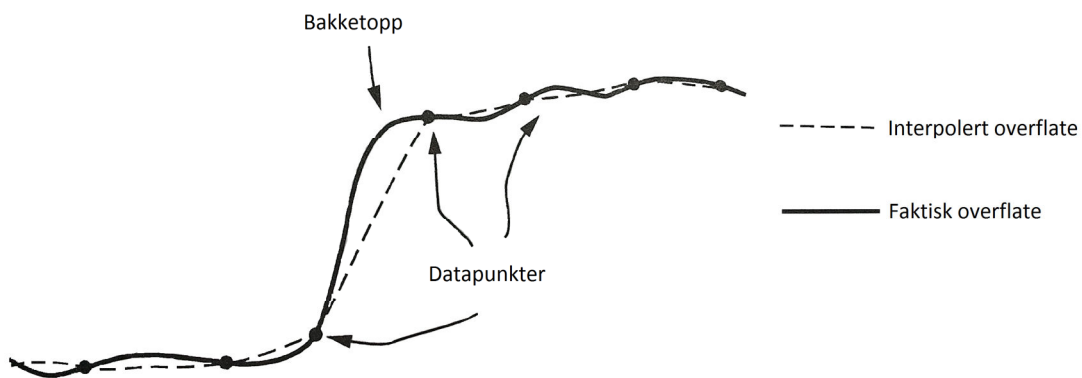
Dermed kan parallaksen beregnes ved simpelthen å subtrahere koordinatene til de korresponderende punktene.

## 3.5 Modelling

Etter at parallaksen er beregnet kan modellering av selve 3D-modellen begynne. På bakgrunn av parallakseverdiene og koordinatene til pikslene i bilderommet beregnes verdiene og koordinatene i objektrommet langs X-, Y- og Z-aksene.

### 3.5.1 Interpolering av Z-verdier

Siden det hoppes over tre piksler for hver piksel som matches vil parallakseverdiene bestå av mange tomme elementer eller nullverdier. Z-verdiene beregnes kun på bakgrunn av de parallakseverdiene der det har blitt utført matching. De Z-verdiene som tilsvare de parallakseverdiene som ikke har blitt matchet blir satt til null. Disse verdiene må beregnes ved å benytte verdiene rundt nullverdiene. Dette gjøres ved interpolering, og her benyttes den samme typen interpolering som ved kalibrering (jf. avsnitt 3.2.3). Figur 3.6 viser et eksempel på forholdet mellom den interpolerte overflaten og den faktiske overflaten. Ved svært bråe endringer viser figuren at den interpolerte overflaten føre til en utjevning av den faktiske overflaten. Dette er ikke regnet som et vesentlig problem her siden den metriske avstanden mellom datapunktene er så liten. Som nevnt utgjør tre piksler kun 0,14 mm med det oppsettet brukt her.



Figur 3.6: Interpolert overflate og den faktiske overflaten i 3D-modellen

### 3.5.2 Koordinater i objektrommet

Før  $Z$ -verdiene interpoleres må disse verdiene beregnes på bakgrunn av parallakseverdiene. Dette gjøres ved å bruke ligning 2.15 der  $x_2 - x_1$  som kjent utgjør parallakseverdiene. Disse parallakseverdiene er i antall piksler og må konverteres til metriske enheter før  $Z$ -verdiene beregnes. Dette gjøres simpelthen ved å multiplisere pikslene med MPP-faktoren, som er oppgitt i tabell 3.1, og dividere på skaleringsfaktoren som beregnes ved ligning 2.1. Som nevnt er høyden mellom kameraet og modelleiren 30 cm, og fokallengden i linsen er oppgitt i tabell 3.2.

Etter at  $Z$ -verdiene har blitt interpolert, som beskrevet i det foregående avsnittet, blir  $X$ - og  $Y$ -koordinatene beregnet ved ligningene 2.2 og 2.3. Koordinatene i bildeplanet må også konverteres til metriske enheter. Når både  $X$ -,  $Y$ -, og  $Z$ -verdier er beregnet utgjør dette en fullverdig 3D-modell.

# 4 Resultater

---

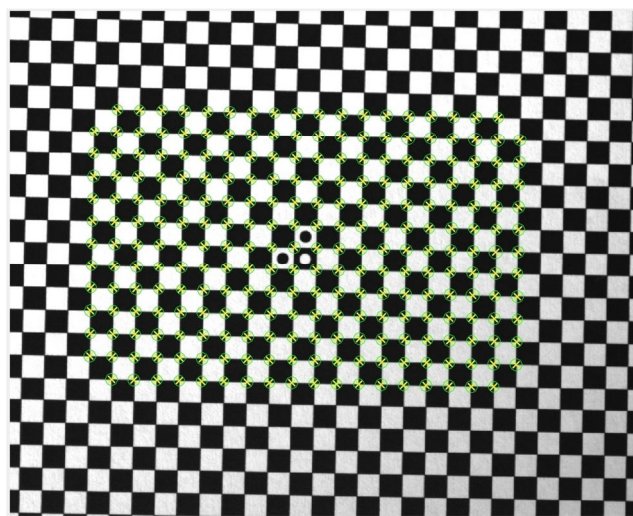
Kapitlet viser og diskuterer resultatene av den metodebruken som ble beskrevet i det foregående kapitlet. Det forklares også eventuelle endringer i metodebruken på bakgrunn av tidligere resultater. Dette kapitlet er strukturert slik at resultatene blir beskrevet kronologisk i henhold til den rekkefølge på prosesseringen av stereobildene. Det vil si at resultatene og verifikasjonen av kalibreringen blir beskrevet først. Deretter blir resultatene fra rektifikasjonen og matchingen omtalt og diskutert. Tilslutt blir selve 3D-modellen fremvist og forklart.

## 4.1 Kalibrering

Her presenteres de resultatene kalibreringen gav. Detaljert spesifikasjoner av kameraet og linsen som ble kalibrert finnes i avsnitt 3.1.1. Det er tatt utgangspunkt i sjakkbrettet som avbildningsobjekt ved kalibrering, men denne kalibreringen vil være lik for alle bilder tatt med det aktuelle kameraet og linsen ved den aktuelle høyde der bildet er i fokus. Disse resultatene vil også dermed være gjeldene for stereobildene av modelleiren siden verken høyden eller fokussettingene er forskjellige.

### 4.1.1 CalDe

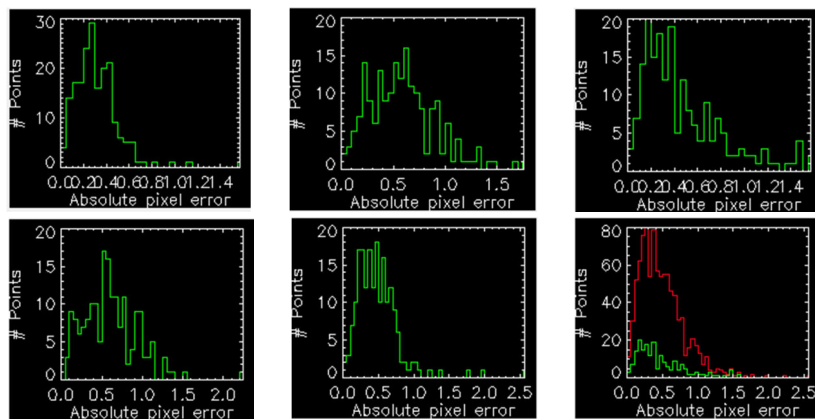
Det ble tatt fem bilder fra ulike innfallsvinkler av sjakkbrettmønstret, men det er imidlertid bare et krav om tre bilder for å kunne fullverdig kalibrere kameraet (jf. avsnitt 3.2.2). I det første bildet detekterte CalDe 193 punkter, eller hjørner, i det andre ble det også detektert 192 punkter, i det tredje 193 punkter, i det fjerde 181 punkter og i det femte bildet 193 punkter. Det er det bildet med færrest detekterte punkter som gir det maksimale antallet punkter som blir brukt under kalibreringsprosessen. Figur 4.1 viser sjakkbrettet samt de detekterte koordinatene til et av bildene. Størrelsen på sirklene vist rundt hjørnene angir den graden av usikkerhet som er knyttet til hvor presist hjørnelokasjonen er detektert. Denne usikkerheten er videre relatert til estimeringen av pikselfeilen for hvert detektert hjørne. Den høyeste sikkerheten er i midten av sirkelen, og det er her CalDe angir koordinatene til hvert hjørne. Koordinatene til disse punktene samt de respektive bildene ble deretter overført til CalLab.



Figur 4.1: Detekterte hjørner på sjakkbrettet i CalDe

### 4.1.2 Callab

Figur 4.2 viser Callab's estimering av pikselfeilhistogrammet til de fem bildene samt det globale pikselfeilhistogrammet for alle bildene. Callab muliggjør manuell fjerning av piksler på bakgrunn av den estimerte pikselfeilen de medfører. Gjennomsnittlig RMS-verdien for pikselfeil til det første bildet er 0,37 piksler, mens for det andre bildet er denne verdien 0,67 piksler, det tredje har RMS-verdien 0,59 piksler, det fjerde har 0,69 piksler og det femte bildet har en gjennomsnittlig RMS-verdi på 0,58. Den globale RMS-verdien for pikselfeil blir da estimert til 0,59 piksler, noe som er å regne som nesten neglisjerbart.



Figur 4.2: De fem lokale pikselfeilhistogrammene samt det globale pikselfeilhistogrammet markert i rødt

Videre ble de intrinsiske parameterne Callab beregnet for kameraet i dette tilfellet som følger:

Fokallengde:

$$f_x = 6646.05 \leftrightarrow 35.22mm$$

$$f_y = 6638.06 \leftrightarrow 35.18mm$$

Prinsipalpunkt:

$$u_0 = 713,509$$

$$v_0 = 637,314$$

Skjevhetsforhold:

$$\gamma = -0,00270604$$

Radielle distorsjonskoeffisienter:

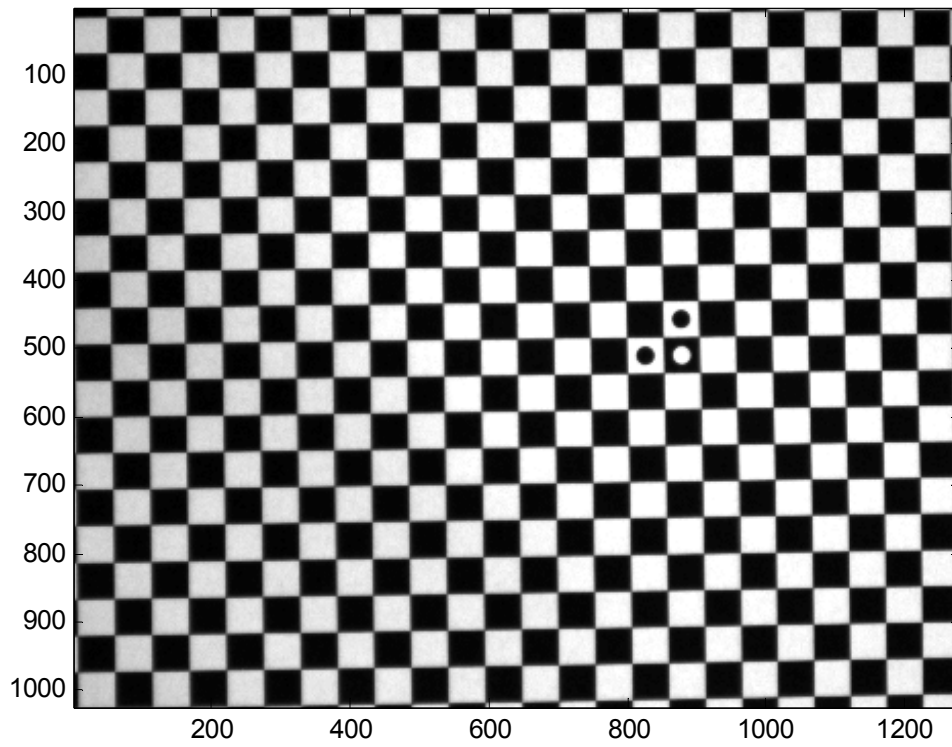
$$k_1 = 0,603666$$

$$k_2 = -50,4027$$

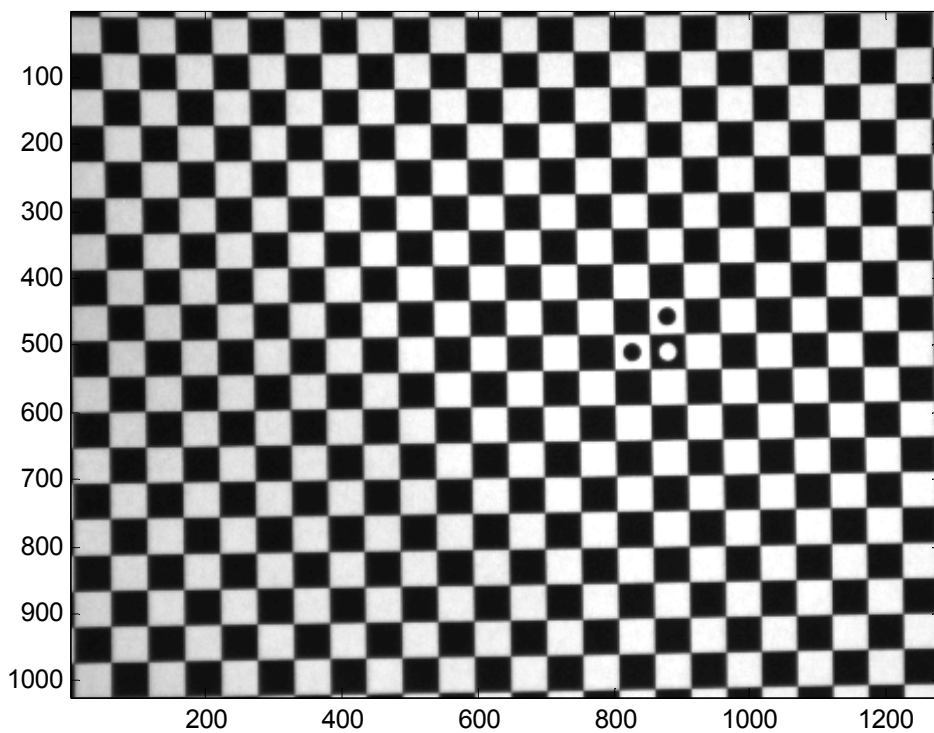
Det er verdt å bemerke at fokallengden Callab estimerte, kun på bakgrunn av de bildene som ble brukt i kalibreringsprosessen, stemmer svært godt med de oppgitte dataene for linsen som ble beskrevet i avsnitt 3.1.1. Callab har også estimert prinsipalpunktet til å være en del forskjøvet fra sitt teoretiske idealpunkt som ville vært, med kameraets oppløsning på  $1280 \times 1024$  piksler, i punktet  $(u_0, v_0) = (640, 512)$ . Dette er en forskyvning på 73 piksler i horisontal retning og 125 piksler i vertikal retning i bildet utfra sitt ideelle punkt. Ellers er skjjevhetsforholdet mellom aksene,  $\gamma$ , tilnærmet lik null, og det er også interessant at den andreordens radielle distorsjonskoeffisienten er vesentlig høyere enn førsteordenskoeffisienten.

### 4.1.3 Implementasjon av kalibreringsresultatene

Figur 4.3 viser det forvrengte bildet av sjakkbrettet før kalibrering, mens figur 4.4 viser bildet etter kalibrering. De visuelle endringene mellom bildet før og etter kalibrering er små. Dette kommer av at linsen i utgangpunktet har svært få aberrasjoner, og dette bekreftes av den lave verdien av den radielle førsteordenskoeffisienten som i all hovedsak bestemmer graden av radiell distorsjon. Det er likefullt utslagsgivende å foreta denne kalibrering siden modellen som skal utvikles bør operere med en metrisk nøyaktighet på godt under 1 mm. Ved nøyere inspeksjon er det dessuten visuelle endringer i kantene av bildet før og etter kalibrering. I det forvrengte bildet er kantene bøyd innover i bildet, men etter kalibrering kan man klart legge merke til at bildet mer "brettet ut". Bildet i figur 4.4 er egentlig større enn det bildet i figur 4.3, men har samme oppløsning. Dette kommer av bildet etter kalibrering er "rettere" enn det før, og dette bekreftes også av de nye koordinatene som ble beregnet ved bruk av ligning 2.5. De numeriske verdiene av disse koordinatene spenner over et større område enn de opprinnelige koordinatene. Konklusjonen er derfor at denne linsedistorsjonen er av typen tønne-distorsjon. Callab har i tillegg estimerte førsteordenskoeffisienten til å være positiv, noe som ytterligere bekrefter tønne-distorsjon.

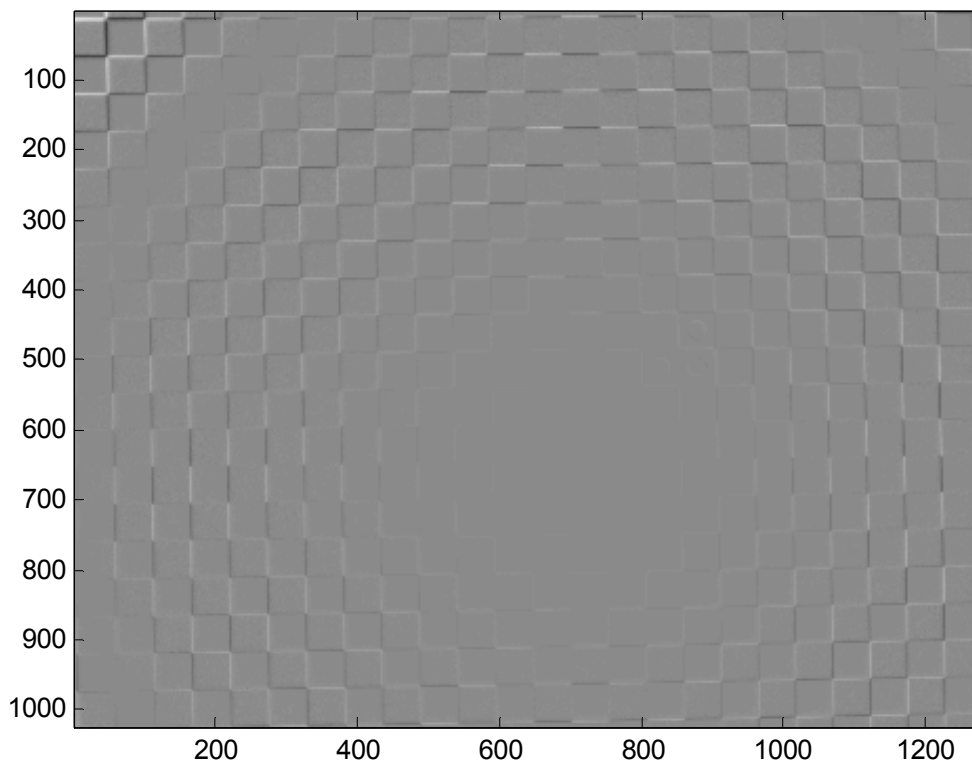


*Figur 4.3: Sjakkbrett før kalibrering*



*Figur 4.4: Sjakkbrett etter kalibrering*

Siden de visuelle endringene mellom figurene 4.3 og 4.4 er vanskelig å få øye på er det tatt med et bilde som viser intensitetsforskjellen mellom de to figurene, og vises i figur 4.5. Området rundt prinsipalpunktet er fri for intensitetsforskjell, hvilket er å forvente siden der ingen linsedistorsjon i dette området. Det virker også som det er en sirkel i ytterkanten av bildet som også er forholdsvis fri for distorsjon. De maksimale intensitetsforskjellene er å finne i hjørnene av bildet, spesielt i det øvre hjørnet til venstre. Dette skyldes at prinsipalpunktet er forskjøvet nedover og til høyre i forhold til sitt teoretiske idealpunkt, det vil si vekk fra det øvre venstre hjørnet.

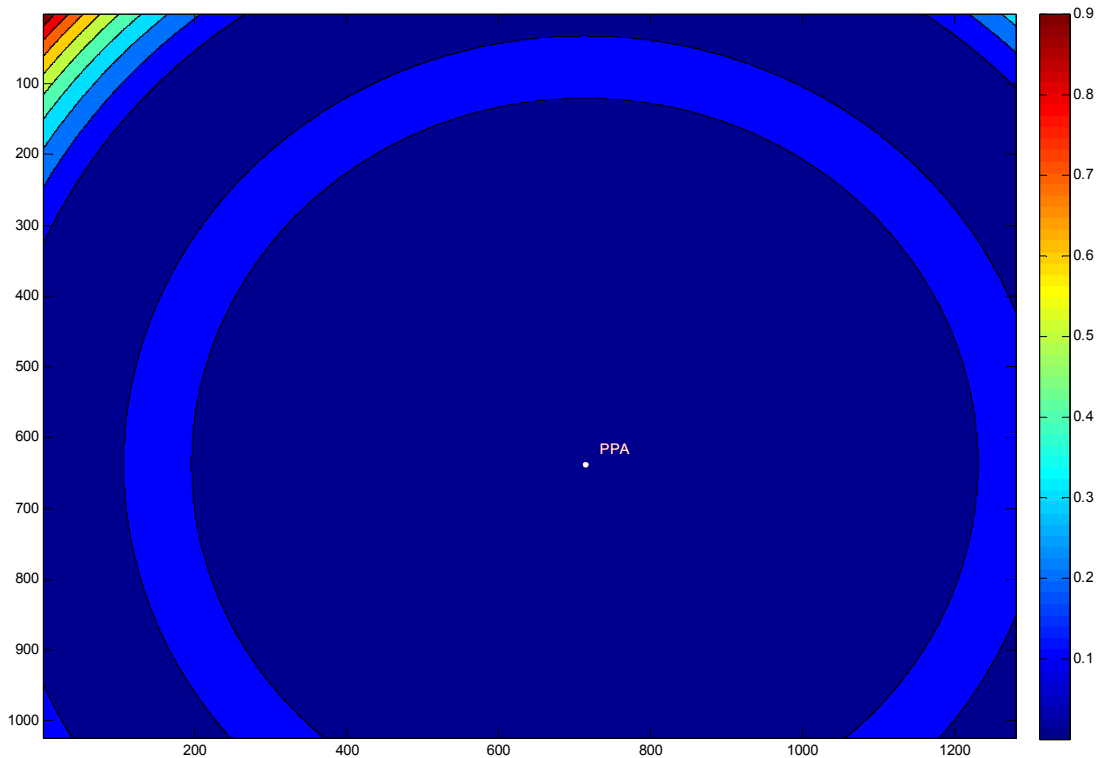


*Figur 4.5: Intensitetsforskjell mellom figur 4.3 og figur 4.4*

Figur 4.6 viser konturplot av det totale pikselskiftet i både x- og y-retning. Det totale pikselskiftet er gitt av  $R = \sqrt{R_x^2 + R_y^2}$ , der  $R_x$  og  $R_y$  er pikselskiftet i x- og y-retning henholdsvis. De numeriske verdiene av pikselskiftet hver enkelt farge tilsvarende er vist til høyre i figuren. I tillegg er PPA plottet i den samme figuren for å vise hvordan disse pikselskiftene er sentrert rundt prinsipalpunktet. PPA er altså origo i alle sirklene i konturplottet. Ved å sammenligne figurene 4.5 og 4.6 fremkommer det at verdiene av pikselskiftene korrelerer naturligvis med verdiene av intensitetsforskjellene før og etter kalibrering. Det høyeste



pikselskiftet ligger i den øvre venstre del og ligger på nesten en hel piksel, mens det er som forventet ingen pikselskift i området rundt PPA.



*Figur 4.6: Konturbildet av pikselskiftet før kalibrering og etter kalibrering. Tallene til høyre er oppgitt i piksler.*

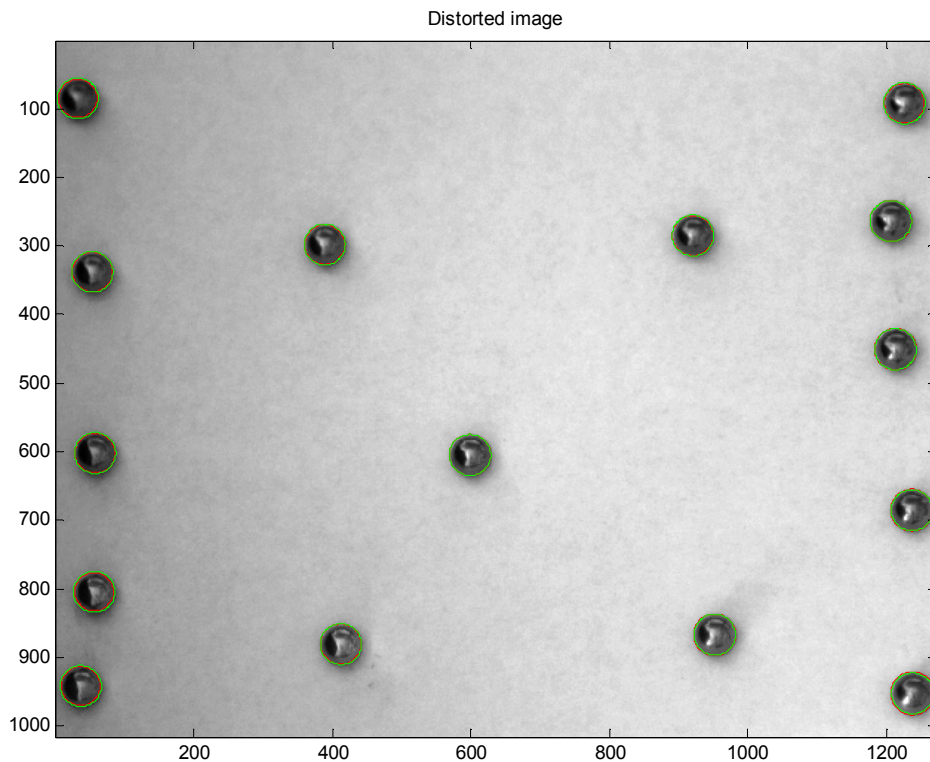
Kalibrering ble også gjort for alle de 12 bildene av modelleiren. Resultatene av alle de kalibrerte bildene er ikke tatt med siden de vil endre seg på nøyaktig likt vis som sjakkbrettet i figur 4.3 og figur 4.4.

#### 4.1.4 Verifikasjon av kalibreringsresultatene

Den eneste parameteren i kalibreringsprosessen som uten videre kan bekrefte at kalibreringen har blitt riktig utført er den estimerte fokallengden. Den verdien Callab har estimert stemmer godt overens med den faktiske verdien på fokallengden. Det anses i tillegg nødvendig å kunne verifisere at kalibreringen har ført til en forbedring av den metriske nøyaktigheten i hele bildet.

Som forklart tidligere i avsnitt 2.2.2 egner kuler seg godt for å kunne avgjøre hvor nøyaktig den geometriske formen er fremstilt i bildet. I dette prosjektet har det blitt brukt 15 lagerkuler, hver med en diameter på 6 mm, til dette formålet.

Lagerkulene har blitt plassert på en jevn flate og avbildet som vist i figur 4.7, og figuren gjengir bildet før kalibrering. De fleste kulene har blitt plassert i ytterkantene av bildet, der distorsjonen er høyest, noen kuler har blitt plassert nærmere midten av bildet, der distorsjonen er noe lavere, og en kule har blitt plassert i midten, der distorsjonen er tilnærmet null. Det antas at formen til denne ene kulen i midten er perfekt avbildet før kalibreringen. Ideen er å bruke denne kulen som referanse for å kunne se på endringene i formen på sirklene før og etter kalibreringen. De grønne sirklene rundt kulene i figur 4.7 representerer idealformen til kulene, det vil si den samme formen som kulen i midten. De røde sirklene rundt kulene representerer den faktiske formen på kulen. De grønne sirklene blir lagt til bildet etter de røde sirklene, så de kulene med idealform har bare grønn sirkel rundt seg mens de sirklene som er på steder med distorsjon vil den røde sirkelen klart skille seg ut fra den grønne sirkelen.

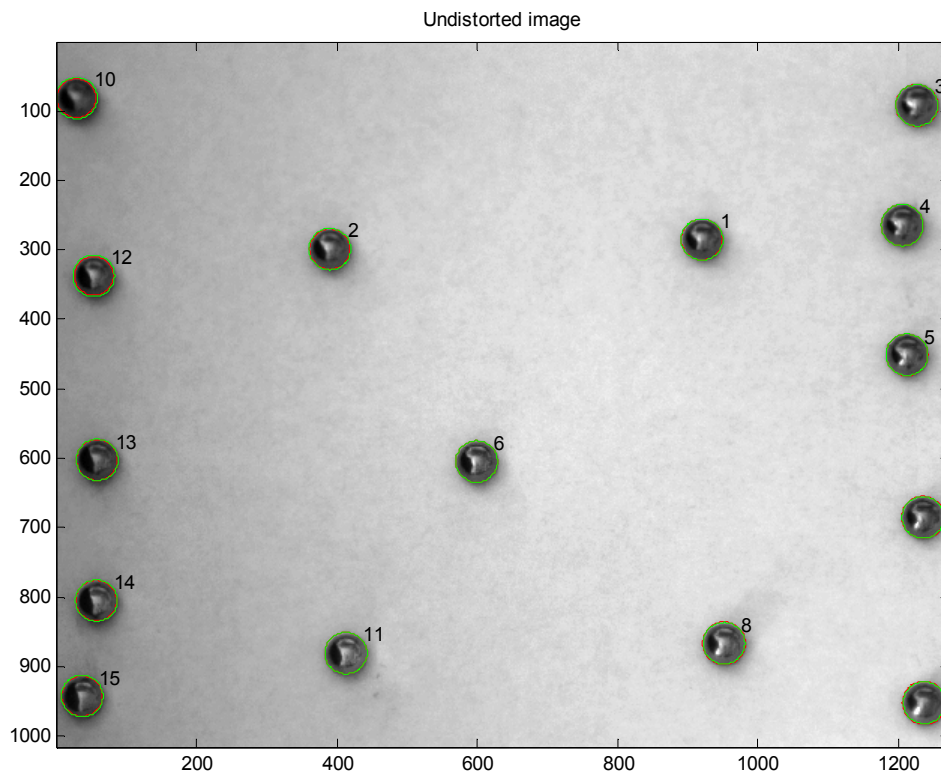


*Figur 4.7: Lagerkuler før kalibrering. Den grønne markeringen tilsvarer den optimale formen på kulen, mens den røde tilsvarer den faktiske formen på kulen.*

For å detektere formen på disse kulene i bildet har det blitt benyttet Hough-transformen, og denne transformen muliggjør lokalisering av sentrum av kulene og deres radius. Denne transformen er nærmere beskrevet i avsnitt 2.2.2. Det har blitt gjort forbehandling av bildet for å framheve kontrasten slik at bakgrunnen

er lysere i forhold til de mørke kulene. Dette forbedrer forhåpentligvis deteksjonen av sirklene.

Figur 4.8 viser den kalibrerte versjonen av den foregående figuren samt nummereringen av de forskjellige kulene. Deteksjonen av kulene har foregått på samme måte som i figur 4.7. Hvis det kalibrerte bildet er en mer korrekt gjengivelse av de geometriske formene til kulene bør de røde sirklene i figuren ha beveget seg nærmere de grønne sirklene. Med noen få unntak er dette tilfellet for de fleste sirklene i figur 4.8.



*Figur 4.8: Lagerkuler etter kalibrering. Den grønne markeringen tilsvare den optimale formen på kulen, mens den røde tilsvare den faktiske formen på kulen*

I både figur 4.7 og 4.8 kan det være vanskelig å se den visuelle forskjellen på de røde og de grønne sirklene. Tabell 4.1 viser derfor blant annet arealforskjellen mellom idealsirkelen og den faktisk detekterte sirkelen før og etter kalibrering for hver av de 15 kulene. Denne arealforskjellen er beregnet ved å ta differensen mellom radiusen i den grønne sirkelen og radiusen i den røde sirkelen og deretter bruke denne radiusen til å regne ut arealet til denne nye sirkelen. Hvis kalibreringen hadde gitt perfekt metrisk nøyaktighet skulle ideelt sett arealdifferansen etter kalibrering være lik null for samtlige kuler. Som tabell 4.1 viser er ikke dette tilfellet, men tabellen viser også at det har blitt en forbedring

av den metriske representasjonen i bildet for nesten alle kuler. Som nevnt tidligere tar kalibreringen ikke høyde for tangentiell distorsjon, og dette vil bidra til at kalibreringen ikke blir perfekt.

Kule nr. (se fig 4.8)	Arealdifferanse, før [pix <sup>2</sup> ]	Arealdifferanse, etter [pix <sup>2</sup> ]	Kompakthet, før	Kompakthet, etter
1	1.0982	1.2958	3.8010	3.8190
2	4.2164	2.6406	3.9529	3.8919
3	2.4595	0.1277	3.8785	3.6072
4	0.0045	0.0044	3.6425	3.6469
5	0.0990	0.0203	3.6087	3.6365
6	0	0	3.6517	3.6561
7	6.2811	4.9732	3.3286	3.3659
8	0.4862	1.3487	3.5573	3.5005
9	0.8636	0.5873	3.5266	3.5523
10	15.4238	11.6551	4.2615	4.1789
11	1.1726	0.0048	3.8061	3.6465
12	3.3278	11.9961	3.9175	4.1873
13	2.6749	1.0813	3.8887	3.8045
14	14.5079	1.8248	4.2410	3.8506
15	9.1255	3.6336	4.1079	3.9350
<b>Gjennomsnitt</b>	<b>4.1161</b>	<b>2.7463</b>	<b>3.8114</b>	<b>3.7519</b>
<b>Varians</b>	<b>24.1732</b>	<b>15.7062</b>	<b>0.0708</b>	<b>0.0543</b>

Tabell 4.1: Arealdifferanse og kompakthet for hver enkelt kule før og etter kalibrering

For enkelte kuler derimot har kalibreringen ført til en forverring av den metriske nøyaktigheten. Dette gjelder kule nr. 1, 8 og 12 (se figur 4.8) og kan skyldes flere faktorer. I verifikasjonsprosessen har det blitt lagt til grunn at alle kulene har nøyaktig samme radius, at de er perfekt sfæriske og at deteksjonen av kulene er feilfri. Det er nok sistnevnte faktor som bidrar til mesteparten av feilene. En feildeteksjon på radiusen på bare noen få piksler fører til stor endring i arealdifferansen. Det mistenkes at dette kan ha skjedd med deteksjonen av kule nr. 12 etter kalibrering.

Tabell 4.1 inkluderer også beregninger av kompaktheten av sirklene både før og etter kalibrering. Kompaktheten beregnes etter ligning 2.7. Det er kule nr. 6 som har den perfekte geometriske formen og derfor også den optimale kompaktheten. Ved perfekt kalibreringen hadde variansen til kompakthetsverdiene vært lik null, og kompaktheten til alle kulene hadde tilnærmet seg verdien av kule nr. 6. Det gjør de ikke, men stort sett er det en forbedring i forhold til verdiene før kalibrering.

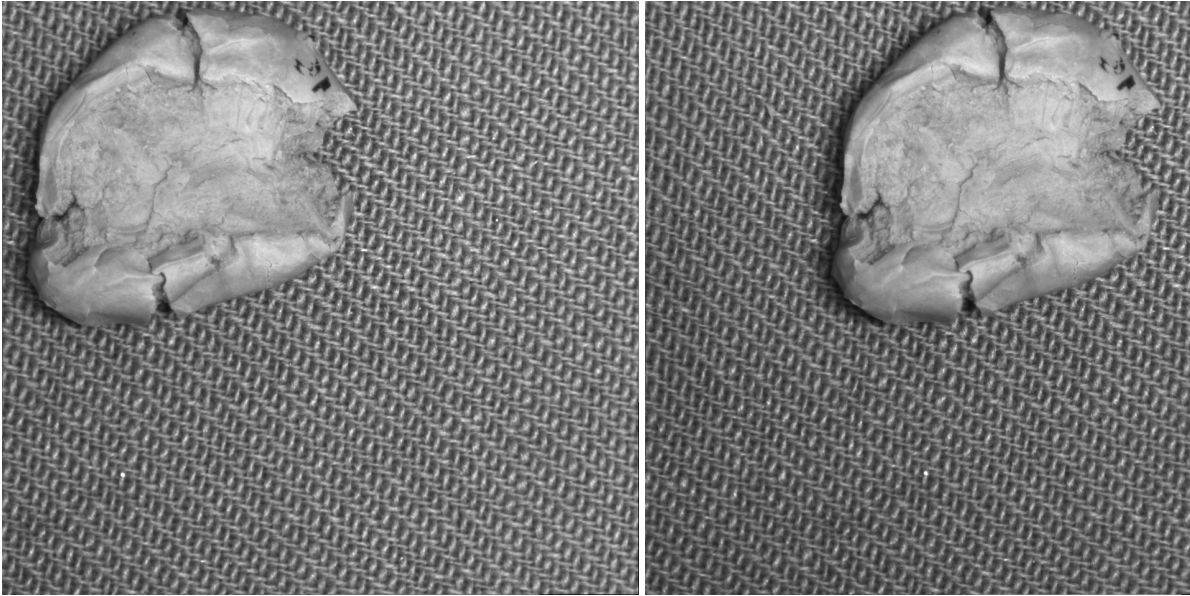
Totalt sett kan det konkluderes med at kalibreringen av linsen har ført til en forbedring av den metriske nøyaktigheten på de aller fleste stedene i bildet. Dette vises nederst i tabell 4.1 der både gjennomsnittet av arealdifferensen etter kalibrering har sunket med  $1,4 \text{ pix}^2$  og snittet av kompakthetsverdiene har nærmet seg idealverdien på 3,65. Variansen til både arealdifferanseverdien og kompakthetsverdiene har også minket etter kalibreringen. Tangentiell distorsjonen fører dessverre til at kalibreringen ikke gir perfekt metrisk nøyaktighet. Distorsjonen i linsen er imidlertid liten i utgangspunktet, men siden 3D-modellen skal operere med en høy grad av nøyaktigheten er det viktig å gjøre distorsjonen så liten som mulig.

## 4.2 Rektifikasjon

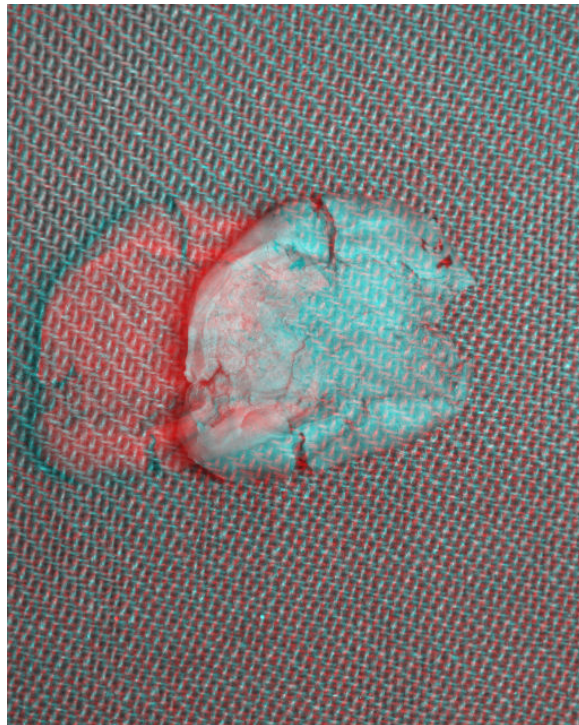
Etter at kalibrering har blitt foretatt av alle 12 bilder av modelleiren må de rektifiseres. Denne bilderektifikasjonen, i motsetning til kalibreringen, må gjøres parvis og kan ikke gjøres separat ved et og et bilde. For å ha varierende basislinje mellom alle de 12 bildene har det første bildet, der modelleiren ligger ytterst til venstre i bildet blitt brukt som referansebilde. Det vil si at rektifikasjonene av alle de andre 11 bildene har blitt gjort opp mot dette referansebildet. For det første av disse 11 bildene vil basislinjen være 1 cm i forhold til referansebildet, for det andre bildet vil basislinjen være 2 cm og for det ellefte bildet vil basislinjen være 11 cm.

Videre i denne rapporten vil det kun bli tatt utgangspunkt i referansebildet og i det tiende av de andre bildene. Dette gir altså en 10 cm lang basislinje. Både referansebildet og bildet som er forskjøvet 10 cm fra referansebildet er vist i figur 4.9. I tillegg er begge stereobildene vist i anaglyfi i figur 4.10. Dette er simpelthen begge stereobildene plassert over hverandre og forskjøvet for å oppnå 3D-effekten.





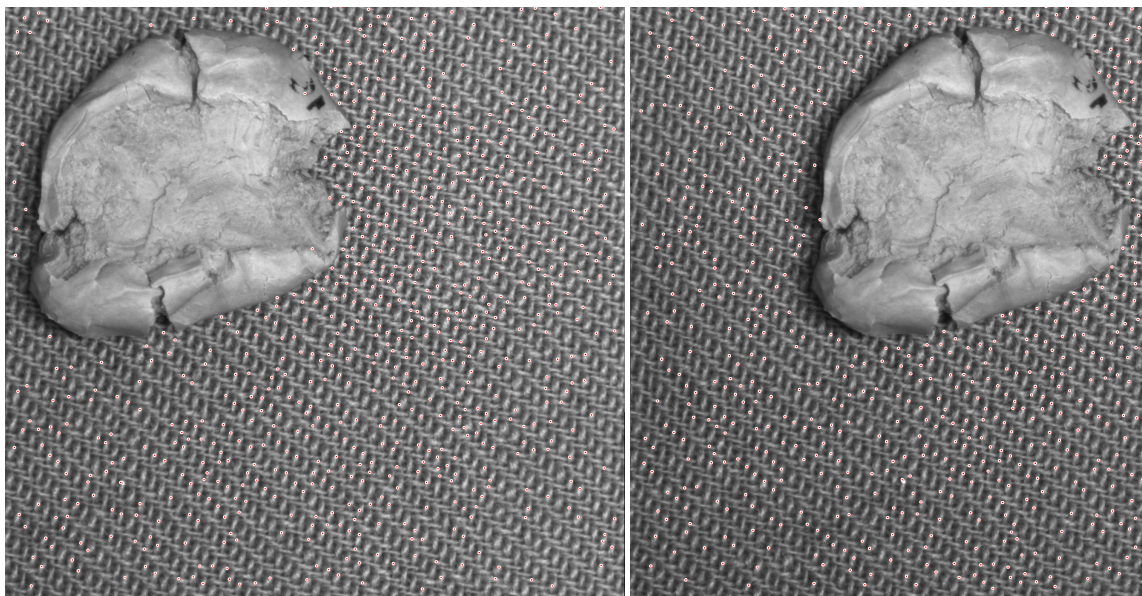
*Figur 4.9: Stereobildene av modelleiren med en basislinje på 10 cm*



*Figur 4.10: Anaglyfisk visning av stereobildene. 3D-effekten kan ses med ordinære 3D-briller farget rød og cyan.*

### 4.2.1 Punktdeteksjon og matching

Deteksjon av punkter gjøres i utgangspunktet som beskrevet i avsnitt 3.3.2. Punkter detektert rundt kantene av bildene blir fjernet fordi de erfaringsmessig gir dårlig transformasjon ved rektifikasjon. Figur 4.11 viser de detekterte punktene er at kantpunktene er blitt fjernet.

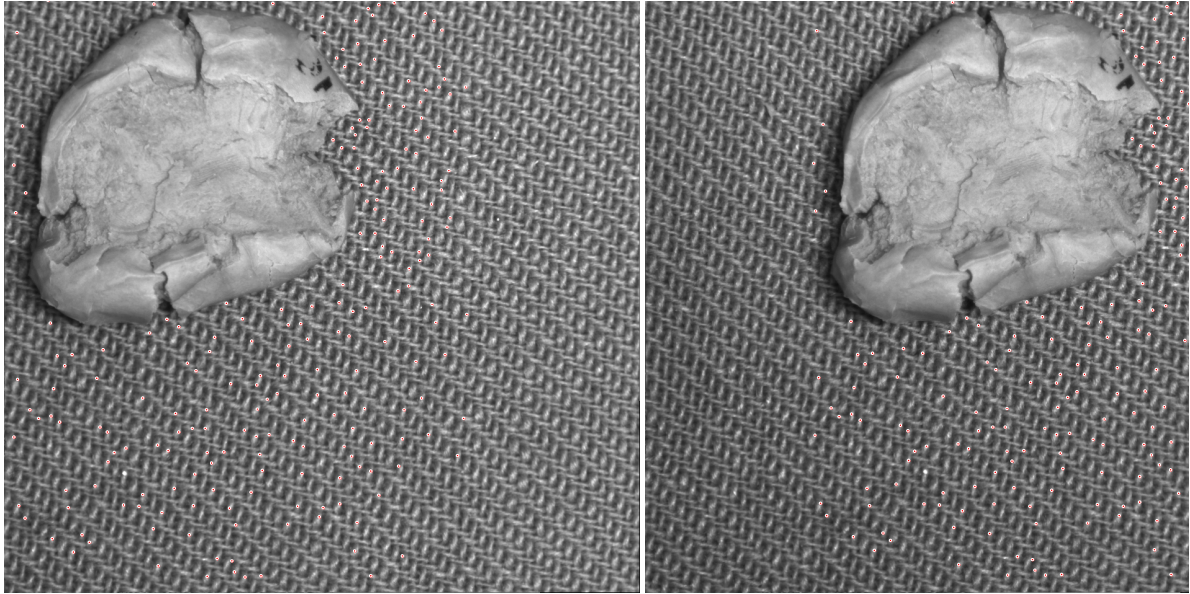


*Figur 4.11: De røde sirklene markerer detekterte punkter i stereobildene. Her er kantpunktene fjernet.*

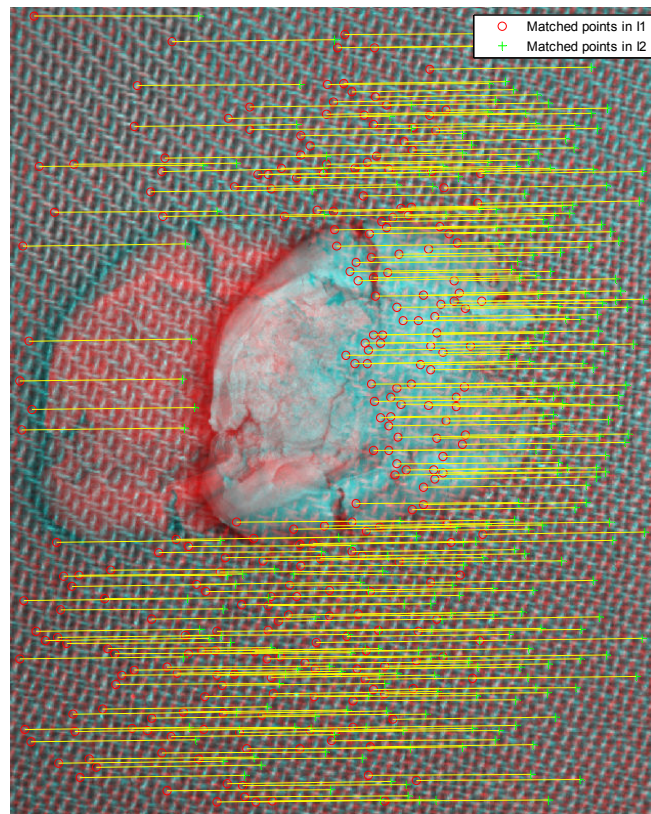
Det er en del punkter i figur 4.11 som ikke har noen korresponderende punkt i det andre bildet. Det er for eksempel til venstre for modelleiren i bildet til høyre, og til høyre for modelleiren i bildet til venstre. Dette kommer av at disse områdene er ute av hverandres synsfelt, dette er med andre ord ikke overlappende områder i stereobildene. Disse punktene vil dermed bli fjernet ved matching. Etter matching beregnes vinkelen og distansen mellom alle punktparene. Som nevnt før er det er lite  $y$ -skift i de korresponderende punktene til stereobildene introdusert av translasjonstrinnet, og derfor aksepteres det noe  $y$ -skiftet i punktparene. Dette skiftet skal derimot ikke være av en urimelig høy verdi og derfor blir punktpar med for høy vinkelverdi fjernet. I tillegg fjernes også punktpar med en urimelig høy eller lav distanse mellom seg. Dette gjøres altså manuelt i tillegg til RANSAC-algoritmen som brukes senere.

Figur 4.12 viser punktene som er igjen etter de overnevnte operasjonene. Figur 4.13 viser de to stereobildene overlagt hverandre og de korresponderende punktparene er forbundet.





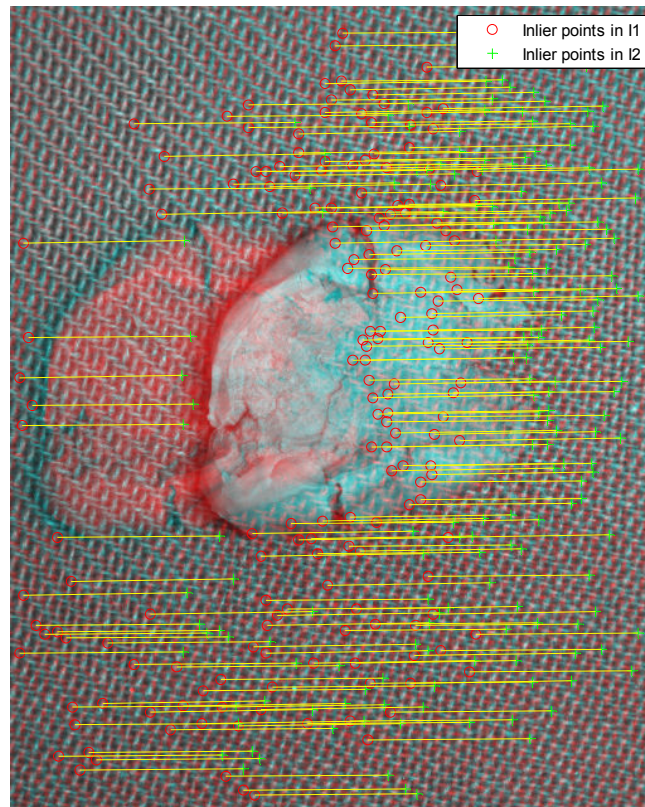
*Figur 4.12: Punktene som er igjen etter fjerning av punkter som ikke kan matches eller er matchet feil*



*Figur 4.13: Korresponderende punkter i anaglyfisk visning. De røde sirklene markerer punktene i bildet til venstre og de grønne tegnene markerer punktene i bildet til venstre*



Figur 4.14 viser punktparene etter RANSAC-algoritmen har blitt benyttet. Flere punktpar har blitt fjernet for å oppnå en best mulig estimering av fundamentalmatrisen. Legg ellers merke til at distansen mellom alle punktparene er forholdsvis like, og at vinkelen til alle punktparene er noe forskjøvet opp i figuren over. Dette bekrefter konklusjonen av den visuelle inspeksjonen av stereobildene før rektifikasjon at translasjonstrinnet forskyver kameraet noe i  $y$ -retning.

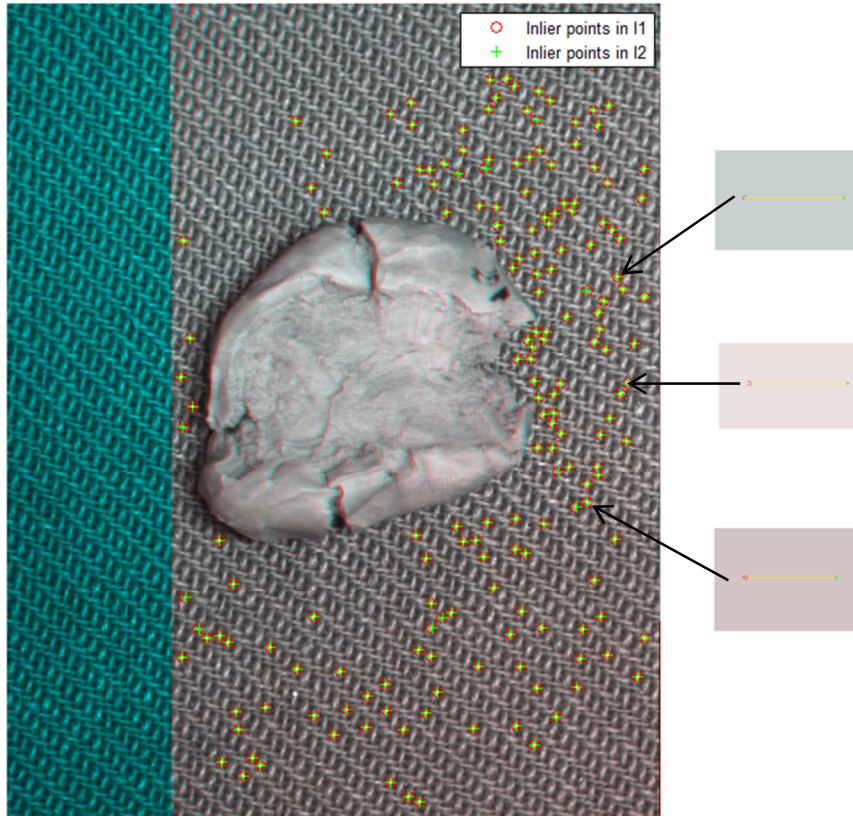


*Figur 4.14: Korresponderende punkter i anaglyfisk visning etter RANSAC*

#### 4.2.2 Perspektivtransformering

På bakgrunn av punktparene i figur 4.14 blir fundamentalmatrisen og homografimatrisene beregnet som beskrevet i avsnitt 3.3.3. Nå kan stereobildene perspektivtransformeres. Resultatet av dette vises i figur 4.15. Siden de nye homografimatrisene kun er forskjellige i perspektivsentrene er de rektifiserte stereobildene nå kun forskjøvet i  $x$ -retning. Stereobildene er nå rektifisert og i tillegg er stereobildene forskjøvet over hverandre. Figuren viser i tillegg tre vilkårlig valgt ut punktpar. Legg merke til at disse punktparene nå fullstendig

rette og har ingen forskyvning lang y-aksen slik som i figur 4.14. Det fargede området i venstre og en liten del nederst i figuren vil bli forskjøvet ut av bildet ved rektifikasjon, og dette skyldes perspektivtransformeringen.



*Figur 4.15: Korresponderende punkter i anaglyfisk visning etter rektifikasjon samt tre utvalgte punktpar. Alle tre punktpar er rette og like lange.*

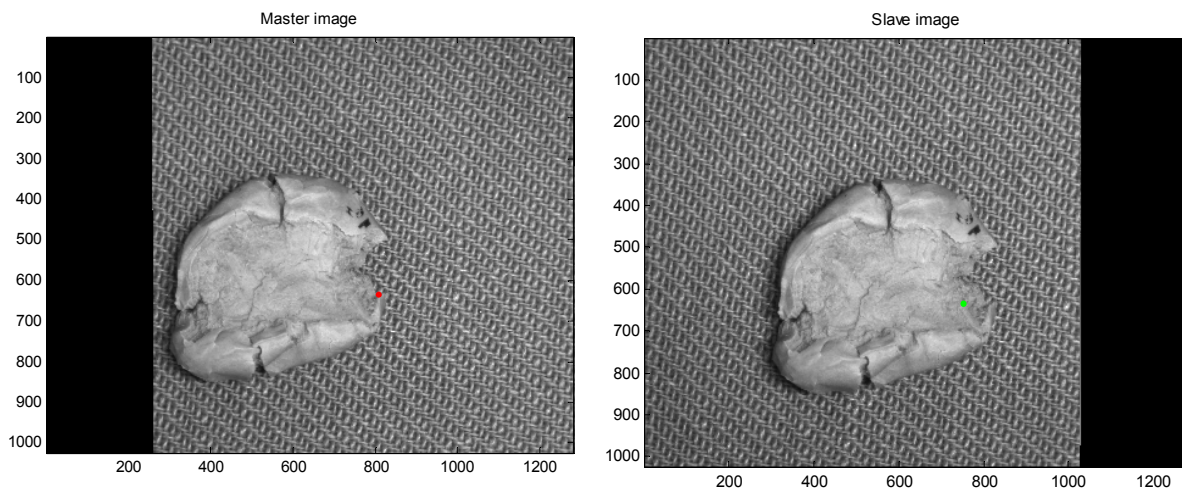
## 4.3 Matching

Etter at stereobildene har blitt rektifisert vil bildematchingen være vesentlig forenklet og mer presis. Matching blir her utført som beskrevet i avsnitt 3.4.2. For hver fjerde piksel i det venstre i figur 4.9 blir det korresponderende punktet i det høyre bildet lokalisert. Matchingen blir kun utført på det område av bildet der modelleiren befinner seg, og dette tilsvarer et område på omtrentlig  $600 \times 600$  piksler. Det vil med andre ord si at det matches ca. 120 000 punkter siden tre

pikslar hoppes over for hver match. Prosesseringstiden for hele bildet er kun på et par minutter ved bruk av en forholdsvis moderne CPU.

### 4.3.1 Korresponderende punkter i stereobildene

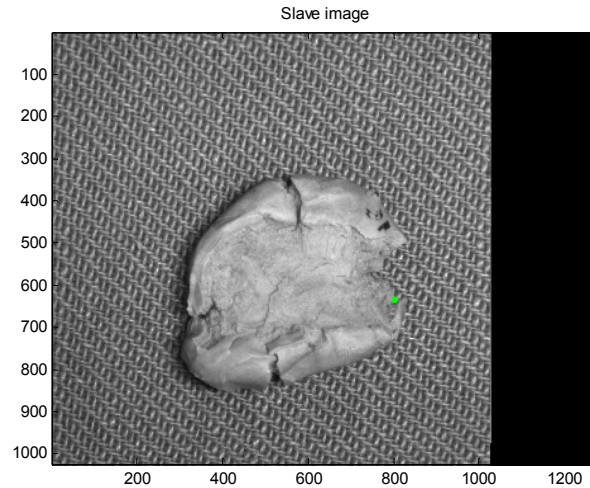
I de aller fleste punktene er matchingen utført perfekt, men det er noen enkelt punkter som ligger til på en slik måte at matchingen blir feilaktig. Dette vil gi store utslag når modellen skal beregnes. Disse feilmatchingene vil framstå som svært høye eller svært dype punkter langs Z-aksen i modellen. Et eksempel på en slik feilmatching er vist i figur 4.16 der det røde punktet tilsvarer det punktet som skal matches i bildet til venstre, og det grønne punktet tilsvarer det matchede punktet i bildet til høyre.



*Figur 4.16: Feilmatching i stereobildene*

For å gi høyere presisjon kan vindustørrelsen økes, men dette vil som nevnt midle ut parallakseverdiene. For å løse dette har det i stedet blitt innført begrensinger i søkeområdet langs kolonnene i det bildet som skal matches. I figur 4.16 var dette søkeområdet satt til 60 pikslar både til venstre og høyre for det punktet som skal matches. Med den lengden på basislinjen brukt her bør ikke det matchede punktet ligge mer enn maksimalt 15 pikslar unna det punktet som matches. I figur 4.17 har søkeområdet langs kolonnene blitt satt til 20 pikslar med samme referansevindu som i figur 4.16. Denne begrensningen gir, som vist i figur 4.17, en klar forbedring i matchingen. Videre i prosjektet har derfor dette søkeområdet blitt begrenset til 20 pikslar istedenfor å søke over alle kolonner.



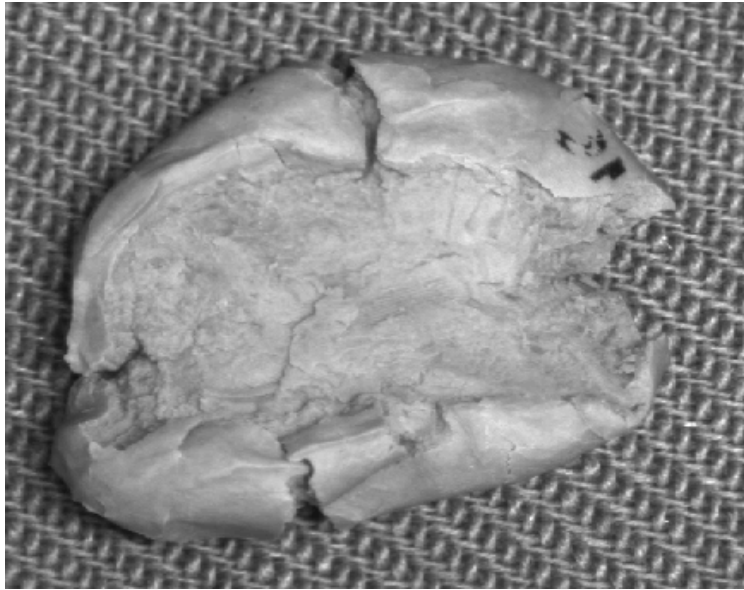


Figur 4.17: Forbedring av matching ved å begrense søkeområdet

### 4.3.2 Parallaxeberegning

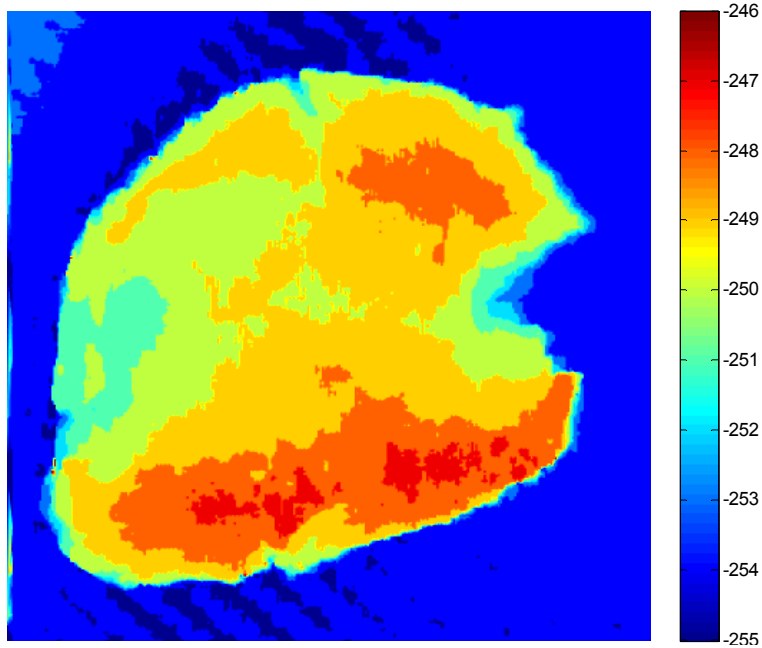
For å beregne stereodispariteten i stereobildene har koordinatene til punktparene blitt trukket fra hverandre etter matching, og disse differanseverdiene utgjør parallaxeverdiene. Koordinatsystem er satt i forhold til hvert sitt respektive PPA-punkt i hvert bilde, som tidligere forklart i avsnitt 3.4.3. Figur 4.18 viser et utsnitt et av stereobildene som tilsvarer det området som har blitt matchet, og figur 4.19 er en skalert visning av parallaxeverdiene i dette området beregnet på bakgrunn av matchingen. Siden Z-verdiene i 3D-modellen vil være proporsjonale med parallaxeverdiene vil figur 4.19 representere et bilde av hvordan modellen vil se ut ovenfra og ned, altså en todimensjonal versjon av 3D-modellen. Rødere områder i figuren representerer eleverte punkter siden parallaxen er høyere i disse områdene, mens blåere områder representerer lavere punkter som altså er lengere unna kameraet. Gulvet modelleiren ligger på vil være det laveste området i bildet og er dermed representert ved blått. Modelleiren har opphøyde kanter i både nedre og øvre del, og er mest representert i den nedre delen ved rødt. Andre karakteristiske områder er i midten helt til høyre av modelleiren. Der er det en gradvis trappeoppbygning i høyden, og dette er representert ved fire nivåer i parallaxeverdi i figuren. Det er også en dyp dal øverst i midten av figuren, og denne dalen er klart representert i parallaxeverdiene ved en vertikal strek i figur 4.19. Legg ellers merke til at forskjellen i parallaxen mellom de laveste og høyeste høyden kun er på 8-9 piksler. Det vil si at hvis matchingen feilmatcher med bare noen få piksler i x-retning er dette av høy betydning for den videre modelleringen.

The selected area that has been matched



*Figur 4.18: Området i bildet som har blitt matchet*

Parallax interpolated



*Figur 4.19: Interpolerte parallaxseverdiene av figur 4.18*

## 4.4 Modellering

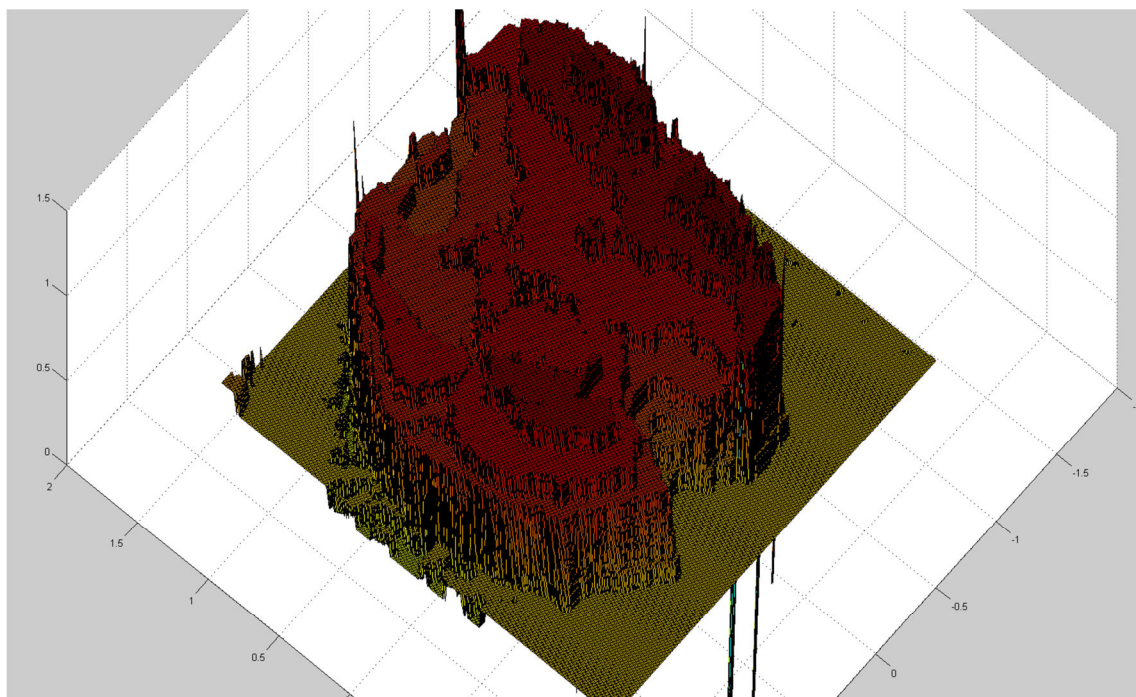
3D-modellen kan nå beregnes på bakgrunn av parallaxseverdiene. Koordinatene i objektrommet og interpoleringen av  $Z$ -verdiene har blitt utført som beskrevet i delkapittel 3.5. Det har vært enkelte utfordringer med å få vist hele figuren i denne rapporten siden Matlab ikke er i stand til å kopiere modellen over til dokumenter. Det har derfor blitt brukt skjermbilder for å vise modellen. Disse figurene gjengir ikke modellen i sin helhet, og derfor har modellene som vist i dette delkapittelet blitt elektronisk vedlagt denne rapporten som Matlab-figurer.

### 4.4.1 Fjerning av ekstrepunkter

Selv med de begrensningene i søkeområdet ved matching (jf. avsnitt 4.3.1) viser det seg likevel at matchingen ved enkelte piksler ikke matcher korrekt. Dette vil som regel vise seg som svært høye eller dype punkter lang  $Z$ -aksen. For å fjerne disse vil det kun aksepteres en viss variasjon rundt avstanden til avbildningsobjektet langs  $Z$ -aksen. Før interpolasjonen av  $Z$ -verdiene vil  $Z$ -verdier som overstiger det aksepterte området settes til null. Deretter vil disse nullverdiene interpoleres på lik måte som resten av nullverdiene. Feilmatchinger som er innenfor det akseptable området vil måtte tas med i modelleringen, men dette utgjør som regel få punkter i forhold til det totale antallet punkter som er matchet.

### 4.4.2 Modellen

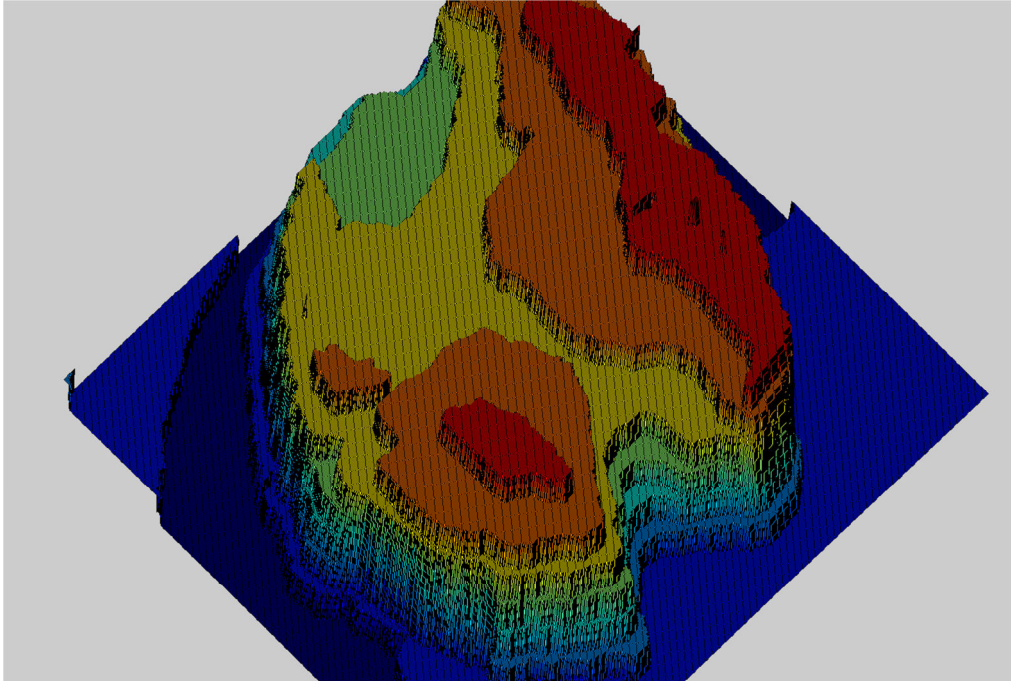
Resultatet av modelleringsprosessen er vist i figur 4.20. 3D-modellen av modelleiren vist fra det øvre høyre hjørnet av figur 4.18. Som det fremkommer av figuren er høydene i modellen delt inn i nivåer. Et slikt nivå utgjør  $\Delta Z$  metriske enheter i høydeforskjell. Med en høyde mellom kameraet og avbildningsobjekt på 30 cm, en MPP-faktor i kameraet på  $5,3 \mu\text{m}$ , en basislinje på 10 cm og en fokallengde i linsen på 35 mm blir  $\Delta Z$  lik 1,17 mm i henhold til ligning 2.17. Det vil si at høydeforskjeller som er mindre enn dette mellom to nivåer vil ikke bli modellert og vil se ut som en konstant høyde i modellen. Derfor vil ikke 3D-modellen være en nøyaktig kopi av det aktuelle avbildningsobjektet. For at en nøyaktig kopi skal gjengis i modellen må betingelsen  $\Delta Z \rightarrow 0$  være oppfylt.



*Figur 4.20: 3D-modellen av modelleiren beregnet på bakgrunn av parallakseverdiene som vist i figur 4.19. Modellen vises fra det nordøstlige hjørnet av figur 4.18. X-, Y- og Z-aksene er gitt i centimeter.*

Koordinatene i objektrommet i figuren over viser at modelleiren er estimert til å være i overkant av 1 cm ved de høyeste punktene og omtrentlig  $2,5 \times 2,5$  cm i lengden og bredden. En maksimal høyde på 1 cm og en  $\Delta Z$  på 1,17 mm gir 8-9 forskjellige nivåer i modellen. Det vil altså si at et enkelt pikselskift i parallakseverdiene utgjør et helt nivå i modellen.

Figur 4.20 viser også at det enkelte feilmatchinger som har oppstått og har gitt utslag i modellen ved forskjellige toppunkter og bunnpunkter. For å øke presisjonen til matchingen kan referansevinduet økes i størrelse. I figur 4.21 er modelleiren blitt matchet med et referansevindu på  $40 \times 40$  piksler. Dette har resultert i så godt som ingen feilmatchinger og dermed heller ingen ekstremverdier langs Z-aksen i modellen. Ulempen med dette er som nevnt at dette fører til midling av parallakseverdiene og høydene i 3D-modellen vil derfor også bli utjevnet. Dette fører til en ukorrekt representasjon av avbildningsobjektet på enkelte steder der det er store variasjoner i høydeforskjellen. Prosesseringstiden vil også være vesentlig høyere ved bruk av et stort referansevindu.



*Figur 4.21: 3D-modellen med referansevindu på  $40 \times 40$  piksler.  
Modellen vises fra det nordøstlige hjørnet av figur 4.18*

I tillegg til de 3D-modellene vist over er det vedlagt flere modelleringer fra de 12 bildene som ble tatt under akvisisjon av stereobildene. 3D-modellene er modellert ved forskjellige basislinjer, fra 2 cm lang basislinje til 10 cm lang basislinje. Det er også tatt med en modell av modelleiren ved 10 cm basislinje der alle pikslene i bildet er matchet. Det er med andre ord ingen piksler som blir hoppet over i denne modellen. Dette er for å kunne sammenligne med modellen i figur 4.20 og deretter kunne evaluere hvor godt interpoleringsmetoden av Z-verdiene fungerer.



# 5 Diskusjon

---

## 5.1 Ytelse

3D-modellene i figur 4.20 og 4.21 har sine begrensinger når det kommer til nøyaktigheten i gjengivelsen av avbildningsobjektet. Disse begrensinger er relatert til høydeoppløsning og presisjonen av matchingen, og dette har vært de største utfordringene i dette prosjektet. Høydeoppløsningen er i all hovedsak begrenset av egenskapene til linsen, der fokallengden og synsfeltet spiller en stor rolle. En høyere høydeoppløsning av modellene i figur 4.20 og 4.21 vil gi en mye mer presis gjengivelse av modelleiren. Som nevnt er også presisjonen på matchingen svært utslagsgivende for modelleringen. Ved feilmatching på kun en piksel fører dette til en forskjell på et helt nivå i 3D-modellen. Det er med andre ord i praksis ingen rom for feilmarginer ved matching. For å øke presisjonen til matchingen har søkeområdet langs kolonnene i det bildet som skal matches blitt begrenset betydelig. I figur 4.21 har også referansevinduet blitt økt vesentlig for å unngå feilmatchinger. Dette gav gode resultater, men ulempen er altså at høydene i modellen får en utjevningseffekt. Bråe høydeforskjeller i avbildningsobjektet vil da muligens ikke bli gjengitt i modellen.

I rektifikasjonsprosessen av stereobildene er det ikke nødvendig å vite de intrinsiske parameterne til kameraet. Derimot kreves det ved beregning av koordinatsystemet at PPA er kjent. Dette punktet kan kun estimeres ved geometriske kalibrering. I tillegg gir kalibrering høyere metrisk nøyaktighet.

Etter rektifikasjon blir det nye PPA-punktet beregnet ved å matche det PPA-punktet gitt ved kalibrering. I disse stereobildene har denne matchingen blitt utført korrekt, men hvis det opprinnelige PPA-punktet er i et svært homogent område kan det oppstå feilmatchingen. Siden hele koordinatsystemet er beregnet på bakgrunn av denne enkeltmatchingen vil dette få store konsekvenser ved parallakseberegningene. En alternativ måte å lokalisere det nye PPA-punktet på er kun å rektifisere det opprinnelige PPA-punktet og sammenligne endringen som oppstår grunnet rektifikasjonen. Det vil dermed være mulig å estimere hvor PPA-punktet vil befinne seg etter rektifikasjon uten å foreta matching.

## 5.2 Modelling av sår

Det overordnede målet med dette prosjektet er å modellere sår i hud. Modelleiren har kun blitt brukt til å utvikle en prototype av 3D-modellen. Derfor ble det på et senere stadium i prosjektet modellert virkelige sår fra pasienter og i tillegg sår fra en modellfot. For å maksimalisere høydeoppløsningen i modellen ble sårene avbildet med så lav høyde som mulig og med så høy basislinje som mulig uten at såret forsvant ut av synsfeltet i stereobildene. Det viste seg derimot at med den linsen som er brukt i dette prosjektet ble høydeoppløsningen i modellen for lav fordi sårene var for grunne i utgangspunktet. Gjengivelsen av sårene ble for unøyaktige til å kunne brukes til visualisering av sårene. Dybdeestimeringen av sårene fungerte derimot godt. Den maksimale målte dybden i sårene stemte overens med maksimal dybden i 3D-modellen.

I tillegg til å redusere høyde og øke basislinjen kan linsen byttes ut med en linse med høyere fokallengde for å øke dybdeoppløsningen. Dette vil øke nøyaktigheten av den visuelle gjengivelsen av såret i 3D-modellen. Problemet ved å øke fokallengden er at synsfeltet minker. Dette medfører at den maksimale basislinjen mellom stereobildene reduseres og høydeoppløsningen reduseres. En løsning for å omgå dette problemet kan være å kombinere flere bilder når stereobildene tas ved hver ende av basislinjen. Hvert stereobilde vil da bestå av en kombinasjon av flere bilder. Dette muliggjør at både fokallengden kan økes og høyden kan reduseres uten at såret er ute av stereobildene.

En annen utfordring ved avbildning av hudsår er kurvaturen rundt såret, det vil si når såret og huden rundt ikke er flatt. Dette oppstår rundt for eksempel arm eller bein. Her vil det altså være mange områder rundt såret som er ved samme høyde som dybden i såret. Disse vil også modelleres ved samme høyde som såret og kan være vanskelig å separere fra selve såret i modellen.

Når hud avbildes i små områder vil det som regel være svært homogene pikselområder i bildet. Dette vil by på utfordringer ved matching og benyttes det et for lite referansevindue vil det resultere i en vesentlig del feilmatchinger. Som nevnt vil en økning av referansevinduet føre til en midling av  $Z$ -verdiene og bråe høydeforskjeller vil ikke kunne oppdages. En mulig måte å gjøre matchingen mer presis uten å øke referansevinduet for mye er å forbehandle stereobildene for å øke kontrasten i bildene.

Ved avbildning av sår kan det hende at kameraet må tiltes slik at bildeplanet ikke ligger parallelt med planet til såret. Dette medfører at skaleringen i  $x$ - og  $y$ -retning vil være forskjellig fordi høydevariabelen i ligning 2.1 vil variere og gi to ulike ligninger. Høydemålingene må dekomponeres i to retninger, og

skaleringfaktorene må beregnes separat basert på de to høydemålingene. Denne tilten av kameraet vil også føre til at det området av planet ved såret som er nærme kamera vil modelleres høyere enn det område som er lengre unna. Det vil med andre ord si at hvis et flatt område modelleres med et tiltet kamera vil man istedenfor å få en flat 3D-modell få en trappemodell der høyden reduseres lengre utover langs y-aksen. Denne tilteeffekten bør korrigeres for eksempel ved å justere høydeberegningene i modellen slik at den tar i bruk forholdet mellom skaleringfaktorene.

Ved akvisisjon av stereobildene av modelleiren var det ingen bevegelse av avbildningsobjektet under kameratranslasjon. Dette vil ført til uønskede pikselforskyvninger, og parameterne for beregninger ved modellering ville blitt gale. Ved akvisisjon av stereobilder av sår hos levende pasienter er bevegelse av såret under kameratranslasjon nesten uunngåelig. En mulig måte å forhindre denne bevegelsen på er å bruke to kamera istedenfor ett. Kameraene vil settes i hver sin ende av basislinjen rettet mot såret, og bildene blir tatt samtidig. Dette gjør at det ikke er nødvendig med translasjon av kameraene under akvisisjon, og det oppstår heller ingen bevegelse av såret.

# 6 Konklusjon

---

Målet med dette prosjektarbeidet har vært å danne en prototype 3D-modell basert på stereoskopiske bilder. Det har vært ønsket at modellen skal kunne gjengi avbildningsobjektet så detaljert som mulig, og at det skal kunne estimeres hva den maksimale dybden i avbildningsobjektet er. Nøyaktig gjengivelse i modellen av avbildningsobjekter er som nevnt avhengig av det oppsettet som brukes. Med den linsen brukt i dette prosjektet har i praksis den maksimale høydeoppløsningen som var mulig blitt oppnådd. Høydeoppløsningen i den resulterende modellen er høyere enn den maksimale oppgitte verdien i oppgaveteksten slik at dette kravet er innfridd. Modellering av modelleiren gav de forventede restantene med det oppsettet brukt i dette prosjektet. Ved modellering av sår fra pasienter viste det seg derimot at matchingen og høydeoppløsningen ble for dårlig for en korrekt gjengivelse av såret i 3D-modellen. Det skyldes at det var for mange homogene området i bildet og at sårene var for grunne med den høydeoppløsningen som linsen gav.

Det var også en føring for oppgaven at prosesseringstiden for hele modelleringen skulle være av en rimelig karakter. Det er matchingsprosessene som opptar mesteparten av prosesseringstiden, og ved å hoppe over hver tredje piksel reduseres prosesseringstiden betraktelig. Prosesseringstiden ved matching kan også reduseres ytterligere ved bruk av parallellprosessering for å utnytte CPU'en maksimalt når store områder skal matches.

I diskusjonskapitlet ble det foreslått en del forbedringer til systemet. De viktigste forbedringene som bør implementeres relateres til økning av høydeoppløsningen og forbedring av matchingen. For å kunne øke høydeoppløsning anbefales det å kombinere flere bilder sammen til et stereobilde ved begge ender av basislinjen. Dette gjør at basislinjen kan økes uten at såret kommer ut av synsfeltet til linsen og vil bidra betraktelig til økning av høydeoppløsningen. Det anbefales videre å forbehandle stereobildene for økning av kontrast siden avbildning av hud som regel gir svært homogene pikselområder i bildet.

Dette prosjektet har kun vært en liten del av et større prosjekt som vil videreføres. Videre arbeid med dette prosjektet vil bli å integrere hyperspektral avbildning sammen med 3D-modelleringen som nevnt i innledningen. I tillegg vil det være mulig å implementere funksjonalitet for beregning av volum og areal av såret i 3D-modellen basert på resultatene i dette prosjektet. Det endelige målet med dette prosjektet bør være å implementere et system som i sanntid

modellerer såret hos pasienten samt kontinuerlig å forbedre nøyaktigheten i modelleringen.

# 7 Bibliografi

---

1. **Mikhail, E.M., Bethel, J.S. og McGlone, J.C.** *Introduction to Modern Photogrammetry*. s.l. : John Wiley & Sons, Inc., 2001. ss. 1-32, 40-43, 88-90, 188-192. 0-471-30924-9.
2. *Photogrammetric reconstruction of high-resolution surface topographies and deformable wing kinematics of tethered locusts and free-flying hoverflies.*  
**Walker, S.M., Thomas, A.L.R. og Taylor, G.K.** 33, s.l. : Journal of the Royal Society Interface, 2008, Vol. 6, ss. 351-366.
3. **Hartley, R. og Zisserman, A.** *Multiple View Geometry in Computer Vision*. s.l. : Cambridge University Press, 2003. 0-521-54051-8.
4. Wikipedia, the free encyclopedia. *Distortion (optics) - Wikipedia, the free encyclopedia*. [Internett] 12 september 2008. [Sitert: 9 april 2012.] Bilder. [http://en.wikipedia.org/wiki/Distortion\\_\(optics\)](http://en.wikipedia.org/wiki/Distortion_(optics)).
5. *Centi-pixel accurate real-time inverse distortion correction.* **de Villiers, J.P., Leuschner, F.W. og Geldenhuysb, R.** 11, Pretoria : SPIE, 2008, Vol. 7266.
6. *A Flexible New Technique for Camera Calibration.* **Zhang, Zhengyou.** 11, Redmond : IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, Vol. 22, ss. 1330–1334.
7. **Boyle, R., M., Sonka og V., Hlavac.** *Image processing, analysis and machine vision*. 3rd. s.l. : Thomson Learning, 2008.
8. **Gonzalez, R. C. og Woods, R.E.** *Digital Image Processing*. s.l. : Pearson Prentice Hall, 2008. ss. 692-738.
9. *A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry.* **Zhang, Z., et al., et al.** 1-2, Cedex : Artificial Intelligence - Special volume on computer vision, 1995, Vol. 78, ss. 87-119. 0249-6399.
10. Locate a template in an image - Simulink. *MathWorks Nordic - MATLAB and Simulink for Technical Computing*. [Internett] MathWorks. [Sitert: 30 Juni 2012.] <http://www.mathworks.se/help/toolbox/vision/ref/templatematching.html>.
11. **Ortiz, Allan.** File:Image rectification.png - Wikipedia, the free encyclopedia. *Wikipedia, the free encyclopedia*. [Internett] 11 juni 2008. [Sitert: 5 juli 2012.] Bilde. [http://en.wikipedia.org/wiki/File:Image\\_rectification.png](http://en.wikipedia.org/wiki/File:Image_rectification.png).

12. *Rectification for Any Epipolar Geometry*. **Oram, Daniel**. Manchester : s.n., 2001.
13. *A compact algorithm for rectification of stereo pairs*. **Fusiello, A., Trucco, E. og Verri, A.** 1, s.l. : Machine Vision and Applications, 2000, Vol. 12, ss. 16-22.
14. EO USB 2.0 CMOS Machine Vision Cameras | Edmund Optics. *Optics, Imaging, and Photonics Technology / Edmund Optics*. [Internett] Edmund Optics. [Siteret: 10 juli 2012.] <http://www.edmundoptics.com/imaging/cameras/usb-cameras/eo-usb-2-0-cmos-machine-vision-cameras/2818>.
15. Compact Fixed Focal Length Lenses | Edmund Optics. *Optics, Imaging, and Photonics Technology / Edmund Optics*. [Internett] Edmund Optics. [Siteret: 10 juli 2012.] <http://www.edmundoptics.com/imaging/imaging-lenses/edmund-optics-designed-lenses/compact-fixed-focal-length-lenses/3070>.
16. **Strobl, K. H., et al., et al.** *DLR CalDe and DLR Callab*. [Programvare] Oberpfaffenhofen : Institute of Robotics and Mechatronics, German Aerospace Center (DLR), 2005. url: <http://www.robotic.dlr.de/callab/>.
17. *More Accurate Pinhole Camera Calibration with Imperfect Planar Target*. **Strobl, K. H. og Hirzinger, G.** 1068 - 1075 , German Aerospace Center (DLR) : IEEE, 2011. 978-1-4673-0062-9.
18. DLR - Institut für Robotik und Mechatronik - Brief Tutorials. *DLR - Institut für Robotik und Mechatronik*. [Internett] DLR. [Siteret: 15 juni 2012.] [http://www.dlr.de/rm/desktopdefault.aspx/tabid-3925/6084\\_read-9196/](http://www.dlr.de/rm/desktopdefault.aspx/tabid-3925/6084_read-9196/) .
19. Interpolate scattered data - MATLAB. *MathWorks Nordic - MATLAB and Simulink for Technical Computing*. [Internett] MathWorks. [Siteret: 11 juni 2012.] <http://www.mathworks.se/help/techdoc/ref/triscatteredinterpclass.html>.
20. Delaunay\_circumcircles.png. *Wikipedia, the free encyclopedia*. [Internett] [Siteret: 15 juni 2012.] Bilde. [http://upload.wikimedia.org/wikipedia/commons/c/c9/Delaunay\\_circumcircles.png](http://upload.wikimedia.org/wikipedia/commons/c/c9/Delaunay_circumcircles.png).
21. *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*. **Fischler, M.A. og Bolles, R.C.** 6, s.l. : Association for Computing Machinery, 1981, Vol. 24, ss. 381 - 395.

# 8 Vedlegg

---

## 8.1 Kodeoversikt

Her gis det en oversikt over de Matlab-scriptene som blir eller har blitt brukt i prosjektet.

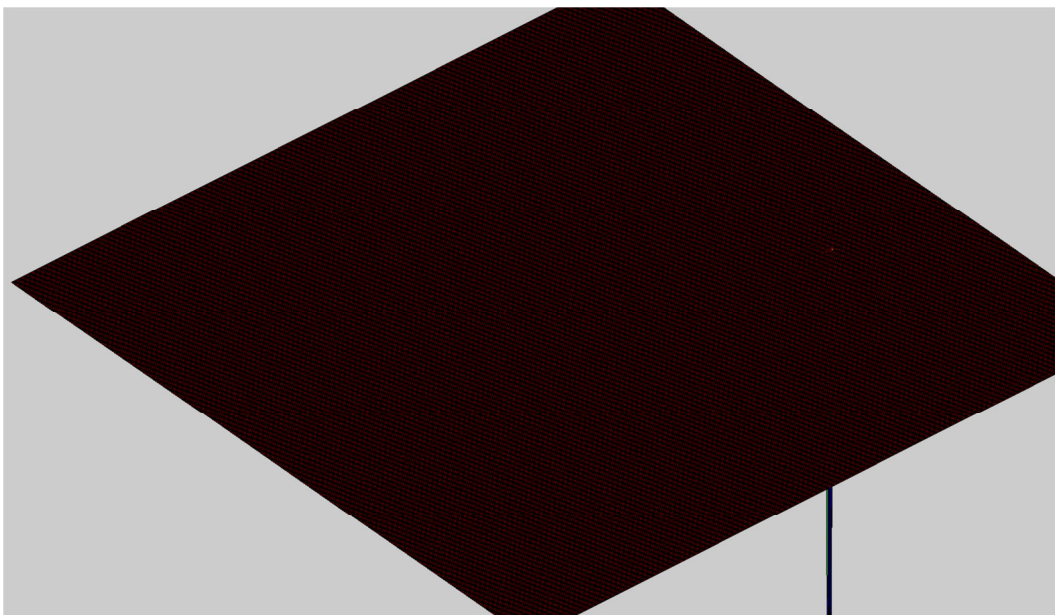
Scriptnavn	Beskrivelse
area_matching.m	Gjør korrelasjonsmatching på hele områder i bildet. Hopper hver tredje piksel for hver match.
area_matching_noskip.m	Gjør korrelasjonsmatching på hele områder i bildet uten å hoppe over punkter.
camera_calibration_callab_matlab.m	Gir resultatene fra kalibrering
correlation.m	Gjør det samme som <i>area_matching</i> , men uten å bruke <i>normxcorr2</i> . Beregner krysskorrelasjonen manuelt. Kan benyttes når stereobildene ikke har epipolar orientering.
correlation_epi.m	Gjør det samme som <i>correlation</i> , men antar epipolar orientering.
make_model.m	Modellerer 3D-modellen basert på parallakseverdiene gitt av matchingen.
point_matching.m	Gjør korrelasjonsmatching for et enkelt punkt. Brukes i <i>area_matching</i> og <i>area_matching_noskip</i>
rectify_images.m	Rektifiserer stereobildene til epipolar orientering.
undistorion.m	Bruker kalibreringsresultatene til å korrigere stereobildene for radiell linsedistorsjon.
verify_cal.m	Evaluerer hvor godt kalibreringen har korrigert for metrisk unøyaktighet.



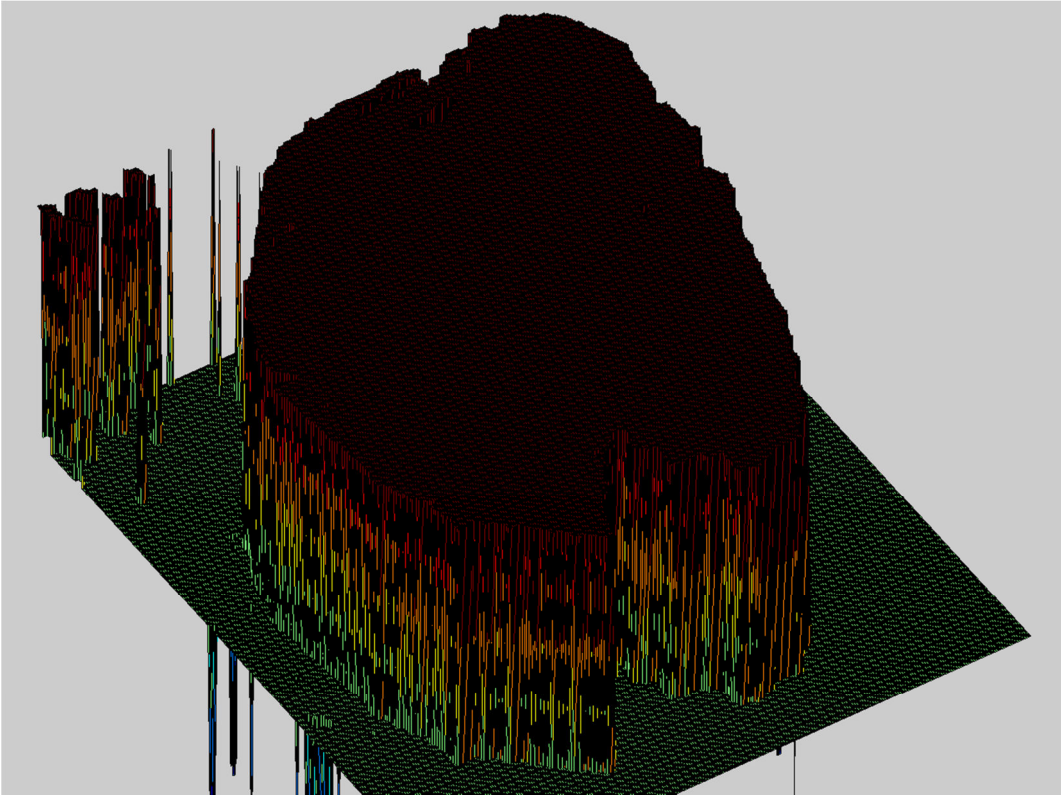
## 8.2 Diverse modeller

Figurene nedenfor viser forskjellige 3D-modeller basert på forskjellige avstander i basislinjen. Alle figurene er basert på de 12 bildene som ble tatt av modelleiren. Den siste figuren viser modellen som gis fra scriptet *area\_matching\_noskip.m* der det ikke foretas noe utelatelse av piksler ved matching. Dette er for å gi muligheten til sammenligne eventuelle endringer som oppstår i modellen når det ikke hoppes over piksler ved matching. Denne modellen er basert på de samme stereobildene som i figur 4.20 og 4.21, det vil si 10 cm lang basislinje.

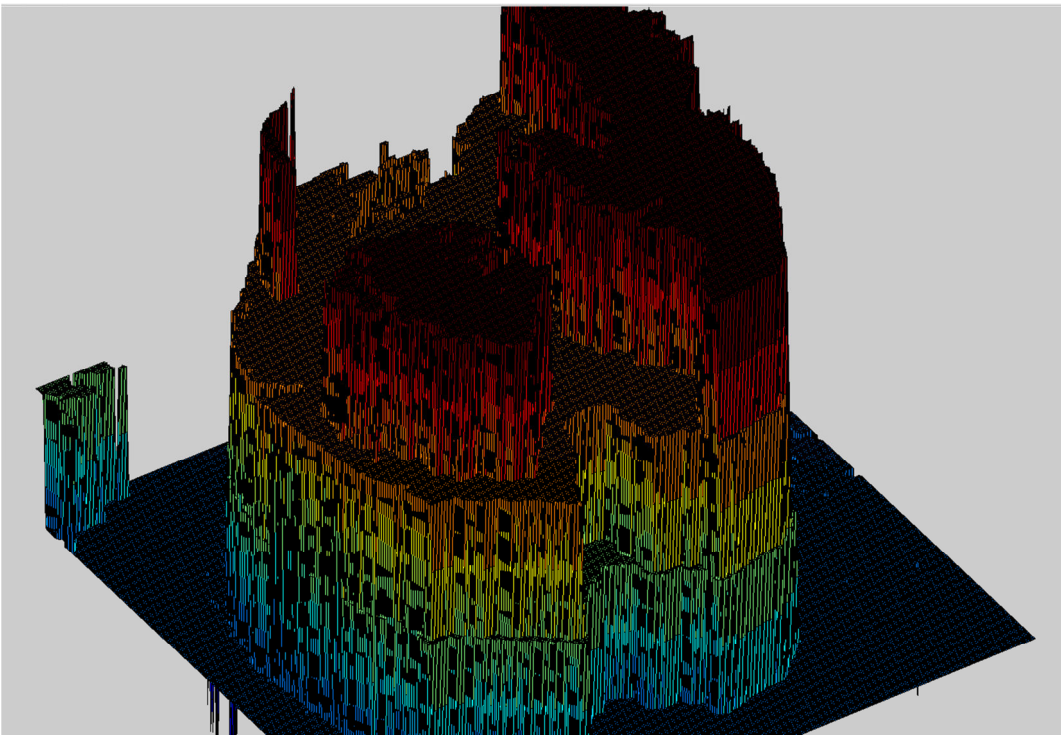
2 cm basislinje:



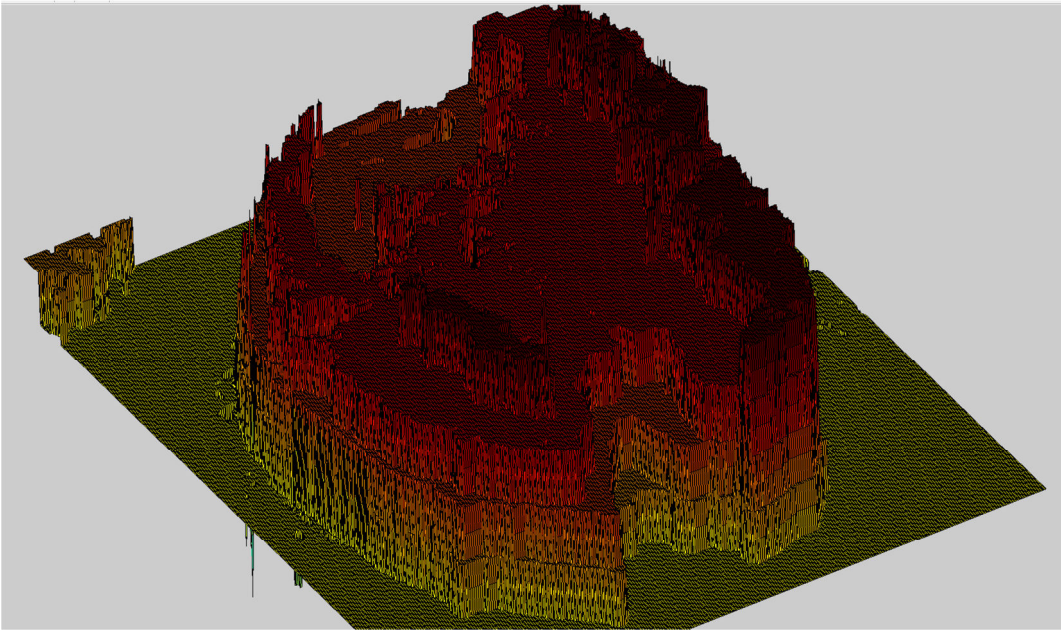
3 cm basislinje:



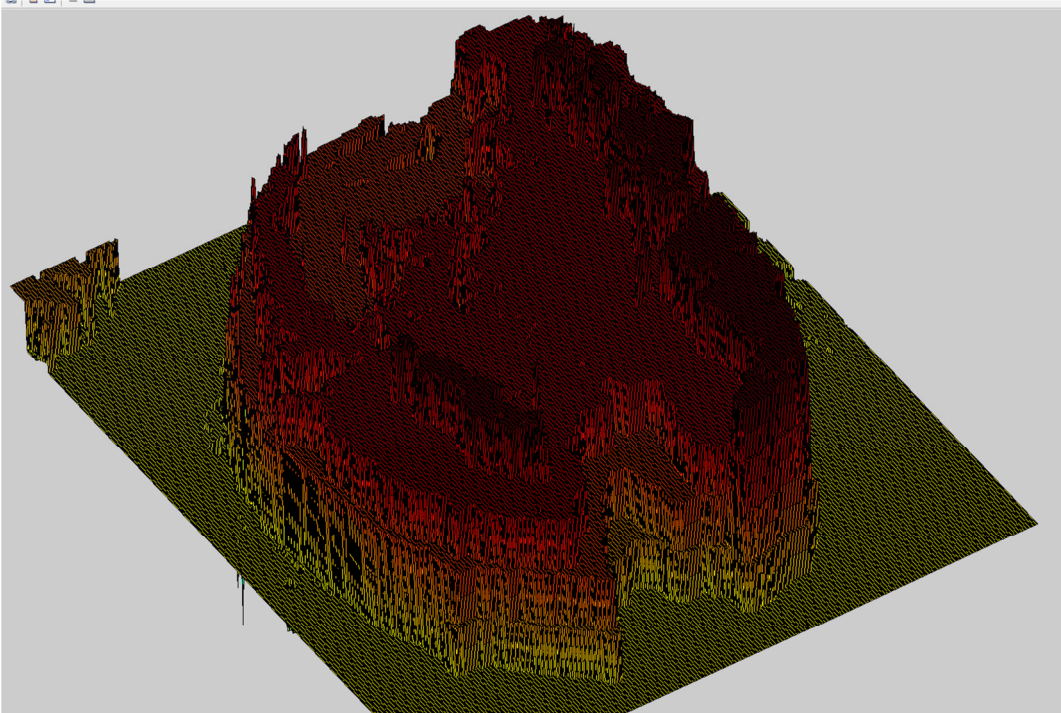
5 cm basislinje:



6 cm basislinje:

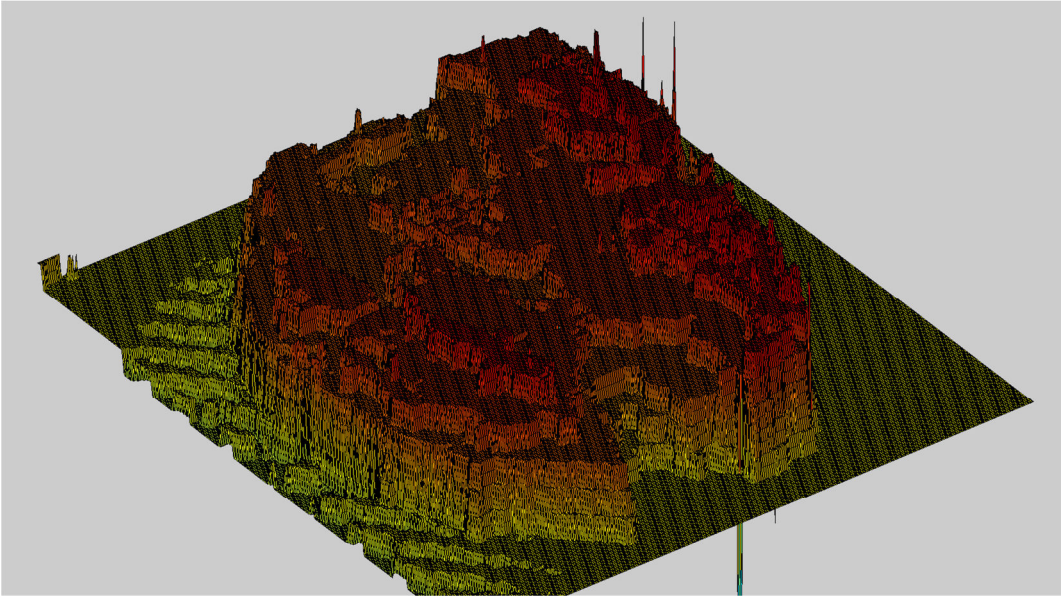


7 cm basislinje:

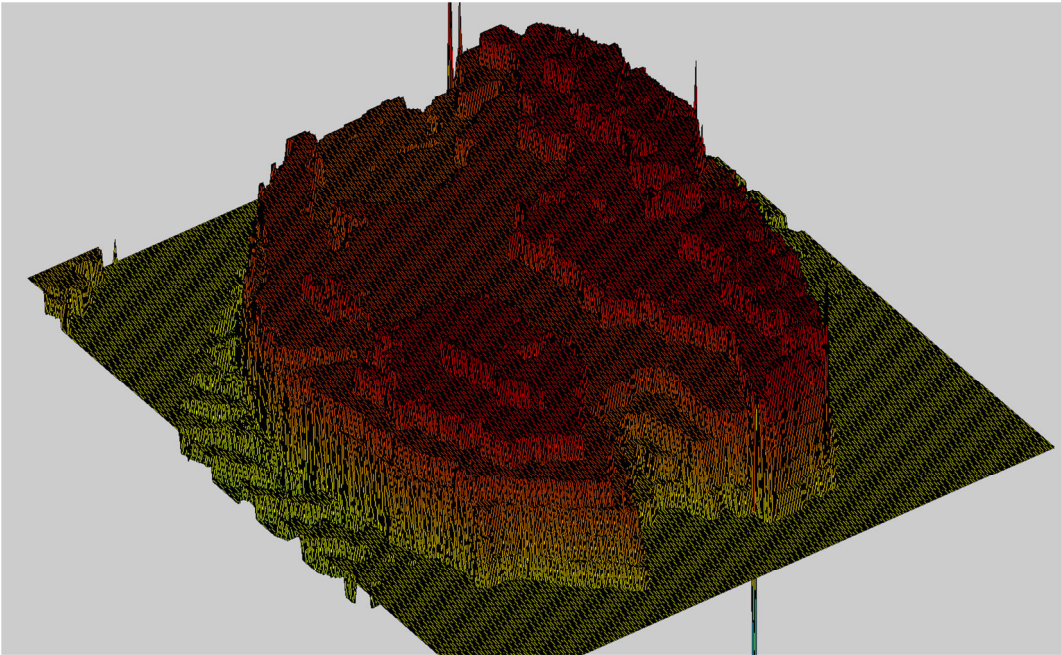




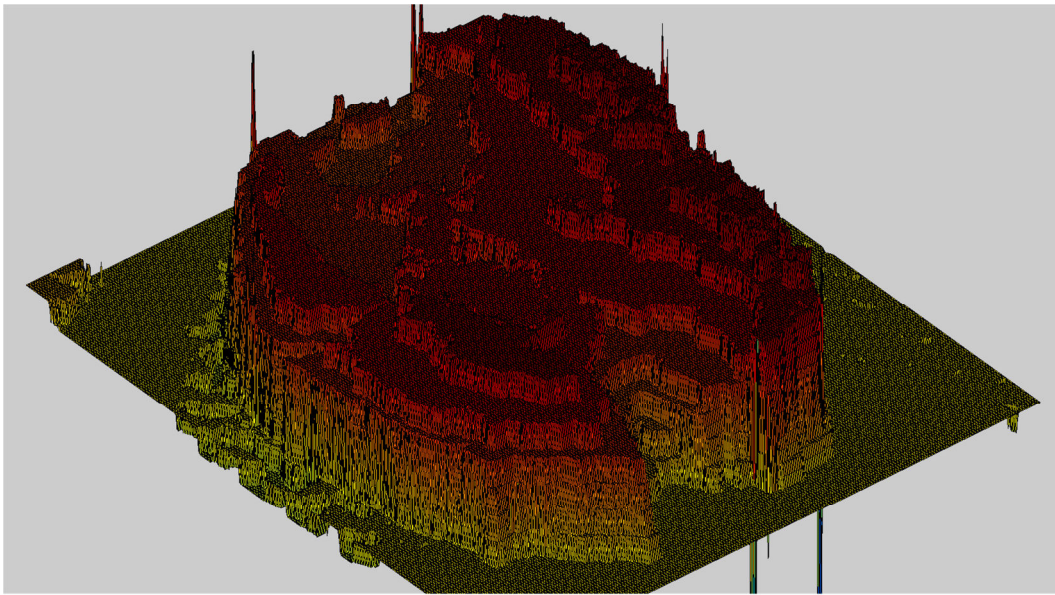
8 cm basislinje:



9 cm basislinje:



10 cm basislinje:



10 cm basislinje, matching av alle piksler:

