2021•2022

Faculteit Industriële Ingenieurswetenschappen

master in de industriële wetenschappen: elektronica-ICT

Masterthesis

Ontwikkeling van een 3D-positiebepalingssysteem op basis van stereovisie met evaluatie op veiligheidshelmdetectie

PROMOTOR :

Prof. dr. ir. Eric DEMEESTER

PROMOTOR :

ing. Niels VANDEKERKHOF

BEGELEIDER :

Mevrouw Yanming WU

Arne Niesten

Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: elektronica-ICT

Gezamenlijke opleiding UHasselt en KU Leuven

UHASSELT

KU LEUVEN

UHASSELT

KU LEUVEN

2021•2022

Faculteit Industriële Ingenieurswetenschappen

master in de industriële wetenschappen: elektronica-ICT

Masterthesis

Ontwikkeling van een 3D-positiebepalingssysteem op basis van stereovisie met evaluatie op veiligheidshelmdetectie

PROMOTOR :

Prof. dr. ir. Eric DEMEESTER

PROMOTOR :

ing. Niels VANDEKERKHOF

BEGELEIDER :

Mevrouw Yanming WU

Arne Niesten

Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: elektronica-ICT

Woord vooraf

Deze masterproef betekent voor mij het einde van mijn masteropleiding tot Industrieel Ingenieur in de Elektronica-ICT aan de Universiteit van Hasselt en de KU Leuven. Tijdens deze opleiding heb ik enorm veel kennis kunnen opdoen en ben ik tot belangrijke inzichten gekomen. Hoewel het niet altijd even makkelijk was, ben ik tevreden met wat ik gedurende deze periode bereikt en bijgeleerd heb. Ook ben ik blij met de mogelijkheid die Aperam Genk mij geboden heeft om mijn opleiding, met een zeer leerrijke en interessante masterproef, te kunnen afsluiten.

Tijdens mijn masterproef bij Aperam Genk heb ik veel ervaring mogen opdoen en heb ik veel bijgeleerd. Het was fijn om eens te mogen proeven van het bedrijfsleven. Zowel op professioneel als op persoonlijk vlak heeft dit mij enorm veel bijgebracht. Kennis en ervaring die ik heel mijn leven met mij mee zal dragen.

Deze masterproef met bijhorende scriptie is tot stand gekomen dankzij de hulp en ondersteuning van meerdere personen, die ik graag zou bedanken.

Om te beginnen zou ik graag mijn externe promotor en begeleider binnen Aperam Genk, Ing. Niels Vandekerkhof willen bedanken voor zijn goede begeleiding en alle mogelijkheden die ik van hem gekregen heb. Verder wil ik ook mijn interne promotor Prof. Dr. Ir. Eric Demeester en mijn interne begeleider PhD. Ing. Yanming Wu bedanken voor hun begeleiding en belangrijke inzichten binnen deze masterproef. Ook zou ik graag Prof. Bart Dreesen willen bedanken voor zijn bijdragen op het vlak van taal en communicatie.

Verder zou ik ook graag een woord van dank uitbrengen naar het hele team binnen Aperam Genk voor alle leerrijk, maar vooral ook leuke momenten en alle ondersteuning die ik van hun gekregen heb. Bedankt voor alle inzichten, conversaties, tips, paraatheid en nog veel meer.

Om af te sluiten zou ik ook graag mijn familie, vriendin en al mijn vrienden willen bedanken voor alle steun gedurende de hele opleiding en mijn uiteindelijke masterproef. Jullie hebben me altijd op alle vlakke ondersteund en stonden altijd voor mij klaar. Dit waardeer ik enorm.

Arne Niesten

Gingelom, 7 juni 2022.

Inhoud

Woord vooraf 1

Lijst van tabellen 5

Lijst van figuren 7

Afkortingen 11

Abstract 13

Abstract (English) 15

1 Introductie 17

1.1 Situering 17

1.1.1 Aperam (Genk) 17

1.1.2 Principe veiligheidskegel 18

1.2 Probleemstelling 19

1.3 Doelstellingen 19

1.3.1 Verduidelijking concept 20

1.4 Overzicht scriptie 21

2 Literatuurstudie 23

2.1 Methodes afstandsbepaling 23

2.2 Stereovisie afstandsbepaling 24

2.2.1 Stereo-opname 26

2.2.2 Beeldverwerking en detectie 29

2.2.3 Stereo matching 40

2.2.4 Afstandsbepaling 44

3 Ontwikkeling 49

3.1 Hardware 49

3.2 Opstelling 50

3.2.1 Camera uitlijning en positionering 52

3.3 Vastleggen stereobeelden 54

3.3.1 Camera instellingen 56

3.4 Stereo kalibratie en rectificatie 58

3.4.1 Reden 58

3.4.2 Toepassing 60

3.5 Beeldverwerking 71
3.6 Objectdetectie 74
3.6.1 Detectieobject 74
3.6.2 Objectdetectiemodel (YOLOv5s) 74
3.6.3 Ontwikkeling objectdetectiemodel 75
3.6.4 Toepassing en gebruik YOLOv5s-helmdetectiemodel 81
3.6.5 Bepaling centrumcoördinaat 84
3.7 3D-positiebepaling 85
3.7.1 Verduidelijking 3D-positie 85
3.7.2 Bepaling 3D-positie 87
3.7.3 Bepaling Z-coördinaat (alternatieve methode) 88

4 Verificatie ontwikkeld 3D-positiebepalingssysteem 89
4.1 Omschrijving testen 89
4.1.1 Testlocatie 89
4.1.2 Testopstelling 90
4.1.3 Testmethode 92
4.2 Registratie en verwerking testresultaten 97
4.2.1 Registratie testresultaten 97
4.2.2 Verwerking testresultaten 97
4.3 Resultaten 99
4.3.1 Resultaten diepte (Z): 1 meter 99
4.3.2 Resultaten diepte (Z): 2 meter 101
4.3.3 Resultaten diepte (Z): 3 meter 103
4.3.4 Resultaten diepte (Z): 4 meter 105
4.3.5 Resultaten diepte (Z): 4,9 meter 107
4.3.6 Algemene bespreking resultaten 109
4.4 Alternatieve testen 110
4.4.1 Video rechte lijn 110
4.4.2 Video beweging 111
4.4.3 Extreme posities 112

5 Conclusie 115

6 Bibliografie 117

Lijst van tabellen

Tabel 1 Weergave tabel resultaten 97
Tabel 2 Weergave tabel resultaten inclusief verwerking (oranje aanduidingen 98
Tabel 3 Resultaten (diepte (Z) 1 meter) 99
Tabel 4 Resultaten (diepte (Z) 2 meter) 101
Tabel 5 Resultaten (diepte (Z) 3 meter) 103
Tabel 6 Resultaten (diepte (Z) 4 meter) 105
Tabel 7 Resultaten (diepte (Z) 4,9 meter) 107
Tabel 8 Maximale afwijkingen in centimeter per coördinaat, binnen de verwerkte resultaten, ten opzichte van de werkelijk coördinaten (positities)..... 109

Lijst van figuren

Figuur 1 Principe veiligheidskegel (2D) 18
Figuur 2 Plaatsing IP-camera's 19
Figuur 3 Verduidelijking concept veiligheidskegel en positiebepaling 20
Figuur 4 Voorbeeld kalibratiepatroon [4] 23
Figuur 5 Basisprincipe stereovisie [7] 25
Figuur 6 Algemene structuur stereo afstandsbepaling 25
Figuur 7 Verduidelijking dispariteit [5] 25
Figuur 8 Verduidelijking cameraposities aan de hand van eigenschappen optische assen 26
Figuur 9 Voorstelling standaard schermparallax 27
Figuur 10 Voorstelling positieve schermparallax 27
Figuur 11 Voorstelling negatieve schermparallax 28
Figuur 12 Juiste uitlijning van de camera's (a) en uitlijning met verticale fouten (b) en (c) [6] 29
Figuur 13 Voorbeelden 'preprocessing' toegepast op een afbeelding van een insect [10] 30
Figuur 14 Voorbeeld annotatiebestand voor trainingsdataset objectdetectie [14] 31
Figuur 15 Visuele weergave resultaat 'Logistic regression' toegepast op een 'supervised' dataset [15] 32
Figuur 16 Visuele voorstelling 'unsupervised' machine learning dataset onderverdeeld in drie clusters (rood, groen, blauw) [15] 32
Figuur 17 Enkele toepassingen machine learning [16] 34
Figuur 18 Verduidelijking ligging deep learning binnen het gebied van machine learning en artificiële intelligentie [16] 34
Figuur 19 Weergave hiërarchie verschillende lagen CNN [14] 35
Figuur 20 Voorstelling architectuur convolutioneel neuraal netwerk [14] 36
Figuur 21 R-CNN objectdetectie systeem overzicht [14], [18] 37
Figuur 22 Voorbeelden objectdetectie aan de hand van R-CNN [18] 37
Figuur 23 Fast R-CNN overzicht [14], [19] 38
Figuur 24 Faster R-CNN architectuur [14], [20] 38
Figuur 25 YOLO architectuur [14], [21] 39
Figuur 26 Voorbeeld 'bounding box' rondom auto [3] 40
Figuur 27 Gedetecteerd object op linker afbeelding (rood) en zoekgebied op de rechter afbeelding (groen) [6] 42
Figuur 28 Bepaling submatrix in rechter afbeelding [6] 42
Figuur 29 Verschil tussen gedetecteerd object op de linker afbeeldingen met submatrix op de rechter afbeeldingen [6] 43
Figuur 30 Berekening horizontale afstand tussen linker en rechter afbeelding [6] 43
Figuur 31 Foto van een object (boom) vastgelegd met twee camera's [6] 44
Figuur 32 Foto van het object (boom) genomen met de linker camera [6] 45
Figuur 33 Foto van het object (boom) genomen met de rechter camera [6] 46

Figuur 34 Afstandsfout veroorzaakt door een plaatsingsfout van 1 pixel [6] 47
Figuur 35 Hikvision DS-2CD2643G2-IZS IP-camera [25] 49
Figuur 36 Cisco CBS250-8PP-D netwerk switch [26] 50
Figuur 37 Infrastructuur rolbrug 50
Figuur 38 Schematische voorstelling opstelling systeem 51
Figuur 39 Stereovisie cameraopstelling 51
Figuur 40 Correct voorbeeld camera-/beelduitlijning [27] 52
Figuur 41 Verduidelijking stereovisie vlakken (pinhole cameramodel) [27] 53
Figuur 42 Paar stereobeelden 54
Figuur 43 Verduidelijking (beeld-)coördinatenstelsel binnen Python 54
Figuur 44 Verduidelijking framesnelheid (fps) [31] 56
Figuur 45 Verduidelijking frame synchronisatie probleem (asynchrone frames) [30] 56
Figuur 46 Enkel veel voorkomende beeldresoluties met bijhorende naam [32] 57
Figuur 47 Verduidelijking brandpuntsafstand [33] 58
Figuur 48 Twee veelvoorkomende types beeldvervorming (links: geen vervorming; midden: negatieve vervorming; rechts: positieve vervorming) [27] 59
Figuur 49 Origineel beeld Hikvision IP-camera, zonder beeldvervorming 59
Figuur 50 Kalibratie patroon [36] 61
Figuur 51 Stereobeelden kalibratie patroon van opstelling 62
Figuur 52 stereobeeld met aanduiding interne hoeken kalibratie patroon (kleuren) 63
Figuur 53 Intrinsieke camera matrix [27] 64
Figuur 54 Berekening representatie punt binnen het coördinatenstelsel van de 2de camera, afhankelijk van de coördinaten van het punt binnen het coördinatenstelsel van de 1ste camera.[27] 65
Figuur 55 rotatie en translatie vector opstelling masterthesis 65
Figuur 56 Verduidelijking stereo rectificatie met ‘0’ als alpha (groene kader) en epipolaire lijnen verticaal uitgelijnd [27] 66
Figuur 57 Theoretisch ‘gerectificeerde camera-/projectiematrices (horizontale stereo variant) [27] 68
Figuur 58 Projectiematrices opstelling masterthesis 68
Figuur 59 Aanduiding gerectificeerd projectiecoördinatenstelsel (groen) binnen beeldcoördinatenstelsel (geel) 69
Figuur 60 Schematische weergave beeldverwerking voor een paar stereobeelden [35] (aangepast) 72
Figuur 61 Toepassing 'remap' functie, met eerder bepaalde parameters, ter vervorming, rectificatie en translatie van beide stereo beelden 73
Figuur 62 Horizontale epipolaire lijnen (geel) aangebracht op de vervormde, gerectificeerde en getransformeerde stereobeelden 73
Figuur 63 Veiligheidshelm, gebruikt voor ontwikkeling detectiemodel 74
Figuur 64 Collage afbeeldingen training en validatie YOLOv5s-helmdetectiemodel 75
Figuur 65 Collage geannoteerde en gelabelde afbeeldingen YOLOv5s-helmdetectiemodel .. 76
Figuur 66 Collage geannoteerde en gelabelde afbeeldingen YOLOv5s-helmdetectiemodel met effecten en veranderingen 77

Figuur 67 Collage YOLOv5s-helmdetectiemodeltoegepast op validatie dataset 80
Figuur 68 Verduidelijking hoekcoördinaten bounding box 82
Figuur 69 Voorbeeld output plot_boxes functie (bounding box, rondom helm, met bijhorend label (tekst 'HELMET') in het groen) (ID-detectie ('0') weergegeven in het blauw)..... 83
Figuur 70 Verduidelijking centrumcoördinaat bounding box..... 84
Figuur 71 Verduidelijking vlakken en coördinaten [45](aangepast) 86
Figuur 72 Testlocatie..... 90
Figuur 73 Stereovisie cameraopstelling 90
Figuur 74 Schematische voorstelling stereovisie opstelling 91
Figuur 75 Testopstelling helmen..... 91
Figuur 76 Visuele verduidelijking locaties (rode stippen) testopstelling (veiligheidshelmen), met aanduiding afstanden binnen de ruimte..... 93
Figuur 77 Verduidelijking positionering 1 (Diepte: 1 meter, horizontale verplaatsing: rechts (32 cm)) 95
Figuur 78 Verduidelijking positionering 2 (Diepte: 2 meter, horizontale verplaatsing: geen (midden))..... 95
Figuur 79 Verduidelijking positionering 3 (Diepte: 5 meter, horizontale verplaatsing: rechts (200 cm))..... 95
Figuur 80 Gebruikte lasterscanner voor nameting en verificatie afstanden en posities..... 96
Figuur 81 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 100cm)..... 99
Figuur 82 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 100cm)..... 100
Figuur 83 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 200cm)..... 101
Figuur 84 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 200cm)..... 101
Figuur 85 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 300cm)..... 103
Figuur 86 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 300cm)..... 103
Figuur 87 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 400cm)..... 105
Figuur 88 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 400cm)..... 105
Figuur 89 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 490cm)..... 107
Figuur 90 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 490cm)..... 107
Figuur 91 Verloop coördinaten (X, Y, Z) tijdens verplaatsing helm in Z-richting met vaste pauzes elke 100cm 111
Figuur 92 Verloop coördinaten (X, Y, Z) tijdens verplaatsing helm in Z-richting met vaste pauzes elke 100 cm en bewegingen over een variabele afstand in de X-richting..... 112

Afkortingen

IP	Internet Protocol
2D	Tweedimensionaal
3D	Driedimensionaal
PoE	Power over Ethernet
RTSP	Real Time Streaming Protocol
RTT	Real Time Transport
FPS	Frames Per Seconde

Abstract

Deze masterproef is uitgevoerd in samenwerking met het Services & Solutions Center van Aperam Genk, dat rollen staal verwerkt conform klanteneisen. Deze rollen worden binnen de productiehal getransporteerd met rolbruggen. Om dit transport zo veilig mogelijk uit te voeren, wordt het principe toegepast van een veiligheidskegel, een gevarenzone in de vorm van een kegel onder de kraanhaak. Deze masterproef beoogt de basis te leggen voor de digitalisering van dit veiligheidshulpmiddel aan de hand van 2 IP-camera's en 3D-positiebepaling.

In een eerste fase zijn de mogelijkheden van stereovisie voor deze toepassing onderzocht. Deze technologie laat toe 3D-informatie uit digitale beelden te extraheren. Met een opstelling bestaande uit twee IP-camera's wordt, na uitvoering van een kalibratiestap met de nodige beeldvervorming en verticale uitlijning, een YOLOv5s-detectiemodel toegepast. Dit om op basis van de locaties van identieke gedetecteerde objecten op beide beelden, in combinatie met eerder bepaalde parameters uit de kalibratiestap, de 3D-posities van objecten binnen de werkelijke ruimte te bepalen.

Testen die uitgevoerd werden ter evaluatie en verificatie van de werking van het 3D-positiebepalingsssyteem, geven slechts beperkte afwijkingen weer. Ten opzichte van de werkelijke posities, zijn de bekomen afwijkingen echter slechts beperkt en blijven ze binnen de door Aperam bepaalde grenzen. Er kan geconcludeerd worden dat het mogelijk is om 3D-posities van objecten voldoende nauwkeurig te bepalen aan de hand van 2 camera's.

Abstract (English)

This master’s thesis was carried out in cooperation with Aperam Genk's Services & Solutions Center, which processes steel coils in accordance with customer requirements. These coils are transported within the production hall using overhead cranes. In order to carry out this transport safely, the principle of a safety cone, a danger zone in the shape of a cone, is applied under the crane hook. This master thesis aims to lay the foundation for the digitalisation of this system using 2 IP cameras and 3D positioning.

In a first phase, the possibilities of stereo vision for this application are investigated. This technology allows 3D information to be extracted from digital images. With a setup consisting of two IP cameras, after performing a calibration step with the necessary image distortion and vertical alignment, a YOLOv5s detection model is applied. This is to determine the 3D positions of objects within real space based on the locations of identical detected objects on both images, in combination with previously determined parameters from the calibration step.

Tests carried out to evaluate and verify the operation of the 3D positioning system show only limited deviations. However, compared to the actual positions, the deviations obtained are limited and remain within the limits defined by Aperam. It can be concluded that it is possible to determine 3D positions of objects with sufficient accuracy using 2 cameras.

1 Introductie

1.1 Situering

Deze masterproef is uitgewerkt in samenwerking met het Service & Solutions Center van Aperam Genk. Hier worden grote rollen roestvast staal van tot wel 27 ton, geproduceerd door Aperam zelf, verwerkt conform eisen van klanten.

Binnen deze productiehal worden de rollen staal getransporteerd door middel van rolbruggen. Om dit transport zo veilig mogelijk uit te voeren wordt als veiligheidshulpmiddel het principe van een veiligheidskegel onder de kraanhaak toegepast. De te bewaren veiligheidsafstand hangt hierbij recht evenredig af van de relatieve hoogte tussen kraanhaak, met eventuele lading, en een persoon.

Aperam wenst dit systeem te digitaliseren om personen binnen de gevarenzone te waarschuwen. Hiervoor is het belangrijk de relatieve positie van een persoon binnen de 3D-ruimte te bepalen ten opzichte van de virtuele 3D-veiligeidskegel.

1.1.1 Aperam (Genk)

Aperam, een jonge beursgenoteerde onderneming die ontstond in 2011 uit een spin-off van Arcelor Mittal, is een wereldleider op het gebied van roestvast staal. Naast een site in Genk zijn er ook productiesites in Frankrijk en Brazilië. Hun Service & Solutions Centers zijn verdeeld over de hele wereld, met klanten verspreid over 20 verschillende landen. Voornamelijk in Europa, maar ook in Azië en Noord- en Zuid-Amerika. De site in Genk is één van de grootste industriële bedrijven in Limburg [1].

Het duurzame product wordt voor verschillende doeleinden gebruikt. Het gaat dan over industriële toepassingen zoals silo's, tankwagens, grote vaten, buizen of trappen gemaakt uit roestvast staal. In huishoudelijke toepassingen is roestvast staal eveneens alomtegenwoordig. Hier gaat het dan bijvoorbeeld over de trommel van een wasmachine, een koelkast of bestek. Zelfs de bollen van het Atomium in Brussel en het Belgische onderzoekscentrum op de Zuidpool zijn gemaakt uit roestvast staal [1].

In Genk worden roestvaste stalen platen en rollen geproduceerd. Hiervoor recycleert Aperam Genk schroot van roestvast staal, komend uit heel Europa. Roestvast staal is dan ook oneindig recycleerbaar. Het staal wordt gegoten in grote slabs, waarvan het gewicht kan oplopen tot wel 35 ton. Hierna wordt het per trein getransporteerd naar het zusterbedrijf in Châtelet, waar het doormiddel van een warmwals in dikte wordt uitgerold. Eenmaal terug in Genk wordt het roestvast staal verder verwerkt in de koudwalserij. Op vraag van klanten wordt het roestvaste staal op maat versneden tot kleinere rollen of platen [1].

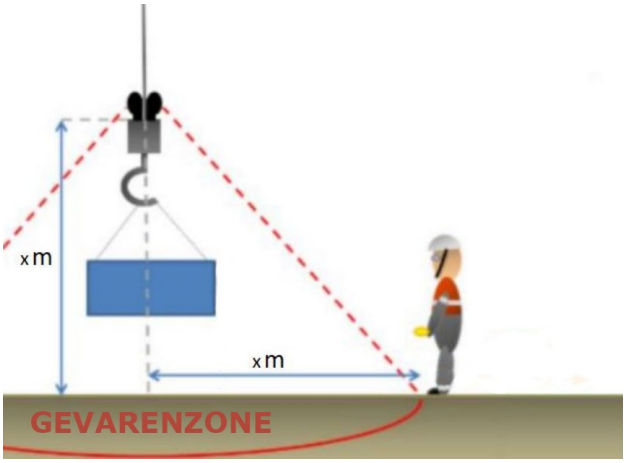
Binnen Aperam spelen de veiligheid en gezondheid van hun medewerkers, maar ook hun buren een belangrijke rol. Hier wordt dan ook heel wat aandacht aan besteed. Aperam is de groenste producent van roestvrij staal in de wereld op het gebied van CO_2-uitstoot. Er wordt zeer veel geïnvesteerd om het modernste en meest milieuvriendelijke staalbedrijf te zijn en te blijven. Hiernaast gebeuren er ook zeer veel investeringen op het gebied van automatisering, nieuwe technologieën en digitalisering. Zo wordt er standaard gebruik gemaakt van robots en volautomatisch gestuurde installaties in het productieproces en labo's. Met behulp van nieuwe wiskundige modellen en een doorgedreven samenwerking met befaamde universiteiten wordt het productieproces continu verbeterd. Enkele investeringen in nieuwe, moderne lijnen zijn hiervan het bewijs [1].

De grootste uitdaging voor Aperam is de toekomst van roestvast staal in Europa te vrijwaren doormiddel van performante fabrieken, innovaties, duurzaamheid en vooral de gezondheid en veiligheid van de medewerkers en stakeholders te garanderen. Binnen Aperam Genk werken er meer dan 1100 medewerkers verspreid over verschillende afdelingen en diensten [1].

Deze masterproef situeert zich in de organisatie van Aperam Genk binnen de 'Industrial Innovation' groep.

1.1.2 Principe veiligheidskegel

Als hulpmiddel voor het zo veilig mogelijk uitvoeren van transport aan de hand van rolbruggen wordt er, zoals reeds eerder vermeld, binnen Aperam gebruik gemaakt van het principe van een veiligheidskegel onder de kraanhaak. Dit is een 3D-gevarenzone in de vorm van een kegel, die de minimaal te bewaren veiligheidsafstand aangeeft. Het principe wordt visueel verduidelijkt op Figuur 1. Hierop is duidelijk te zien dat de straal van de gevarenzone, en dus ook de minimaal te bewaren veiligheidsafstand, recht evenredig afhangt van de relatieve hoogte tussen kraanhaak en persoon.



Figuur 1 Principe veiligheidskegel (2D)

De persoon afgebeeld op Figuur 1 bevindt zich buiten de gevarenzone en staat dus ook verder dan de minimale te bewaren veiligheidsafstand. Volgens het principe van de veiligheidskegel zou deze persoon dus (theoretisch) veilig zijn voor de risico's veroorzaakt door transport met een rolbrug.

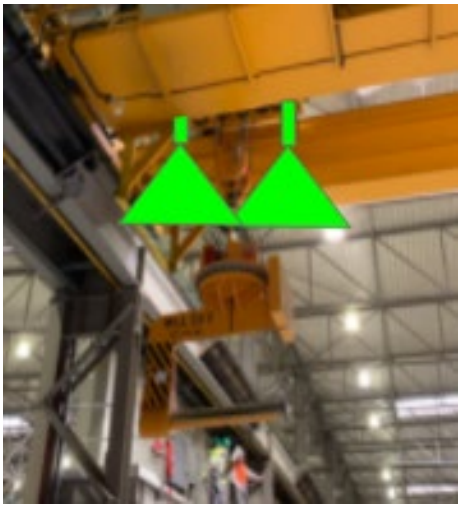
Het is belangrijk om te vermelden dat het principe van de veiligheidskegel geen absoluut betrouwbare veiligheid garandeert. Het is slechts een hulpmiddel ter verbetering van de veiligheid.

1.2 Probleemstelling

Om het transport met rolbrug op een zo veilig mogelijke manier uit te kunnen voeren, maakt Aperam als veiligheidshulpmiddel gebruik van het principe van een veiligheidskegel onder de kraanhaak, zoals reeds verduidelijkt in deel 1.1.2. De minimaal te bewaren veiligheidsafstand hangt hierbij recht evenredig af van het relatieve hoogteverschil tussen de kraanhaak en een persoon.

Momenteel dienen personen zelf een inschatting te maken van hun minimaal te bewaren veiligheidsafstand ten opzichte van de kraanhaak van een rolbrug, met eventuele lading. Aperam wenst dit systeem nu te digitaliseren om automatisch personen binnen de gevarenzone te waarschuwen. Het bieden van een zo hoog mogelijke veiligheid van medewerkers is voor Aperam namelijk van groot belang.

Op termijn is het de bedoeling het gedigitaliseerde veiligheidshulpmiddel op meerdere/alle rolbruggen binnen Aperam geïnstalleerd zal worden. Aperam wenst hierdoor de kost zo laag mogelijk te houden en te werken op basis van visie. Dit aan de hand van 2 IP-camera's, geïnstalleerd op een vaste positie op de kat boven de kraanhaak, zoals verduidelijkt op Figuur 2. De kat van een rolbrug is de module die een horizontale verplaatsing van de kraanhaak over de rolbrug mogelijk maakt.



Figuur 2 Plaatsing IP-camera's

Het veiligheidshulpmiddel zal niet enkel gebruikt worden wanneer er zich een lading aan de kraanhaak bevindt. Gezien de kraanhaak op zichzelf ook een gevaar vormt wordt ervoor gekozen om in deze situatie ook aan waarschuwing te doen.

1.3 Doelstellingen

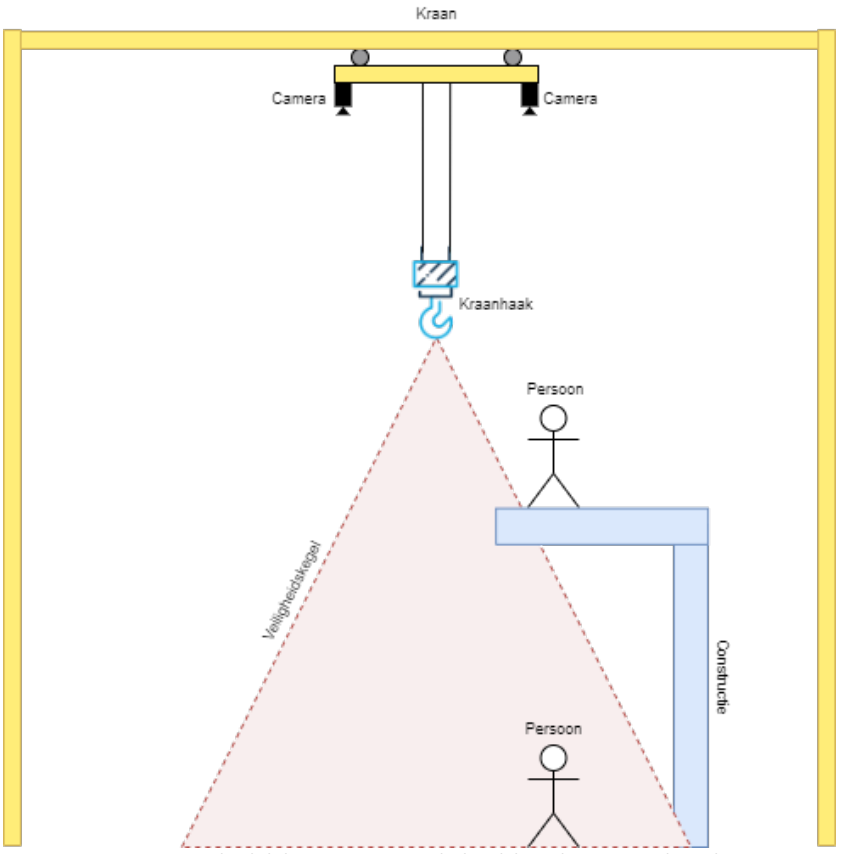
Deze masterproef beoogt de basis te leggen voor de digitalisering van dit veiligheidshulpmiddel aan de hand van 2 IP-camera's. Dit door de ontwikkeling van een systeem voor het detecteren van personen en het bepalen van hun positie binnen de 3D-ruimte.

Het detecteren van personen en het bepalen van hun positie binnen de 3D-ruimte is namelijk van groot belang. Allereerst zal er aan de hand van het relatieve hoogteverschil tussen een gedetecteerd persoon en de kraanhaak, de minimaal te bewaren veiligheidsafstand bepaald worden. Verder is het ook belangrijk om de positie van een gedetecteerd persoon binnen de 3D-ruimte te kennen om na te kunnen gaan of deze persoon zich binnen de gevarenzone bevindt, en dus gewaarschuwd moet worden, of niet.

1.3.1 Verduidelijking concept

Figuur 3 verduidelijkt het gehele concept. Aan de hand van 2 IP-camera's, gemonteerd op een vaste positie op de kat van een rolbrug boven de kraanhaak, dienen personen gedetecteerd te worden en hun positie, binnen de 3D-ruimte, bepaald te worden (doel masterproef). Dit om in eerste instantie, afhankelijk van het relatieve hoogteverschil tussen een gedetecteerd persoon en de kraanhaak, de minimaal te bewaren veiligheidsafstand te kunnen bepalen. Deze is zoals reeds eerder vermeld recht evenredig met het relatieve hoogteverschil. Verder zal er afhankelijk van de positie van een gedetecteerd persoon binnen de 3D-ruimte en de minimaal te bewaren veiligheidsafstand nagegaan kunnen worden of deze persoon zich binnen de gevarenzone bevindt of niet, en indien nodig gewaarschuwd moet worden.

Op Figuur 3 zijn 2 personen zichtbaar. Eéntje bevindt zich op een constructie op een zekere hoogte en de andere bevindt zich op het niveau van de grond. Zoals te zien bevindt de persoon op de constructie zich buiten de gevarenzone, en dus verder dan de minimaal te bewaren veiligheidsafstand. Gezien het kleine relatieve hoogteverschil tussen kraanhaak en deze persoon is de minimaal te bewaren veiligheidsafstand hier niet zo groot. De andere persoon bevindt zich op het niveau van de grond. Zoals te zien is er een groot relatief hoogteverschil tussen deze persoon en de kraanhaak. Dit heeft een grotere minimaal te bewaren veiligheidsafstand tot gevolg. Gezien de positie van deze persoon zich binnen de gevarenzone bevindt, dient deze persoon dan ook gewaarschuwd te worden. Dit geheel toont nogmaals het belang aan van het detecteren en het bepalen van de 3D-positie van een personen.



Figuur 3 Verduidelijking concept veiligheidskegel en positiebepaling

Aperam wenst enkel personen te waarschuwen die zich, wat hoogteniveau betreft, onder de kraanhaak bevinden (positief relatief hoogteverschil). Personen die zich op gelijke hoogte of boven de kraanhaak bevinden dienen algemeen niet gewaarschuwd te worden.

1.4 Overzicht scriptie

Hoofdstuk 1 verduidelijkt de situering, probleemstelling en doelstellingen van deze masterproef. Dit met extra verduidelijking omtrent het principe en het belang van de te behalen doelstellingen.

Hoofdstuk 2 gaat dieper in op de bestaande literatuur. Hierin wordt er onderzoek gedaan naar bestaande oplossingen en methodes voor het behalen van de doelstellingen van deze masterproef en wordt er extra uitleg en duiding gegeven bij de gebruikte principes. Dit om deze te begrijpen en er inzichten in te verkrijgen.

Hoofdstuk 3 bespreekt de effectieve ontwikkeling van het 3D-positiebepalingssysteem op basis van stereovisie. Er wordt dieper ingegaan op de hardware, gebruikte technieken, keuzes die gemaakt dienden worden en nog veel meer. De uitwerking en werking van het geheel worden hierin grondig besproken met waar nodig extra duiding en uitleg.

Hoofdstuk 4 van deze scriptie gaat in op de testen en testresultaten ter evaluatie en verificatie van de werking van het ontwikkelde 3D-positiebepalingssysteemsysteem op basis van stereovisie. Er wordt dieper ingegaan op de resultaten en waar nodig extra duiding gegeven.

Hoofdstuk 5 van deze scriptie bevat de finale conclusie. Het geheel wordt binnen dit hoofdstuk geëvalueerd. Dit hoofdstuk biedt ook een blik op eventuele verdere stappen.

2 Literatuurstudie

Binnen deze literatuurstudie, die uitgevoerd werd aan het begin van de masterproef periode om inzichten te krijgen en de meest fundamentele onderdelen te begrijpen, wordt de basis van afstandsbepaling, met de focus op stereovisie afstandsbepaling, besproken. Dit geeft een theoretisch kijk op het concept en gebruikte methodes binnen bestaande literatuur in combinatie met extra uitleg ter verduidelijking.

2.1 Methodes afstandsbepaling

Binnen de literatuur bestaan er twee algemene methodes voor het bepalen van afstanden, namelijk een actieve methode en een passieve methode [2].

Bij de actieve methode wordt er gebruik gemaakt van signalen om de afstand tot het doel te bepalen. Vaak gaat het hier dan om systemen waarbij de zogezegde Time-of-Flight (ToF) van een laserstraal, echo of radiosignaal gemeten wordt door een specifieke sensor, om hieruit de afstand tot het doel te berekenen. Het grote ongemak hierbij is de mogelijke verwarring van echo's van eerdere of latere signaalpulsen en ook de beperkte nauwkeurigheid, zeker voor langere afstanden. Het nauwkeurigheidsbereik ligt meestal tussen de één en vier meter. Tegenwoordig bestaan er al oplossingen voor dit soort problemen door bijvoorbeeld te werken met frequentie modulatie om verschillende pulsen te kunnen onderscheiden [2].

Sensor gebaseerde systemen zijn vaak het nauwkeurigst. Voor lange afstanden is het mogelijk om gebruik te maken van RADAR of LIDAR. Al is het hierbij wel belangrijk om te vermelden dat deze zeer kostelijk kunnen zijn en dat voornamelijk bij LIDAR de data verwerking complex kan zijn [3].

Bij de passieve methode wordt er eerder gekeken naar de positie van een object om aan afstandsbepaling te kunnen doen. Deze systemen zijn vaak gebaseerd op camera's en computer visie technieken. De passieve systemen kunnen nogmaals opgedeeld worden in twee subtypes, namelijk monovisie en stereovisie [2].

Bij monovisie systemen wordt er gebruik gemaakt van één camera en referentiepunten, met gekende grootte, in beeld om afstanden te bepalen [2]. Vaak is er op voorhand een kalibratie stap nodig waarbij de link wordt gelegd tussen afstanden in de werkelijke ruimte en pixels op beeld [4]. Een voorbeeld van een kalibratiepatroon met gekende groottes is te zien op Figuur 4.

Figuur 4 Voorbeeld kalibratiepatroon [4]

Stereovisie is een computer visie systeem dat gebaseerd is op stereoscopische technieken om afstanden te bepalen. Het systeem maakt gebruik van twee camera's, vanuit enigszins verschillende gezichtspunten, als één camera om de illusie van diepte te creëren. Dit biedt de mogelijkheid om, met een grote nauwkeurigheid, afstanden te bepalen [2]. Beide beelden dienen wel op hetzelfde moment genomen te worden. Het principe is vergelijkbaar met het menselijke zicht. Beide ogen zien een gelijkaardig beeld, maar met een kleine verplaatsing. Deze verplaatsing is omgekeerd evenredig met de afstand tussen de ogen en het object zelf [5]. Voor het bepalen van de afstand tot een bepaald punt of object is het belangrijk dat die punt/object zichtbaar is vanuit beide camerastandpunten.

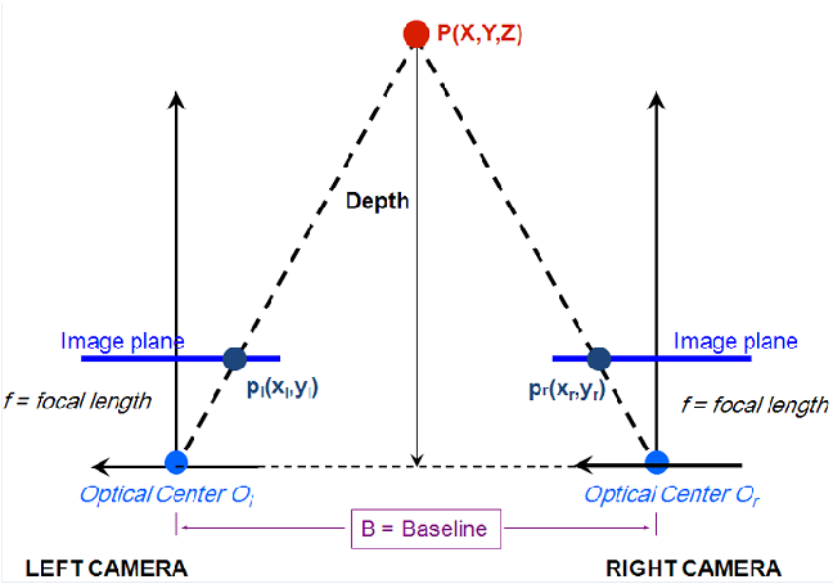
Gezien er binnen het project er gebruik gemaakt wordt van twee (IP) camera's zal de verdere focus van de literatuurstudie hierop liggen. Andere technieken zouden zeker ook bruikbaar kunnen zijn, maar vallen buiten de scope van deze masterproef.

2.2 Stereovisie afstandsbepaling

Stereoscopie is een techniek die wordt gebruikt voor het opnemen en weergeven van stereoscopische (3D) beelden. Er kan een illusie van diepte gecreëerd worden door gebruik te maken van twee beelden die op enigszins verschillende posities zijn genomen [6].

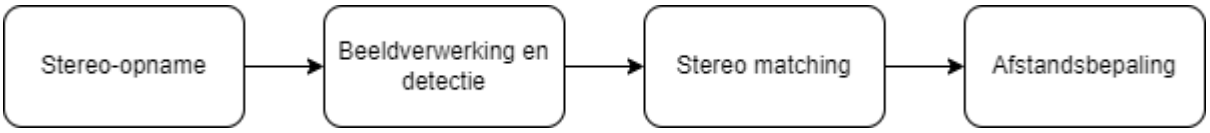
Wat belangrijk is om te weten bij stereovisie afstandsbepaling is dat de afstand tussen het 2D-vlak van de camera's en het hieraan evenwijdige 2D-vlak, waarin het gedetecteerde object van belang zich bevindt op een bepaalde (X, Y) positie, bepaald wordt. Een horizontale of verticale verplaatsing binnen hetzelfde 2D-vlak (X, Y) zou dus geen invloed mogen hebben op de afstandsbepaling in diepte.

Figuur 5 verduidelijkt de basis van stereovisie. Een object op positie P (X, Y, Z) is zichtbaar op beide camera's en valt dus binnen het overlappend gebied van beide beelden, wat nodig is om aan stereovisie positiebepaling te kunnen doen. De 'baseline', ookwel gekend als afstand tussen beide camera's, en de brandpuntsafstanden van beide camera's zijn gekend of kunnen bepaald worden doormiddel van een kalibratiestap. Dit door gebruik te maken van een kalibratiepatroon (ookwel gekend als dambordpatroon), zoals afgebeeld op Figuur 4. Afhankelijk van de gekozen camera positie en positionering dient er ook op basis van deze kalibratiestap aan beeld vervorming, translatie en rotatie gedaan te worden. Afhankelijk van de posities van het object van belang, P (X, Y, Z) en het verschil in positie (X, Y) (blauwe stip) op beide 2D-beelden (blauwe lijnen) van de camera's, kan de diepte ('Depth' op figuur) (Z) en bijhorende X en Y coördinaten (van het object) berekend worden. Zoals reeds eerder vermeld gaat het bij het bepalen van de diepte (Z) niet over de afstand tot het object zelf, maar over de Z coördinaat (afstand) tot het vlak, dat evenwijdig is aan het cameravlak, waarin het object van belang zich op een bepaalde (X, Y) positie bevindt.



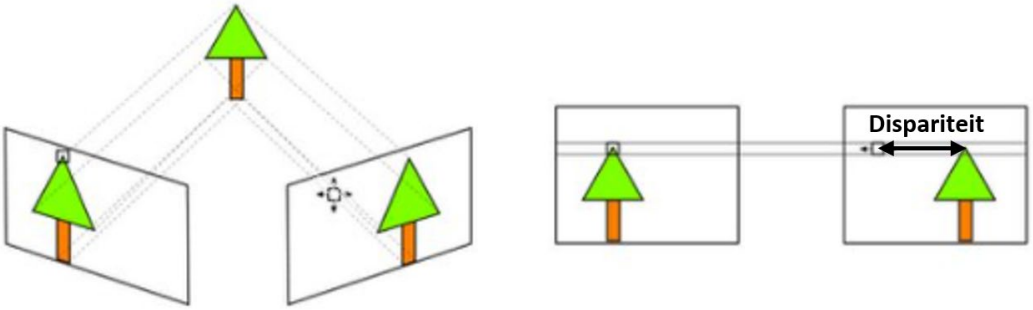
Figuur 5 Basisprincipe stereovisie [7]

Omtrent het bepalen van afstanden op basis van stereovisie zijn er online tal van bronnen terug te vinden. Over het algemeen wordt er overal gebruik gemaakt van dezelfde structuur, zoals afgebeeld op Figuur 6. Natuurlijk zijn er overal wel verschillen in aanpak. Vaak hebben deze verschillen te maken met het detectiemodel voor het detecteren objecten, het toepassen van stereo matching of niet, het toepassen van effectieve afstandsbepaling, etc.



Figuur 6 Algemene structuur stereo afstandsbepaling

Figuur 6 toont dat er vaak eerst een stereo-opname gemaakt wordt, waarop er, al dan niet voorafgaand door beeldverwerking, een detectiemodel wordt toegepast. Hierna wordt het gedetecteerde object, de gedetecteerde persoon, etc., indien nodig, onderworpen aan stereo matching. Op basis van pixelinformatie van het object zelf en eventuele naburige pixels op het beeld van één camera, wordt de locatie (X, Y) van een gedetecteerd object van belang op het andere beeld gezocht. Dit is een kerntechnologie binnen de computervisie. Het wordt veel gebruikt in de wereld van het autonoom rijden, *augmented reality* en robotsturingen. Het doel van het bepalen van de locatie (X, Y) van een object van belang op een paar stereobeelden is de mogelijkheid tot het bepalen van de dispariteit. Dispariteit wordt gedefinieerd als de horizontale (of eventuele verticale) verplaatsing tussen een paar corresponderende pixels (object van belang) in het linker- en rechterbeeld [8]. Figuur 7 verduidelijkt dit.



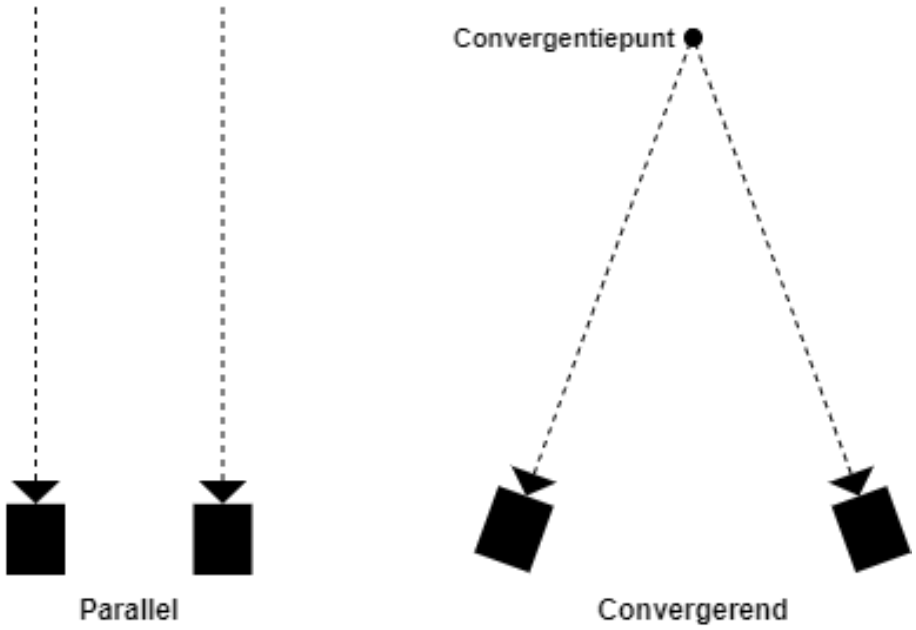
Figuur 7 Verduidelijking dispariteit [5]

De laatste stap binnen het algemene proces van stereo afstandsbepaling is het bepalen van de effectieve afstand. Hiervoor wordt er gebruik gemaakt van de dispariteit van een object van belang tussen beide beelden, de eigenschappen van de camera's en opstelling, en de relaties tussen de zijden en hoeken van driehoeken. Ookwel driehoeksmeetkunde of trigonometrie genoemd.

De reeds eerder besproken onderdelen, gevisualiseerd op Figuur 6, worden in de volgende delen in meer detail besproken. Dit met extra uitleg ter verduidelijking van het nut en de werking ervan.

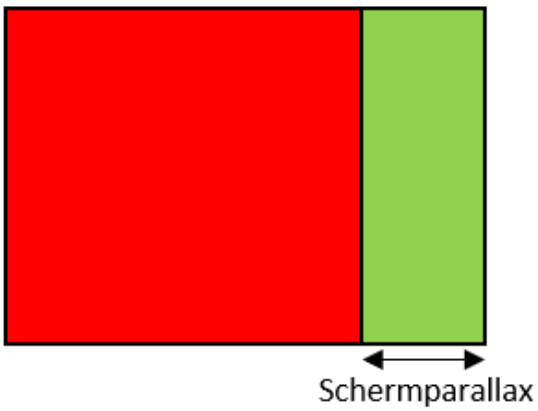
2.2.1 Stereo-opname

De eerste stap binnen het algemene proces van stereo afstandsbepaling is het maken van een stereo-opname. Afhankelijk van de geometrie van de optische assen, bestaan er twee mogelijkheden: parallel of convergerend. Het gebruik van divergerende assen (tegenovergestelde van convergerend), wordt gezien de toepassing zelden tot nooit gebruikt. Er moet namelijk een overlap zijn tussen het zichtveld van beide camera's om aan stereo afstandsbepaling te kunnen doen. Het menselijke visueel systeem werkt op basis van convergerende assen. Hierbij zijn beide assen gericht op het object van belang en kruisen ze hier (convergentiepunt). Als het object zich op een zeer verre afstand bevindt, is er bijna geen oogconvergentie en kan er gesproken worden over een parallel verloop van beide optische assen [5]. Figuur 8 verduidelijkt dit.



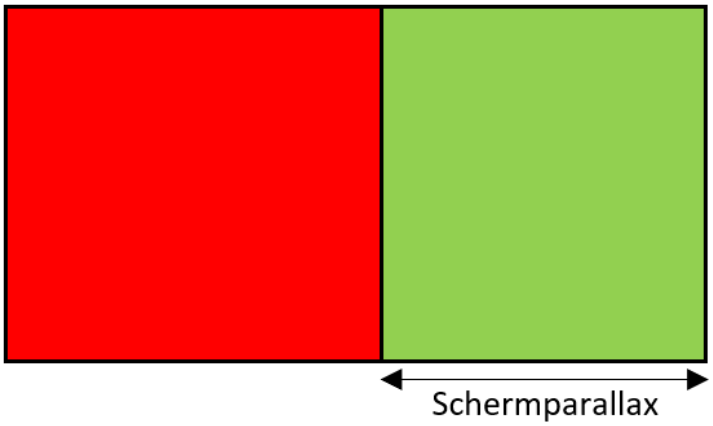
Figuur 8 Verduidelijking cameraposities aan de hand van eigenschappen optische assen

De keuze voor welke methode hangt voornamelijk van de toepassing en omstandigheden af. Beide methodes hebben hun voor- en nadelen. Het toepassen van de convergerende methode voorbij de convergentieafstand (afstand tussen camera's en het convergentiepunt) bevordert positieve schermparallax. Schermparallax is het verschil in afstand tussen twee 2D-beelden wanneer deze gelijktijdig geprojecteerd worden op een scherm zonder het gebruik van een stereobril [9]. Dit wordt voorgesteld op Figuur 9, waarbij één 2D-projectie in het rood weergegeven wordt en één 2D-projectie in het groen.



Figuur 9 Voorstelling standaard schermparallax

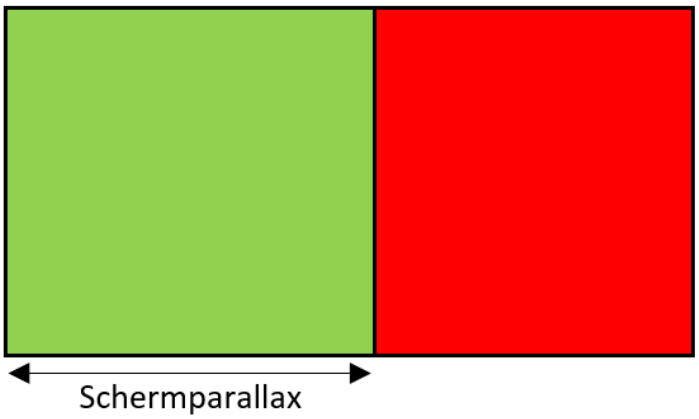
Positieve schermparallax is de positieve variant hiervan. Deze treedt op wanneer we het object van belang verder dan het convergentiepunt, en dus de convergentieafstand plaatsen. De schermparallax, zoals te zien op Figuur 9 zal toenemen en zal er nu uitzien zoals op Figuur 10. Het geprojecteerde 3D-beeld van het object van belang, bestaande uit twee 2D-beelden van het object, dat zich normaal in het convergentiepunt moet bevinden, valt grotendeels weg en wordt meer en meer weergegeven als twee aparte 2D-objecten. De effectieve 3D-projectie van de twee 2D-beelden lijkt zich eigenlijk achter het scherm te bevinden, gezien de foute convergentieafstand ten opzichte van de effectieve afstand tot het object van belang [9].



Figuur 10 Voorstelling positieve schermparallax

In het bovenstaande geval zal het gebruik van de convergerende methode dus een foutief resultaat opleveren. Het gebruik van de parallelle methode is hier aangeraden. Dit is dus onmiddellijk al een nadeel van de convergerende methode. De parallelle methode heeft natuurlijk ook zijn nadelen. Zo zullen camera's die parallel fotograferen convergeren naar oneindig. Dit heeft als gevolg dat alle gefotografeerde objecten een negatieve schermparallax zullen hebben en dus eigenlijk de illusie creëren van uit het scherm te springen [9].

Negatieve schermparallax treedt op wanneer we het object van belang zich voor het convergentiepunt bevindt, in tegenstelling tot positieve schermparallax. Het lijkt alsof de 3D-projectie van het object, bestaande uit twee 2D-beelden, zich voor het scherm bevindt. De schermparallax zoals te zien op Figuur 9 evolueert doormiddel van de negatieve schermparallax naar een beeld dat eruitziet zoals op Figuur 11. Het gaat hier nog steeds om hetzelfde object van belang. Het is duidelijk zichtbaar dat het linker- en rechter beeld ten opzichte van Figuur 9 zijn omgewisseld. Dit geeft onmiddellijk dan ook een nadeel van de parallelle methode [9].

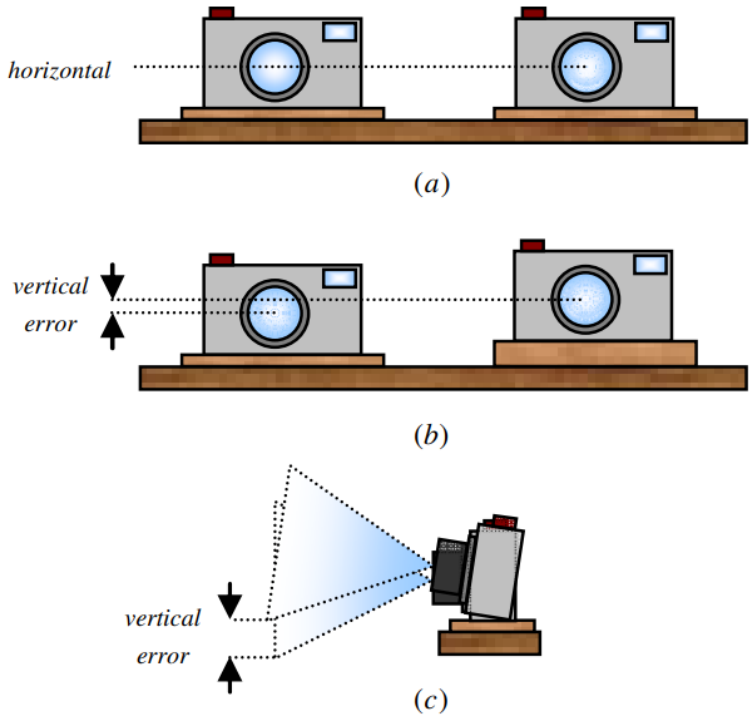


Figuur 11 Voorstelling negatieve schermparallax

Over het algemeen wordt de parallelle benadering gebruikt voor het maken van opnames van een landschapsarchitectuur, sport of alle situaties waarin voorwerpen op de voorgrond zich niet dichtbij en op een variabele afstand van de camera's bevinden. De schermparallax is hier namelijk zo goed als oneindig. Bij het maken van een opname van een object op de voorgrond dat dichtbij de camera's staat, wordt er meestal gebruik gemaakt van de convergerende methode. Het nadeel aan deze methode is dat het convergentiepunt en dus ook de convergentieafstand op voorhand al bepaald moeten worden, bij het gebruik van een vaste cameraopstelling. Het gebruik van de convergerende methode is dus binnen vele toepassingen van afstandsbepaling niet geschikt [9]. Binnen de literatuur omtrent afstandsbepaling wordt er meestal gebruik gemaakt van de parallelle methode.

Naast de methode voor het richten van camera's voor het maken van een stereo-opname zijn ook de uitlijning van de camera's en de synchronie van opname tussen de twee camera's van belang. Dit om volgende stappen binnen het proces van afstandsbepaling te vereenvoudigen en om nauwkeurige resultaten te bekomen.

Stereoscopische beelden kunnen worden gemaakt met een paar camera's, net zoals onze eigen ogen. De belangrijkste beperkingen bij het maken van een paar stereoscopische beelden zijn dat de camera's uitgelijnd moeten worden, zoals wordt verduidelijkt op Figuur 12, en dat de opname(s) synchroon gemaakt moeten worden op hetzelfde moment. Voor het oplossen van uitlijningsproblemen wordt er vaak gebruik gemaakt van een kalibratiestap, met beeldvervorming, -translatie, -rotatie en -rectificatie tot gevolg. Het synchroon of gelijktijdig maken van een opname is vooral van belang bij bewegende objecten [2] [6]. Het bepalen van de afstand tot een bewegend object, bij een niet synchrone opname, zal foutieve resultaten opleveren. Door beweging in de tussentijd bij een niet synchrone stereo opname, zal er een extra verschil in dispariteit kunnen optreden, wat een foutieve afstandsbepaling oplevert.



Figuur 12 Juiste uitlijning van de camera's (a) en uitlijning met verticale fouten (b) en (c) [6]

Bij het gebruik van de converterende methode voor het positioneren van de camera's, worden deze onder een bepaalde hoek, in de richting van het object van belang, naar elkaar toe gedraaid. Bij het effectief bepalen van de afstand aan de hand van driehoeksmeetkunde of trigonometrie zal er rekening met deze draaiingshoeken gehouden moeten worden of een correctie aangebracht moeten worden aan de hand van een kalibratie- en beeldverwerkingsstap. Ook zal er rekening gehouden moeten worden met de eigenschappen en instellingen van de camera. Het gaat hier dan voornamelijk over eigenschappen van de lens. De afstand tussen de camera's zal eveneens in rekening gebracht moeten worden. Het is dus belangrijk deze te kennen of te achterhalen. Voor het bepalen van de tussenafstand tussen de camera's is het voornamelijk belangrijk dat het object van belang zichtbaar is op beeld en dat er rekening gehouden wordt met de afstand tussen de camera's en het object zelf. Bij een grote afstand tussen camera's en object is het aangeraden de camera's wat verder van elkaar te plaatsen voor een nauwkeurig afstandsbepaling, dan bij het bepalen van een aftand van een object dichtbij de camera's. Buiten dat het object van belang zichtbaar moet zijn op beide camera's is er geen officiële leidraad voor de tussenafstand. De tussenafstand wordt in rekening gebracht bij het bepalen van de effectieve afstand tussen camera's en het object van belang bij het toepassen ervan.

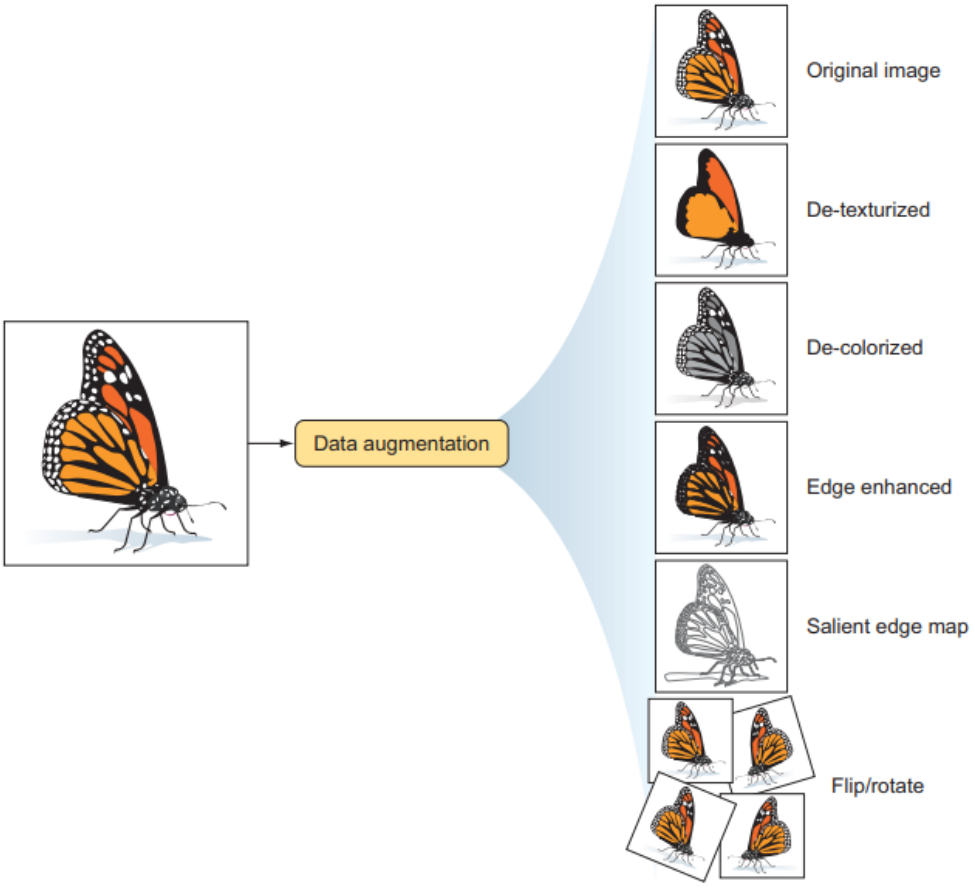
2.2.2 Beeldverwerking en detectie

Een volgende stap na het maken van een stereo-opname is het toepassen van eventuele beeldverwerking en detectie. Binnen de literatuur worden er hiervoor veel verschillende methodes toegepast. Zowel voor het verwerken en bewerken van beelden, het zogenaamde 'preprocessing', als voor het detecteren van objecten, personen, etc..

Het detecteren van objecten, personen, etc. is een belangrijke stap binnen de afstandsbepaling. Deze stap gaat namelijk vooraf aan het bepalen van de effectieve afstand. De snelheid van het gehele afstandsbepalingssysteem is dus ook sterk afhankelijk van deze stap. Het is belangrijk, voor de prestaties van het gehele systeem, dat deze stap zo nauwkeurig en efficiënt mogelijk gebeurt [2]. Hierdoor wordt het effectieve gebruik, trainen en testen van een detectiemodel vaak voorafgegaan door het zogenaamde 'preprocessing'.

Binnen machine learning projecten wordt in het algemeen meestal een ‘preprocessing’ of ‘cleaning’ stap uitgevoerd. Dit, opschonen en voorbereiden van data, is een belangrijke stap voor het effectieve uitbouwen van een model. Het doel van deze stap is om data klaar te maken voor het machine-learning model, zodat het gemakkelijker te analyseren en computationeel te verwerken is. Gebaseerd op de toepassing en de nodige vereisten zal het dus vaak nodig zijn om aan data ‘preprocessing’ te doen, vooraleer een model te trainen, testen of effectief te gebruiken [10].

Bij het toepassen van data ‘preprocessing’ op beelden gaat het vaak over het aanpassen van het beeldformaat, geometrische transformaties, kleurentransformaties, omzetten van kleur naar een grijsschaal, etc.. Enkele voorbeelden hiervan zijn te zien op Figuur 13 [10].



Figuur 13 Voorbeelden 'preprocessing' toegepast op een afbeelding van een insect [10]

Binnen de literatuur zijn er heel wat voorbeelden terug te vinden waar er ‘preprocessing’ wordt toegepast vooraleer een effectief detectiemodel toe te passen. Zaarane A. en anderen [2], maken in hun onderzoek gebruik van *edge detection*. Dit is een techniek voor het vinden van randen van objecten in een afbeelding. Salman Y. D. en anderen [11], maken in hun onderzoek gebruik van het verminderen van de resolutie en het omzetten van de RGB-afbeelding naar een afbeelding bestaande uit alleen maar grijswaarden [2]. Hsu en Wang [12], maken in hun onderzoek gebruik van het verkleinen van afbeeldingen tot een bepaalde grote, om de computationele snelheid te verhogen [2]. Dit zijn slechts enkele voorbeelden van manieren om aan ‘preprocessing’ te doen, om de snelheid van het latere detectiemodel te verhogen. Natuurlijk, zijn er nog vele andere methodes, maar het is niet relevant deze allemaal te bespreken. Het toepassen en de keuze van bepaalde soorten ‘preprocessing’ hangen voornamelijk af van de toepassing, de input en het detectiemodel zelf.

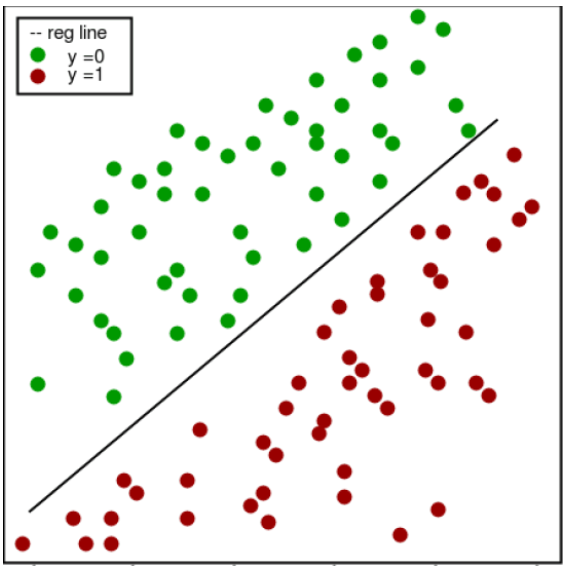
Naast het toepassen van ‘preprocessing’ om het model vlot te kunnen trainen, testen en te laten functioneren is het detecteren van objecten, personen, etc. afhankelijk van de toepassing ook van belang. Het detecteren van objecten, personen, etc. gebeurt, in het geval van stereoscopische afstandsbepaling, aan de hand van een detectiemodel. Dit is een computermodel dat erop getraind is om bepaalde instanties van objecten, personen, etc. van een bepaalde klasse binnen een beeld te detecteren [13]. Het voordeel van een objectdetectienetwerk ten opzichte van een beeldclassificatienetwerk is dat een objectdetectienetwerk zowel de klasse van een object kan weergeven, als de coördinaten ervan binnen het beeld. Bij een beeldclassificatienetwerk, ligt de focus meer op het aangeven of een bepaald object, persoon, etc. zich op beeld bevindt, zonder rekening te houden met de coördinaten [14]. Voor het verdere verloop zal een persoon ook als een detecteerbaar object gezien worden en zal er gezien de toepassing aandacht besteed worden aan het objectdetectienetwerk.

Objectdetectie valt binnen de wereld van het machinaal leren onder de ‘*supervised*’ problemen. Dit houdt in dat het objectdetectiemodel getraind moet worden aan de hand van gelabelde voorbeelden. Elke afbeelding in de trainingsdataset vereist een bestand dat de grenzen en klassen van een of meer objecten bevat. Een voorbeeld van zo een bestand is te zien op Figuur 14. Voor het labelen van deze objecten binnen afbeeldingen zijn er verschillende open-source hulpmiddelen beschikbaar om annotaties te maken voor het toepassen van objectdetectie [14].

```
<?xml version="1.0"?>
<annotation>
<folder>data</folder>
<filename>200328654-001.jpeg</filename>
<path/>
<source>
<database>Unknown</database>
</source>
<size>
<width>979</width>
<height>645</height>
<depth>3</depth>
</size>
<segmented>0</segmented>
<object>
<name>Lion cub</name>
<pose>Unspecified</pose>
<truncated>0</truncated>
<occluded>0</occluded>
<difficult>0</difficult>
<bndbox>
<xmin>194</xmin>
<ymin>176</ymin>
<xmax>545</xmax>
<ymax>606</ymax>
</bndbox>
</object>
<object>
```

Figuur 14 Voorbeeld annotatiebestand voor trainingsdataset objectdetectie [14]

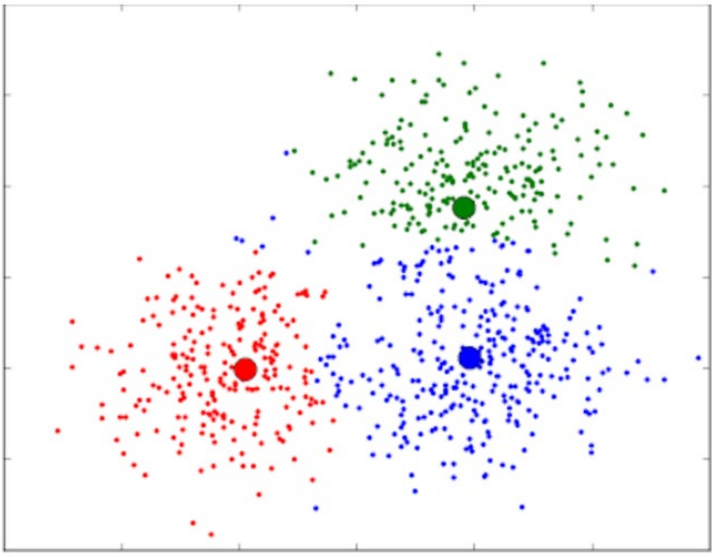
Figuur 15 verduidelijkt ‘supervised’ machine learning. Er is een visuele weergave van het resultaat van *‘logistic regression’* te zien, toegepast op een ‘supervised’ dataset. De zwarte doorlopende lijn geeft de scheiding tussen twee verschillende klassen weer. ‘Supervised’ machine learning is één van de twee situaties voor het toepassen van machine learning, waarin we de uitkomst van een bepaalde input op voorhand al kennen. ‘Supervised’ machine learning wordt gebruikt om twee soorten problemen op te lossen, namelijk classificatie en regressie. Bij classificatie gaat het over plaatsen van bepaalde inputs bij specifieke categorieën, zoals het classificeren van objecten of het herkennen van spraak. Bij regressie gaat het eerder over een oneindig aantal categorieën met een oneindig aantal waardes. Een voorbeeld hierbij is het maken van een voorspelling van hoeveel iemand voor een product zal betalen of het maken van een weersvoorspelling [15].



Figuur 15 Visuele weergave resultaat 'Logistic regression' toegepast op een 'supervised' dataset [15]

Enkele veel voorkomende 'supervised' machine learning algoritmes zijn: Lineaire en logistische regressie, 'Naïve Bayes', 'Support vector machines', Beslissingsbomen en willekeurige bossen en Artificiële neurale netwerken [15].

De tegenhanger van 'supervised' machine learning en tweede categorie van machine learning is '*unsupervised*' machine learning. Hierbij wordt er niet gewerkt met vooraf gedefinieerde categorieën, maar worden er oplossingen voor clusterproblemen gezocht. Data wordt aan de hand van gemeenschappelijke kenmerken in bepaalde clusters verdeeld [15]. Een visuele voorstelling hiervan is te zien op Figuur 16.



Figuur 16 Visuele voorstelling 'unsupervised' machine learning dataset onderverdeeld in drie clusters (rood, groen, blauw) [15]

Op basis van bepaalde kenmerken kan het model dus een voorspelling doen. K-means is een bekend machine learning algoritme voor het clusteren van gegevens. Hierbij is het van belang om duidelijk aan te geven in hoeveel clusters de data verdeeld moet worden. Te veel clusters zullen zorgen voor een overdreven complex en onnauwkeurig model, maar aan de andere kant zullen te weinig clusters ervoor zorgen dat data zonder gemeenschappelijke kenmerken samen gegroepeerd wordt [15].

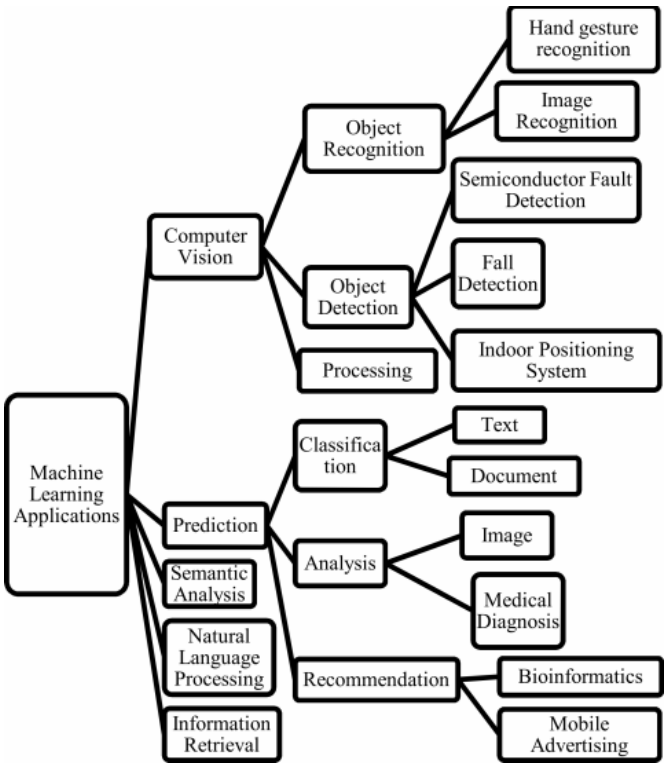
Naast het clusteren van data kan ‘unsupervised’ machine learning ook gebruikt worden voor het dimensioneel reduceren van een dataset, wanneer deze te veel ‘features’ of eigenschappen bevat. Naarmate het aantal eigenschappen van dataset toeneemt zal er namelijk een grotere trainingsdataset nodig zijn en stijgt de kans om een model te overtrainen. Bij het overtrainen van een model zal het model wel correcte voorspellingen uitvoeren op gekende trainingsdata, maar slechte voorspellingen uitvoeren op onbekende data. ‘Unsupervised’ machine learning algoritmes kunnen gegevens analyseren en minder/niet-gerelateerde attributen vinden die weggelaten kunnen worden, om het model te vereenvoudigen, zonder informatie te verliezen. Hoofdcomponentenanalyse (Principal component analysis) (PCA) is hier een populair voorbeeld van [15].

Om verder te gaan op het reeds eerder besproken objectdetectiemodel wordt er typisch gebruik gemaakt van machine learning, ookwel gekend als machinaal leren. Machine learning is een subcategorie van artificiële intelligentie (AI). Artificiële intelligentie heeft betrekking op het even intelligent maken van machines als het menselijk brein. In de informatica wereld wordt deze term gebruikt wanneer een machine in staat zijn de omgeving waar ten menen en acties te ondernemen om doelen te behalen. Informeel wordt de term gebruikt wanneer een machine functies kan uitvoeren die mensen associëren met menselijke acties, zoals leren en problemen oplossen. Gezien leren binnen artificiële intelligentie zo een vitaal aspect vormt is er een aparte subcategorie genaamd ‘machine learning’ (machinaal leren) voor gemaakt [16].

Machine learning is de toepassing van kunstmatige intelligentie op computers zodat ze kunnen leren, van de aan hen verstrekte gegevens, om hun eigen beslissingen te nemen, die vergelijkbaar zijn met menselijke beslissingen. Dit geeft computers de mogelijkheid om te leren en voorspellingen te doen op basis van gegevens, informatie en interacties met de echte wereld [17].

Algoritmen voor machinaal leren kunnen zelf beslissingen nemen zonder expliciet te zijn geprogrammeerd om dit te doen. Beslissingen worden gemaakt aan de hand van wiskundige modellen, getraind en ontwikkeld aan de hand van aangeboden data, beter bekend als de training dataset. Bij machine learning algoritmes is het nodig om aan het systeem kenmerken mee te geven zodat het hierop kan leren. Dit proces heet ‘Feature Engineering’ [17].

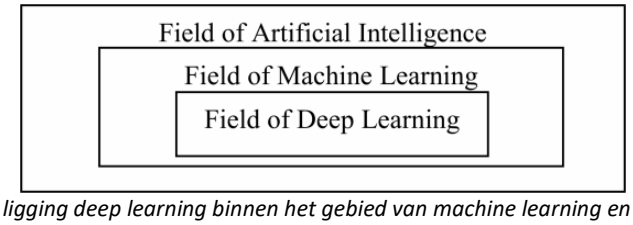
Machine learning kent zeer veel toepassingen. Figuur 17 verduidelijkt enkelen hiervan.



Figuur 17 Enkele toepassingen machine learning [16]

In functie van het detecteren van personen wordt er gebruik gemaakt van een objectdetectiemodel, zoals reeds eerder besproken. Binnen machine learning moeten de kenmerken van het objectdetectiemodel worden gedefinieerd aan de hand van verschillende methodes en vervolgens met behulp van een techniek zoals 'Support Vector Machines' (SVM's) worden geclassificeerd [17]. Binnen machine learning bestaan er voor het detecteren van objecten heel wat algoritmes met elk hun specifieke eigenschappen. Enkel voorbeelden hiervan zijn beslissingsbomen, 'Naïve Bayes', 'Bayesian Network' (BN), 'k-Nearest Neighbour' , 'Sequential Minimal Optimization' (SMO), 'AdaBoost' en 'Bagging' met vergelijking van parameters zoals tijd en nauwkeurigheid [16]. Hiernaast bestaan er ook nog andere methodes zoals 'Scale-Invariant Feature Transform' (SIFT), 'Histogram of Oriented Gradients' (HOG) toepassingen en het 'Viola-Jones' objectdetectieframework [17].

Deep learning is op zijn beurt dan weer een subcategorie van machine learning [16]. Dit wordt visueel verduidelijkt op Figuur 18. De laatste decennia wordt er enorm veel onderzoek gedaan naar machine learning. Dit leidt tot hoge verwachtingen van machines. Deep learning is een poging in deze richting [16].



Figuur 18 Verduidelijking ligging deep learning binnen het gebied van machine learning en artificiële intelligentie [16]

Deep learning is een subcategorie van machine learning die gebruik maakt van een aanpak, bestaande uit meerder lagen, om kenmerken van hoog niveau te extraheren uit gegevens die aangeleverd worden. Voor deep learning is het niet nodig dat kenmerken handmatig aangeleverd worden, maar in plaats daarvan probeert het gegevens om te zetten in een abstracte representatie. Simpel gezegd leert het van voorbeelden en worden deze gebruikt voor toekomstige toepassingen. Deep learning kan worden vergeleken met artificiële neurale netwerken (ANN), zoals aanwezig in het menselijke brein. De meeste deep learning methodes

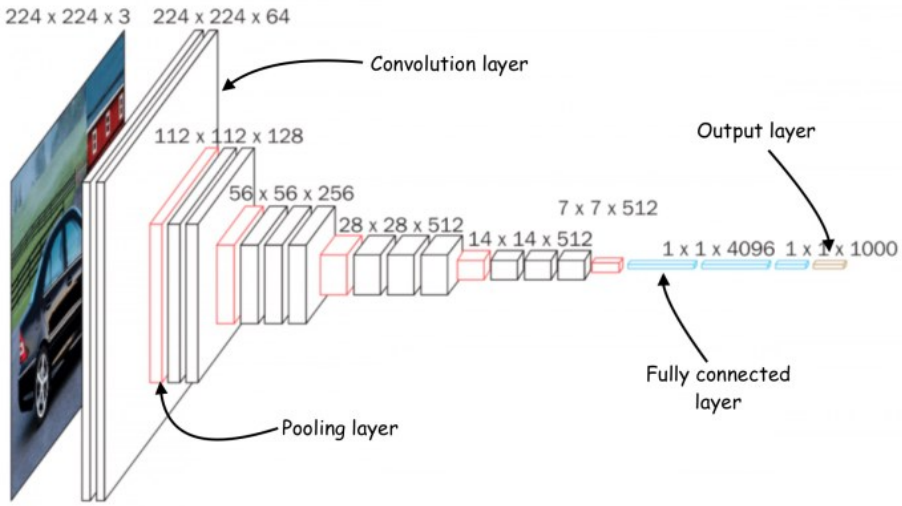
maken dan ook gebruik van zulke neurale netwerken om resultaten te bereiken. Zo goed als alle deep learning modellen vereisen enorme rekenvermogens en grote hoeveelheden gelabelde gegevens om kenmerken uit de te gegevens te halen en hiervan te leren. Enkele dagelijkse toepassingen van deep learning zijn nieuwsaggregatie, detectie van frauduleus nieuws, visuele herkenningen, taalverwerking, etc. [17].

In functie van het detecteren van personen aan de hand van deep learning wordt er meestal gebruik gemaakt van '*Convolutional Neural Networks'* (CNN's). In het Nederlands bekend als convolutioneel neurale netwerken. Het is een van de belangrijkste componenten van de meeste op deep learning gebaseerde computervisietoepassingen. CNN is een type neuraal netwerk dat patronen in een multidimensionale ruimte efficiënt kan vastleggen. Dit maakt het dan ook geschikt voor afbeeldingen, hoewel het ook wordt gebruikt om andere soorten gegevens te verwerken [14].

Elk convolutioneel neuraal bestaat uit een of meer convolutionele lagen. Dit is een softwarecomponent die betekenisvolle waarden extraheert uit het invoerbeeld. Elke convolutielaag bestaat uit verschillende filters. Dit zijn matrices die een afbeelding onderzoeken en de gewogen som van pixelwaarden op verschillende locaties registreren. Elk filter extraheert verschillende attributen en heeft verschillende waarden. De typische output van een convolutielaag is een set "kenmerkkaarten". Door verschillende convolutielagen te combineren, kan er een hiërarchie van visuele patronen geïdentificeerd worden. De onderste laag produceert bijvoorbeeld een kaart met verticale en horizontale randen, hoeken en andere eenvoudige patroonkenmerken. Volgende lagen zullen complexere patronen, zoals rasters en cirkels herkennen. Naarmate er dieper in het netwerk gegaan wordt, zullen de lage ingewikkelde en complexe objecten zoals personen, planten, etc. detecteren [14]. Dit wordt verduidelijk op Figuur 19.

Figuur 19 Weergave hiërarchie verschillende lagen CNN [14]

De meeste convolutionele neurale netwerken maken gebruik van verschillende pooling-lagen. Dit zijn lagen om de grootte van kenmerkkaarten geleidelijk te verkleinen, met behoud van de meest befaamde delen. Het meest belangrijke type pooling laag dat in CNN's wordt gebruikt, genaamd 'Max-pooling', behoudt de maximale waarde van een groepering pixels. Als je bijvoorbeeld een pooling laag met een grootte van '2' gebruikt, neemt deze groeperingen van 2x2 pixels uit iedere kenmerkenkaart, die door de voorgaande laag zijn geproduceerd, en behoudt het de hoogste waarde. Het toepassen van deze methode behoudt de meest relevante kenmerken, maar halveert de grootte. De pooling-laag stelt CNN's in staat om hun functionaliteit te generaliseren en ze minder gevoelig te maken voor bewegende objecten tussen verschillende frames. Ten slotte wordt de uitvoer van de convolutielaag afgevlakt tot een eendimensionale matrix die een numerieke weergave is van alle kenmerken van de afbeelding. Deze matrix wordt toegevoegd aan een reeks van volledig geconnecteerde lagen van artificiële neuronen die kenmerken omzetten in een soort van output die van het netwerk wordt verwacht [14]. Een voorstelling van de architectuur van een convolutioneel neuraal netwerk is te zien op Figuur 20.

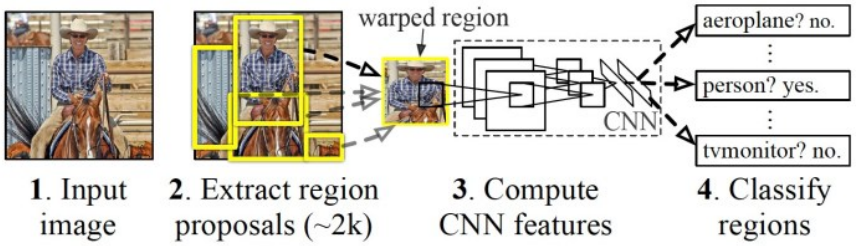


Figuur 20 Voorstelling architectuur convolutioneel neuraal netwerk [14]

Eén van de meest voorkomende taken voor convolutionele neurale is beeldclassificatie, waarbij het netwerk een afbeelding als invoer neemt en een lijst met waarschijnlijkheidswaarden als uitvoer teruggeeft. Stel dat men een neuraal netwerk traint om 100 klassen van objecten te detecteren, zal de uitvoerlaag 100 numerieke uitgangen hebben, die elk hun waarschijnlijkheid aangeeft dat een bepaalde afbeelding tot deze klasse behoort [14].

Enkele voorbeelden van neurale netwerkarchitecturen voor objectdetectie zijn:

R-CNN deep learning model



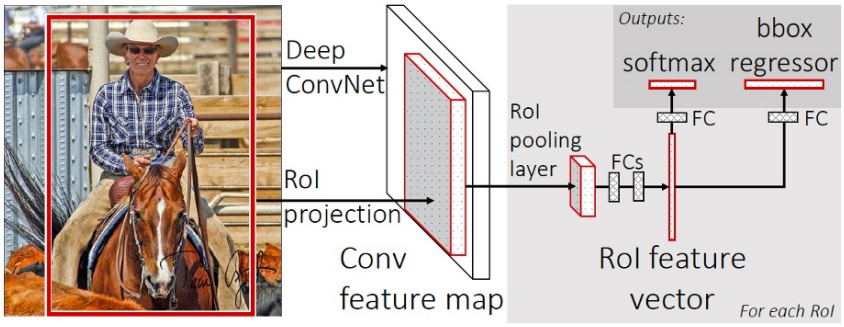
Figuur 21 R-CNN objectdetectie systeem overzicht [14], [18]

Het *'Region-based Convolutional Neural Network'* (R-CNN), waarvan de architectuur is afgebeeld op Figuur 21, bestaat uit drie belangrijke componenten. Ten eerste maakt een regioselector gebruik van selectief zoeken. Dit is een algoritme dat gebieden van pixels in een afbeelding vindt die objecten kunnen vertegenwoordigen. Vervolgens worden deze regio's met mogelijke objecten vervormd tot een, op voorhand doorgegeven, grootte en doorgegeven aan een convolutioneel neuraal netwerk. Het CNN zal deze voor elke regio afzonderlijk verwerken en kenmerken extraheren via een reeks convolutiebewerkingen. Het CNN maakt gebruik van volledig aangesloten lagen om kenmerkkaarten te coderen naar een ééndimensionale vector van numerieke waarden. Uiteindelijk zal een classificatie machine learning model de gecodeerde kenmerken van het CNN converteren naar een reeks klassen. Er zal hierbij één aparte uitgangsklasse zijn die voorzien is voor de achtergrond, die geen overeenkomst vertoont met objecten en dus over blijft [14], [18]. Figuur 22 geeft een voorbeeld weer van de output van het R-C neurale netwerk (R-CNN).

Figuur 22 Voorbeelden objectdetectie aan de hand van R-CNN [18]

Het gebruik van R-CNN heeft enkele nadelen. Zeker voor een realtime toepassing. Ten eerste moet het model voor elk beeld ongeveer 2000 afzonderlijke regio's genereren en bijsnijden, wat tijd in beslag neemt. Ten tweede moet het model voor elke van de 2000 regio's de kenmerken apart berekenen. Dit is rekenintensief en vertraagt het hele proces, wat RCNN ongeschikt maakt voor realtime objectdetectie. Het is lastig de snelheid te verbeteren en berekeningen te integreren, gezien het model bestaat uit drie afzonderlijke componenten [14], [18].

Fast R-CNN



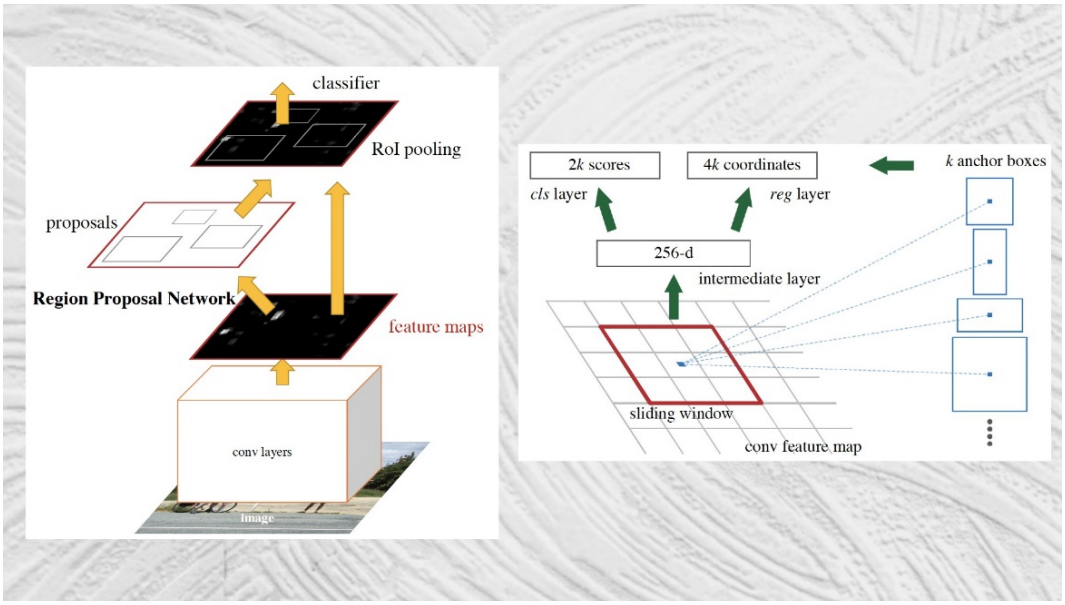
Figuur 23 Fast R-CNN overzicht [14], [19]

Fast R-CNN, waarvan een overzicht is weergegeven op Figuur 23, is en vernieuwde versie van R-CNN, dat reeds eerder besproken werd. Binnen deze versie zouden een aantal problemen van zijn vorige versie opgelost zijn. Fast R-CNN brengt kenmerken extractie en regioselectie samen in één machine learning model [14], [19].

Fast R-CNN ontvangt het beeld, samen met een set van interessante regio's en geeft een lijst van bounding boxes en klassen van de in beeld gedetecteerde objecten. Eén van de belangrijkste innovaties binnen Fast R-CNN is het gebruik van een pooling laag voor de interessante regio's. Dit is een activiteit waarbij een CNN-functiekaart en interessegebieden voor een afbeelding worden gebruikt om voor elk gebied overeenkomstige kenmerken te leveren. Dit biedt Fast R-CNN de mogelijk tot het in één keer extraheren van kenmerken voor alle interessante regio's op beeld, in tegenstelling tot R-CNN, waar elke regio afzonderlijk verwerkt wordt. Het resultaat is een toename in snelheid [14], [19].

Echter blijft er binnen R-CNN nog één probleem onopgelost. Met Fast R-CNN beeldregio's nog steeds afzonderlijk worden geëxtraheerd en als input aan het model geleverd worden. Dit maakt Fast R-CNN geen ideale keuze voor realtime objectdetectie [14], [19].

Faster R-CNN

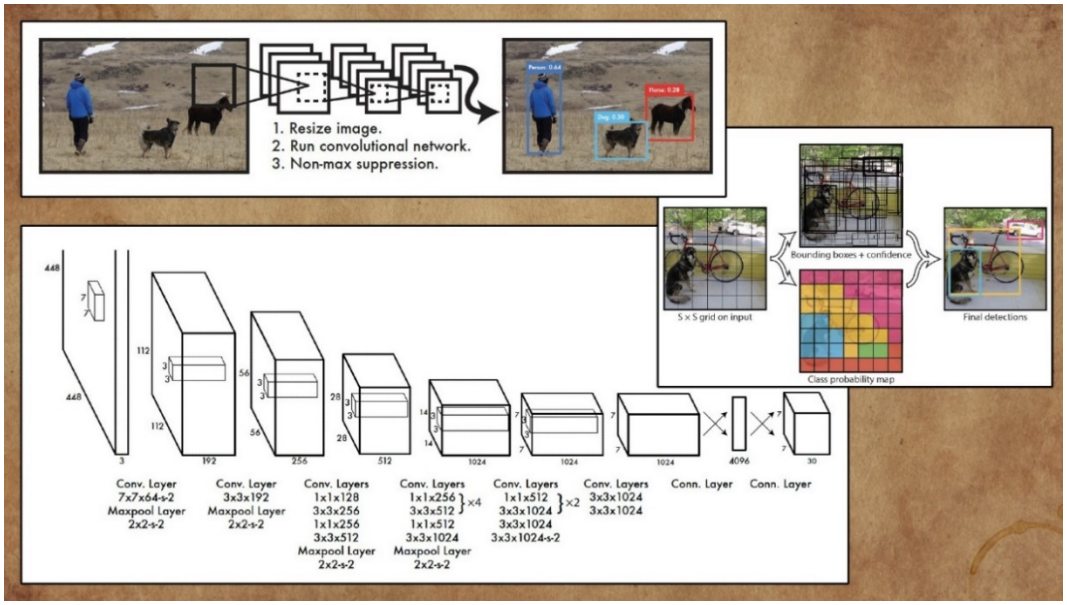


Figuur 24 Faster R-CNN architectuur [14], [20]

Faster R-CNN lost het laatste probleem, zoals reeds eerder besproken bij Fast R-CNN, op. Er wordt binnen Faster R-CNN gebruik gemaakt van een mechanisme dat regio extractie integreert in het objectdetectienetwerk. Faster R-CNN neemt een afbeelding als input en retourneert een lijst van objectklassen, met hun bijhorende bounding boxes (detectieregio's op beeld). De architectuur van Faster R-CNN is grotendeels gelijk aan die van Fast R-CNN. De grootste vernieuwing is het *'Region Proposal Network'* (RPN). Regio voorstel netwerk in het Nederlands. Dit is een component die een, door het convolutioneel neuraal netwerk geproduceerde, kenmerkenkaart neemt en Deze gegevens worden vervolgens doorgegeven aan de gebiedsclusterlaag van belang. De rest van het gele proces is vergelijkbaar met het Fast R-CNN-proces [14], [20].

Door de integratie van de regiodetectie binnen de hoofdarchitectuur van het neurale netwerk, bereikt Faster R-CNN een bijna realtime objectdetectiesnelheid [14], [20].

YOLO



Figuur 25 YOLO architectuur [14], [21]

'You Only Look Once' (YOLO), een familie van neurale netwerken die de snelheid en nauwkeurigheid van objectdetectie verbeterd aan de hand van deep learning [14], [21]. Het deelt afbeeldingen op in bepaalde zones volgens een rasterstructuur. Elke cel binnen deze gritstructuur is verantwoordelijk voor het detecteren van objecten binnen zijn eigen cel [22].

Een van de belangrijkste innovaties van YOLO is de integratie van de gehele workflow voor objectdetectie en classificatie in één enkel netwerk. In tegenstelling tot het afzonderlijk extraheren van kenmerken en regio's combineert YOLO alles in één enkele beweging door slechts één netwerk. Vandaar ook de naam van het netwerk. YOLO is in staat om realtime objectdetectie uit te voeren aan bepaalde streaming framerates [14], [21].

Voor het effectief kunnen toepassen van stereovisie afstandsbepaling worden er binnen de literatuur verschillende methodes gebruikt om aan objectdetectie te doen. Het toepassen van objectdetectie, kan gelijktijdig op beide camera's, maar ook slechts op één camera. Dit wordt later verder besproken in het gedeelte 2.2.3 Stereo matching.

Zaarane A. en anderen [2], maken in hun onderzoek gebruik van een objectdetectie algoritme voor het detecteren van auto's. Eerst passen ze een hypothesegeneratiestap toe, gevolgd door een hypotheseverificatiestap. In de stap van hypothesegeneratie worden potentiële locaties van voertuigen gegenereerd. Deze generatie is gebaseerd op het matchen van voertuigsjablonen met beelden aan de hand van cross correlatie. In de verificatiestap worden de gegenereerde hypotheses geverifieerd door twee bewerkingen uit te voeren, namelijk kenmerkextractie en classificatie. In een laatste stap wordt er met behulp van '*AdaBoost classifier'* een tweedimensionale discrete wavelet transformatie uitgevoerd om kenmerken te extraheren uit de gegenereerde hypotheses en hier vervolgens classificatie mee toe te passen [2].

Karthika K. en anderen [3], maken in hun onderzoek gebruik van het YOLO-objectdetectie algoritme, dat reeds eerder werd besproken. De invoer van hun YOLO-algoritme is een camerabeeld. Het object dat op beeld wordt gedetecteerd, wordt visueel aangegeven met een begrensde box, een zogenaamde 'bounding box'. De klasse-labels en bijhorende hoekcoördinaten van een detectie zijn de output van het netwerk [3]. Ter verduidelijking is er een voorbeeld van een 'bounding box' te zien op Figuur 26. Vaak wordt deze ook voorgesteld samen met het label van de bijhorende klasse, dat in dit geval 'auto' zou kunnen zijn.

Figuur 26 Voorbeeld 'bounding box' rondom auto [3]

Over het algemeen wordt er in de literatuur gebruik gemaakt van verschillende methodes om aan objectdetectie te doen. Vaak zijn er zelfs verschillende opties mogelijk. De keuze kan gebaseerd worden op eigenschappen van het detectiemodel zoals snelheid en nauwkeurigheid. Binnen elk model opzich zijn er ook meestal meerdere instellingen mogelijk. Er wordt hier dan gesproken over parameters. Binnen het toepassen van stereovisie afstandsbepaling is het voornamelijk belangrijk dat een object van belang gedetecteerd wordt en de positie op beeld verkregen kan worden. Afhankelijk van de toepassing zouden ook snelheid en nauwkeurigheid een belangrijke rol kunnen spelen.

2.2.3 Stereo matching

Om in een volgende fase de effectieve afstand, tot een gedetecteerd object van belang, te kunnen berekenen is het, zoals reeds eerder vermeld, belangrijk om de dispariteit tussen het object van belang op het linker- en rechterbeeld te kennen. Hierbij vormt stereo matching een belangrijke, maar niet noodzakelijke, stap.

Het doel van stereo matching is om aan de hand van een paar stereobeelden de dispariteit van pixels te kunnen berekenen voor elke pixel in het referentiebeeld, waarbij dispariteit wordt gedefinieerd als de horizontale verplaatsing tussen een paar corresponderende pixels in het linker- en rechterbeeld [8]. Dit werd reeds eerder al verduidelijkt op Figuur 7. Voor het zo snel en correct mogelijk kunnen uitvoeren van deze stap is een juiste uitlijning van de positionering van de camera's, in combinatie met eventuele kalibratie en beeldverwerking, van groot belang. De juiste positionering werd eerder besproken in het deel 2.2.1 Stereo-opname.

Voor het bepalen van de effectieve afstand is het belangrijk om de dispariteit van een gedetecteerd object van belang tussen het linker- en rechterbeeld te kennen. Dit kan gedaan worden op basis van twee methodes. Het is hierbij van belang dat het op het linker- en rechterbeeld om hetzelfde object gaat. Ook al zijn er meerdere objecten van belang gedetecteerd op beeld [2].

Een eerste methode is het bepalen van de locatie van een object op beide beelden. Hierbij is het dus ook effectief nodig om het objectdetectie algoritme toe te passen op beide beelden. Dit om hierna eventueel stereo matching algoritmes toe te passen om een match te vinden tussen gedetecteerde objecten en dus ook zeker te zijn dat het om hetzelfde object gaat, om vervolgens de dispariteit te bepalen. Een nadeel aan deze methode is dat het toepassen van de objectdetectie en stereo matching algoritmes op beide beelden tijd en rekenkracht in beslag neemt [2].

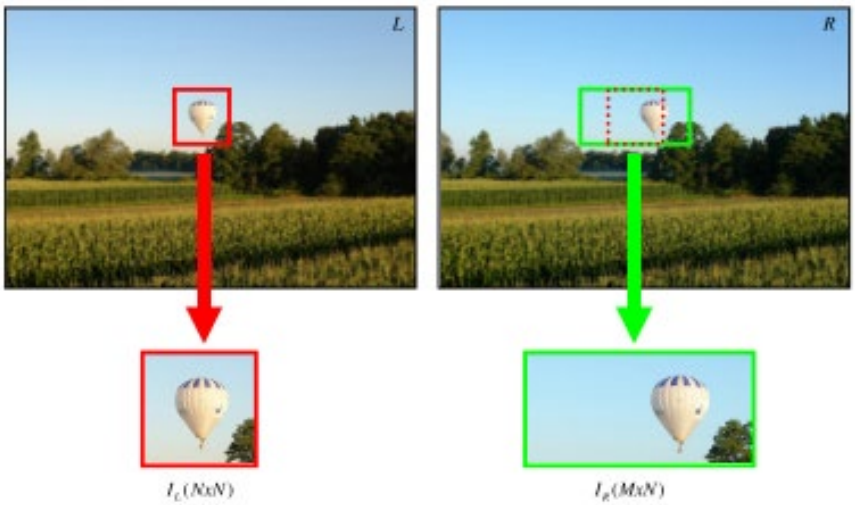
Een tweede methode is het toepassen van objectdetectie voor het object van belang op slechts één van beide beelden, om vervolgens het gedetecteerde object van belang te matchen met hetzelfde object op het andere beeld. Objectdetectie wordt hier dus slechts toegepast op één van beide beelden, wat tijd en rekenwerk bespaart [2].

Het ‘matchen’ zelf gebeurd meestal aan de hand van het toepassen van ‘cross-correlatie’, kruiscorrelatie in het Nederlands. Kruiscorrelatie geeft aan in welke mate elementen op elkaar lijken. De kruiscorrelatiefunctie varieert in principe tussen’+1’ en ‘-1’. Het beste resultaat van kruiscorrelatie wordt aangegeven wanneer de functies waardes aangeeft dichtbij ‘+1’. De best gedetecteerde match is op deze manier dan ook te bekomen. Het is het punt waar de kruiscorrelatiefunctie zijn hoogste waarde aanneemt, boven een bepaalde minimale drempelwaarde. Indien er geen match werd gevonden zal de kruiscorrelatie een laag resultaat aangeven, onder een bepaalde drempelwaarde. In dat geval werd er geen match gevonden en kan het zijn dat het object van belang niet op beide beelden zichtbaar is [2]. Bij een correcte uitlijning van de camera’s zal de kruiscorrelatie functie slechts in één richting uitgevoerd moeten worden, gezien het object van belang zich dan op hetzelfde niveau zou moeten bevinden. Dit bespaart heel wat tijd en rekenkracht en toont het belang en voordeel van een correcte uitlijning.

Voor het toepassen van kruiscorrelatie bestaan er verschillende methodes. Sommige zijn minder gevoelig wat belichting betreft dan anderen, sommige kunnen toegepast worden op geroteerde of herschaalde beelden, etc..

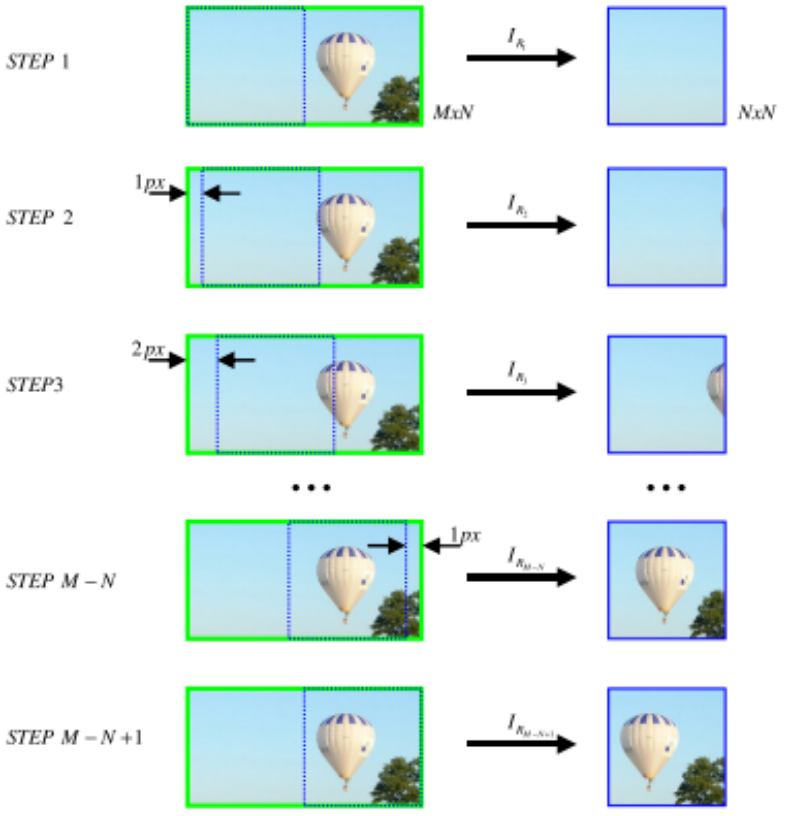
Een methode die in de literatuur terug te vinden is, is het gebruik van een optimaliseringsalgoritme dat zoekt naar minimale verschillen tussen beide beelden [6].

Om een voorbeeld te bespreken. Stel dat de vierkant matrix I_L een gedetecteerd object op de linker afbeelding voorstelt, zoals te zien is op Figuur 27, met een luchtballon als object van belang. De eigenschappen van de stereoscopische afbeeldingen zijn gekend. Beide objecten hebben een horizontale verschuiving ten opzichte van mekaar ondergaan bij het vergelijken van beide afbeeldingen. Aan de hand van deze informatie kan het zoekgebied op de rechter afbeelding bepaald worden. Dit zoekgebied wordt gedefinieerd als matrix I_R en is ook te zien op Figuur 27. Bij een correcte uitlijning van beide camera’s zullen de verticale dimensies van I_R en I_L hetzelfde zijn, maar zal de horizontale dimensie van I_R groter zijn [6].



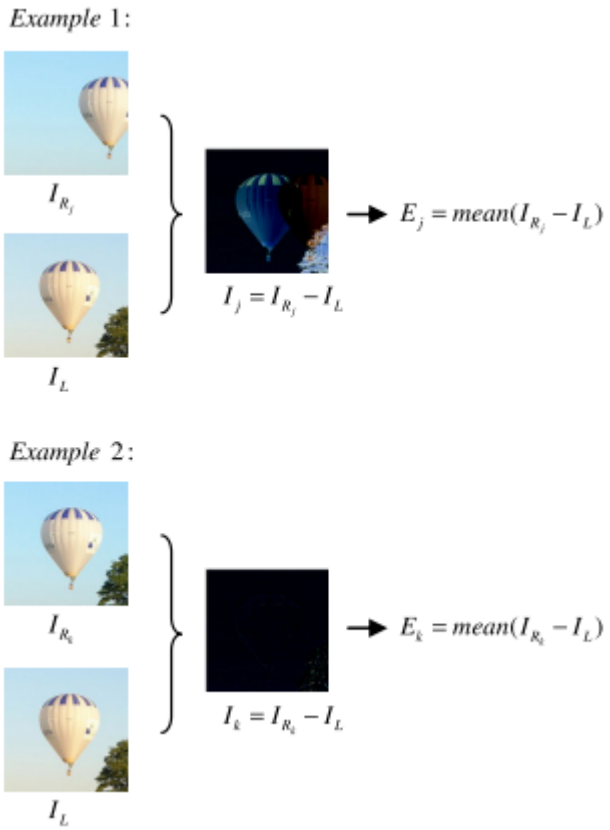
Figuur 27 Gedetecteerd object op linker afbeelding (rood) en zoekgebied op de rechter afbeelding (groen) [6]

Een volgende stap is het vinden van een locatie binnen het zoekgebied I_R, waar het beeld het best overeenkomt met matrix I_L. Dit kan gedaan worden door het aftrekken van matrix I_L van alle mogelijke submatrices (met dezelfde grootteorde van matrix I_L) binnen matrix I_R. Wanneer de grootte van matrix I_L NxN is en de grootte van matrix I_R MxN, dan moeten er, bij M>N, M–N+1 submatrices gecontroleerd worden [6]. Dit wordt voorgesteld op Figuur 28.



Figuur 28 Bepaling submatrix in rechter afbeelding [6]

Het resultaat van elke aftrekking is een nieuwe matrix genaamd I_i, die weergeeft hoe gelijkend de afgetrokken matrices van afbeeldingen zijn. Des te meer gelijkend de afgetrokken matrices van de afbeeldingen waren, des te lager het absolute gemiddelde van de matrix I_i. Figuur 29 toont twee voorbeelden van aftrekkingen. Bij het eerste voorbeeld (bovenste) matchen de afbeeldingen niet, terwijl in het tweede voorbeeld (onderst) beide afbeeldingen bijna perfect matchen [6].

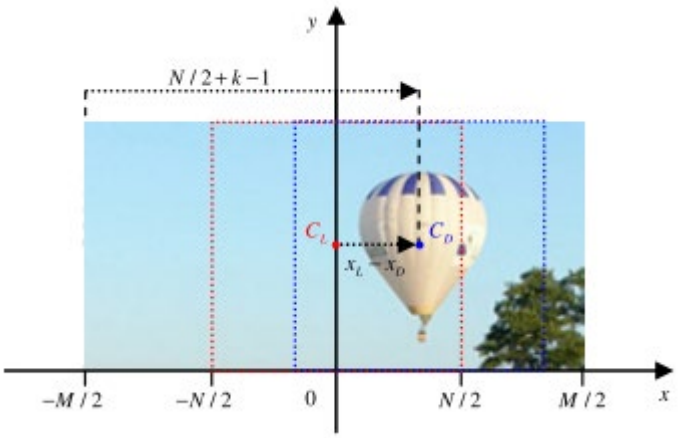


Figuur 29 Verschil tussen gedetecteerd object op de linker afbeeldingen met submatrix op de rechter afbeeldingen [6]

Matrix I_K, met het laagste gemiddelde van al zijn elementen, stel de locatie voor waar matrix I_L het best matcht, en dus overeenkomt, met matrix I_R. Dit is ook de plaats van het object van belang in het rechter beeld. Aan de hand van deze overeenkomst in het linker- en rechterbeeld kan de afstand, ookwel gekend als dispariteit, tussen de beelden berekend worden volgens Vergelijking 1 [6]. Figuur 30 geeft duiding bij deze vergelijking.

$$X_L - X_D = \frac{N}{2} + k - 1 - \frac{M}{2}$$

Vergelijking 1 Berekening dispariteit [6]



Figuur 30 Berekening horizontale afstand tussen linker en rechter afbeelding [6]

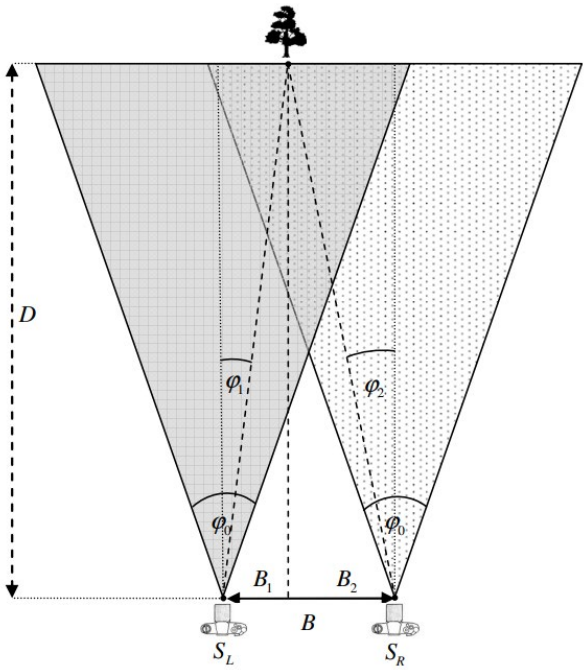
Dit voorbeeld geeft een goed overzicht van veelgebruikte methodes binnen de literatuur. Het toegepaste principe is overal gelijkaardig en gebaseerd op het nagaan van de dispariteit tussen overeenstemmende gedetecteerde objecten. Op linker- en rechterbeeld binnen het gebied van stereo visie afstandsbepaling.

2.2.4 Afstandsbepaling

Stereoscopische beelden stellen ons in staat de afstand te berekenen tussen de camera's en het gekozen object [6]. Dit op basis van het verschil in afstand tussen het object van belang op linker- en rechter beeld, ookwel gekend als dispariteit, in combinatie met de afstand tussen de twee camera's en de brandpuntsafstand van de camera's [23]. Het bepalen van de effectieve afstand vormt een laatste stap binnen het proces van stereovisie afstandsbepaling.

Ter verduidelijking wordt er een voorbeeld gegeven en besproken. Hierbij wordt er gebruik gemaakt van de relaties tussen de zijden en hoeken van driehoeken, ook wel driehoeksmeetkunde of trigonometrie genoemd.

Stel dat, zoals weergegeven op Figuur 31, het rechter beeld wordt genomen op positie S_R en het linker beeld op positie S_L. B stelt de afstand voor tussen de twee camera's en φ_0 is de horizontale beeldhoek van de camera's. De positie van het voorwerp (afstand 'D') kan hieruit worden berekend door enkele meetkundige afleidingen uit te voeren. [6]



Figuur 31 Foto van een object (boom) vastgelegd met twee camera's [6]

De afstand 'B' kan uitgedrukt worden als de som van de afstanden 'B_1' en 'B_2', als de optische assen van de camera's evenwijdig zijn. φ_1 en φ_2 stellen de hoeken voor tussen de optische assen van de cameralenzen en het gekozen object [6]. De formule voor het berekenen van de afstand 'B' ziet eruit als volgt:

$$B = B_1 + B_2 = D \times \tan\varphi_1 + D \times \tan\varphi_2$$

Vergelijking 2 Afstand tussen camera's [6]

Wanner we deze vergelijking herwerken naar de afstand 'D', dan ziet deze er al volgt uit:

$$D = \frac{B}{\tan\varphi_1 + \tan\varphi_2}$$

Vergelijking 3 Herwerking Vergelijking 2 naar afstand 'D' [6]

Door gebruik te maken van Figuur 32, Figuur 33 en basis trigonometrie, een tak van de wiskunde die zich bezighoudt met de relaties tussen de zijden en hoeken van driehoeken en de verhoudingen ervan, zijn de volgende verbanden te vinden:

$$\frac{x_1}{\frac{x_0}{2}} = \frac{\tan\varphi_1}{\tan\frac{\varphi_0}{2}}$$

Vergelijking 4 Toepassing van basis trigonometrie op Figuur 32 [6]

$$\frac{-x_2}{\frac{x_0}{2}} = \frac{\tan\varphi_2}{\tan\frac{\varphi_0}{2}}$$

Vergelijking 5 Toepassing van basis trigonometrie op Figuur 33 [6]

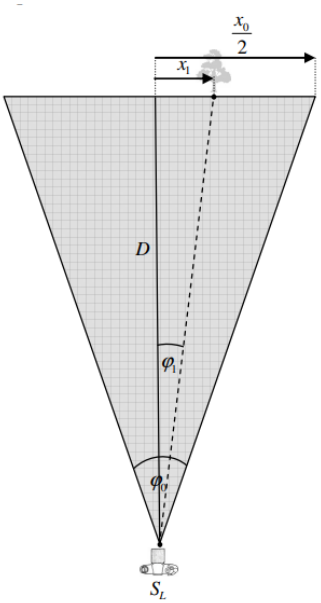
Hierbij stelt 'x' het horizontale aantal pixels voor, verduidelijkt op Figuur 32 en Figuur 33 [6].

Als de afstand tussen de camera's (B), het aantal horizontale pixels (x_0), de kijkhoek van de camera (ϕ_0) en het horizontale verschil tussen hetzelfde voorwerp op beide beelden (X_L-X_D) bekend zijn, dan kan de afstand tot het voorwerp (D) op volgende manier bepaald worden [6]:

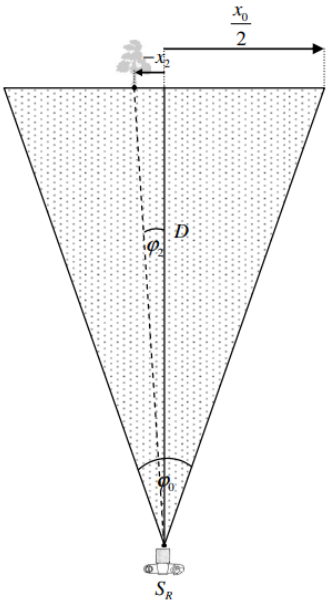
$$D = \frac{B \times X_0}{2 \times \tan\frac{\varphi_0}{2} \times (X_L - X_D)}$$

Vergelijking 6 Afstand 'D' camera's tot object [6]

Onderstaande figuren Figuur 32 en Figuur 33 stellen schematisch foto's voor, genomen met de linker en rechter camera.



Figuur 32 Foto van het object (boom) genomen met de linker camera [6]



Figuur 33 Foto van het object (boom) genomen met de rechter camera [6]

De nauwkeurigheid van de berekende positie (afstand D) hangt af van verschillende variabelen. De plaats van het voorwerp in het rechter plaatje kan tot op één pixel nauwkeurig worden bepaald (zie Figuur 34). Elke pixel komt overeen met een beeldhoek, waarvan de grootte afgeleid kan worden uit Vergelijking 7. φ_0 stelt hierbij de horizontale beeldhoek van de camera voor en x_0 de beeld resolutie (in pixels) [6].

$$\Delta\varphi = \frac{\varphi_0}{X_0}$$

Vergelijking 7 Verhouding tussen pixel en beeldhoek [6]

Figuur 34 toont de beeldhoek $\Delta\varphi$ van één pixel die resulteert in een afstandsfout ΔD [6]. Uit Figuur 34 kunnen we volgend verband halen:

$$\frac{\tan\varphi}{\tan(\varphi - \Delta\varphi)} = \frac{\Delta D + D}{D}$$

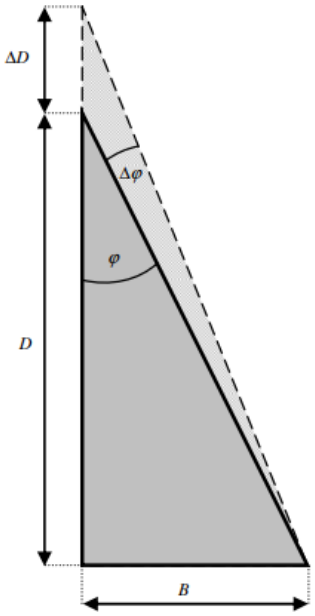
Vergelijking 8 Toepassing basis trigonometrie op Figuur 34 met aanwezigheid van afstandsfout ΔD [6]

Uit bovenstaande Vergelijking 8 kunnen we de afstandsfout ΔD halen. De vergelijking ziet er als volgt uit:

$$\Delta D = \frac{D^2}{B} \times \tan\Delta\varphi$$

Vergelijking 9 Berekening afstandsfout ΔD [6]

De effectieve fout kan in werkelijkheid groter zijn door optische fouten of onjuistheden (vervormingen, aberratie enz.) [6].



Figuur 34 Afstandsfout veroorzaakt door een plaatsingsfout van 1 pixel [6]

Binnen de literatuur worden er nog andere voorbeelden aangehaald voor het bepalen van afstanden aan de hand van stereovisie, maar deze berusten vrijwel allemaal op dezelfde principes uit de driehoeksmeetkunde of trigonometrie.

Een eenvoudige andere methode die terug te vinden is in bestaande voorbeelden, maakt enkel gebruik van de dispariteit, brandpuntsafstand uitgedrukt in pixels en de afstand tussen beide camera's. De gebruikte formule voor de berekening van de diepte (Z-coördinaat) ziet er als volgt uit:

$$Diepte\ (Z) = \frac{(afstand\ tussen\ camera's) \times (brandpuntsafstand\ [pixels])}{dispariteit}$$

Vergelijking 10 Eenvoudige berekening diepte (Z)

Op basis van gegevens uit de afstandsbepaling (brandpuntsafstand lenzen, afstand tussen camera's en de dispariteit op beeld) kunnen naast de diepte tussen het vlak van de camera's en het vlak van het gedetecteerde object (Z) ook nog de horizontale (X) en verticale verplaatsing (Y) ten opzichte van het middelpunt (midden tussen beide camera's) bepaald worden. Dit resulteert in de mogelijkheid van het geheel tot het bepalen van de X-coördinaat (horizontale verplaatsing), Y-coördinaat (verticale verplaatsing) en Z-coördinaat (afstand tussen vlak camera's en vlak gedetecteerd object) van een gedetecteerd object van belang binnen de 3D-ruimte op basis van stereovisie.

3 Ontwikkeling

In dit deel van de scriptie wordt de effectieve ontwikkeling van het 3D-posititiebepalingssysteem op basis van stereovisie uitgewerkt en waar nodig verder verduidelijkt. Het geeft een beeld van alle stappen die ondernomen worden om tot het werkelijke resultaat en doel van deze masterproef te komen in combinatie met extra uitleg ter verduidelijking.

Wat software betreft werd er, voor de ontwikkeling van het 3D-positiebepalingssysteem, gebruik gemaakt van Python (versie 3.9.13) met bijhorende functiebibliotheken (OpenCV, Pandas, Numpy, time, os, etc.). Als programmeeromgeving werd er gebruik gemaakt van Visual Studio Code. Meer uitleg omtrent de software en gebruikte functies is terug te vinden binnen de volgende delen.

3.1 Hardware

De hardware voor deze masterproef werd aangeleverd door Aperam zelf. Gezien Aperam het systeem op termijn wil installeren op de meeste/alle rolbruggen, wenst Aperam de kost van het geheel zo laag mogelijk te houden. Aperam telt namelijk heel wat rolbruggen binnen hun S&S Center en productiesite te Genk, maar ook in alle andere wereldwijde vestigingen. Hierdoor werd er door Aperam Genk gekozen om te werken met 2 IP-Camera's.

Een Internet Protocol-camera, algemeen bekend als een IP-camera, is een digitale videocamera, vergelijkbaar met een webcam, die gegevens verzendt en ontvangt via een netwerk of het internet. In tegenstelling tot een gewone webcam is het een op zichzelf staand apparaat met een uniek IP-adres dat alleen een netwerkverbinding nodig heeft om afbeeldingen te verzenden. IP-camera's maken op dezelfde manier verbinding met een netwerk als elk ander standaard netwerkapparaat zoals een laptop, tablet of printer. Een IP-camera legt een afbeelding vast zoals een digitale camera, comprimeert het bestand en verzendt het over het netwerk. IP-camera's kunnen worden gebruikt op een bekabeld netwerk dat is aangesloten op een modem of breedbandrouter via een Ethernet-kabel, of draadloos via een WiFi-router [24].

Aperam leverde voor de masterproef 2 Hikvision IP-camera's aan. Het ging hierbij over het DS-2CD2643G2-IZS model, dat afgebeeld wordt op Figuur 35 en waarvan de specificaties terug te vinden zijn in *Bijlage A* van deze scriptie. De camera's hebben vrij te kiezen instellingen via hun netwerkinterface en ondersteunen Power over Ethernet (PoE), dat gebruikt kan worden voor het voeden van de camera's.

Figuur 35 Hikvision DS-2CD2643G2-IZS IP-camera [25]

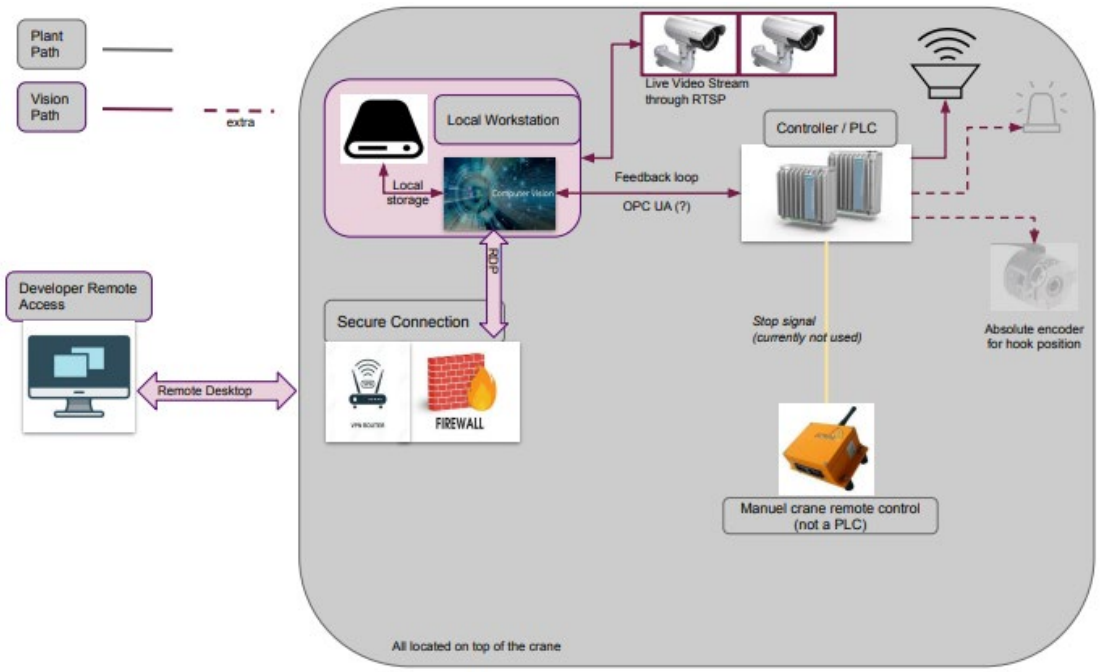
Om beide IP-camera's van voeding en een connectie te kunnen voorzien werd er door Aperam ook een ethernet switch aangeleverd. Dit is een apparaat dat in staat is om op het gebied van netwerkinfrastructuur verschillende netwerksegmenten, zoals ethernet verbindingen in dit geval, met elkaar te verbinden tot één homogeen netwerk (OSI-datalinklaag). Voor deze masterproef ging het om het CBS250-8PP-D model van de Cisco Business 250 reeks, zoals afgebeeld op Figuur 36. Dit is een slimme ethernet switch met 8 gigabit ethernet poorten die Power over Ethernet (PoE) ondersteund [26]. De ethernet poorten kunnen dus gebruikt worden voor zowel communicatie als voeding.

Figuur 36 Cisco CBS250-8PP-D netwerk switch [26]

Naast het aanleveren van hardware zoals de IP-camera's en een ethernet switch werden ook de nodige ethernetkabels, voeding en allerlei andere kleinigheden zoals schroeven, boutjes, etc. voorzien.

3.2 Opstelling

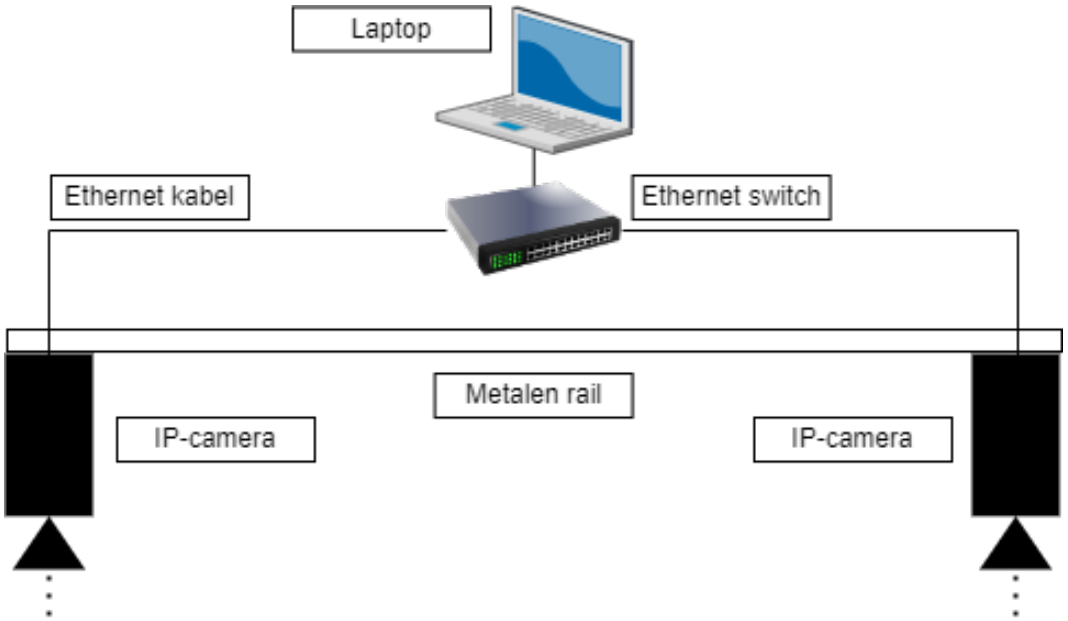
In eerste instantie was het de bedoeling dat de ontwikkeling van het 3D-positibepalingssysteem onmiddellijk zou gebeuren aan de hand van hardware geïnstalleerd op een rolbrug binnen het Service & Solutions Center. De infrastructuur zou er dan uitzien zoals te zien op het schema op Figuur 37. Voor de uitwerking van het 3D-positiebepalingssysteem zou er dan ook gebruik gemaakt kunnen worden van een desktop PC op de rolbrug, met aanwezigheid van een grafische kaart (GPU) van het merk NVIDIA, ter snellere verwerking van de beelden.



Figuur 37 Infrastructuur rolbrug

Helaas kon door een vertraging binnen Aperam de hardware (IP-camera's, ethernet switch, desktop PC, externe wifi-antenne, etc.) niet tijdig, binnen de periode van deze masterproef, op de rolbrug geïnstalleerd worden om hier gebruik van te kunnen maken.

Voor de ontwikkeling van het 3D-positiebepalingssysteem op basis van stereovisie werd er dus gebruik gemaakt van een tijdelijke opstelling, zoals schematisch verduidelijkt op Figuur 38 en afgebeeld op Figuur 39. Een verschil ten opzichte van de installatie op de kraan is dat de IP-camera's bij deze opstelling platliggen in horizontale richting ten opzichte van een hangende positie aan de kat van de rolbrug, zoals reeds eerder verduidelijkt in deel 1.2. Verder word er bij deze opstelling gebruik gemaakt van een eigen laptop (HP EliteBook 850 G4), zonder grafische kaart, en zal er waarschijnlijk een verschil zijn in de afstand tussen beide IP-camera's. Naast een verschil in computerkracht en waarschijnlijk ook snelheid, hebben deze verschillen geen invloed op de ontwikkeling en functionaliteit van het geheel.



Figuur 38 Schematische voorstelling opstelling systeem

Figuur 39 Stereovisie cameraopstelling

Verder in deze scriptie worden geregeld de namen linker en rechter camera gebruikt, met bijhorend linker en rechter beeld. De bepaling van links en rechts gebeurde voor deze masterproef vanuit het perspectief van een persoon die, al kijkend naar de camera's, voor de camera's staat. Figuur 39 verduidelijkt dit. De linker camera bevindt zich op deze figuur links en de rechter camera rechts.

3.2.1 Camera uitlijning en positionering

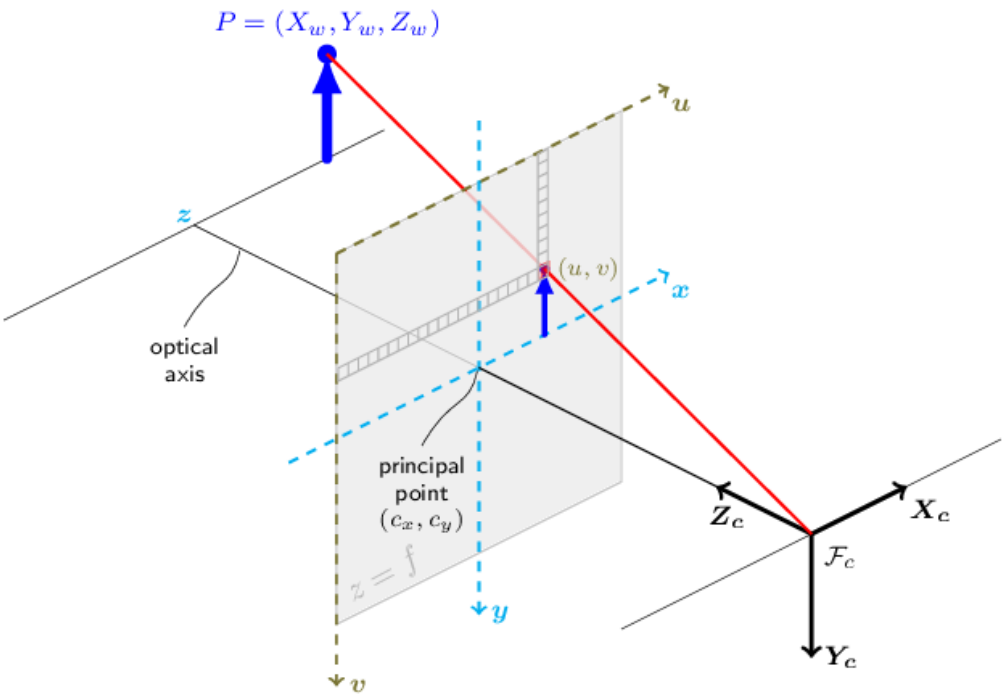
Zoals te zien op het schema op Figuur 38 en de voorstelling van de opstelling op Figuur 39 zijn beide IP-camera's vastgemaakt aan de uiteindes van een metalen rail. Deze rail werd voorzien door Aperam ter fixatie van de opstelling. De optische assen van beide camera's, aangegeven met de korte stippellijn onder elke camera op Figuur 38, bevinden zich bij deze opstelling 71 cm van elkaar. Er is gekozen voor een parallelle cameraopstelling, gezien een object zich niet altijd op een vaste positie voor de camera's zal bevinden en er dus geen vast convergentiepunt of convergentieafstand is. Uit de literatuurstudie blijkt dat dit ook de meest voorkomende en gebruikte cameraopstelling voor stereovisie afstandsbepaling is. De theoretische achtergrond rond camera plaatsing werd reeds eerder aangehaald in deel 2.2.1. Bij de ontwikkeling van het 3D-positiebepalingssysteem aan de hand van de tijdelijke opstelling werd er rekening gehouden met de mogelijke toepassing ervan onder de rolbrug.

Bij het fysiek positioneren van de camera's is er zo veel mogelijk rekening gehouden met de parallelle plaatsing ervan. De camera's werden zowel in horizontale als verticale richting zo nauwkeurig mogelijk manueel uitgelijnd ten opzichte van elkaar. Ook werd er rekening gehouden met de rotaties van beide camera's ten opzichte van elkaar. Dit gezien stereovisie afstandsbepaling, en dus ook positiebepaling, enkel toegepast kan worden binnen het gebied van overlap van beide beelden. De grootte van dit gebied is optimaal zo groot mogelijk. Vandaar ook de horizontale camera plaatsing. In ideale omstandigheden is er hierdoor ook enkel een verschil in horizontale verplaatsing tussen beide beelden. Dit vergemakkelijkt in een volgende fase ook de berekening van de dispariteit. Fouten op het gebied van camera positionering worden in een volgende fase, aan de hand van kalibratie en beelduitlijning (translatie, rotatie) verwijderd. Dit om 2 beelden te hebben die zowel horizontaal als verticaal perfect met mekaar uitgelijnd zijn (parallelle optische assen) en ook wat rotatie betreft geen verschil ten opzichte van elkaar hebben. Enkel een horizontaal verschil op beeld ten gevolge van het verschil in horizontale plaatsing mag zichtbaar zijn. Op basis hiervan kan dan ook de dispariteit (in horizontale richting) bepaald worden. Figuur 40 toont een duidelijk voorbeeld van 2 correct uitgelijnde beelden met enkel een horizontale verplaatsing. De rode horizontale lijnen op beeld zijn epipolaire lijnen. Dit zijn snijpunten van het epipolaire vlak met het beeldvlak. Gezien de correcte camerapositionering, met eventuele beelduitlijning, lopen deze onderling parallel en verticaal uitgelijnd met elkaar op beide beelden. Dit toont een correcte parallelle positie van beide optische assen aan. De epipolaire lijnen op gelijke hoogte op beide beelden verbinden ook dezelfde punten binnen de 3D-ruimte me elkaar.

Figuur 40 Correct voorbeeld camera-/beelduitlijning [27]

Naast een zo nauwkeurig mogelijke uitlijning van beide camera's werden beide camera's ook waterpas en op gelijke afstand van de muur geplaatst. Dit gezien de afstand (Z) tot het vlak, waarin het object van belang zich bevindt, bepaald wordt in combinatie met de verplaatsing in horizontale (X) en verticale (Y) richting ten opzichte van het centrumpunt van het vlak (projectiecoördinatenstelsel). Dit centrumpunt bevindt zich in het midden tussen de snijpunten van beide optische assen met het vlak en heeft dus als X- en Y-coördinaat 0. De waterpas positie van de uitgelijnde camera's zorgt ervoor dat de oriëntatie van het vlak, waarin het object van belang zich bevindt en dat loodrecht staat op de parallelle optische assen, zich in een optimale en makkelijke positie bevindt voor de latere verificatie van het systeem aan de hand van testen. Het vlak zou door de waterpas en gelijke (evenwijdige) plaatsing evenwijdig zijn aan 2 muren binnen de ruimte. Dit vergemakkelijkt het uitvoeren van testen ter verificatie van het systeem. Het objectvlak is evenwijdig aan het cameravlak.

Deze plaatsing van de camera's heeft geen invloed op de werking van het systeem, maar er dient wel rekening gehouden te worden met het feit dat de afstand tot het vlak, waarin het object van belang zich op een bepaalde (X, Y) (projectiecoördinatenstelsel) positie bevindt, bepaald wordt, en niet de afstand van het midden tussen beide camera's tot het object zelf. Tenzij het object zich in het centrumpunt (hoofdpunt) (coördinaat (0,0)) bevindt natuurlijk. Dit principe wordt verduidelijkt op Figuur 41 aan de hand van slechts 1 camera. Linksonder wordt het coördinatenstelsel van de camera's duidelijk weergegeven. Dit is ook de positie van de eigenlijke camera. De Z-as stelt de optische as van de camera voor. Het grijze vlak, dat het projectiebeeldvlak voorstelt, staat duidelijk loodrecht op deze optische as, met als snijpunt van beide het middelpunt (hoofdpunt), waar het later besproken gerectificeerde projectieassenstelsel (projectiecoördinatenstelsel) (lichtblauwe assen), gevestigd is. Dit wordt aangegeven als *'principal point'* met als coördinaten (Cx, Cy). Het punt P stelt de positie van het object van belang voor, dat zich op een bepaalde (X, Y) positie binnen het vlak, dat zich loodrecht op de optische as op een afstand Z van de camera's bevindt. De projectie van dit 3D-punt op het 2D-projectievlak wordt weergegeven met de kleine donkerblauwe peil, bij het punt (u, v) binnen het projectievlak met projectiecoördinatenstelsel.



Figuur 41 Verduidelijking stereovisie vlakken (pinhole cameramodel) [27]

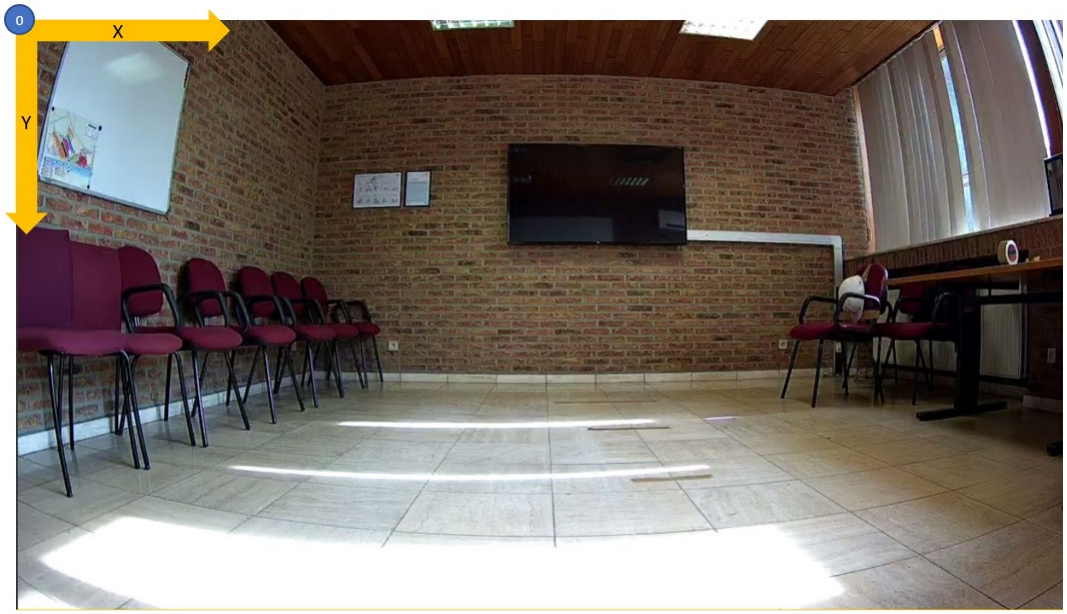
3.3 Vastleggen stereobeelden

Voor het kunnen toepassen van 3D-positiebepaling op basis van stereovisie zijn er in een eerste fase een paar stereobeelden nodig. Dit zijn 2 beelden, met liefst een zo groot mogelijke overlap, vanuit een ander perspectief. Het kan hier gaan over live beelden, video's of een paar foto's. Een goed voorbeeld van een stereobeeld is het zicht van menselijke ogen. Deze bekijken de omgeving, met een grote overlap tussen beide beelden, vanuit een ander perspectief. Figuur 42 geeft een duidelijk voorbeeld van een paar stereo beelden. Deze werden genomen met de cameraopstelling, gebruikt binnen deze masterproef, zoals reeds eerder aangehaald in deel 3.2. Het verschil in perspectief bij deze opstelling, bedraagt een verschil in horizontale verplaatsing van 71 cm tussen de optische assen van beide camera's.

Figuur 42 Paar stereobeelden

Beide beelden (stereo) worden in Python verkregen door gebruik te maken van functies uit de Python OpenCV functiebibliotheek. OpenCV (Open Source Computer Vision Library) is een open source computer visie en machine learning softwarebibliotheek. Deze werd ontwikkeld om een gemeenschappelijke infrastructuur voor computervisietoepassingen aan te bieden en hiermee het gebruik van computervisie, voor commerciële doeleinden, te vergemakkelijken en versnellen [28].

Het beeldcoördinatenstelsel op beeld (2D) ziet er binnen Python uit, zoals verduidelijkt op Figuur 43, met als maximale waarden voor X en Y de beeldresolutie.



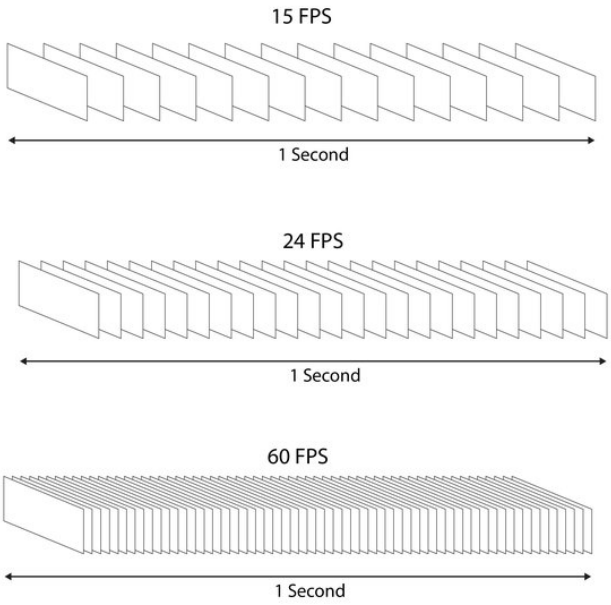
Figuur 43 Verduidelijking (beeld-)coördinatenstelsel binnen Python

Voor live beelden van de IP-camera's wordt er naast het gebruik van functies uit de OpenCV functiebibliotheek gebruik gemaakt van het 'Real Time Streaming Protocol' (RTSP). Het Real Time Streaming Protocol is een applicatielaag protocol (OSI-model) dat wordt gebruikt om de levering van gegevens met realtime-kenmerken te regelen. RTSP biedt een uitgebreid raamwerk dat de realtime, on-demand levering van gecontroleerde gegevens zoals audio en video mogelijk maakt. De gegevensbron kan zowel een live datafeed zijn, als een opgeslagen datafeed. Het protocol is in staat tot het regelen van meerdere sessies voor gegevenslevering, biedt een manier om leveringskanalen zoals UDP, Multicast UDP, TCP te selecteren en is in staat een leveringsmechanisme, op basis van Real Time Transport (RTT), te selecteren [29].

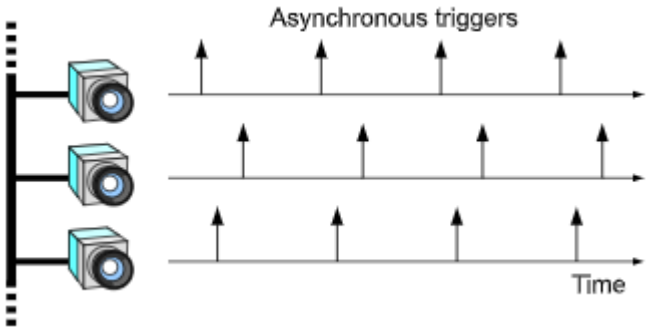
Naast het gebruik van live stereobeelden voor het toepassen van 3D-positiebepaling kan er ook gebruik gemaakt worden van een stereo video-opname of een enkel paar stereo afbeeldingen. Voor een volgende fase, waarin de effectieve 3D-postie en afstanden bepaald worden, is het wel van belang enkele beeld-, camera- en opstellingsparameters te kennen of kunnen berekenen. Dit wordt later verder verduidelijkt.

Wat belangrijk is voor het zo nauwkeurig mogelijk bepalen van 3D-positites op basis van stereovisie, bij bewegende objecten of beelden, is de synchronie tussen beide beelden. Een beeld bestaat uit een aaneenschakeling van frames. Dit zijn momentopnames. Door deze momentopnames aan een bepaalde snelheid na elkaar te tonen wordt een bewegend beeld verkregen. De snelheid van het na elkaar tonen van frames wordt de framesnelheid genoemd en wordt uitgedrukt in het aantal frames per second (fps). Een beeld is dus eigenlijk niet continu, maar bestaat uit een aaneenschakeling van momentopnames aan een bepaalde snelheid. Dit wordt verduidelijkt op Figuur 44, waar elk wit vlak 1 momentopname voorstelt.

De Hikvision IP-camera's die binnen deze masterproef gebruikt worden bieden een maximale framesnelheid van 30 frames per seconde aan. Dit aan een maximale resolutie van 1688 x 1520. Bij het synchroon lopen van beide beelden, en bijgevolg dus ook momentopnames, is alles in orde en zullen er, wat beeldsynchronisatie betreft, geen onnauwkeurigheden veroorzaakt worden binnen de positiebepaling. Bij het niet synchroon lopen van beide beelden, en bijgevolg ook momentopnames, daarentegen, kunnen er wel onnauwkeurigheden en afwijkingen op de uiteindelijke resultaten bekomen worden. In de tijd tussen het maken van een momentopname met beide camera's voor het maken van een stereo momentopname, kan er een beweging plaatsgevonden hebben. Dit zou zowel de verplaatsing van een object, als een verplaatsing van de camera's kunnen zijn. Dit verschil, dat klein of groot kan zijn afhankelijk van de hoeveelheid beweging en de framesnelheid, zal een verschil tussen beide beelden opleveren bij een paar stereobeelden. Beide momentopnames van elke camera komen namelijk niet van op hetzelfde moment in tijd. Dit zal een effect hebben op de uiteindelijke resultaten van het 3D-positiebepalingsmodel. Het niet synchroon lopen van 2 camera's, en dus ook momentopnames; wordt verduidelijkt op Figuur 45. Er is tussen camera 1 en camera 2 een duidelijk verschil in timing merkbaar. In deze tussentijd is er een kans op beweging, en dus een kans op een onjuist stereobeeld. De oorzaak van het niet synchroon binnenkomen van beide beelden kan te maken hebben met het niet synchroon lopen van beide camera's, vertragingen over het netwerk, het niet synchroon ophalen van beide beelden, etc [30].



Figuur 44 Verduidelijking framesnelheid (fps) [31]

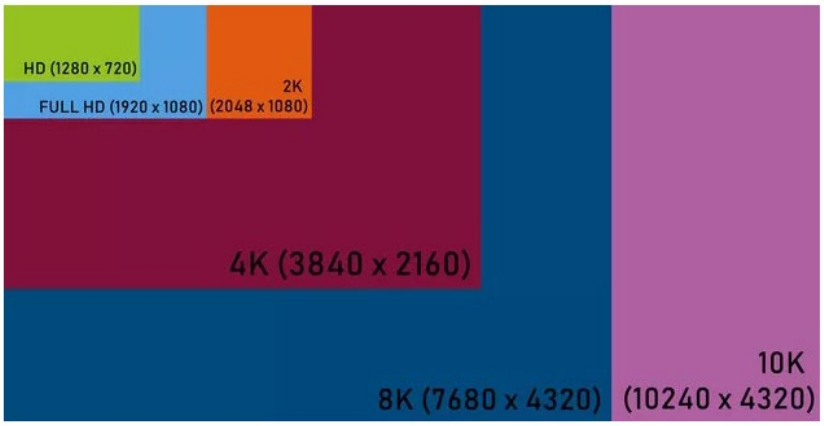


Figuur 45 Verduidelijking frame synchronisatie probleem (asynchrone frames) [30]

3.3.1 Camera instellingen

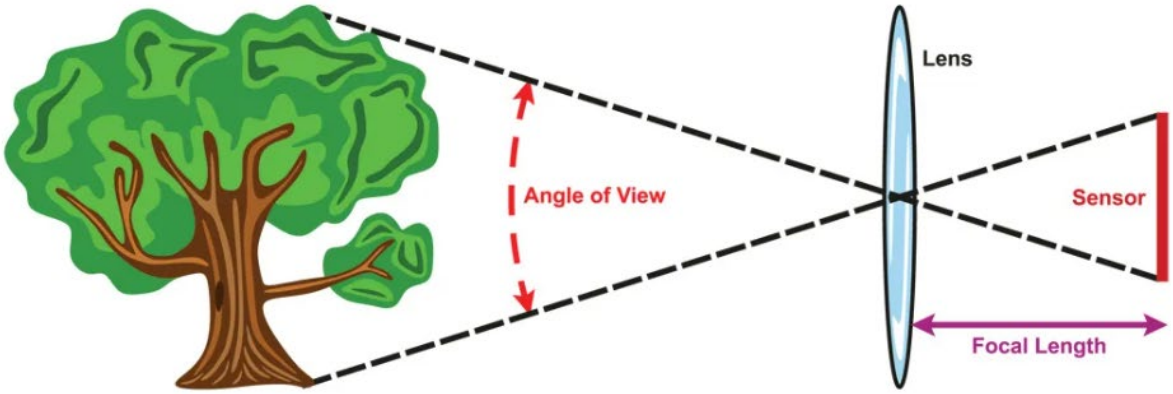
Binnen deze masterproef wordt er voor het verkleinen van eventuele fouten door het niet synchroon lopen van beide live beelden, gewerkt met een zo hoog mogelijke framesnelheid van beide camera's. Namelijk 30 frames per seconde (fps). De tijd tussen elk frame, bij elke camera, is hierbij slechts 0,0333 seconden. Het verschil in tijd tussen frames van beide camera's, bij een asynchrone werking, zou dus ook maximum 0,0333 seconden kunnen bedragen bij gelijke instellingen. Bij het ophalen van de camerabeelden via het RTS-protocol voor het vormen van een stereopaar, wordt van elke camera het laatst mogelijke frame (momentopname) gebruikt. Het verschil in tijd tussen het ophalen van beide laatste frames zou, gezien de verwerking, bij het gebruik van deze methode een invloed kunnen hebben op de resultaten. Ook vertragingen binnen het (lokale) netwerk en door de camera's zelf zijn jammer genoeg niet uitsluitbaar. De verwerkingscapaciteiten en verwerkingssnelheid van de gebruikte computer spelen hierbij ook een belangrijke rol. Gezien het gebruik van het uiteindelijke 3D-positiebepalingssysteem, binnen de digitalisatie van het veiligheidshulpmiddel, wordt er verkozen om te werken met zo real time mogelijke beelden (live). Het synchroon lopen van beide beelden speelt enkel een belangrijke rol bij bewegende objecten en/of beelden. Bij statische objecten en beelden speelt het verschil in timing tussen beide momentopnames van een stereobeeld weinig tot geen rol.

Verder wordt er binnen deze masterproef gebruik gemaakt van beelden met een gelijke resolutie en bijhorende beeldverhouding. De resolutie van een beeld geeft aan uit hoeveel pixels per rij en kolom een beeld is opgebouwd. Elke pixel bestaat uit één enkele kleur en dus één blokje aan informatie. De verhouding tussen het aantal pixels in breedte (rij) en het aantal pixels in hoogte (kolom) wordt de beeldverhouding genoemd [32]. Op Figuur 46 zijn enkele veel voorkomende resoluties te zien met hun bijhorende naam. In een volgende fase van het 3D-positiebepalingssysteem worden afstanden bepaald op basis van coördinaten (pixels) van detecties op beeld. Indien beide beelden geen gelijke resolutie hebben worden deze automatisch verschaald naar een gelijke, zelf in te stellen, resolutie. Voor de ontwikkeling van het geheel aan de hand van de tijdelijke testopstelling wordt er gebruik gemaakt van een resolutie van 1280 x 720 (HD) met een bijhorende beeldverhouding van 16:9 (breedte:hoogte). Andere resoluties, met bijhorende beeldverhoudingen zijn zeker ook mogelijk, maar kunnen een invloed hebben op de verwerkingssnelheid van het geheel. De zone van overlap tussen beide beelden is in ideale omstandigheden best zo groot mogelijk. In het verder verloop van de ontwikkeling en uitwerking van de scriptie wordt er rekening gehouden met de gelijke beeldresolutie van 1280 x 720 voor beide beelden. Zo zal het latere objectdetectiemodel ook op deze beeldresolutie getraind worden.



Figuur 46 Enkel veel voorkomende beeldresoluties met bijhorende naam [32]

Wat brandpuntafstanden van de camera's betreft wordt er, gezien de variabele brandpuntsafstanden van beide Hikvision IP-camera lenzen (2.8 – 12 mm [25]), die niet manueel in te stellen zijn, één maal individueel (automatisch) gefocust binnen de ruimte waar deze opengesteld stonden. Dit alvorens het uitvoeren van de volgende stap binnen de ontwikkeling van het geheel, namelijk de kalibratie stap. De brandpuntsafstand is de afstand, in millimeters, tussen het optische centrum van de cameralens en de beeldsensor, waar alle beeldinformatie wordt vastgelegd. Als het licht de voorkant van de lens binnenkomt, wordt het gebogen en vervormd door de lens zodat het convergeert naar één bepaald punt. Het optische centrum of brandpunt genoemd [33]. Dit wordt verduidelijkt op Figuur 47, waar het rode vlak met bijschrift 'Sensor' de beeldsensor voorstelt en de afstand, aangeduide met een paarse pijl en bijschrift 'Focal Length' de brandpuntsafstand. Zoals te zien op Figuur 47 beïnvloedt de brandpuntsafstand ook de kijkhoek van de camera. Dit wordt verduidelijkt door de rode pijl met bijschrift 'Angle of View'. Hoe groter de brandpuntsafstand, hoe dichter bij het beeld en hoe kleiner de beeldhoek. Hoe kleiner de brandpuntsafstand, hoe verder het beeld en hoe groter de beeldhoek [33].



Figuur 47 Verduidelijking brandpuntsafstand [33]

Optimaal hebben beide camera's binnen stereovisie dezelfde brandpuntsafstand, maar dit is geen vereiste gezien er hier binnen de kalibratiestap rekening mee gehouden kan worden [34]. Dit is het geval binnen deze masterproef. Het uitvoeren van de eenmalige autofocus stap geeft in normale omstandigheden bij een gelijk zicht van beide camera's geen groot verschil in brandpuntsafstand.

3.4 Stereo kalibratie en rectificatie

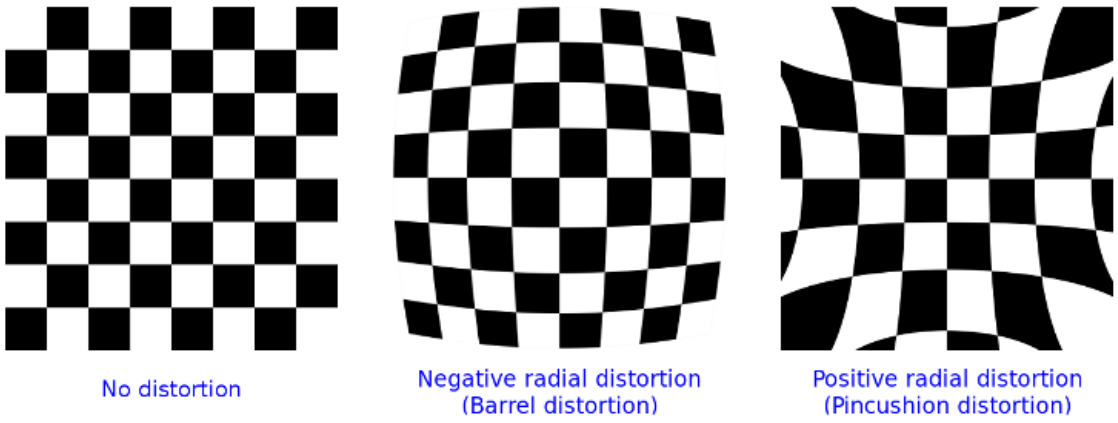
Een volgende stap binnen de uitwerking van het 3D-positiebepalingssysteem is het toepassen van stereo kalibratie en stereo rectificatie. Dit ter bepaling van enkele belangrijke parameters voor het oplossen van lensvervormingen en onnauwkeurigheden bij de parallelle camera uitlijning en het leggen van verbanden tussen de werkelijke 3D-ruimte en de 2D-camerabeelden.

In deel 3.2 werd de cameraopstelling met bijhorende positionering en uitlijning al grondig besproken. Er werd hierin uitgelegd dat de camera's langs elkaar (horizontaal) gepositioneerd staan met een tussenafstand van 71 cm tussen beide optische assen. Wat positionering betreft werd er gekozen voor een parallelle opstelling, met parallel verloop van beide optische assen. De camera's werden hierbij zo nauwkeurig mogelijk verticaal, en wat rotaties betreft, uitgelijnd. Als beide camera's correct verticaal uitgelijnd zijn, dan zal het gedetecteerde object zich verticaal (Y-coördinaat) op dezelfde hoogte bevinden op beide beelden. Dit vergemakkelijkt de latere berekening van diepte (Z-coördinaat) tot het vlak, en de positie van het gedetecteerde object binnen dit vlak (X- en Y-coördinaat), gezien er wat berekening van de dispariteit betreft, enkel nog naar het verschil in X-coördinaat (horizontaal, in pixels op beeld) gekeken moet worden. Dichtbijgelegen objecten, zullen hierbij een groter verschil in pixels (dispariteit) hebben [35].

3.4.1 Reden

Het manueel uitlijnen van beide camera's biedt geen zekerheid op een correcte positionering. Ook beeldvervormingen door de lens dienen opgelost te worden, om aan een correcte 3D-positiebepaling te kunnen doen. Dit gezien krommingen op beeld een vertekend beeld kunnen geven. Zeker wanneer een object zich binnen een zone bevindt, waar veel van deze krommingen aanwezig zijn. Het werkelijke beeld wordt hier samengedrukt of uitgetrokken, waardoor er een verschil is tussen pixels op beeld en afstanden in de werkelijke 3D-ruimte, binnen het beeld. Figuur 48 toont 2 voorbeelden van veelvoorkomend types beeldvervorming. Het gaat hier over negatieve beeldvervorming (midden) en positieve beeldvervorming (rechts). Hierbij is een duidelijke vervorming (uitrekking en samendrukking) te zien ten opzichte van het originele patroon (links). De witte en zwarte vierkanten met gelijke grote, te zien op het meest linkse

vlak, komen door de vervorming niet meer overeen op het middelste en meest rechtse vlak. Pixels zijn uitgetrokken of samengedrukt ten opzichte van het originele beeld (links) en afstanden ten opzichte van het werkelijke vlak (links) kloppen niet meer. Binnen de stereovisie 3D-positiebepaling, waar informatie uit pixels en afstanden gehaald wordt, vormt dit geen gunstig effect. Wat duidelijk merkbaar is, is dat de beeldvervormingen toenemen naarmate men van het centrum van een beeld naar de buitenkant toe gaat. De hoeveelheid beeldvervorming en het type hangt af van de lens van de camera.



Figuur 48 Twee veelvoorkomende types beeldvervorming (links: geen vervorming; midden: negatieve vervorming; rechts: positieve vervorming) [27]

Bij het observeren van een frame, te zien op Figuur 49, gemaakt door één van de Hikvision IP-camera's van de opstelling, is duidelijk negatieve beeldvervorming zichtbaar. De scheiding tussen de achterste muur van de ruimte en het plafond (bruin) vertoont een duidelijke kromming, terwijl dit in werkelijkheid een rechte horizontale lijn is.

Figuur 49 Origineel beeld Hikvision IP-camera, zonder beeldvervorming

Een laatste en misschien wel belangrijkste reden voor het toepassen van stereo kalibratie en stereo rectificatie is het leggen van het verband tussen de werkelijke 3D-ruimte en beide 2D-beelden. Ook zullen er uit deze stappen heel wat belangrijke parameters omtrent de opstelling, camera's, beeldvlakken, assenstelsels, coördinatenstelsels, werkelijke ruimte, etc gehaald kunnen worden.

3.4.2 Toepassing

Het softwarematig oplossen van de beeldvervorming en onjuistheden binnen de camera positionering gebeurt aan de hand van verschillende stappen. Binnen deze stappen worden enkele belangrijke parameters bepaald, die later gebruikt worden voor de het effectief oplossen van de beeldvervorming (kalibratie) en camera uitlijning (rectificatie). Het gaat hier over intrinsieke parameters, extrinsieke parameters, projectiematrices, roratiematrices, translatiematrices, etc. Deze worden later in meer detail besproken.

Het is belangrijk te weten dat deze parameter slechts één keer bepaald moeten worden bij een bepaalde opstelling met gelijke instellingen. Het effectief softwarematig oplossen van de beeldvervorming en camera uitlijning gebeurt in een latere fase, frame per frame, afhankelijk van enkele parameters. Bij elke aanpassing, wat opstelling of beeldinstellingen betreft, zullen deze parameters voor deze specifieke situatie opnieuw bepaald moeten worden.

De functies gebruikt binnen de volgende stappen maken gebruik van het homogene coördinatensysteem. Dit is een coördinatenstelsel dat wordt gebruikt binnen de projectieve meetkunde. Het gebruik ervan maakt het mogelijk om punten op oneindig te vertegenwoordigen door eindige coördinaten en vereenvoudigt de formules in vergelijking met zijn cartesische tegenhangers. Men kan een homogene vector verkrijgen door een 1 toe te voegen aan de n-dimensionele cartesische vector. De omzetting van homogene vector naar cartesische vector gebeurt door het delen van alle elementen van de homogene vector door het laatste element [27].

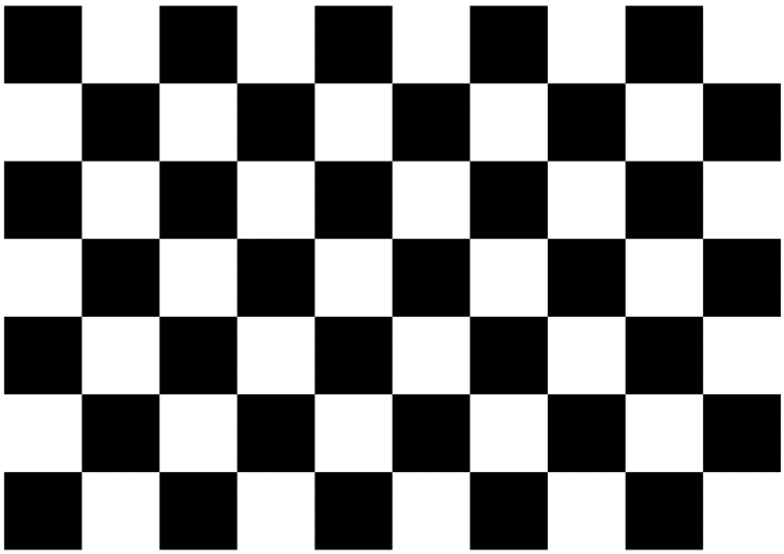
Extra duiding omtrent de gebruikte functies, hun parameters en de algemene werking is terug te vinden op de OpenCV documentatie pagina omtrent ‘Camera Calibration and 3D Reconstruction’ [27].

Binnen de volgende stappen wordt het gehele proces van stereo kalibratie en rectificatie verduidelijkt.

STAP 1: Bepaling object- en beeldpunten

In een eerste stap dienen enkele object- en beeldpunten bepaald te worden. Afhankelijk van deze punten kan het verband gelegd worden tussen een 3D-punt in de werkelijke ruimte (objectpunt) en het 2D-punt op beeld (beeldpunt) [36].

Voor het bepalen van deze punten wordt er gebruik gemaakt van kalibratie patroon, zoals afgebeeld op Figuur 50. Dit patroon dient vlak binnen beeld gehouden te kunnen worden. Binnen de ontwikkeling van het 3D-positiebepalingssysteem werd het kalibratie patroon afgeprint en werd het op een vlak oppervlak geplakt. Een kleine afwijking in vlakte, zal het perspectief verstoren en kan in een volgende fase een fout opleveren. Het is ook belangrijk om de grootte van 1 vierkant te kennen. Dit om later het verband te kunnen leggen tussen afstanden op beeld en afstanden in werkelijkheid.



Figuur 50 Kalibratie patroon [36]

Met het kalibratie patroon op beeld van beide camera's dienen er enkele stereo afbeeldingen gemaakt worden. Hierbij is het belangrijk dat de grootte van het kalibratiepatroon niet gewijzigd wordt. Het patroon moet ook volledig zichtbaar zijn op beide beelden. Tijdens het maken van een paar stereo afbeeldingen mag het patroon ook niet bewegen. Het is dus belangrijk om synchrone stereo afbeeldingen te maken of te zorgen voor een statische positie van het patroon. Dit om ervoor te zorgen dat er geen verschil in werkelijke positie is tussen beide beelden bij een paar stereobeelden [36]. Voor de hoeveelheid stereobeelden die gemaakt moeten worden, met het kalibratie patroon op beeld, is er geen regel. Binnen de ontwikkeling van het systeem werden er telkens een 10-tal stereobeelden gemaakt, waarbij het kalibratiepatroon op een statische plek opgesteld stond. Wat belangrijk is, is het veranderen van positie en rotatie van het patroon op beeld na elk paar stereobeelden. Het patroon moet onder verschillende hoeken en op verschillende afstanden staan omdat er verschillende perspectieven nodig zijn voor het uitvoeren van een zo goed mogelijke kalibratie en latere rectificatie [36]. Ook hier werd binnen de ontwikkeling rekening mee gehouden.

Figuur 51 toont enkele voorbeelden van stereobeelden, gemaakt met de stereovisie opstelling, waar het kalibratiepatroon duidelijk zichtbaar is op ieder stereobeeld en er rekening gehouden werd met de verschillende rotaties en verplaatsingen. Per camera, werd ieder beeld apart opgeslagen in een folder. Op deze stereobeelden is er ook een verschil in camera uitlijning merkbaar. Bij het bekijken van de lampen aan het plafond in de werkelijke 3D-ruimte op beide beelden valt het op dat er een verschil in hoogte op beeld merkbaar is.

Figuur 51 Stereobeelden kalibratie patroon van opstelling

Aan de hand van de stereobeelden van het kalibratie patroon van beide camera's kunnen de object- en beeldpunten bepaald worden. Hiervoor wordt er in eerste instantie gebruik gemaakt van de OpenCV 'findChessboardCorners()' functie. Deze functie gaat per paar stereobeelden op zoek naar de interne hoeken van het kalibratie patroon op beeld. Dit zijn de hoeken van de zwarte/witte vierkanten, aan de binnenkant van het gehele patroon. De hoeken aan de rand van het patroon tellen niet mee [27]. Dit wordt apart uitgevoerd voor het linker en rechter beeld.

Voor de objectpunten, waar het gaat over de locaties van 3D-punten (hoeken) in de werkelijke ruimte, gaat het over dezelfde 3D-punten. Dit gezien het om een stereobeeld van dezelfde scene met statische plaatsing van het kalibratie patroon gaat. Binnen de objectpunten wordt er rekening gehouden met de werkelijke groottes van de vierkanten, die zoals reeds eerder aangegeven gekend moeten zijn of gemeten dienen te worden. De beeldpunten hiertegen verschillen wel van waarden. Er is namelijk een verschil in camera positie tussen beide camera's en hierdoor ook een verschil in perspectief. Het kalibratiepatroon, met bijhorende hoeken, bevindt zich op ieder beeld van het stereobeeld (links/rechts) op een andere positie. Het principe van de object- en beeldpunten wordt ook verduidelijkt op van Figuur 51. Het kalibratie patroon bevindt zich voor elk paar stereobeelden op dezelfde positie in de werkelijke 3D-ruimte (objectpunten), maar het kalibratie patroon bevindt zich op beide 2D-beelden op een andere positie (beeldpunten). Hierdoor worden de beeldpunten ook apart beschouwd voor beide beelden.

Voor het bepalen van de effectieve beeldpunten, wordt er na het detecteren van de hoeken ook nog verder gebruik gemaakt van de OpenCV ‘cornerSubPix’ functie. Deze functie wordt gebruikt voor het vinden subpixel-accurate posities van de interne hoeken van het kalibratie patroon [27]. Dit zijn veel exactere locaties van de hoeken op beeld. Dit met als doel de nauwkeurigheid van de kalibratie- en rectificatiestap te verhogen.

Figuur 52 geeft een visuele weergave van de gedetecteerde interne hoeken van het kalibratie patroon op beeld weer (kleuren), die gebruikt worden voor het bepalen van de object- en beeldpunten. Binnen het Python programma worden deze stereobeelden, met aanduidingen ook weergegeven. Voor het weergeven van de stereobeelden met aanduidingen, wordt er voor ieder beeld (links/rechts) gebruik gemaakt van de OpenCV ‘drawChessboardCorners’ en ‘imshow’ functies. Dit ter manuele evaluatie van de gevonden interne hoeken op de kalibratie patronen. Indien er een stereobeeld tussen zit, dat zichtbaar foutieve resultaten oplevert, dient zowel het linker- als rechterbeeld van dit stereobeeld verwijdert te worden uit de gebruikte folders.

Figuur 52 stereobeeld met aanduiding interne hoeken kalibratie patroon (kleuren)

Zowel de objectpunten (algemeen) als beeldpunten (voor linker en rechter beeld) worden opgeslaan in een arrays en van hieruit verder verwerkt/gebruikt.

STAP 2: Stereo kalibratie: Bepaling parameters

Binnen deze stap, die volgt op de bepaling van de beeld- en objectpunten, worden enkele belangrijke parameters bepaald voor het verwijderen van de beeldvervorming. Ook de rotatie en translatie vectoren, die rotaties en verplaatsingen tussen beide camerabeelden aangeven worden hieruit verkregen. Voor het bepalen van deze parameters wordt er gebruik gemaakt van de object- en beeldpunten en de OpenCV ‘stereoCalibrate’ functie. De OpenCV ‘stereoCalibrate’ functie bepaalt de parameters voor het kalibreren van de stereo cameraopstelling. Dit door het berekenen van de intrinsieke parameters voor elke van de 2 camera’s en de extrinsieke parameters tussen de 2 camera’s [27].

Als input van de functie worden de objectpunten, beeldpunten van de linker camera en beeldpunten van de rechter camera meegegeven. Deze werden bepaald in stap 1. Ook de beeldresolutie, van 1280 x 720 in dit geval, wordt aan de functie meegegeven. Op basis hiervan worden, zoals reeds eerder vermeld, de intrinsieke parameters voor elke van de 2 camera’s en de extrinsieke parameters (rotatie en translatie) tussen de 2 camera’s berekend [27]. Verder kunnen er ook nog specifieke *flags* ingesteld worden met bepaalde extra instellingen of vereisten. Binnen de ontwikkeling van het 3D-positiebepalingssysteem, werden er geen instellingen op basis van flags aangepast.

De output van deze functie bevat heel wat matrices en vectoren met verschillende parameters voor het oplossen van de beeldvervorming en het kalibreren van de stereo cameraopstelling. De gebruikte output van deze functie binnen de ontwikkeling van het 3D-positiebepalingssysteem bedraagt [27]:

- intrinsieke matrix van de linker camera
- intrinsieke matrix van de rechter camera
- vector van vervormingscoëfficiënten van de linker camera
- vector van vervormingscoëfficiënten van de rechter camera
- rotatie matrix
- translatie vector

Verder kunnen er ook nog andere resultaten verkregen worden, zoals de essentiële en fundamentele matrices, maar deze worden verder niet besproken binnen deze scriptie, gezien ze niet gebruikt werden. Extra documentatie hieromtrent is altijd terug te vinden via de OpenCV functiedocumentaties [27].

De intrinsieke cameramatrices van de camera's (A), zoals afgebeeld op Figuur 53, projecteren 3D-punten in het camera-coördinatenstelsel naar het 2D-pixelcoördinatenstelsel (beeld). Deze matrices zijn samengesteld uit de brandpuntsafstanden (fx, fy), uitgedrukt in pixeleenheden, en het hoofdpunt (cx, cy) dat zich gewoonlijk dicht bij het beeldcentrum bevindt (snijpunt projectievlak en optische as). De matrices zijn niet afhankelijk van de bekeken 3D-ruimte, maar van de camera's met bijhorende instellingen zelf. Eenmaal bepaald, kunnen deze matrices voor onbepaalde tijd worden gebruikt, zolang de brandpuntsafstand ongewijzigd blijft. Als het camerabeeld met een bepaalde factor wordt geschaald, moeten alle instellingen met diezelfde factor worden geschaald [27].

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Figuur 53 Intrinsieke camera matrix [27]

De vectoren van vervormingscoëfficiënten van beide camera's bevatten de nodige gegevens voor het oplossen van de beeldvervorming ten gevolge van de cameralens. Zo een vector ziet er theoretisch uit als volgt: (k1,k2,p1,p2[,k3[,k4,k5,k6[,s1,s2,s3,s4[,τx,τy]]]]). Dit voor 4, 5, 8, 12 of 14 elementen [27].

Verder berekend de functie de transformatie tussen de 2 camera's die een stereopaar vormen. Indien men de poses van een object ten opzich van de eerste en tweede camera berekent, respectievelijk (R1, T1) en (R2, T2), voor een stereocamera waarbij relatieve positie en oriëntatie vastliggen, dan hebben deze poses een bepaald verband met elkaar. Dit betekent dat als de relatieve posities en oriëntaties (R, T) van de 2 camera's bekend zijn, (R2 en T2) kunnen worden berekend als (R1, T1) wordt gegeven. De functie berekend R en T zodat R2 gelijk is aan het product van R en R1, en zodat T2 gelijk is aan het product van R en T1 plus T [27]. Men kan dus een 3D-punt binnen het coördinatenstelsel van de eerste camera berekenen, wanneer men de coördinaten van het punt binnen het coördinatenstelsel van de tweede camera kent [27]. Dit wordt verduidelijkt op Figuur 54.

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix}$$

Figuur 54 Berekening representatie punt binnen het coördinatenstelsel van de 2de camera, afhankelijk van de coördinaten van het punt binnen het coördinatenstelsel van de 1ste camera.[27]

De rotatie en translatie vector zien er voor de gebruikte stereovisie opstelling binnen deze masterthesis uit zoals te zien op Figuur 55. De eerste waarde binnen de translatie vector, namelijk 710,1699764, geeft de baseline, afstand tussen de optische assen van beide camera's bij hun start (camera), weer in mm. Deze afstand komt overeen met de reeds eerder manueel gemeten afstand van 71 cm tussen beide optische assen (centrum camera). Dit bevestigt dat het niet nodig is, om deze afstand manueel te meten voor gebruik binnen de code. De 2de en 3de waardes binnen deze vector geven het verschil, in mm, in verticale afstand (hoogte) (Y) en in diepte (Z) weer tussen beide camera's. Dit verschil is respectievelijk 2,88280459 mm (Y) en 5,69010539 mm (Z). Deze kleine fouten bevestigen een vrij nauwkeurige manuele camera uitlijning, maar bevestigen ook een zeer kleine imperfectie in de uitlijning. De zeer kleine waardes van de rotatiematrix duiden op een min of meer parallel verloop van beide optische assen. Gezien de kleine afwijkingen binnen de rotatie en translatie vectoren, passen we in een volgende stap ook nog stereo rectificatie toe.

```
Rotation Vector:
[[ 0.99990614  0.00957543 -0.00979913]
 [-0.0093067   0.99958911  0.02711088]
 [ 0.01005471 -0.02701714  0.9995844 ]]
Translation Vector:
[[710.1699764 ]
 [  2.88280459]
 [  5.69010539]]
```

Figuur 55 rotatie en translatie vector opstelling masterthesis

De reeds eerder besproken matrices en vectoren bevatten, zoals reeds eerder besproken, heel wat informatie over de camera's en hun intrinsieke en extrinsieke parameters, maar ook over de beeldvervorming en camera positionering. Deze gegevens worden binnen volgende stappen gebruikt voor het softwarematig uitlijnen van beide camera's en het oplossen van beeldvervormingen van beide camera's, maar ook binnen het effectieve bepalen van posities, gezien heel wat verbanden tussen de werkelijke 3D-wereld en het 2D-beeldvlak van beide camera's en nu gelegd zijn.

STAP 3: Stereo rectificatie: Bepaling parameters

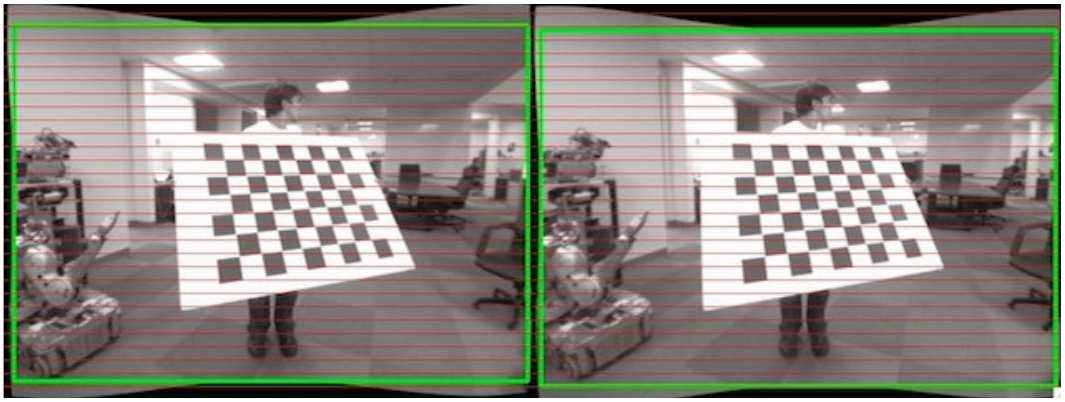
Binnen een volgende en 3de stap van het gehele kalibratie en rectificatieproces worden de rectificatie matrices van de beeldvlakken van beide camera's (virtueel) berekend (parameters). Dit aan de hand van de OpenCV 'stereoRectify' functie. Dit met als doel deze zo perfect mogelijk met elkaar uit te lijnen zodat ze binnen hetzelfde vlak vallen en zo veel mogelijk overlap hebben. Hierdoor worden alle epipolaire lijnen parallel en wordt het bepalen van de dispariteit in een volgende fase eenvoudiger [27]. Er zal slechts rekening gehouden moeten worden met een verandering van één coördinaat op beeld voor een gedetecteerd object op beide beelden. Binnen de stereovisie 3D-positiebepaling, worden afstanden en posities bepaald, ten opzichte van het cameravlak. Dit dient slechts één vlak te zijn en dat is dan ook de reden voor het uitvoeren van een stereo rectificatie stap.

Als input van de functie worden de, door de ‘stereoCalibrate’ functie berekende matrices en vectoren, meegegeven. Het gaat hier over:

- intrinsieke matrix van de linker camera
- intrinsieke matrix van de rechter camera
- vector van vervormingscoëfficiënten van de linker camera
- vector van vervormingscoëfficiënten van de rechter camera
- rotatie matrix
- translatie vector

Hiernaast worden ook de beeldresolutie (1280 x 720) en een *alpha* waarde meegegeven. Ook binnen deze functie werden er geen instellingen op basis van flags aangepast.

Alpha is een waarde die de schaalverdeling voor het uiteindelijke beeld instelt. Indien deze waarde leeg gelaten wordt of op ‘-1’ ingesteld wordt, dan wordt er gebruik gemaakt van de standaard schaling door de functie. In andere gevallen dient deze waarde tussen ‘0’ en ‘1’ ingesteld te worden. Het meegeven van een waarde van ‘0’ betekent dat de gerectificeerde beelden zodanig worden vergroot en verschoven, dat alleen geldige pixels zichtbaar zijn. Het meegeven van een waarde van ‘1’ betekent dat de gerectificeerde beelden zodanig worden verscherpt en verschoven zodat alle pixels van de originele beelden, van de camera’s, behouden blijven. Geen pixels van het bronbeeld gaan hierbij ‘verloren’. Elke tussenliggende waarde geeft een tussenresultaat tussen beide gevallen. Binnen de ontwikkeling van het 3D-positiebepalingssysteem aan de hand van de opstelling wordt er gebruik gemaakt van een waarde van ‘0’ voor alpha. Hiermee worden de beelden zodanig ingezoomd en verschoven dat enkel de geldige pixels zichtbaar zijn op beeld [27]. Gezien de beeldhoek van beide camera’s, vormt dit inzoomen geen probleem. Het zicht blijft voldoende ruim en er is voldoende overlap tussen beide beelden binnen het stereo paar. Het toepassen hiervan zorgt er ook voor dat een object op beeld, niet buiten beeld kan vallen ten gevolge van het toepassen van rectificatie. Figuur 56 verduidelijkt het gebruik van een alpha waarde van ‘0’ binnen de ‘stereoRectify’ functie op een paar gekalibreerde stereobeelden. Bij het meegeven van een waarde ‘0’ voor alpha wordt het beeld binnen de groene kaders (geldige pixels) gebruikt en valt de rest weg. Bij het meegeven van de waarde ‘1’ zal het volledige beeld (ook buiten groene kaders) gebruikt worden. Inclusie de zwarte banden, ten gevolge van het oplossen van de beeldvervormingen door de lens. Bij zeer sterke beeldvervormingen, zou het kunnen dat er nog steeds een stukje zwarte band binnen beeld valt. Dit is dus geen fout. Op beeld is het parallelle en horizontale verloop van de epipolaire lijnen duidelijk zichtbaar, wat duid op een correcte parallelle en verticale uitlijning [27].



Figuur 56 Verduidelijking stereo rectificatie met ‘0’ als alpha (groene kader) en epipolaire lijnen verticaal uitgelijnd [27]

De output van de ‘stereoRectify’ functie bevat heel wat parameters (vectoren en matrices) met informatie over de camera’s, beide beelden, hun verband en de opstelling. Meer specifiek bevat, de gebruikte output binnen dit 3D-positiebepalingssysteem op basis van stereovisie, de volgende informatie [27]:

- 3 x 3 rectificatietransformatie (rotatiematrix) voor de linker camera
- 3 x 3 rectificatietransformatie (rotatiematrix) voor de rechter camera
- 3 x 4 projectiematrix in het nieuwe (gerectificeerde) coördinatenstelsels voor de linkse camera
- 3 x 4 projectiematrix in het nieuwe (gerectificeerde) coördinatenstelsels voor de rechtse camera

Verder kunnen er ook nog andere resultaten verkregen worden, zoals de Q-dieptematrix en matrices met geldige pixels, maar deze worden verder niet besproken binnen deze scriptie, gezien ze niet gebruikt werden. Extra documentatie hieromtrent is altijd terug te vinden via de OpenCV functiedocumentaties [26].

De rectificatietransformatie matrices (rotatiematrices) brengen punten in het niet-gerectificeerde coördinatenstelsel van de camera’s over op punten in het gerectificeerde coördinatenstelsel van de camera’s. In meer technische termen betekend dit dat er een wijziging van de basis van het niet-gerectificeerde coördinatenstelsel van een camera naar het gerectificeerde coördinatenstelsel van de camera [27].

De projectiematrices projecteren punten uit het gerectificeerde coördinatenstelsel van een camera, naar het gerectificeerde werkelijke beeld van een camera. Deze matrices zien er in theorie uit zoals afgebeeld op Figuur 57 [27]. Het gaat hier om de horizontale stereo variant, die gebruikt werd binnen deze ontwikkeling. De verticale stereovariant wordt verder niet besproken, maar werkt op een gelijkaardige manier, maar dan aan de hand van Y-coördinaten in plaats van X-coördinaten en ziet er theoretisch dus ook lichtjes anders uit.

De beelden van de eerste en tweede camera zijn ten opzichte van elkaar verschoven, voornamelijk langs de x-as (er kan een lichte verticale verschuiving zijn). In de gerectificeerde beelden zijn de overeenkomstige epipolaire lijnen van de linker- en rechtercamera horizontaal en hebben ze gelijke y-waarden. P1 en P2 zien er hierbij uit zoals afgebeeld op Figuur 57. Dit zijn de effectieve nieuwe ‘gerectificeerde’ camera matrices. Hierbij stelt ‘f’ de brandpuntsafstand in pixels voor en ‘Tx’ de horizontale verplaatsing tussen beide camera’s. ‘cx1’ en ‘cx2’ zijn gelijk indien de hoofdpunten (‘principal popints’) van beide camera’s dezelfde pixelcoördinaten in de gerectificeerde weergave hebben. Dit kan zo ingesteld worden aan de hand van de ‘CALIB_ZERO_DISPARITY’ flag die standaard gebruikt wordt [27]. Dit om later op correcte wijze de dispariteit op beeld te kunnen gebruiken, gezien er geen verschil is tussen de locaties van beide gerectificeerde coördinatensystemen van beide camera’s op beeld.

$$\mathbf{P1} = \begin{bmatrix} f & 0 & cx_1 & 0 \\ 0 & f & cy & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{P2} = \begin{bmatrix} f & 0 & cx_2 & T_x * f \\ 0 & f & cy & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Figuur 57 Theoretisch 'gerectificeerde camera-/projectiematrices (horizontale stereo variant) [27]

De projectiematrices zien er binnen de gebruikte stereo opstelling van deze masterthesis uit zoals te zien op Figuur 58. Wanneer deze gegevens vergeleken worden met de theoretische voorstelling op Figuur 57 dan valt er hier heel wat informatie uit te halen.

```
P LEFT IMAGE:
[[780.47659603   0.         632.10225677   0.        ]
 [  0.         780.47659603 441.22934532   0.        ]
 [  0.           0.           1.           0.        ]]
P RIGHT IMAGE:
[[7.80476596e+02 0.00000000e+00 6.32102257e+02 5.54293403e+05]
 [0.00000000e+00 7.80476596e+02 4.41229345e+02 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 1.00000000e+00 0.00000000e+00]]
```

Figuur 58 Projectiematrices opstelling masterthesis

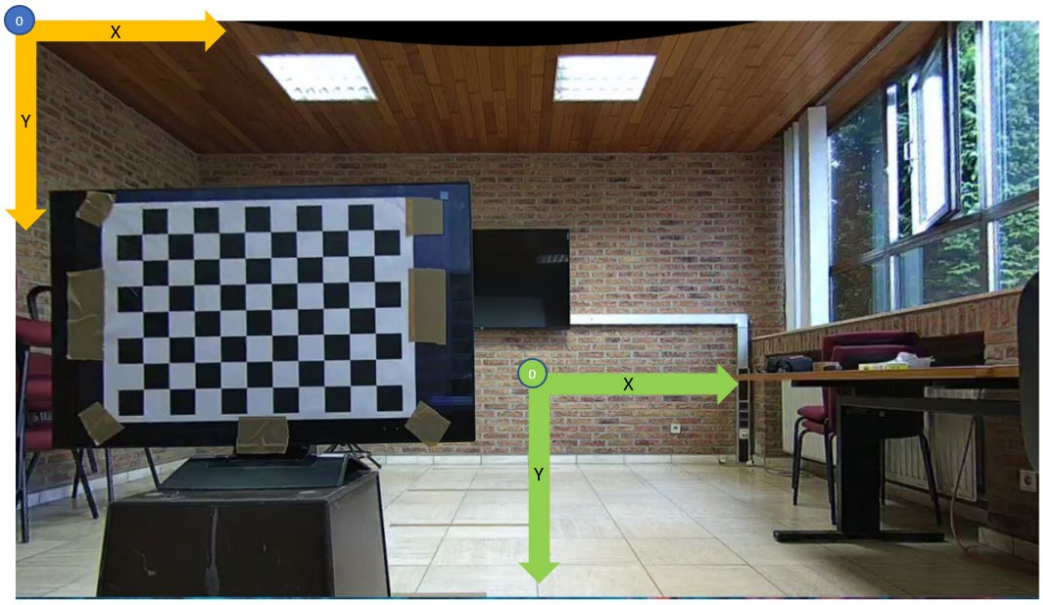
Informatie uit projectiematrices (geen homogene coördinaten):

- brandpuntsafstand (f): 780,47659603 pixels
- X-positie hoofdpunt linker camera: 632,10225677 pixels
- X-positie hoofdpunt rechter camera: 632,10225677 pixels
- Y-positie hoofdpunt linker camera: 441,22934532 pixels
- Y-positie hoofdpunt rechter camera: 441,22934532 pixels
- horizontale verplaatsing tussen beide camera's (optische assen) (Tx): 710,19862 mm

Uit deze gegevens valt af te leiden dat beide beelden, bij het toepassen van de kalibratie- en rectificatie parameters (vectoren en matrices), in een volgende stap, een gelijke brandpuntsafstand en hoofdpunt op beeld (beeldcoördinatenstelsel) zullen hebben. Hiernaast wordt de horizontale verplaatsing tussen beide camera's, uit bovenste informatie, ook nogmaals bevestigd.

Naast het beeldcoördinatenstelsel zal er in een volgende fase ook gebruik gemaakt worden van het gerectificeerde (projectie) coördinatenstelsel. Dit stelsel werd op basis van de kalibratie- en rectificatieparameters, perfect uitgelijnd binnen beide beelden. Het nulpunt van het gerectificeerde (projectie) coördinatenstelsel bevindt zich binnen het beeldcoördinatenstelsel (van gekalibreerde en gerectificeerde beelden) op het hoofdpunt (cx, cy). Dit hoofdpunt is gelijk voor beide gekalibreerde en gerectificeerde stereobeelden. In het geval van het systeem binnen deze masterproef bedraagt het hoofdpunt (nulpunt) als beeldcoördinaten, verkregen uit bovenstaande rectificatiegegevens, X = 632 en Y= 441. Dit punt (groen bolletje met '0'), met bijhorende verduidelijking van het gerectificeerde (projectie) coördinaten- en assenstelsel (groen) wordt verduidelijkt op Figuur 59. Hierbij werd er een schatting gemaakt van de positie

van het hoofdpunt binnen het beeldcoördinatenstelsel. De groene pijlen geven het gerectificeerde projectiecoördinaten- en assenstelsel weer, met als hoofd- en nulpunt het punt (632, 441) binnen het beeldcoördinaten- en assenstelsel (gele pijlen).



Figuur 59 Aanduiding gerectificeerd projectiecoördinatenstelsel (groen) binnen beeldcoördinatenstelsel (geel)

De parameters, verkregen door het uitvoeren van de stereorectificatie stap, worden binnen een volgende fase, in combinatie met enkele kalibratie parameters, effectief gebruikt voor het vervormen, transformeren, roteren van ieder frame van elke camera apart. Verder worden de parameters ook gebruik voor het gerectificeerde projectie coördinaten- en assenstelsel en uiteindelijk ook in een volgende fase voor het effectief bepalen van afstanden en werkelijke 3D-coördinaten. Het verband tussen een positie op twee 2D-beelden en binnen de werkelijke 3D-ruimte werd namelijk gelegd binnen de vorige stappen.

STAP 4: Opstellen vervormings-, rectificatie- en transformatiekaart

Nu dat alle parameters reeds bepaald zijn, wordt er in een laatste fase van het stereo kalibratie- en rectificatieproces een vervormings-, rectificatie- en transformatiekaart opgesteld. Dit gebeurt apart per camera. Deze kaart bevat alle nodige parameters en gegevens voor het verkrijgen van onverstoorde en uitgelijnde (stereo)beelden [37]. Het effectief toepassen van deze vervormings-, rectificatie- en transformatiekaart gebeurt pas in een volgende fase van de ontwikkeling en wordt later verder besproken.

Voor het opstellen van de vervormings-, rectificatie- en transformatiekaart wordt er gebruik gemaakt van de OpenCV ‘initUndistortRectifyMap’ functie. Deze berekent de beeldvervormingen, -rectificaties en -transformaties en geeft het resultaat weer in de vorm van kaarten, die later gebruikt kunnen worden voor ‘remappen’ van beelden. Het onvervormde beeld ziet eruit als het oorspronkelijke beeld, alsof het is opgenomen met perfect uitgelijnde en gekalibreerde camera. In geval van een stereocameraopstelling wordt er binnen de functie rekening gehouden met de eerder bepaalde projectiematrices [37]. Het opstellen van deze kaarten, gebeurt zoals reeds eerder vermeld per camera apart. Het is bij het opstellen van de kaarten dus wel belangrijk dat er rekening gehouden wordt met de stereovisie opstelling. Dit is dus mogelijk binnen de functie (projectiematrices meegeven aan functie).

Het nieuwe camerabeeld zal anders georiënteerd worden in de coördinatenruimte door het toepassen van deze kaarten (later besproken). Dit volgens de parameter R van de camera. Het toepassen hiervan zorgt voor het horizontale verloop van de epipolaire lijnen op beide beelden over dezelfde Y-coördinaat [37].

In technische aspecten berekend de functie de kaarten voor het latere toepassen van het inverse mapping algoritme, dat door de 'remap' functie wordt gebruikt. Dit wil zeggen dat de functie voor elke pixel (u, v), in het gecorrigeerde en gerectificeerde beeld, de overeenkomstige coördinaten in het bronbeeld (oorspronkelijke beeld van de camera) berekent [37]. Zoals reeds eerder vermeld wordt de functie per camera apart aangeroepen.

Als input van de 'initUndistortRectifyMap' functie worden reeds eerder bepaalde parameters (kalibratie en rectificatie) meegegeven. Het gaat hier, per camera apart, over [37]:

- intrinsieke matrix
- vector van vervormingscoëfficiënten
- 3 x 3 rectificatietransformatie (rotatiematrix)
- 3 x 4 projectiematrix in het nieuwe (gerectificeerde) coördinatenstelsels

Verder wordt aan deze functie ook nog de beeldresolutie (1280 x 720) meegegeven in combinatie met het type kaart. CV_16SC2 in dit geval. Extra uitleg hieromtrent is terug te vinden binnen de OpenCV functiedocumentatie [37].

De output van deze functie bedraagt 2 kaarten. De functie berekent namelijk voor elke pixel (u, v), in het gecorrigeerde en gerectificeerd beeld, de overeenkomstige coördinaten in het bronbeeld (oorspronkelijke beeld) [37]. De eerste kaart bevat alle x-waarden en de 2de kaart alle y-waarden.

Na het toepassen van deze functie voor beide camera's, met elk hun specifieke parameters die eerder bepaald werden, worden alle kaarten, met hun specifieke informatie, weggeschreven naar 1 Xml-bestand. Dit is een simpel tekst gebaseerd bestandsformaat voor het structureel weergeven van informatie. De data binnen dit bestand (4 kaarten aan gegevens) wordt in een latere fase binnen de ontwikkeling gebruikt voor het toepassen van vervormingen, rectificaties en transformaties voor elk opgehaald frame van elke camera.

Zoals reeds eerder vermeld dienen alle bovenste stappen slechts 1 keer uitgevoerd te worden per cameraopstelling met gelijke instellingen. Deze instellingen en de positionering van de opstelling dienen dus tijdens het gebruik van deze parameters, met bijhorende kaarten, constant te blijven. Bij het uitvoeren van een aanpassing, dienen alle eerder vermelden stappen opnieuw herhaald te worden ter herkalibratie en -rectificatie van, het stereovisie systeem. Het niet herhalen van deze stappen zal in een latere fase onnauwkeurigheden of onjuiste gegevens opleveren. Ook al lijken beide beelden in orde te zijn.

3.5 Beeldverwerking

Binnen deze fase van de ontwikkeling van het 3D-positiebepalingssysteem op basis van stereovisie wordt de toepassing van de reeds eerder besproken kaarten, aan het eind van deel 3.4.2, uitgelegd. Deze kaarten werden opgesteld aan de hand van heel wat parameters die bepaald werden tijdens de stereo kalibratie en stereo rectificatie stappen. Ook deze parameters werden in deel 3.4.2 al besproken. De kaarten zelf werden alle 4 (2 specifieke kaarten per camera) weggeschreven naar een Xml-bestand. Van hieruit worden deze gegevens binnen de code dan ook terug opgehaald en gesorteerd per camera en per coördinaat. Het kalibratie gedeelte werd binnen de code apart gehouden van de andere code. Dit gezien de bepaling van al de parameters en het opstellen van de kaarten slecht één keer moet gebeuren voor elke onveranderde opstelling, wat onderlinge camera posities en instellingen betreft, zoals ook reeds eerder besproken in deel 3.4.2.

De kaarten bevatten voor elke pixel (u, v) in het gecorrigeerde en gerectificeerd beeld, de overeenkomstige coördinaten in het bronbeeld (oorspronkelijke beeld) [37]. Op deze manier bevatten ze heel wat gegevens voor het vervormen, rectificeren en transformeren van de beelden. Dit voor het verkrijgen van onverstoorde en uitgelijnde (stereo)beelden. Meer details hieromtrent werden reeds eerder al besproken. De eerste kaart bevat alle x-waarden en de 2de kaart alle y-waarden.

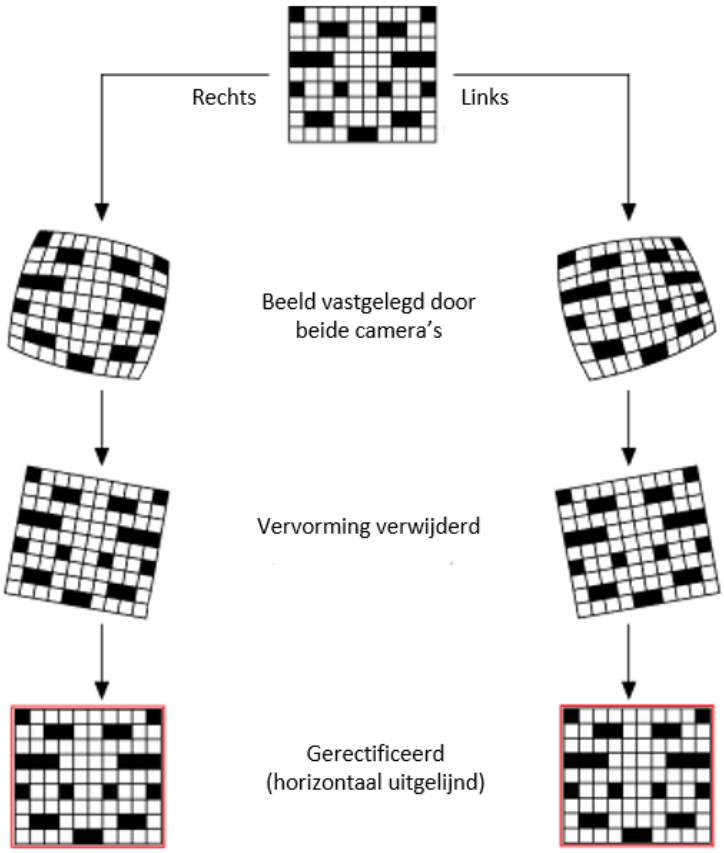
Voor het toepassen van de vervormingen, rectificaties en transformaties op beeld, aan de hand van gegevens uit de kaarten, wordt er gebruik gemaakt van de OpenCV ‘remap’ functie [37]. Zoals reeds eerder vermeld zal het toepassen van deze functie apart per camera en per cameraframe moeten gebeuren.

Als input van de ‘remap’ functie worden de volgende parameters gebruikt [37]:

- origineel camera frame
- eerste kaart (x-waarden)
- tweede kaart (y-waarden)

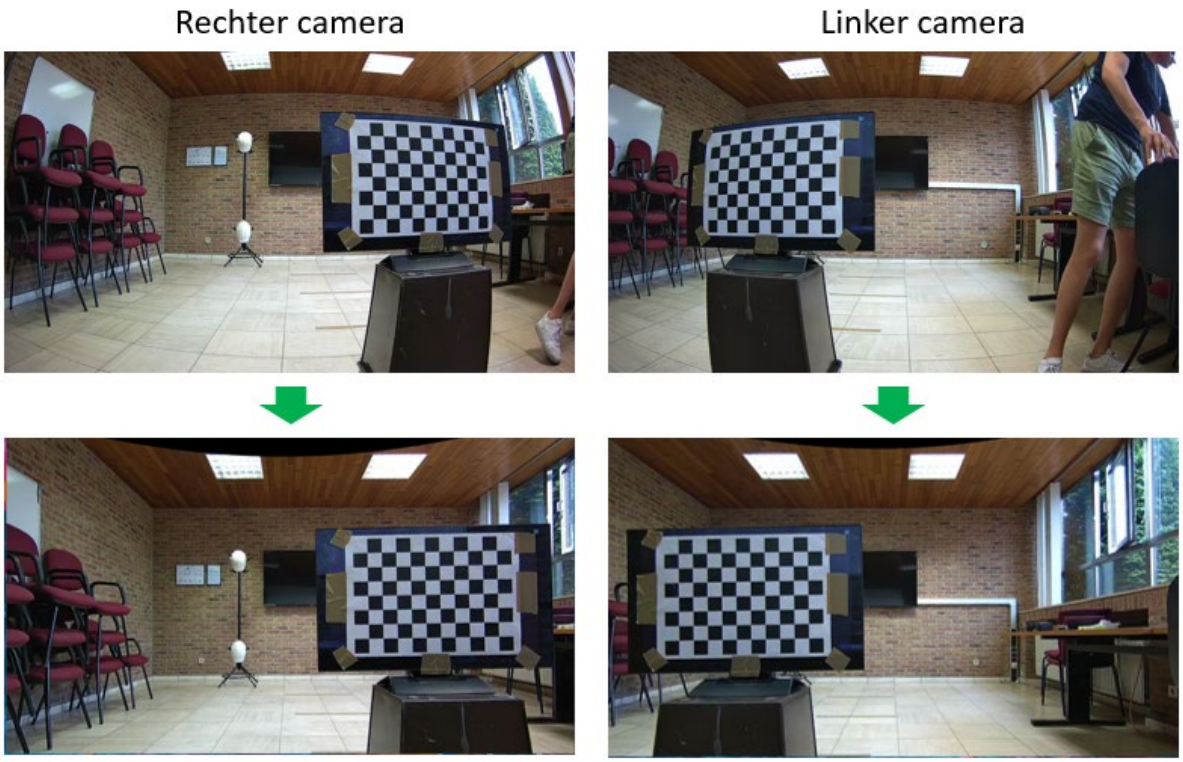
Verder wordt aan deze functie ook nog een interpolatiemethode ‘INTER_LINEAR’ (bilineaire interpolatie) en pixel extrapolatie methode (grenstype) ‘BORDER_CONSTANT’ meegegeven. Extra uitleg hieromtrent is terug te vinden binnen de OpenCV functiedocumentatie [37].

De output van deze functie bedraagt de vervormde, gerectificeerde en getransformeerde input frame, op basis van alle reeds eerder bepaalde parameters. Schematisch ziet dit eruit zoals weergegeven op Figuur 60. In een eerste fase valt de scene te zien (beeld met zwarte en witte blokjes bovenaan). Deze scene wordt vastgelegd door een rechter en linker camera. Beide beelden van de scene zijn duidelijk vervormd en niet uitgelijnd. In een volgende fase wordt de beeldvervorming, ten gevolge van de eigenschappen van de lens, verwijderd. De lijnen op beeld zijn nu weer allemaal recht. Beide beelden zijn hier echter nog niet uitgelijnd ten opzichte van elkaar. Dit gebeurt binnen de laatste fase. Beide beelden worden hier gerectificeerd (horizontaal uitgelijnd). Deze vormen nu een correct stereo paar, dat gebruikt kan worden voor het bepalen van 3D-positites en -afstanden.



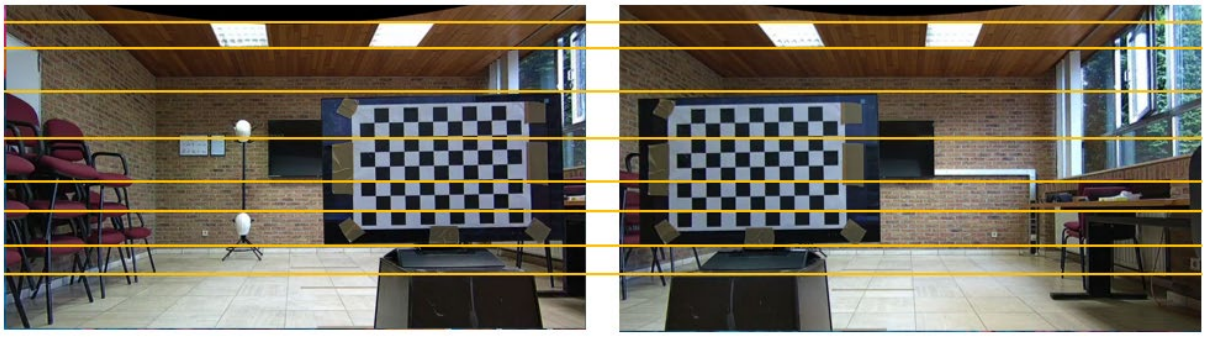
Figuur 60 Schematische weergave beeldverwerking voor een paar stereobeelden [35] (aangepast)

Figuur 61 geeft 2 originele frames weer, gemaakt aan de hand van de stereovisie cameraopstelling, gebruikt binnen deze masterproef, zoals verduidelijkt in deel 3.2. Beide frames (1 per camera) werden als parameter meegegeven aan hun specifieke 'remap' functie (specifieke parameters per camera en bijhorende beelden). De vervormde, gerectificeerde en getransformeerde frames worden onder de originele frames weergegeven. Zoals op beide frames te zien werden beeldvervorming, ten gevolge van de eigenschappen van de lens, verwijderd en werden beide beelden gerectificeerd (horizontaal uitgelijnd binnen hetzelfde vlak. Deze vormen nu een correct stereo paar, dat gebruikt kan worden voor het bepalen van 3D-posities en -afstanden. De kleine zwarte zones, te zien aan de bovenkant van beide vervormde, gerectificeerde en getransformeerde frames is te wijten aan het oplossen van beeldvervormingen ten gevolge van de lens. De lens van beide Hikvision camera's, geeft namelijk een zeer ruim beeld, met ten gevolge redelijk wat vervorming. Tussen de originele frames en de vervormde, gerectificeerde en getransformeerde frames is er op het beeld zelf ook nog een verschil merkbaar. Op de originele frames is er een persoon op beeld zichtbaar en op de verwerkte beelden niet. Dit heeft puur te maken met het maken van de beelden, maar heeft voor de rest geen invloed. Buiten dat deze persoon niet aanwezig is op de uiteindelijke beelden is er geen verschil. De scene zelf bleef ongewijzigd. Zoals reeds eerder aangegeven gebeurt de benoeming van de camera's (linker en rechter) vanuit het perspectief van een persoon, kijkend naar de camera's.



Figuur 61 Toepassing 'remap' functie, met eerder bepaalde parameters, ter vervorming, rectificatie en translatie van beide stereo beelden

Wanneer er op beide vervormde, gerectificeerde en getransformeerde beelden enkele epipolaire lijnen getrokken worden, zoals te zien op Figuur 62, valt er duidelijk te zien dat punten binnen de werkelijke 3D-wereld zich op beeld op een gelijke hoogte bevinden. Enkel wat horizontale positie op beeld betreft is er een verschil merkbaar (dispariteit). Dit toont een correcte vervorming, rectificatie en transformatie van beide beelden aan.



Figuur 62 Horizontale epipolaire lijnen (geel) aangebracht op de vervormde, gerectificeerde en getransformeerde stereobeelden

Beide beelden vormen nu een correct paar stereobeelden, dat in combinatie met eerder bepaalde parameters, gebruikt kan worden voor het bepalen van 3D-positites en- afstanden.

3.6 Objectdetectie

Om het in een latere fase kunnen toepassen van effectieve 3D-positiebepalingssysteem, dienen er één of meerdere objecten te zijn, waarvan de 3D-posities en bijgevolg ook afstanden bepaald dienen te worden. Op beeld zal er dus aan objectdetectie gedaan moeten worden. De uitwerking hiervan, binnen deze masterproef, wordt binnen dit deel van de scriptie omschreven, met bijhorende extra uitleg ter verduidelijking. Natuurlijk zouden alternatieve methodes, dan de gebruikte methode binnen deze masterproef, zeker ook mogelijk kunnen zijn. Het onderzoeken hiervan valt buiten deze scope van deze masterthesis.

3.6.1 Detectieobject

Gezien de uiteindelijke toepassing van het 3D-positiebepaliungssysteem op basis van stereovisie, binnen het S&S Center van Aperam Genk (doel Aperam), werd er als object gekozen voor een veiligheidshelm. Figuur 63, toont een voorbeeld van een veiligheidshelm, die gebruikt werd binnen de ontwikkeling van het detectiemodel en latere verificatie van het 3D-positiebepalingssysteem. De keuze voor een veiligheidshelm werd gemaakt gezien er, binnen het S&S Center van Aperam Genk, personen gedetecteerd dienen te worden, op basis van hun veiligheidshelmen. Dit gezien de camera's loodrecht, ten opzichte van het vlak van de vloer, geïnstalleerd zullen worden, aan de kat van de rolbrug, met als gevolg dat van een persoon enkel het bovenaanzicht (helm) zichtbaar zal zijn. Natuurlijk zal het objectdetectiemodel, indien nodig, hertraind moeten worden naar deze specifieke locatie en toepassing, gezien de andere omstandigheden (belichting, beeld, type helmen, …). Het huidige detectiemodel werd namelijk ontwikkeld aan de hand van de huidige opstelling, binnen een kantoorruimte van Aperam, met lokale omstandigheden. Het ontwikkelen van een zo performant mogelijk objectdetectiemodel voor veiligheidshelmen, valt buiten de scope van deze masterthesis. Het gebruikte (helm-) detectiemodel, waarvan de ontwikkeling verder wordt besproken, dient voornamelijk ter verdere ontwikkeling en verificatie van het 3D-positiebepalingssysteem op basis van stereovisie, binnen de gebruikte opstelling en omstandigheden.

Figuur 63 Veiligheidshelm, gebruikt voor ontwikkeling detectiemodel

3.6.2 Objectdetectiemodel (YOLOv5s)

Als objectdetectiemodel werd er binnen deze masterproef, omtrent de ontwikkeling van een 3D-positiebepalingssysteem op basis van stereovisie, gekozen voor een YOLOv5(s)-objectdetectiemodel. Dit is een objectdetectiemodel uit de YOLO-reeks van objectdetectie algoritmes. Versie 5 (v5) is een geoptimaliseerde versie van eerdere objectdetectiemodellen uit de YOLO-reeks. Deze 5de versie werd uitgebracht in 2020. YOLOv5-objectdetectiemodellen staan gekend om hun realtime detecties aan hoge snelheden met grote nauwkeurigheid [22], [38]. Vandaar ook de keuze hiervoor binnen deze masterproef. De YOLOv5-familie maakt ook gebruik van het PyTorch framework en ondersteund dit ook [22]. PyTorch is een open source framework voor machine learning, dat veel gebruikt wordt binnen 'deep learning', dat het proces van onderzoek naar implementatie versnelt [39]. Binnen de YOLOv5-familie werd er gekozen voor het YOLOv5s algoritme. Deze keuze werd gemaakt op basis van de hoge

prestaties wat snelheid en nauwkeurigheid betreft [38]. Het gebruik van een ander objectdetectiemodel zou zeker ook mogelijk kunnen zijn, maar werd niet geëvalueerd binnen deze masterproef. Meer informatie omtrent YOLO is terug te vinden in deel 2.2.2.

3.6.3 Ontwikkeling objectdetectiemodel

Binnen dit deel van de masterthesis wordt de effectieve ontwikkeling van het YOLOv5s-objectdetectiemodel besproken. Dit voor het detecteren van veiligheidshelmen, aan de hand van de stereovisie opstelling, gebruikt binnen deze masterproef, binnen de omstandigheden van de later besproken ontwikkel- en testlocatie. De uitwerking hiervan gebeurt aan de hand van verschillende stappen.

STAP 1: Verzamelen data

In een eerste stap binnen de ontwikkeling van het YOLOv5s-helmdetectiemodel dient er test- en validatiedata verzameld te worden. Dit ter creatie van een dataset ter verdere training van het model en verificatie van de correcte werking.

Aan de hand van de reeds ontwikkelde stereovisie opstelling werden er afbeeldingen van de helm gemaakt. Hierbij werd er gebruik gemaakt van de vervormde, gerectificeerde en getransformeerde beelden. Dit gezien het uiteindelijke detectiemodel in een latere fase ook op deze beelden toegepast zal worden, en er vermeden wil worden dat er vervormde helmen, ten gevolge van beeldvervorming door de lens, in de dataset zitten. Het is hierbij belangrijk om te vermelden dat het gaat om afbeeldingen van dezelfde resolutie als de camera's. 1280 x 720 in dit geval. Het helmdetectiemodel wordt dan ook ontwikkeld op basis van deze resolutie. Het veranderen van de resolutie van de camera's zal, naast het feit dat er opnieuw aan kalibratie en rectificatie gedaan moet worden, ook een invloed hebben op de latere toepassing van het detectiemodel.

Wat positie van de helm op beeld betreft werd er gezorgd voor voldoende variatie. Dit zowel op het vlak van positie, afstand als rotatie. Ook waren er afwijkingen binnen de omstandigheden (belichting, helderheid, …) Enkele afbeeldingen zijn te zien op de collage, weergegeven op Figuur 64.

Figuur 64 Collage afbeeldingen training en validatie YOLOv5s-helmdetectiemodel

Binnen deze stap werd de eigenlijke dataset gevormd. In dit geval gaat het over allerlei afbeeldingen van de veiligheidshelm, afgebeeld op Figuur 63.

STAP 2: Labelen data

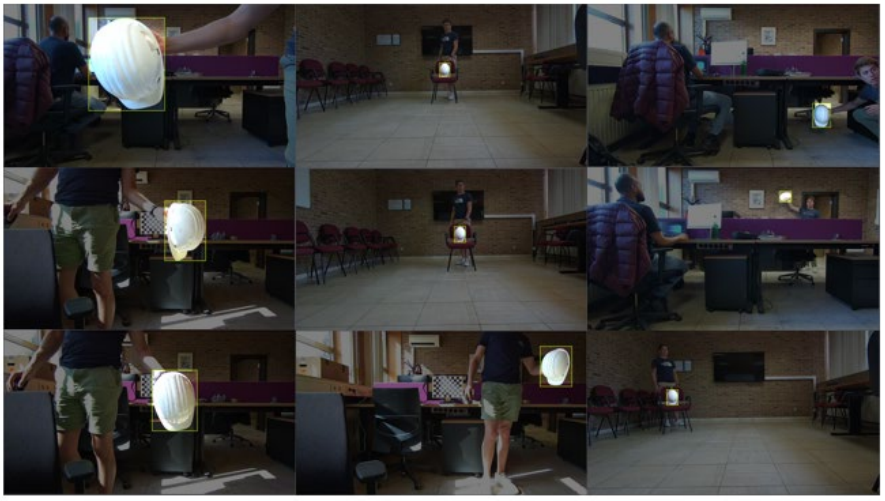
In een volgende stap, binnen de ontwikkeling van het YOLOv5s-helmdetectiemodel dient de verzamelde train- en verificatiedata gelabeld te worden. Dit om te weten waar op beeld (coördinaten) het object zich bevindt en om welk label (klasse) het gaat. In dit geval gaat het over slechts 1 klasse, genaamd helm ('HELMET' binnen het model). Naast het detecteren van objecten is het YOLOv5s-model dus ook in staat tot het classificeren (identificeren) van objecten. Meer theoretische uitleg omtrent het labelen, werd eerder besproken in deel 2.2.2. Gezien het gaat over gelabelde data spreken we over 'supervised' machinaal leren. Ook dit werd reeds eerder besproken en uitbundig uitgelegd in deel 2.2.2.

Binnen de ontwikkeling van het YOLOv5s-helmdetectiemodel, werd er gekozen om de verzamelde data (afbeeldingen) te uploaden op Roboflow en van hieruit te labelen, uit te breiden en te splitsen. Hierover later meer informatie.

Roboflow is een online platform dat ontwikkelaars in staat stelt hun eigen computervisie-applicaties te maken, ongeacht hun vaardigheden of ervaring. Het biedt de mogelijk tot het omzetten van onbewerkte afbeeldingen naar op maat getrainde computervisiemodellen met alle nodige tussenstappen en biedt hiernaast alle tools die nodig zijn om ze te implementeren voor gebruik binnen specifieke toepassingen. Roboflow ondersteunt zowel objectdetectie- als classificatiemodellen [40].

Binnen Roboflow diende alle geüploade afbeeldingen manueel geannoteerd en gelabeld te worden. Dit houdt in dat op iedere afbeeldingen de helm aangeduid diende te worden, door het tekenen van een vierkant, waarbinnen de helm op beeld zich bevindt, en deze dan ook gelabeld diende te worden.

De uiteindelijk gelabelde afbeeldingen zien er dan binnen Roboflow uit zoals te zien op de collage, afgebeeld op Figuur 65.



Figuur 65 Collage geannoteerde en gelabelde afbeeldingen YOLOv5s-helmdetectiemodel

STAP 3: Uitbreiden en splitsen dataset

In een 3de stap wordt de dataset uitgebreid en gesplitst in een train en validatie set. De train set voor het effectieve trainen (ontwikkelen) van het model en de validatie set ter verificatie en validatie van de werking van het getrainde model. Ook deze stap werd uitgevoerd binnen Roboflow.

Alvorens de gelabelde data (afbeeldingen) te splitsen in een train en validatie set. Biedt Roboflow de optie om de gelabelde dataset uit te breiden. Binnen de gratis versie kan deze uitbreiding tot een factor 3, binnen de betalende versie tot een factor 30. Het uitbreiden van de dataset houdt in dat er op basis van de gelabelde data (afbeeldingen) nieuwe (gelabelde) data gegenereerd wordt. Deze wordt vervolgens toegevoegd aan de origineel gelabelde dataset. Dit ter uitbreiding van de hoeveelheid data. Het toepassen van deze stap kan extra variëteit aan omstandigheden (belichting, rotaties, kleuren, ...) toevoegen aan de originele dataset.

Binnen deze ontwikkeling werd de dataset met een factor 3 (x 3) uitgebreid. Op de extra afbeeldingen werden verschillende effecten en veranderingen toegepast. Het gaat hier om:

- spiegeling: horizontaal en verticaal
- rotatie: Tussen -5° en +5°
- kleurtint: Tussen -16° en +16°
- helderheid: Tussen -19% en +19%

Deze factoren werden bepaald door het subjectief beoordelen en afstellen van de waarden. Naast de eerder benoemde en gebruikte opties biedt Roboflow ook nog tal van andere opties aan. Deze werden niet gebruikt, gezien ze op het eerste zicht geen meerwaarde zouden bieden aan variatie binnen de gelabelde dataset (gelabelde afbeeldingen).

Door het toepassen van deze effecten en veranderingen werd de gelabelde dataset uitgebreid van 296 naar 888 afbeeldingen. Enkele voorbeelden van de 888 gelabelde afbeeldingen met eventuele effecten en veranderingen zijn te zien op de collage, afgebeeld op Figuur 66.

Figuur 66 Collage geannoteerde en gelabelde afbeeldingen YOLOv5s-helmdetectiemodel met effecten en veranderingen

Nadat de gelabelde dataset uitgebreid werd tot 888 gelabelde afbeeldingen, werd de dataset gesplitst in een train en validatie set. Voor de validatie set werden er 24 willekeurige gelabelde afbeeldingen uit de dataset gehaald en in een validatie set gestoken. De resterende 864 gelabelde afbeeldingen vormen de training set.

Roboflow biedt de optie tot het exporteren van de ontwikkelde dataset met train en validatie afbeeldingen. Dit kan zowel naar een zip bestand als in de vorm van code voor het rechtstreeks gebruik van de dataset vanuit Roboflow. Het exporteren zelf kan ook in verschillende formaten. Het gaat hier over verschillende JSON-formaten zoals COCO en createML tot een CSV RetinaNet Keras formaat. Binnen deze masterproef werd er gekozen om de dataset te exporteren in het YOLO v5 PyTorch formaat als een zip bestand, maar ook in de vorm van code voor later rechtstreeks gebruik vanuit Roboflow.

De gehele dataset met alle aanpassingen blijft te allen tijde beschikbaar voor gebruik, aanpassingen, etc. binnen Roboflow.

STAP 4: Trainen YOLOv5s-helmdetectiemodel

Binnen de 4de stap van de ontwikkeling van het YOLOv5s-helmdetectiemodel wordt het effectieve trainen en valideren van het model besproken.

Voor het trainen van het model werd er gebruik gemaakt van Google Colab. Binnen Google Colab, de online Python-omgeving van Google, kan iedereen Python-code schrijven en uitvoeren vanuit een browser. Vooral geschikt voor machine learning en data-analyse. Technisch gezien is Colab een gehoste Jupyter-notebookservice die niet zelf geïnstalleerd dient te worden. Het biedt (beperkt) gratis toegang tot enkele handige computerhulpmiddelen, waaronder GPU's [41]. De keuze voor Google Colab werd gemaakt, gezien het gebruiksgemak en de mogelijkheid tot het gebruik van GPU's en TPU's, voor een snellere training van het model ten opzichte van eigen laptop.

Voor het effectieve trainen van het model werd in eerste instantie de yolov5 GitHub repository gecloned [42]. Deze bevat alle data en functies voor het trainen van een YOLOv5s model, maar ook andere modellen. Naast het clonen van deze repository, werden ook de nodige functiebibliotheken zoals Torch, OS, IPython en Roboflow geïmporteerd. Verder werd er hier ook aangegeven dat er, indien mogelijk, gebruik gemaakt wordt van een GPU of TPU voor het trainen van het model. De output van de cel gaf aan dat er gebruik gemaakt werd van een 'NVIDIA Tesla T4' GPU, wat een voordeel biedt, gezien het trainen gebeurt op basis van grafische beelden.

In een volgende fase werd de link gelegd tussen Google Colab en de dataset binnen Roboflow. Voor het leggen van de link tussen beide werd de reeds eerder verkregen code vanuit Roboflow gebruikt. Deze bevat een API-key en andere informatie zoals gebruikersnaam, projectnaam en versie. De dataset werd op deze manier geïmporteerd en gedownload binnen Google Colab.

De volgende fase is het effectieve trainen van het model. Hiervoor wordt er gebruik gemaakt van het 'train.py' bestand uit de YOLOv5 GitHub repository [42]. Bij het aanspreken van het train.py bestand, ter training, worden de volgende parameters meegegeven:

- img: 1280 (breedte in pixels afbeeldingen)
- epochs: 50 (aantal itteraties)
- data: {dataset.location}/data.yaml (locatie dataset binnen Google Colab)
- weights: yolov5s.pt (parameters model)

Wat andere parameters betreft werden er geen wijzigingen doorgevoerd en gebruik gemaakt van de standaard ingestelde waarden. Het gebruikte oproepen van de train.py code, met parameters, ziet er als volgt uit:

!python train.py --img 1280 --epochs 50 --data {dataset.location}/data.yaml --weights yolov5s.pt

Resultaten omtrent het trainen van het model zijn terug te vinden in *Bijlage B* van deze scriptie, maar worden verder niet besproken.

STAP 5: Validatie YOLOv5s-helmdetectiemodel

Aan de hand van de eerder afgesplitste validatie set (24 gelabelde afbeeldingen) uit de gehele dataset (888 gelabelde afbeeldingen) wordt de werking van het model binnen deze 5de stap gevalideerd. Dit ter controle en evaluatie van de werking en nauwkeurigheid van het YOLOv5s-hemldetectiemodel.

Voor het valideren wordt er gebruik gemaakt van het detect.py programma uit de YOLOv5 GitHub repository [42]. Dit python programma biedt alle functionaliteiten en functies aan voor het gebruik van het YOLOv5s model. Het detect.py programma wordt binnen een volgende fase van de ontwikkeling ook effectief gebruikt voor het detecteren van veiligheidshelmen binnen de stereovisie 3D-positiebepalings toepassing.

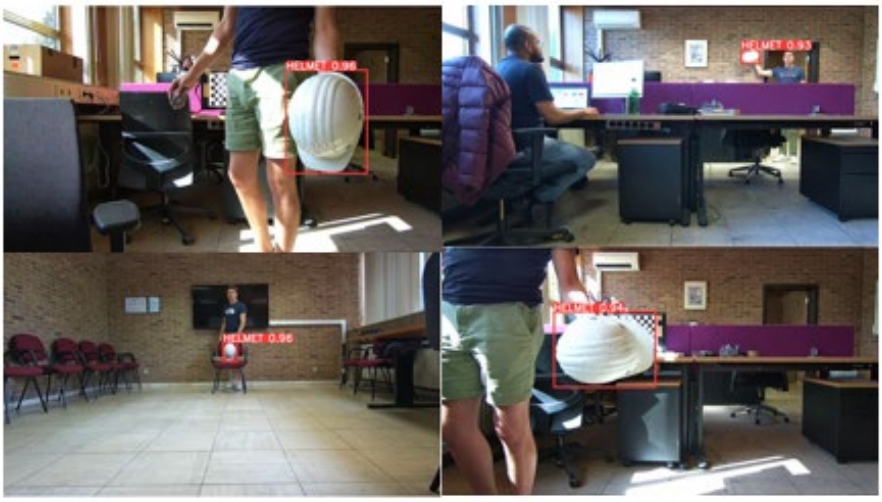
Aan het detect.py programma worden de volgende parameters meegegeven:

- weights: runs/train/exp/weights/best.pt (model)
- img: 1280 (breedte in pixels afbeeldingen)
- conf: 0.2 (20%) (minimale zekerheid detectie)
- source: {dataset.location}/valid/images (locatie validatie set)

Wat andere parameters betreft werden er geen wijzigingen doorgevoerd en gebruik gemaakt van de standaard ingestelde waarden. Het gebruikte oproepen van de detect.py code, met parameters, ziet er als volgt uit:

!python detect.py --weights runs/train/exp/weights/best.pt --img 1280 --conf 0.2 --source {dataset.location}/valid/images

De afbeeldingen uit de validatie set worden verwerkt door het model en opgeslaan in een folder binnen Google Colab. Rond iedere helm werd een bounding box getekend, met hierboven het bijhorende label en score van zekerheid. Deze zekerheid score (waarde tussen 0 en 1, inclusief 1) geeft de zekerheid van de voorspelling van het model aan. Enkele van de resultaten (afbeeldingen) worden weergegeven op de collage, afgebeeld op Figuur 67.



Figuur 67 Collage YOLOv5s-helmdetectiemodeltoegepast op validatie dataset

De laagste zekerheid score die behaald werd, binnen het gebruik van de validatie dataset bedraagt 0.93. Dit betekent in dit geval dat het model hier 93% zeker was dat het om een (veiligheids-)helm ging. Andere waardes lagen allemaal tussen de 0.93 en 0.96. Dit wijst op een nauwkeurige werking van het YOLOv5s-helmdetectiemodel.

Gezien het YOLOv5s-helmdetectiemodel voornamelijk gebruikt wordt voor de verdere ontwikkeling en verificatie van het 3D-positiebepalingsststeem aan de hand van stereo visie, werd er geen verder onderzoek gedaan of optimalisaties uitgevoerd. Het model kan altijd aangepast of hertraind worden aan de hand van de dataset binnen Roboflow die altijd uitbreidbaar of aanpasbaar is.

De code voor het trainen van het model, met nodige documentatie, werd naast Google Colab ook lokaal opgeslagen binnen het Python project, waarin de ontwikkeling van het 3D-positiebepalingssysteem gebeurt.

STAP 6: Exporteren YOLOv5s-helmdetectiemodel

Binnen deze 6de en laatste stap wordt het getrainde YOLOv5s-helmdetectiemodel geëxporteerd vanuit Google Colab voor lokaal gebruik.

Het model wordt geëxporteerd als een '.pt' (PyTorch) bestand. Dit gezien YOLOv5, zoals reeds eerder vermeld in deel 3.6.2, gebruik maakt van het PyTorch framework en het gebruik hiervan ook ondersteunt.

3.6.4 Toepassing en gebruik YOLOv5s-helmdetectiemodel

Binnen dit deel van de masterthesis wordt de toepassing van het eerder ontwikkelde YOLOv5s-helmdetectiemodel, binnen de masterproef omtrent het bepalen van 3D-posities op basis van stereovisie, besproken.

Het model werd, zoals reeds eerder vermeld, vanuit Google Colab geëxporteerd als een '.pt' (PyTorch) bestand. Binnen de code wordt dit bestand dan ook gebruikt voor het uitvoeren van (helm-)detecties. Het bestand is makkelijk aanpasbaar. Zo kan er op eenvoudige manier gebruik gemaakt worden van een ander detectiemodel. Het model dient wel een model uit de YOLOv5-familie te zijn. Dit gezien de verdere code en functies voor het toepassen van het model hierop gebaseerd zijn.

Binnen een eerste fase, voor het uitvoeren van detecties, wordt er gecontroleerd of de gebruikte computer 'CUDA' ondersteunt. Dit aan de hand van een functie ('torch.cuda.is_available()') uit de Python Torch functiebibliotheek. CUDA is een parallel computerplatform en programmeermodel ontwikkeld door NVIDIA. Door gebruik te maken van de kracht van GPU-versnellers, kunnen ontwikkelaars hiermee hun applicaties versnellen. Gezien de ontwikkeling door NVIDIA, kan CUDA enkel gebruikt worden op bepaalde types NVIDIA GPU's [43]. De eerder besproken computer, die gebruikt werd binnen de ontwikkeling van het 3D-positiebepalings systeem biedt (jammer genoeg) geen ondersteuning tot het gebruik van een GPU en dus bijgevolg ook geen ondersteuning tot het gebruik van CUDA. In dit geval wordt er dus gebruikt gemaakt van de CPU voor het verwerken van de beelden en het detecteren van helmen.

Voor het toepassen van het YOLOv5s-helmdetectiemodel wordt er gebruik gemaakt van enkele bestaande functies uit een YOLOv5 Custom Object Detection GitHub repository [44]. Deze werden lokaal opgeslaan binnen de project folder ter gebruik.

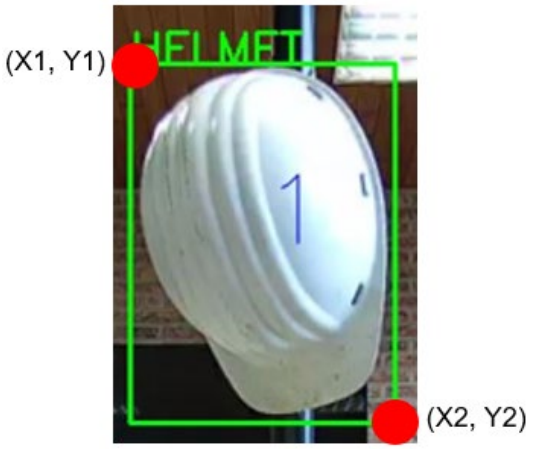
STAP 1: Toepassen YOLOv5s-helmdetectiemodel

In een eerste stap wordt het YOLOv5s-hemldetectiemodel effectief toegepast op de vervormde, gerectificeerde en getransformeerde frames. Dit voor de (stereo) beelden van elke camera apart. Het model wordt dus tweemaal toegepast. Eénmaal voor het rechter beeld van het paar stereobeelden en éénmaal voor het linker beeld van het paar stereobeelden. Binnen deze toepassing wordt er dus geen gebruik gemaakt van het principe van stereo matching. Meerdere helmen zouden zich namelijk op gelijke Y-positie op beeld kunnen bevinden. Ook is de achtergrond binnen de ruimte van de opstelling te variabel wat beeld betreft. Deze factoren zouden het correct gebruik van stereo matching sterk bemoeilijken.

Voor het toepassen van het model wordt er gebruik gemaakt van de YOLOv5 Custom Object Detection 'score_frame' functie [44]. Deze functie neemt een enkel frame als input en geeft het frame een score op basis van het YOLOv5-model. Aan deze functie worden volgende parameters meegegeven:

- frame: frame (linker of rechter camera)
- model: gebruikte detectiemodel (YOLOv5) (YOLOv5s-helmdetectiemodel (.pt))
- device: beschikbare hardware voor detectie (CUDA (NVIDIA GPU) / CPU) (CPU in geval gebruikte computer)

De output van deze functie bedraag de labels van de (eventuele) detectie(s), met hierbij de hoekcoördinaten van de linker tophoek en rechter onderhoek van de bounding boxes, in de vorm van homogene coördinaten. Dit zijn de hoekcoördinaten (homogeen) van de rechthoeken die de locatie van de helm (op beeld) aangeven. Deze hoekcoördinaten worden verduidelijkt op Figuur 68. Ook is hier het label ('HELMET'), dat aangegeven wordt boven de bounding box, goed zichtbaar. Het effectieve tekenen en weergeven van de bounding box met label wordt in stap 2 besproken. De homogene coördinaten dienen voor het plotten van de bounding box namelijk eerst omgezet te worden naar coördinaten binnen het gebruikte beeldcoördinatenstelsel. Figuur 68 dient enkel ter verduidelijking van de 2 homogene hoekcoördinaten.



Figuur 68 Verduidelijking hoekcoördinaten bounding box

De output van de functie is in de vorm van arrays. Eén array bevat alle labels (enkel 'HELMET' in dit geval). Een tweede array, bestaat op zichzelf ook uit arrays. Voor elke detectie, met label, een array met de 4 eerder besproken hoekcoördinaten (homogeen binnen het beeldcoördinatenstelsel) en als laatste en 5de waarde ook nog de zekerheid score. Deze geeft, zoals reeds eerder besproken, aan (waarde tussen 0 en 1, inclusief 1) hoe zeker het model is dat het om een helm gaat. Binnen code wordt de eerste array met de labels gekoppeld aan een variabele 'labels' en de tweede array met de hoekcoördinaten en zekerheid score aan de variabele 'cord' (coördinaten). Dit ter verder gebruik.

Beide arrays worden opgesteld op basis van de positie van een detectie op beeld. Aan de hand van de positie van een detectie binnen de array, wordt er aan elke detectie een ID toegewezen. Dit gebeurt niet binnen de functie, maar binnen de code zelf. Het model wordt op beide beelden op dezelfde manier toegepast, dit zorgt voor gelijke ID's van detecties, op basis van de posities van detecties op beeld. Hierbij is het belangrijk dat een helm zichtbaar is op beide beelden. Gezien beide beelden, door hun verschil in positie en perspectief een zone hebben waar geen overlap tussen beide beelden is (1 specifieke kant per camera) werd er gedacht aan het verwijderen van deze zones op beeld. Binnen Aperam werd deze optie besproken, maar de keuze gemaakt om beide zones te behouden voor eventuele latere doeleinden. Dit met als gevolg dat het noodzakelijk is dat een helm zichtbaar is op beide beelden, voor het uitvoeren van (correcte) 3D-positiebepaling op basis van stereovisie.

Zoals reeds eerder verduidelijkt gebeurt dit apart per vervormde, gerectificeerde en getransformeerde beeld van een paar stereobeelden (linker en rechter beeld). De functie wordt dus tweemaal opgeroepen met parameters van eenmaal de linker en andermaal de rechter camera.

STAP 2: Weergeven bounding boxes met label op frame (beeld)

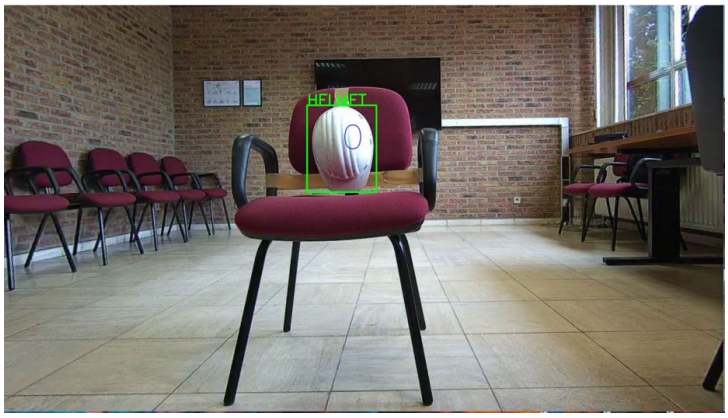
Binnen een 2de stap worden de effectieve bounding boxes, met bijhorend label (enkel 'HELMET' in dit geval, op elke beeld van een paar stereobeelden (linker / rechter camera) weergegeven. Volgende functie, die verantwoordelijk is voor het weergeven van de bounding boxes met bijhorend label op beeld, wordt dus ook per beeld (frame) en per camera apart uitgevoerd.

Als functie wordt er gebruik gemaakt van de YOLOv5 Custom Object Detection 'plot_boxes' functie [44]. Deze functie neemt een frame en reeds eerder bepaalde resultaten ervan (labels en homogene hoekcoördinaten) als invoer en plot de bounding boxes en het label op elk frame.

Als input aan de functie worden de volgende parameters meegegeven:

- results: reeds eerder bepaalde resultaten uit 'score_frame' functie (array van labels en array van arrays met bijhorende homogene hoekcoördinaten en zekerheid score)
- frame: frame (linker of rechter camera)
- model: gebruikte detectiemodel (YOLOv5) (YOLOv5s-helmdetectiemodel (.pt))
- minimale zekerheid (optioneel): minimale zekerheid score voor plotten bounding box met bijhorend label

De output van deze functie bedraagt het ingevoerde frame, met hierop de geplotte bounding boxes met bijhorend label. Een voorbeeld hiervan is te zien op Figuur 69.



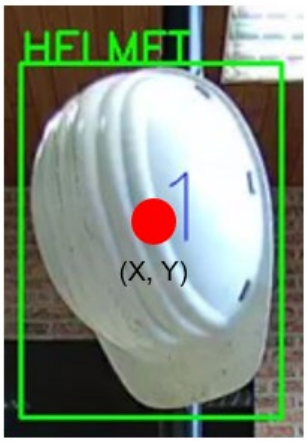
Figuur 69 Voorbeeld output plot_boxes functie (bounding box, rondom helm, met bijhorend label (tekst 'HELMET') in het groen) (ID-detectie ('0') weergegeven in het blauw)

Indien er een minimale zekerheid score meegegeven werd aan de functie, zal er binnen de functie eerst gekeken worden of de zekerheid score van een detectie zich boven deze waarde bevindt of hier gelijk aan is. Indien de zekerheid score zich onder dit minimale niveau bevindt, wordt er geen bounding box op beeld geplot met bijhorend label. In andere gevallen wel.

Voor het plotten van een bounding box op beeld (beeldcoördinatenstelsel) worden de homogene coördinaten van de 2 meegeven hoekpunten (verduidelijkt op Figuur 68), binnen de functie, omgezet naar werkelijke coördinaten op beeld (beeldcoördinatenstelsel). Elke homogene X-waarde (homogene X-coördinaat) wordt vermenigvuldig met de breedte (in pixels) van ieder frame (1280 in dit geval). Elke homogene Y-waarde (homogene Y-coördinaat) wordt vermenigvuldig met de hoogte (in pixels) van ieder frame (720 in dit geval). Op basis van deze coördinaten binnen het beeldcoördinatenstelsel, worden de bounding boxes geplot met bijhorend label als tekst erboven. Binnen de bounding box wordt ook het ID van de detectie als tekst weergegeven.

3.6.5 Bepaling centrumcoördinaat

Verdergaand op bovenstaande stap 2 uit deel 3.6.4 wordt er voor iedere geplotte bounding box, de centrumcoördinaat binnen het beeldcoördinatenstelsel bepaald. Dit ter bepaling van de locatie van het object (veiligheidshelm in dit geval) aan de hand van slechts 1 paar coördinaten (X, Y) (centrum detectie) op beeld (beeldcoördinatenstelsel). De centrumcoördinaat van een bounding box wordt verduidelijkt op Figuur 70. Binnen de code gebeurt dit apart per detectie ID en apart per bounding box (eentje met ID x op het linker beeld en de andere met hetzelfde ID x op het rechter beeld). Dit gezien er binnen de volgende fase, onmiddellijk na het bepalen van de centrumcoördinaten van de bounding boxes per detectie ID, er per detectie ID de 3D-positie bepaald wordt. Kort samengevat wordt er dus per detectie ID voor beide bounding boxes op het stereobeeld (één op het linker beeld en één op het rechter beeld) de centrum coördinaten bepaald. Binnen de code wordt het bepalen van de centrum coördinaat per bounding box, met specifieke ID, apart per beeld uitgevoerd. Dit ID, per ID, gezien er onmiddellijk aan 3D-positiebepeling per ID gedaan wordt.



Figuur 70 Verduidelijking centrumcoördinaat bounding box

Als eerste stap, voor het bepalen van de centrumcoördinaat van een bounding box, per ID en per stereobeeldbeeld apart (links, rechts), wordt de gehele array uit stap 1 van deel 3.6.4, met hierin per detectie een array met de coördinaten van de 2 hoekpunten en de zekerheid score, gefilterd op basis van dezelfde gebruikte minimale zekerheid score als binnen stap 2 van deel 3.6.4. Het totaal aantal getekende bounding boxes per beeld, kom hierdoor ook overeen met het aantal arrays, met coördinaten van hoekpunten en de zekerheid score, binnen de bovenliggende array. Dit filteren wordt ook gelijkaardig uitgevoerd binnen de 'plot_boxes' functie uit stap 2.

Wanneer deze array van arrays bekeken wordt voor beide stereo beelden, komt het aantal arrays binnen de array overeen met het aantal detectie ID's op beeld.

Per detectie ID apart op het paar stereobeelden, worden vervolgens de centrumcoördinaten van de 2 bijhorende bounding boxes (één met ID op linker beeld en één met zelfde ID op rechter beeld) bepaald.

Hiervoor worden in eerste instantie de homogene coördinaten van de 2 meegeven hoekpunten (verduidelijkt op Figuur 68) omgezet naar werkelijke coördinaten op beeld (beeldcoördinatenstelsel). Elke homogene X-waarde (homogene X-coördinaat) wordt vermenigvuldig met de breedte (in pixels) van ieder frame (1280 in dit geval). Elke homogene Y-waarde (homogene Y-coördinaat) wordt vermenigvuldig met de hoogte (in pixels) van ieder frame (720 in dit geval). Dit levert de werkelijke coördinaten op beeld van beide hoekpunten op binnen het beeldcoördinatenstelsel.

Gebaseerd op de coördinaten van de 2 hoekpunten, verduidelijkt op Figuur 68, binnen het beeldcoördinatensysteem wordt er aan de hand van volgende berekening de centrumcoördinaat berekend:

$$(X1 + (X2 - X1)/2, Y1 + (Y2 - Y1)/2)$$

Vergelijking 11 Berekening centrumcoördinaat bounding box op basis van de coördinaten van 2 hoekpunten (linkerbovenhoek en rechter onderhoek) (beeldcoördinatenstelsel)

Bij de X-coördinaat van de linkerbovenhoek van de bounding box, wordt de helft van de breedte van de bounding box (in pixels) opgeteld. Hetzelfde gebeurt voor de Y-coördinaat, alleen wordt er hier gebruik gemaakt van de helft van de hoogte (in pixels) van de bounding box. Het gebruik van Vergelijking 11 levert de centrumcoördinaat per bounding box op.

De centrumcoördinaten (beeldcoördinatenstelsel) van de 2 bounding boxes met dezelfde detectie ID, worden binnen een volgende fase gebruikt voor het effectief bepalen van de 3D-positie aan de hand van stereovisie. Deze 3D-positiebepaling gebeurt apart per detectie ID. Vandaar, zoals reeds eerder vermeld, het apart bepalen van de centrumcoördinaten van de bounding boxes per detectie ID.

3.7 3D-positiebepaling

Binnen deze laatste fase van de ontwikkeling van het 3D-positiebepalingssysteem op basis van stereovisie worden de effectieve 3D-posities van objecten (veiligheidshelmen in dit geval) binnen de werkelijke ruimte bepaald. Dit ten opzichte van de stereovisie opstelling. Het bepalen van de 3D-positie gebeurt, net zoals het berekenen van de centrumcoördinaten van de bounding boxes, per ID.

Resultaten worden binnen Python gekoppeld aan variabelen ter mogelijke verdere verwerking en worden ook uitgeprint binnen de terminal, met verduidelijking van hun ID op beeld.

3.7.1 Verduidelijking 3D-positie

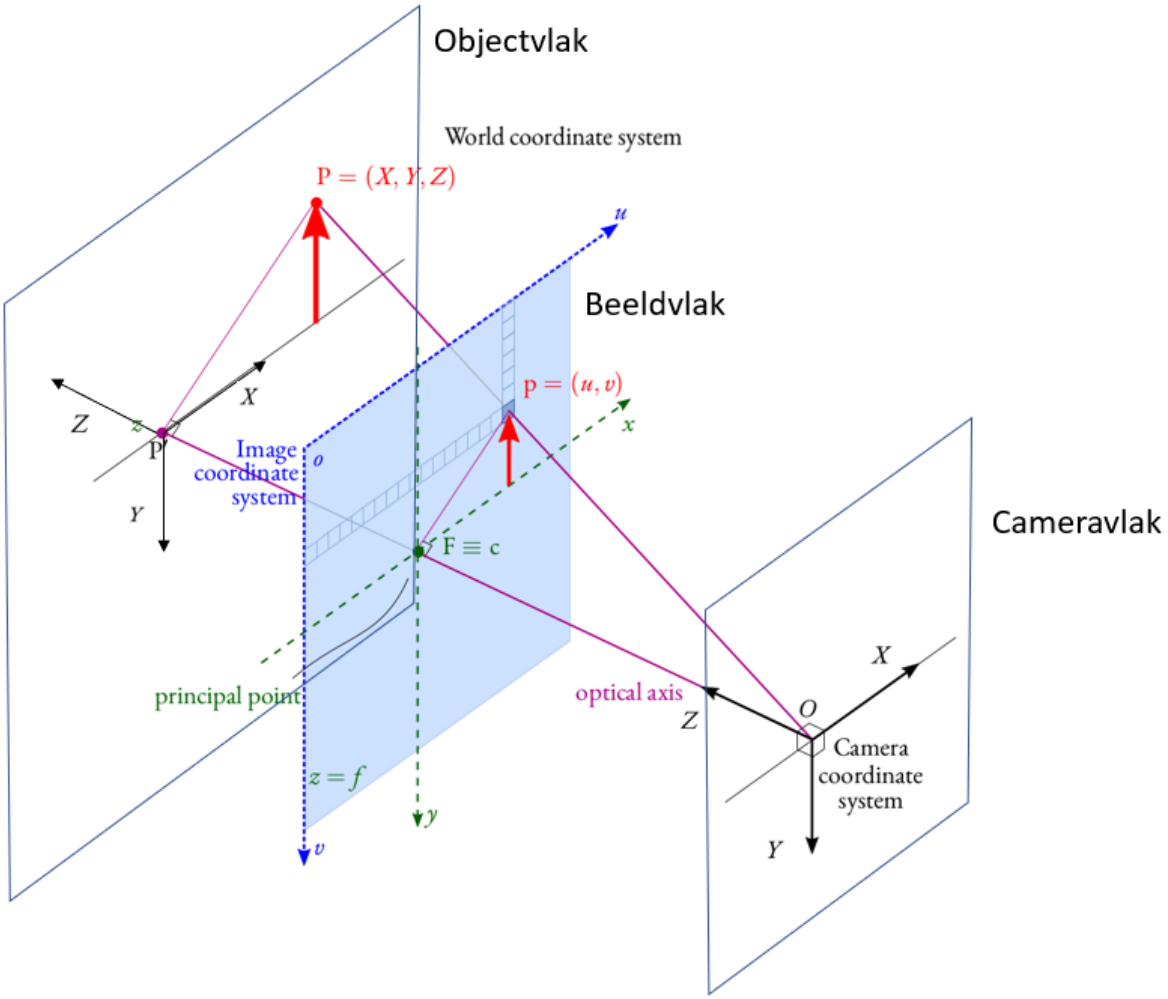
De uiteindelijk 3D-positie van een object wordt weergegeven in de vorm van werkelijke coördinaten (X, Y, Z) (uitgedrukt in cm) binnen de 3D-ruimte (werkelijk coördinatenstelsel). Dit ten opzichte van het middelpunt van de stereovisie opstelling (middelpunt tussen beide camera lenzen). Onderstaande Figuur 71 verduidelijkt deze coördinaten en later gebruikte vlakken. Het rode punt P stelt hierbij een object voor dat zich bevindt op een 3D-positie met coördinaten (X, Y, Z) binnen het werkelijk coördinatenstelsel ('World coordinate system'). Dit ten opzichte van de stereovisie opstelling, aangeduid op Figuur 71 als 1 geheel (linksonder). Binnen deze figuur wordt er rekening gehouden met de gerectificeerde camerabeelden (verticale uitlijning en parallel verloop beide optische assen, met enkel een horizontale verplaatsing tussen beide beelden. Hierover later meer verduidelijking. Beide camera's bevinden zich hierdoor softwarematig binnen hetzelfde vlak (cameravlak) (enkel een horizontaal verschil in positie).

Het objectvlak, beeldvlak en cameravlak zijn evenwijdig aan elkaar en staan loodrecht op de optische as (gecombineerde as voor stereovisie opstelling, vertrekkend vanuit het middelpunt tussen beide camera's (centrumpunt cameravlak) en parallel aan de gecorrigeerde optische assen van beide camera's). Het snijpunt van deze vlakken met de (gemeenschappelijke) optische as van de stereovisie opstelling wordt het centrumpunt van ieder vlak genoemd. Het punt P, waar een zogenaamd object zich bevindt, ligt binnen het objectvlak, dat evenwijdig is met het cameravlak, op een bepaalde (X, Y) positie ten opzichte van het centrumpunt van dit vlak. De afstand tot dit vlak, vanaf het cameravlak (centrumpunt), wordt aangegeven door de Z-coördinaat. Gezien beide vlakken evenwijdig aan mekaar zijn, is deze afstand overal gelijk,

voor dezelfde vlakken. Bijgevolg hebben de 3D-coördinaten (positie) (in cm) die bepaald worden binnen het werkelijke coördinatenstelsel, ten opzichte van de (gerectificeerde) stereovisie opstelling de volgende betekenis (ook verduidelijkt op Figuur 71):

- Z: Afstand tussen cameravlak en objectvlak (vlak waarin object zich op een bepaalde (X, Y) positie bevindt)
- X: Horizontale verplaatsing van object binnen het objectvlak, ten opzichte van het centrumpunt van het objectvlak
- Y: Verticale verplaatsing van het object binnen het objectvlak, ten opzichte van het centrumpunt van het objectvlak

Extra verduidelijking en oriëntaties van de assenstelsels zijn aangegeven op Figuur 71. Binnen het beeldvlak op deze figuur wordt er gebruik gemaakt van het projectiecoördinatenstelsel en niet van het beeldcoördinatenstelsel. Beide coördinatenstelsels en hun verschil worden besproken aan het einde van deel 3.4.2 bij de bespreking van Figuur 59, waar deze ook visueel verduidelijkt worden.



Figuur 71 Verduidelijking vlakken en coördinaten [45](aangepast)

3.7.2 Bepaling 3D-positie

Binnen de code wordt de 3D-positie berekend per ID. Hiervoor wordt er gebruik gemaakt van de eerder bepaalde projectiematrices van beide camera's, uit stap 3 van deel 3.4.2, en de centrumcoördinaten (beeldcoördinatenstelsel) van de bijhorende bounding boxes (één op het linker beeld met ID x en één op het rechter beeld met ID x).

Deze 4 parameters worden, binnen de code, meegegeven aan de OpenCV 'triangulatePoints' functie. Deze functie reconstrueert driedimensionale punten (in homogene coördinaten) uit hun waarnemingen met de stereovisie opstelling [27]. Als input worden aan de functie de volgende parameters meegegeven [27]:

- 3x4 projectiematrix van de linker camera
- 3x4 projectiematrix van de rechter camera
- 2x1 centrumcoördinaat bounding box linker beeld (beeldcoördinatenstelsel)
- 2x1 centrumcoördinaat bounding box rechter beeld (beeldcoördinatenstelsel)

De projectiematrices projecteren de 3D-punten binnen het werkelijke coördinatenstelsel van de ruimte naar beide beelden (beeldcoördinatenstelsel) [27].

De output van de functie is een 4x1 array die de homogene coördinaten (3D-positie) (X, Y, Z) van het object van belang (veiligheidshelm binnen deze ontwikkeling) bevat in combinatie met een 4^{de} waarde die de omzettingsfactor weergeeft. Deze homogene punten zijn allemaal uitgedrukt in het werkelijke coördinatenstelsel [27]. Dit ziet er theoretisch als volgt uit:

$$[X, Y, Z, N]$$

Voor het bekomen van de werkelijke coördinaten (niet homogeen), wordt elke coördinaat (X, Y, Z) binnen de code gedeeld door de 4^{de} waarde uit deze array. De omzettingsfactor 'N' in het geval van bovenstaand voorbeeld. De werkelijke coördinaten binnen de 3D-ruimte zijn hierbij:

$$[X/N, Y/N, Z/N]$$

Het resultaat is een 3x1 array die de werkelijke coördinaten van het object binnen de 3D-ruimte bevat. Dit in centimeters [cm]. Gezien de projectiematrix van de linker camera als eerst meegegeven werd aan de OpenCV 'TriangulatePoints' functie, en deze via de OpenCV 'stereoRectify' functie bepaald werd, zijn de coördinaten niet weergegeven ten opzichte van de gehele stereo opstelling (midden). Binnen de OpenCV functiedocumentatie staat er namelijk dat indien de projectiematrices van de OpenCV 'stereoRectify' functie gebruikt werden, de geretourneerde punten voorgesteld worden ten opzichte van het gerectificeerde coördinatenstelsel van de eerste (linker) camera [27]. Voor het verkrijgen van de werkelijke 3D-positie ten opzichte van het midden van de stereo cameraopstelling, wordt er hierdoor bij de X-coördinaat de helft van de totale afstand tussen beide camera's (in cm) opgeteld. De afstand tussen beide camera's wordt verkregen vanuit de reeds eerder bepaalde translatie vector uit de kalibratie stap. Aan de Y- en Z-coördinaat dienen geen aanpassingen doorgevoerd te worden. Dit gezien beide gerectificeerde coördinatenstelsel van de camera's uitgelijnd werden in deze richtingen. Hun enigste onderlinge verschil is een translatie (verplaatsing) over de tussenafstand in de X-richting.

Het resultaat hiervan is de 3D-positie (X, Y, Z) van een object binnen de werkelijke ruimte ten opzichte van het midden van de gehele stereovisie cameraopstelling.

3.7.3 Bepaling Z-coördinaat (alternatieve methode)

Binnen de code wordt de Z-coördinaat uit deel 3.7.2 ook nog bepaald aan de hand van een andere alternatieve methode. Dit ter bewijs van de alternatieve methode en ter verificatie van de eerder bepaalde Z-coördinaat uit deel 3.7.2. Ook deze methode wordt per detectie ID apart uitgevoerd op basis van de centrumcoördinaten van beide bijhorende bounding boxes.

Deze methode maakt gebruik van de dispariteit tussen een object op beide beelden. Binnen deze masterproef gaat het om het verschil in pixels tussen de locatie van de helm (centrum bounding box) op beide beelden. Dit in de X-richting gezien er enkel in deze richting een verschil in positie aanwezig kan zijn, ten gevolge van het verschil in horizontale positie tussen beide camera's. Extra uitleg en verduidelijking omtrent de dispariteit is terug te vinden in deel 2.2 (inleiding stereovisie afstandsbepaling).
Binnen de code wordt de dispariteit op beeld berekend door het berekenen van het verschil tussen de X-coördinaat van het centrum van de bounding box uit het rechter beeld en de X-coördinaat van het centrum van de bounding box uit het linker beeld. In formule vorm ziet dit er als volgt uit:

Dispariteit = (X-coördinaat centrum bounding box rechter beeld) - (X-coördinaat centrum bounding box linker beeld)

Vergelijking 12 Berekening dispariteit

Het resultaat hiervan zal bij correcte toepassing altijd positief zijn. Indien niet kan het zijn dat linker- en rechter beeld omgekeerd werden of dienen te worden.

Naast de dispariteit van het object op beide beelden wordt er ook nog gebruik gemaakt van de afstand tussen beide camera's (in cm), verkregen uit de translatie vector uit de kalibratie stap (in mm), besproken in deel 3.4.2. Met hiernaast ook nog de brandpuntsafstand uitgedrukt in pixels, verkregen uit de projectiematrices, zoals ook besproken in deel 3.4.2.

Op basis van deze gegevens kan de Z-coördinaat die de afstand (in cm) van het cameravlak tot het objectvlak, waarbinnen het object zich op een (X, Y) positie bevindt, berekend worden. Hiervoor wordt er gebruik gemaakt van volgende formule:

Afstand(Z) = ((afstand tussen beide camera's (cm)) * (brandpuntsafstand in pixels)) / dispariteit

Vergelijking 13 Berekening afstand (Z) (Z-coördinaat) in cm

Het resultaat hiervan is Z-coördinaat die de afstand (in cm) van het cameravlak tot het objectvlak, waarbinnen het object zich op een (X, Y) positie bevindt, aangeeft. Deze waarde zou voor elke detectie ID moeten overeenkomen met de reeds eerder berekende Z-coördinaat uit deel 3.7.2.

Wanneer alle 3D-positites van alle gedetecteerde veiligheidshelmen (ID's) boven de minimale zekerheidsscore bepaald zijn van een paar stereobeelden, worden alles stappen vanaf het vastleggen van een paar stereobeelden binnen de code herhaald. Dit met uitzondering van de kalibratie en rectificatie stap. Deze dient slechts eenmaal per ongewijzigde cameraopstelling (inclusief instellingen) uitgevoerd te worden ter bepaling van de parameters. De herhaling van het geheel, met het weergeven van het paar stereobeelden met detecties en hun 3D-positites binnen de werkelijke 3D-ruimte voor de stereovisie cameraopstelling, geeft een live output van het 3D-positiebepalingssysteem. De snelheid van het geheel (aantal frames en positiebepalingen per seconde) hangen hierbij af van de verwerkingssnelheid van het geheel door de gebruikte computer. Bij de gebruikte HP-laptop ging het hier ongeveer om een 2 frames per seconde (stereobeelden) met bijhorende 3D-positiebepalingen van alle detecties op het paar stereobeelden.

4 Verificatie ontwikkeld 3D-positiebepalingssysteem

Binnen dit hoofdstuk wordt de verificatie van de werking van het ontwikkelde 3D-positiebepalingssysteem op basis van stereovisie besproken. Het gaat hier om de verificatie van de werking van het gehele systeem (3D-positiebepalingsysteem). Niet om verificatie van individueel gemaakte keuzes binnen de ontwikkeling. Deze keuzes werden namelijk al onderbouwd en deels geëvalueerd binnen de ontwikkeling zelf, besproken in hoofdstuk 3. Het vergelijken van de invloed van verschillende keuzes valt buiten de scope van deze masterproef.

De verificatie van de werking van het ontwikkelde 3D-positiebepalingssysteem op basis van stereovisie wordt uitgevoerd aan de hand van testen. Deze ter vergelijking van de resultaten van het 3D-positiebepalingssysteem ten opzichte van de werkelijk 3D-posities, geverifieerd aan de hand van manuele metingen en positioneringen. Hoe deze testen uitgevoerd werden, wordt besproken in het 1ste deel van dit hoofdstuk. De registratie en verwerking van de test resultaten wordt besproken in het 2de deel van dit hoofdstuk. De verwerkte resultaten van de testen worden vervolgens weergegeven, gevisualiseerd en besproken in het 3de deel van dit hoofdstuk. Deze resultaten stellen ons in staat om in een volgend hoofdstuk een finale conclusie te kunnen trekken, omtrent de ontwikkeling en werking van het systeem.

Voor het uitvoeren van testen ter verificatie van het ontwikkelde stereovisie 3D-positiebepalingssysteem wordt er gebruik gemaakt van het eerder ontwikkelde YOLOv5s-helmdetectiemodel. Dit model en de ontwikkeling ervan, werd reeds eerder besproken binnen deel 3.6. Zoals daar ook vermeld is het model eenvoudig hertrainbaar en binnen de code eenvoudig aanpasbaar.

Aperam wenst voor het gebruik van het systeem, ter gebruik binnen de digitalisering van het veiligheidskegel veiligheidshulpsysteem, een maximale afwijking van 50 cm ten opzichte van werkelijke posities te bekomen. Bij waardes hoger dan deze waarde, kan het systeem niet goedgekeurd en gebruikt worden.

4.1 Omschrijving testen

Binnen dit deel van de verificatie van de werking van het ontwikkelde 3D-positiebepalingssysteem worden de testen zelf besproken. Het gaat hier over de testlocatie, testopstelling, testmethode, etc. De resultaten van de testen worden binnen het volgende deel weergegeven, gevisualiseerd en besproken.

4.1.1 Testlocatie

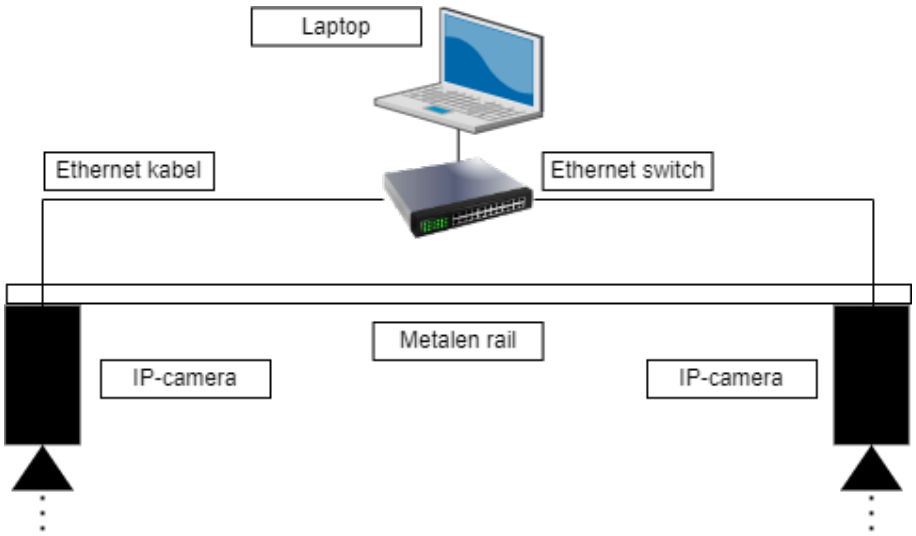
Voor het uitvoeren van de testen werd er binnen Aperam gebruik gemaakt van een beschikbare vergaderzaal. Alle voorwerpen, zoals tafels, stoelen, elektronica, etc. werden binnen deze ruimte aan de kant geschoven om zo veel mogelijk open ruimte ter beschikking te hebben. De test locatie wordt weergegeven op Figuur 72 Testlocatie. Meer details omtrent de testlocatie (zoals grootte) worden later nog besproken binnen het gedeelte omtrent de test methodes.

Figuur 72 Testlocatie

4.1.2 Testopstelling

Voor het uitvoeren van de testen binnen de beschikbare vergaderzaal van Aperam, werd er gebruik gemaakt van de eerder besproken stereovisie opstelling binnen deel 3.2. Deze opstelling ziet eruit zoals afgebeeld op Figuur 73 en schematisch weergegeven op Figuur 74. Zoals vermeld binnen deel 3.2 werd de opstelling zo nauwkeurig mogelijk uitgelijnd binnen de ruimte (waterpas plaatsing, zelfde afstand van de muur, etc.). Dit om het testen te kunnen vergemakkelijken. Het cameravlak, beeldvlak, objectvlak, zouden hierbij naast evenwijdig aan elkaar, ook evenwijdig horen te zijn met de muur achter en muur voor de cameraopstelling. Ook de vierkante tegels op de grond, met gekende grootte, zouden dan gebruikt kunnen worden. De zo correct mogelijke uitlijning van de opstelling ten opzichte van de gebruikte testruimte (testlocatie) vormt een belangrijke factor binnen het latere verifiëren van de werkelijke 3D-positites. Ook de camera's onderling werden zo nauwkeurig mogelijk manueel uitgelijnd. Dit om manueel een zo een correct mogelijke parallelle stereovisie opstelling te bekomen, ter vergemakkelijking van het kalibratie en rectificatieproces. Posities van de opstelling en camera's onderling werden opgemeten en gecorrigeerd op basis van een rolmeter en een waterpas.

Figuur 73 Stereovisie cameraopstelling



Figuur 74 Schematische voorstelling stereovisie opstelling

Als detectieobject werd er zoals reeds eerder al vermeld gebruik gemaakt van een veiligheidshelm. Dit gezien het ontwikkelde en gebruikte YOLOv5s-detectiemodel hierop getraind werd. Binnen de testopstelling werd er gebruik gemaakt van 2 helmen. Deze werden wat plaatsing betreft recht boven elkaar geïnstalleerd. Gezien de zo nauwkeurig mogelijke uitlijning van de opstelling, zouden deze zich hierdoor binnen hetzelfde objectvlak, op een andere (X, Y) positie moeten bevinden. De Z-afstand van beide helmen zou normaalgezien (zonder fouten binnen de manuele uitlijning (onzekerheid testen)) gelijk dienen te zijn. Een verduidelijking van de opstelling van de 2 helmen wordt getoond op Figuur 75. Er werd hierbij gekozen voor een verschil in verticale positie. Theoretisch gezien zou er dus enkel een verschil in Y-coördinaat (positie) mogen zijn.

Figuur 75 Testopstelling helmen

Binnen de later besproken effectieve testen, ter verificatie van de werking van het 3D-postitiebepalingssysteem op basis van stereovisie, bevindt de onderstaande helm zich altijd op dezelfde gefixeerde positie binnen de helm testopstelling. De bovenste helm werd op variabele hoogtes geplaatst. Dit gezien deze, voor elke later besproken testpositie, op beide camerabeelden van het stereopaar zichtbaar dient te zijn, en dit anders niet het geval was. Hierover meer duiding bij het bespreken van de effectieve testen.

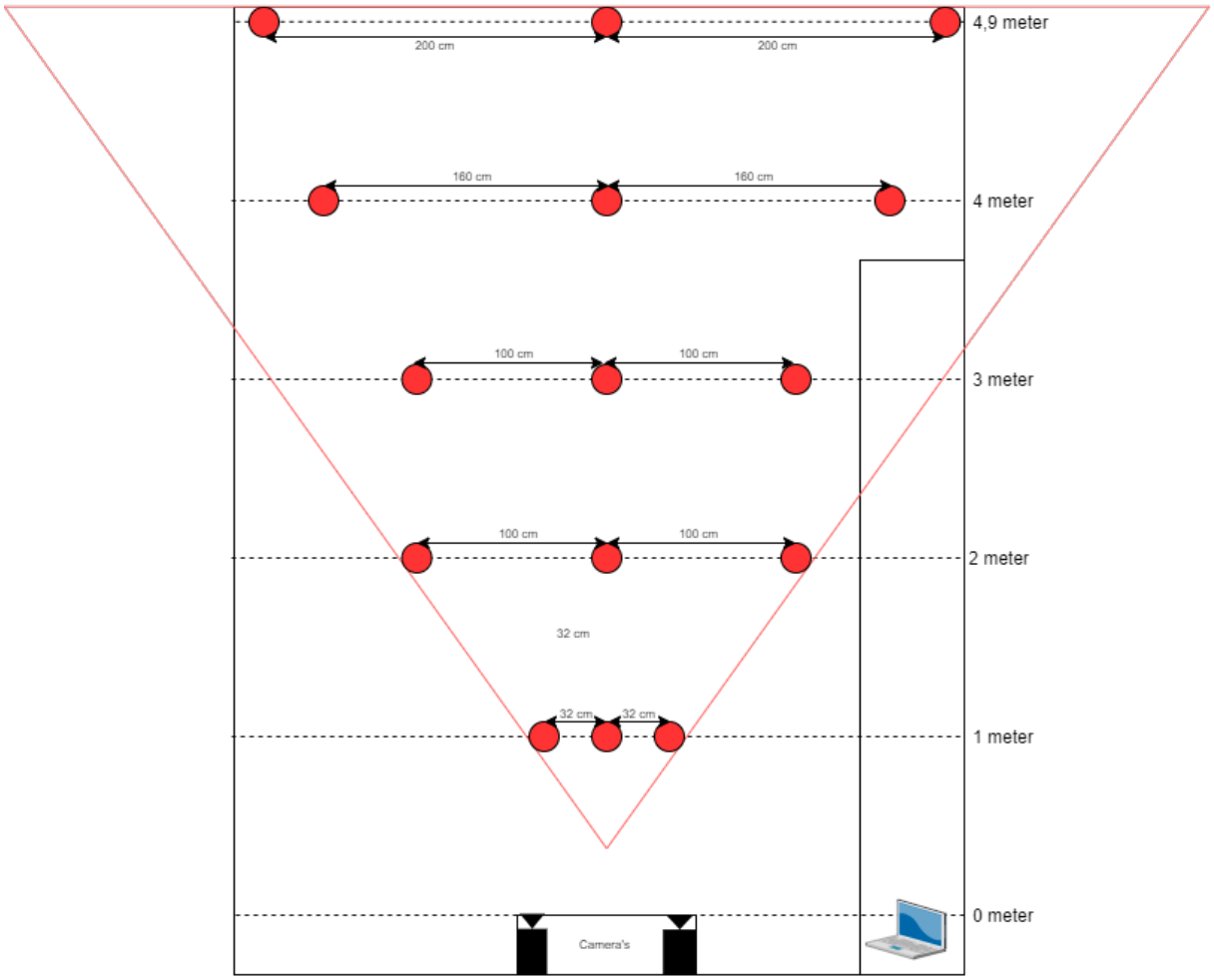
Binnen de ontwikkeling van het geheel werd er, zoals in deel 3.2 aangegeven, ook al gebruik gemaakt van de testopstelling. Zo werden binnen de uitleg omtrent de ontwikkeling (hoofdstuk 3) de parameters van de testopstelling, kalibratie en rectificatie besproken (projectiematrices, translatievectoren, translatiematrices, etc.). Ook werd er voor het trainen van het YOLOv5s-detectiemodel gebruik gemaakt van afbeeldingen, gemaakt met de gekalibreerde en gerectificeerde cameraopstelling.

Voor het uitvoeren van de later besproken testen, werd er al aan kalibratie en rectificatie van het geheel gedaan. Dit zoals besproken binnen deel 3.4. De parameters van deze specifieke testopstelling (rotatie matrices, translatie vectoren, projectie matrices, etc.) werden hier ook al weergegeven en besproken. Deze worden namelijk gebruikt voor het vervormen en uitlijnen van de camerabeelden en het berekenen van de 3D-positie.

4.1.3 Testmethode

Op basis van de uitgelijnde, gekalibreerde en gerectificeerde stereovisie cameraopstelling (beelden), werden er testen, ter verificatie van de werking van het 3D-positiebepalingssysteem op basis van stereovisie, uitgevoerd. Hiervoor werd er gebruik gemaakt van de besproken helm testopstelling, weergegeven op Figuur 75 en besproken in de bijhorende alinea erboven.

Binnen de kantoorruimte, waar de testen uitgevoerd werden, werd deze testopstellingen (helmen) op verschillende statische locaties ten opzichte van de stereovisie cameraopstelling geplaatst. Deze locaties (rode stippen) binnen de kantoorruimte worden verduidelijkt op onderstaande Figuur 76. Ter verificatie van de werking van het systeem worden resultaten van het ontwikkelde 3D-positiebepalingssysteem op basis van stereovisie in een volgende fase vergeleken met de werkelijke 3D-posities van de helmen (2) binnen de ruimte (rode stippen). Beide resultaten ten opzichte van de stereovisie cameraopstelling. Alle resultaten, zowel de werkelijke, als de bepaalde door het 3D-positiebepalingsststeem worden bijgehouden in een Excel-bestand, om deze van hieruit te kunnen verwerken en visualiseren? Dit om hieruit conclusies te kunnen trekken.



Figuur 76 Visuele verduidelijking locaties (rode stippen) testopstelling (veiligheidshelmen), met aanduiding afstanden binnen de ruimte

Voor het bepalen van de locaties werd er rekening gehouden met verschillende factoren. Bepalingen en variaties van de locaties worden in onderstaand delen besproken. Dit in combinatie met een deeltje omtrent extra duiding bij de verificatie van de werkelijke posities en een deeltje omtrent de onzekerheid van deze testmethode voor het in een volgende fase kunnen vergelijken van werkelijke en bepaalde 3D-posities ten opzichte van de stereovisie cameraopstelling.

Variatie diepte (Z-waarde)

Om te beginnen werd er gekozen voor een vaste variatie in diepte (Z-waarde 3D-positie), van telkens één meter, ten opzichte van de cameraopstelling. Deze afstanden worden aangegeven aan de rechter kant van de grote rechthoek, die de kantoorruimte in 2D (bovenaanzicht) voorstelt. Het gaat hier over afstanden van (0 meter), 1 meter, 2 meter, 3 meter, 4 meter en 4,9 meter. Gezien de limieten van de vergaderzaal, wat grootte betreft, was een afstand van 5 meter niet haalbaar. De lijnen, die deze afstanden aangeven, lopen parallel met elkaar en parallel ten opzichte van de muren voor en achter de stereovisie cameraopstelling. Binnen de werkelijke ruimte werden er op deze afstanden ook markeringen aangebracht met tape. Deze afstanden werden binnen de ruimte, ten opzichte van de stereovisie cameraopstelling, opgemeten aan de hand van een rolmeter. Rekening houdend met het parallelle verloop. Gezien de zo nauwkeurig mogelijke uitlijning van de stereovisie cameraopstelling, wat onderlinge positionering en gelijke afstand tot de muur achter de opstelling betreft, zouden deze lijnen en markeringen in theorie ook parallel dienen te lopen, ten opzichte van de stereovisie cameraopstelling. Naast het

zo nauwkeurig mogelijk manueel nameten van deze afstanden en het zo nauwkeurig mogelijk manueel uitlijnen van beide camera's ten opzichte van elkaar, kon deze plaatsing (met parallel verloop) natuurlijk niet met 100% zekerheid geverifieerd worden. Eventuele incorrecties binnen deze uitlijning en het bepalen van de afstanden vormen een onzekerheid voor de verificatie van het 3D-positiebepalingssysteemop basis van deze afstanden.

Variatie horizontale verplaatsing (X-waarde)

Naast het toepassen van een variatie in diepte (uiteindelijke Z-waarde) werd er ook gekozen om de statische positie van de testopstelling (helmen) te variëren in horizontale richting (uiteindelijke X-waarde). Dit telkens over een variabele afstand. Voor het bepalen van deze variabele afstand, werd er rekening gehouden dat de testopstelling (helmen) zich binnen het gebied van overlap van beide vervormde, gerectificeerde en getransformeerde beelden van beide camera's bevindt. Ook werd er rekening gehouden met de limieten van de ruimte. Zo kon de testopstelling in sommige gevallen niet verder naar links of rechts verplaatst worden. Er werd telkens gekozen voor een gelijke verplaatsing in de horizontale richting (links en rechts) ten opzichte van het midden. Deze variaties in verplaatsing werden apart per diepteafstand bepaald. Per diepteafstand werd de helm 1 maal in het midden gepositioneerd voor een test.

Voor het bepalen van het midden (X=0 (centrumpunt)) (werkelijke 3D-coördinatenstelsel) werd er gekeken naar de lijnen van de tegels op de vloer. De vierkante tegels (40 cm x 40 cm) liggen namelijk in een vast patroon op de grond. Parallel aan de eerder bepaalde lijnen die de diepteafstanden vanaf het (uitgelijnde) cameravlak aangeven. De opstelling werd gepositioneerd met het middelpunt boven een lijn tussen tegels op de vloer. Gezien de zo nauwkeurig mogelijke manuele positionering van de camera's zou deze lijn theoretisch parallel lopen aan optische assen en bijgevolg ook loodrecht staan op het uiteindelijke cameravlak. Beide camera's werden namelijk manueel ook zo waterpas mogelijk geplaatst. Ook hier vormt de uitlijning van de positionering een onzekerheid voor de verificatie van het geheel.

Naast het midden werd de testopstelling (helmen) ook nog naar links en rechts verplaats. Dit over een gelijke vaste afstand per diepteafstand. Zoals reeds eerder uitgelegd. Alle 3 de posities per diepteafstand worden ook verduidelijkt op Figuur 76. Hierop staan ook de afstanden van de verplaatsingen (links en rechts) ten opzichte van de middenpositie aangegeven. De rode driehoek geeft het gebied van overlap tussen beide vervormde, gerectificeerde en getransformeerde beelden weer. Bij kleine diepte waardes vormde dit de limiet voor verplaatsing in horizontale richting.

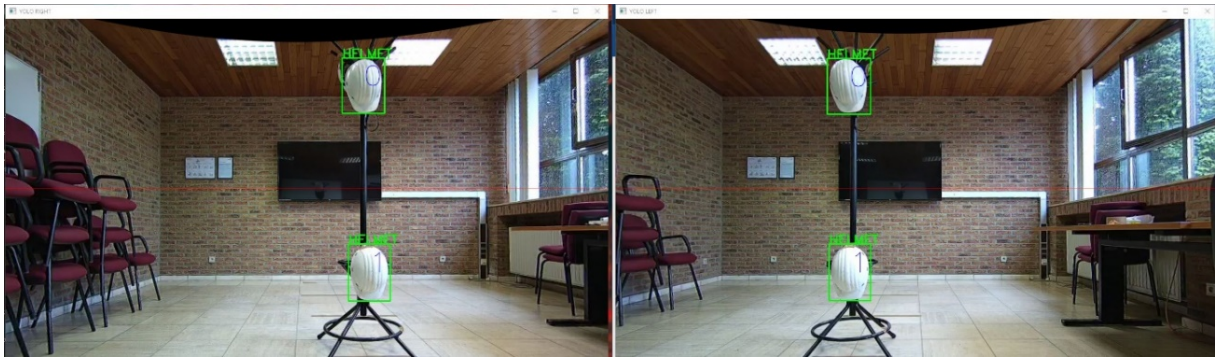
Variatie verticale verplaatsing (Y-waarde)

Naast het verplaatsen van de testopstelling (helmen) in diepte (Z-waarde) en in horizontale afstand (X-waarde) werd er ook nog een noodzakelijke verplaatsing gedaan van de bovenste helm van de testopstelling. Dit om ervoor te zorgen dat deze zich bij iedere positie van de testopstelling binnen beeld bevindt. Deze verticale verplaatsing van de bovenstaande helm was hetzelfde voor iedere horizontale positie bij een bepaalde diepteafstand. Per dieptepositie werd deze zo hoog mogelijk geplaatst. Dit tot aan het hoogst mogelijke punt van de testopstelling, wat het geval was voor een diepte van 4 meter en 4,9 meter. De onderste helm bevond zich altijd op dezelfde positie. Ten gevolge van de correcte uitlijning van beide camera's onderling en ten opzichte van de ruimte zoals eerde besproken, zou deze zich dus theoretisch overal op dezelfde Y-waarde moeten bevinden.

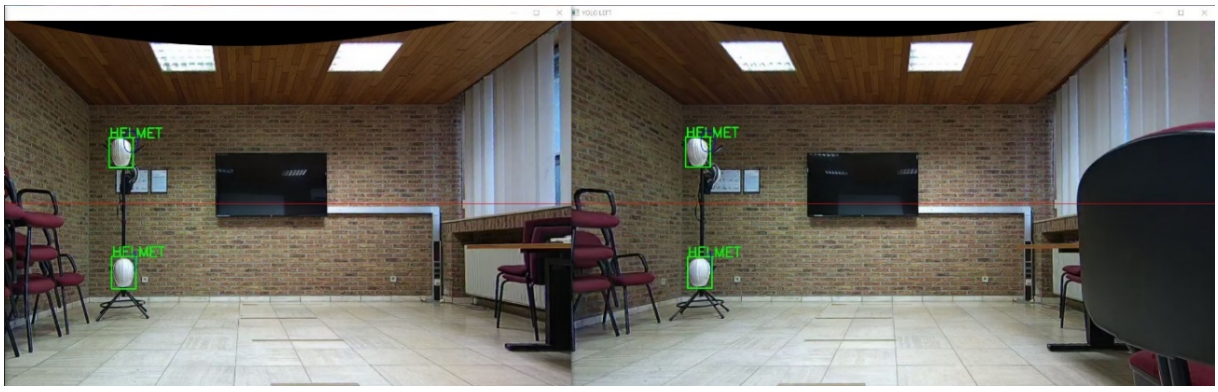
Verduidelijking posities

Onderstaande figuren verduidelijken enkele van de gebruikte posities vanuit het vervormde, gerectificeerde en getransformeerde beeld van beide camera's. Op deze figuren zijn duidelijk de detecties met bijhorende ID's zichtbaar. Dit in combinatie met verplaatsingen binnen de ruimte in X, Y en Z positie. Locaties worden aangegeven in het bijschrift van de figuren. Verplaatsingen naar links en rechts worden, zoals reeds eerder besproken, bekeken vanuit het perspectief van iemand die al kijkend naar de opstelling, voor de opstelling staat. Vanuit het perspectief van de camera's dienen deze termen omgedraaid te worden. Dit is een keuze die aan het begin van de ontwikkeling gemaakt werd.

Figuur 77 Verduidelijking positionering 1 (Diepte: 1 meter, horizontale verplaatsing: rechts (32 cm))



Figuur 78 Verduidelijking positionering 2 (Diepte: 2 meter, horizontale verplaatsing: geen (midden))



Figuur 79 Verduidelijking positionering 3 (Diepte: 5 meter, horizontale verplaatsing: rechts (200 cm))

Verificatie werkelijke posities

Alle bovenstaande werkelijke posities binnen de 3D-ruimte (werkelijk 3D-coördinatenstelsel (3.7.1)) (inclusief individuele posities helmen apart) werden nagemeten en geverifieerd aan de hand van het manueel nameten van de afstanden met een rolmeter (nauwkeurigheid 1 mm) en het gebruik van een laser afstandmeter (nauwkeurigheid 1 mm). Dit ten opzichte van de cameraopstelling (binnen het werkelijke 3D-coördinatenstelsel (0, 0, 0)), uitgaande van de eerder besproken positionering, met uitlijning, binnen de 3D-testruimte. De gebruikte laser afstandsmeter wordt weergegeven op Figuur 80.

Figuur 80 Gebruikte lasterscanner voor nameting en verificatie afstanden en posities

Onzekerheden testmethode

Zoals reeds eerder aangegeven bevat de gebruikte testmethode een onzekerheid. Deze berust namelijk op het feit dat beide camera's perfect waterpas gepositioneerd staan en perfect uitgelijnd zijn ten opzichte van de muur achter de camera's, en dus bijgevolg ook de testruimte (tegels, andere muur, etc.). Een 100% zekerheid kan echter niet bereikt worden. Dit gezien beide camera's manueel waterpas gepositioneerd werden en aan de hand van een manuele meting uitgelijnd werden. Ook op het bepalen van de werkelijke posities, aan de hand van de rolmeter en laserafstandsmeter, kan altijd een onzekerheid zitten.

Beide factoren, in combinatie met eventuele andere factoren, kunnen een invloed hebben op de uiteindelijke resultaten. Gezien het voornamelijk gaat om een verificatie van de werking van het 3D-positiebepalingssysteem op basis van stereovisie, dient er rekening gehouden te worden met deze factoren bij het trekken van conclusies omtrent de werking van het systeem op basis van de resultaten.

4.2 Registratie en verwerking testresultaten

4.2.1 Registratie testresultaten

Zowel de werkelijke 3D-posities ten opzichte van de stereovisie cameraopstelling, bepaald aan de hand van manuele metingen, als resultaten van het 3D-positiebepalingssysteem op basis van stereovisie, ten opzichte van de stereovisie cameraopstelling, werden tijdens het uitvoeren van de testen bijgehouden in tabellen in een Excel-bestand. Zo een tabel ziet eruit zoals Tabel 1. Alle afstanden en posities werden hierbij in centimeter ingevuld, wat de standaard eenheid is voor de bepaalde 3D-posities aan hand van het 3D-positiebepalingssysteem op basis van stereovisie.

Links in het witte vakje onder 'AFSTAND' werd telkens de gebruikte diepte afstand (Z), waarbinnen er verplaatsingen uitgevoerd werden in X- (en Y-) richting, ingevoerd. Het gaat hier over de afstanden van 1 meter, 2 meter, 3 meter, 4 meten en 4,9 meter (in cm). In totaal werd het geheel dus 5 keren ingevuld. Dit voor elke vaste verandering in diepte (Z) van 1 meter.

Vervolgens werden de werkelijke en bepaalde 3D-posities (werkelijke coördinatenstelsel) in de vakjes onder de groene titels ingevoerd. Elke titel heeft betrekking op een helm, binnen een bepaalde positie. De gebruikte term 'LAAG' binnen de groene titels, heeft betrekking tot de onderste helm van de 2 gebruikte helmen. De gebruikte term 'HOOG' op de bovenste helm. Verder benadrukken de andere termen 'LINKS', 'MIDDEN' en 'RECHTS' de eventuele horizontale verplaatsingen van de gehele helmen testopstelling (enkel verplaatsing bij 'LINKS' en 'RECHTS'.

In de 7 witte vakjes onder elke groene titel werden op de eerste rij de werkelijke X, Y en Z posities ten opzichte van de stereovisie cameraopstelling ingevuld. Dit binnen de juiste kolom (X, Y, Z). De rij waarin de werkelijke waardes ingevuld werden, wordt aan de linkerkant aangeduide met de term 'WERK.'. In de rij hieronder, voorafgegaan met de term 'METING' aan de linkerkant, worden de waardes, verkregen van het 3D-positiebepalingsssyteem op basis van stereovisie, ingevuld. In het enkele vakje helemaal onder elke tabel, werd telkens de verkregen Z-waarde, via de reeds eerder besproken 'alternatieve methode', ingevuld.

Tabel 1 Weergave tabel resultaten

AFSTAND			LAAG-LINKS				LAAG-MIDDEN				LAAG-RECHTS		
			X	Y	Z		X	Y	Z		X	Y	Z
	WERK.					WERK.				WERK.			
	METING					METING				METING			
			HOOG-LINKS				HOOG-MIDDEN				HOOG-RECHTS		
			X	Y	Z		X	Y	Z		X	Y	Z
	WERK.					WERK.				WERK.			
	METING					METING				METING			

4.2.2 Verwerking testresultaten

Op basis van alle verzamelde resultaten, werd er voor elke coördinaat van elke verplaatsing van elke helm voor elk verschil in diepte, de absolute afwijking in centimeters tussen de werkelijke coördinaat (afstand) en bepaalde coördinaat (afstand) (3D-positiebepalingssysteem) berekend. Deze waarden worden aangegeven in de 3 vakjes rechts van de oranje gemarkeerde term 'AFW [cm]' in de kolom van de bijhorende coördinaat. Deze vakjes zijn te zien op Tabel 2.

Hiernaast wordt dit verschil telkens ook nog procentueel [%] uitgedrukt ten opzichte van de werkelijke positie. Deze waarden worden weergegeven in de 3 vakjes, per bijhorende coördinaat, onder het verschil in cm, rechts van de aanduiding 'AFW [%]'.

Tabel 2 Weergave tabel resultaten inclusief verwerking (oranje aanduidingen

AFSTAND														
		LAAG-LINKS				LAAG-MIDDEN				LAAG-RECHTS				
		X	Y	Z		X	Y	Z		X	Y	Z		
	WERK.				WERK.				WERK.					
	METING				METING				METING					
	AFW [cm]				AFW [cm]				AFW [cm]					
	AFW [%]				AFW [%]				AFW [%]					
		HOOG-LINKS				HOOG-MIDDEN				HOOG-RECHTS				
		X	Y	Z		X	Y	Z		X	Y	Z		
	WERK.				WERK.				WERK.					
	METING				METING				METING					
	AFW [cm]				AFW [cm]				AFW [cm]					
	AFW [%]				AFW [%]				AFW [%]					

Bij elke X-waarde voor de middelste positie is er duidelijk een grijs vakje zichtbaar bij het procentuele verschil. Dit gezien deze procentuele waardes niet berekend kunnen worden. Voor de middelste positie werd er, zoals beschreven in de testen, gebruik gemaakt van het midden (Z-as) van de opstelling. Hier bedraagt de werkelijke X-waarde 0. De procentuele resultaten (afwijkingen) worden weergegeven ten opzichte van deze waarde. Het procentueel weergeven van een waarde ten opzichte van '0' is niet mogelijk. Dit valt in een latere fase ook te zien bij het weergeven en bespreken van de resultaten.

Zoals reeds eerder vermeld werden deze tabellen ingevuld per vaste verplaatsing in diepte van 1 meter. Bovenstaande Tabel 2, werd in totaal dus 5 keren ingevuld, met bijhorende verwerking.

Deze resultaten werden van hieruit verder gevisualiseerd in grafieken. Hiervoor werd er gekozen voor staafdiagrammen. Dit ter makkelijke vergelijking van onderlinge resultaten binnen een vaste variatie in diepte. Een eerste grafiek bevat alle procentuele afwijkingen ten opzichte van de werkelijke posities. Een tweede grafiek bevat de afwijkingen in centimeters. Dit telkens per helm, per test locatie (links, midden, rechts) en per vaste variatie in diepte (1 meter).

Het is belangrijk beide grafieken weer te geven voor het trekken van eventuele conclusies. Het enkel tonen van de grafiek met de afwijkingen getoond in percentages [%] zou een vertekend beeld kunnen geven. Dit zeker voor kleine afstanden. Zo zou het verschil in X positie slechts 1 centimeter kunnen zijn bij een werkelijke positie van X = 2 en een bepaalde positie X = 1. Procentueel zou dit een verschil van 50% geven, wat een groot verschil lijkt te zijn. In werkelijkheid bestaat dit verschil slechts uit 1 centimeter. Wat een zeer kleine afwijking is. Dit voorbeeld verduidelijkt het belang van het weergeven en vergelijken van beide resultaten. Voor het trekken van eventuele conclusies dienen beide grafieken dus ook in rekening gebracht te worden.

Deze grafieken werden apart opgesteld per variatie in diepte (1 meter). Dit gezien een gecombineerde grafiek een vertekend beeld zou kunnen geven, gezien de verschillen in posities en verplaatsingen. Onderlinge vergelijkingen tussen verschillende situaties binnen verschillende dieptes, kunnen een foutief beeld geven en leiden tot foutieve conclusies. Gegevens omtrent de onderste helm in de middelste positie zouden wel voor verschillende posities vergeleken kunnen worden. Dit gezien het vaste verschil in positie. Dit geldt ook voor enkele andere situaties met gelijke verplaatsingen. Ook onderlinge afwijkingen tussen verschillende coördinaten binnen verschillende posities kunnen ook niet vergeleken worden.

Zeker als het gaat om de procentuele afwijkingen. Ook hier werd er namelijk gebruik gemaakt van verschillende posities. De procentuele afwijkingen werden bepaald ten opzichte van andere werkelijke afstand. Onderlinge vergelijkingen zouden dus ook een foutief beeld kunnen geven. Gezien het voornaamste doel van verificatie van de werking van het systeem, vormt de moeilijke onderlinge vergelijking geen probleem. Op basis van de uitgevoerde testen en visualisaties kunnen de nodige conclusies getrokken worden omtrent de algemene werking en nauwkeurigheid.

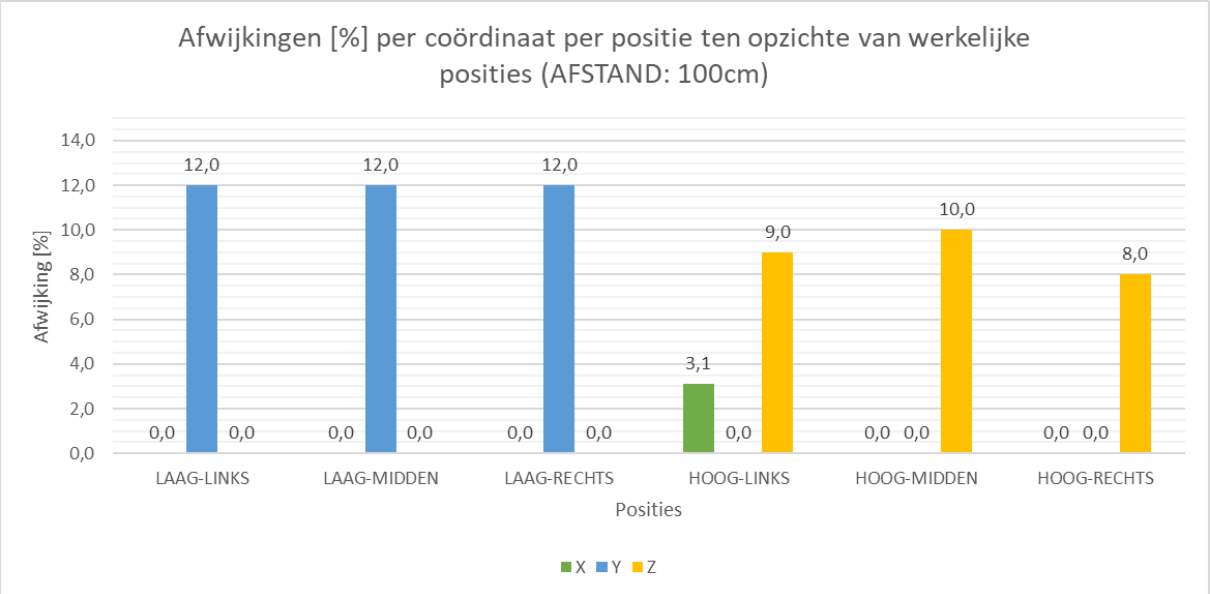
4.3 Resultaten

In onderstaande delen worden de resultaten, volgens eerder besproken tabellen en grafieken, weergegeven en besproken. Dit per elke vaste variatie in diepte (1 meter, 2 meter, 3 meter, 4 meter, 4.9 meter). In het hierop volgende deel worden de resultaten ook nog algemeen samen besproken.

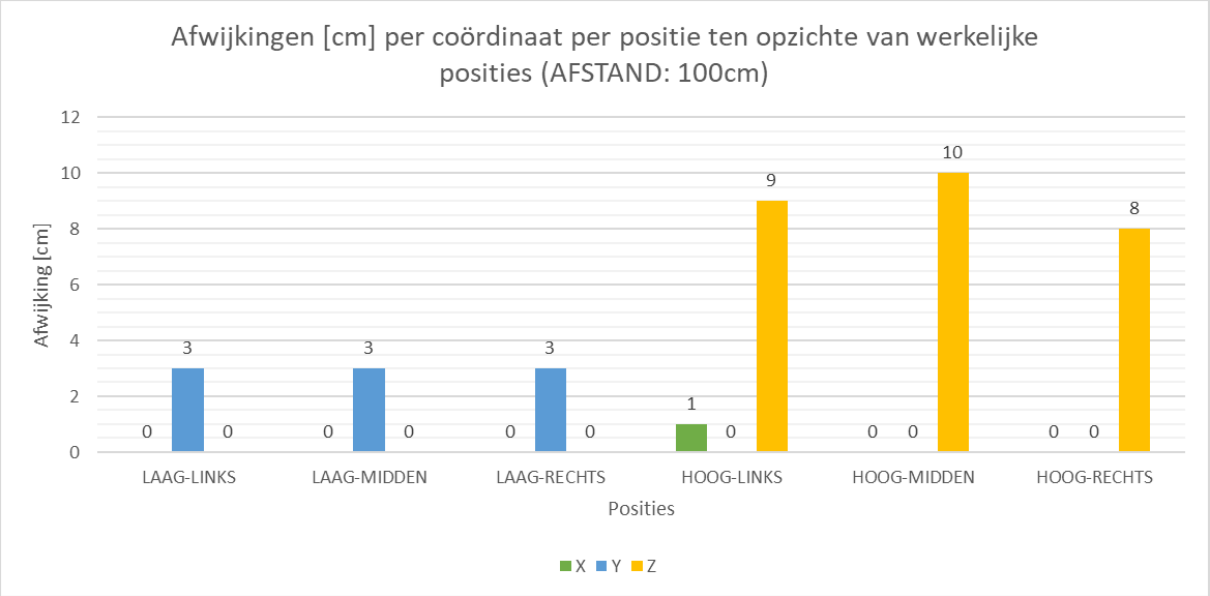
4.3.1 Resultaten diepte (Z): 1 meter

Tabel 3 Resultaten (diepte (Z) 1 meter)

AFSTAND			LAAG-LINKS				LAAG-MIDDEN				LAAG-RECHTS		
100			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	32	25	100	WERK.	0	25	100	WERK.	-32	25	100
		METING	32	22	100	METING	0	22	100	METING	-32	22	100
					100				100				100
		AFW [cm]	0	3	0	AFW [cm]	0	3	0	AFW [cm]	0	3	0
		AFW [%]	0,0	12,0	0,0	AFW [%]	##########	12,0	0,0	AFW [%]	0,0	12,0	0,0
			HOOG-LINKS				HOOG-MIDDEN				HOOG-RECHTS		
			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	32	-34	100	WERK.	0	-34	100	WERK.	-32	-34	100
		METING	33	-34	109	METING	0	-34	110	METING	-32	-34	108
					109				110				108
		AFW [cm]	1	0	9	AFW [cm]	0	0	10	AFW [cm]	0	0	8
		AFW [%]	3,1	0,0	9,0	AFW [%]	##########	0,0	10,0	AFW [%]	0,0	0,0	8,0



Figuur 81 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 100cm)



Figuur 82 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 100cm)

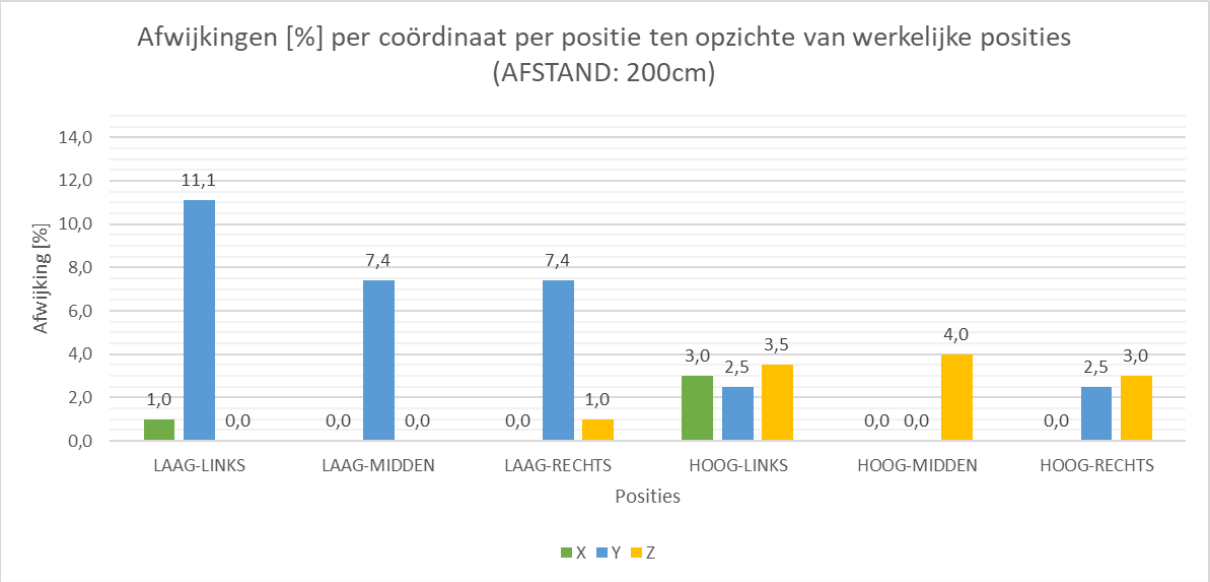
Bovenstaande Tabel 3, Figuur 81 en Figuur 82 geven de resultaten van de testen, uitgevoerd bij een diepte (Z) van 1 meter, weer. Bij het evalueren van de procentuele afwijkingen zijn er relatief hoge waardes zichtbaar. Een overeenkomstige procentuele afwijking van de Y-positie van de onderste helm van 12%. Voor de bovenste helm is er een procentuele afwijking in de Z-positie van gemiddeld 9% zichtbaar. Het vergelijken van deze procentuele afwijkingen met de afwijkingen in centimeter, verduidelijkt dat het voor de procentuele afwijking van 12% slecht gaat om een verschil van 3 centimeter. Dit verschil zou te wijten kunnen zijn aan de kijkhoek van de camera's op de 3D-helm of een gevolg van een eventuele onnauwkeurigheid binnen de positionering van de stereovisie cameraopstelling. Voor de procentuele afwijking van gemiddelde 9% voor de Z-positie van de bovenste helm, gaat het om een verschil in centimeter van ongeveer 9 cm. Ook dit verschil zou te maken kunnen hebben met de kijkhoek van de camera of een gevolg kunnen zijn van onnauwkeurigheden binnen de camera. Het kleine verschil van slechts 1 centimeter in horizontale positie van de bovenste helm bij een verplaatsing naar links is hoogstwaarschijnlijk te wijten aan een onnauwkeurigheid binnen de positionering of zou ook eventueel te maken kunnen hebben met de kijkhoek van de camera's.

Verdere evaluatie van meerdere resultaten zal een beter zicht bieden op de hier besproken afwijkingen. Deze zouden zich namelijk kunnen voortzetten binnen andere verplaatsingen. Binnen het verschil in diepte van slechts 1 meter, ten opzichte van de stereovisie cameraopstelling, waren er, zoals reeds eerder besproken, slechts beperkte verplaatsingen mogelijk.

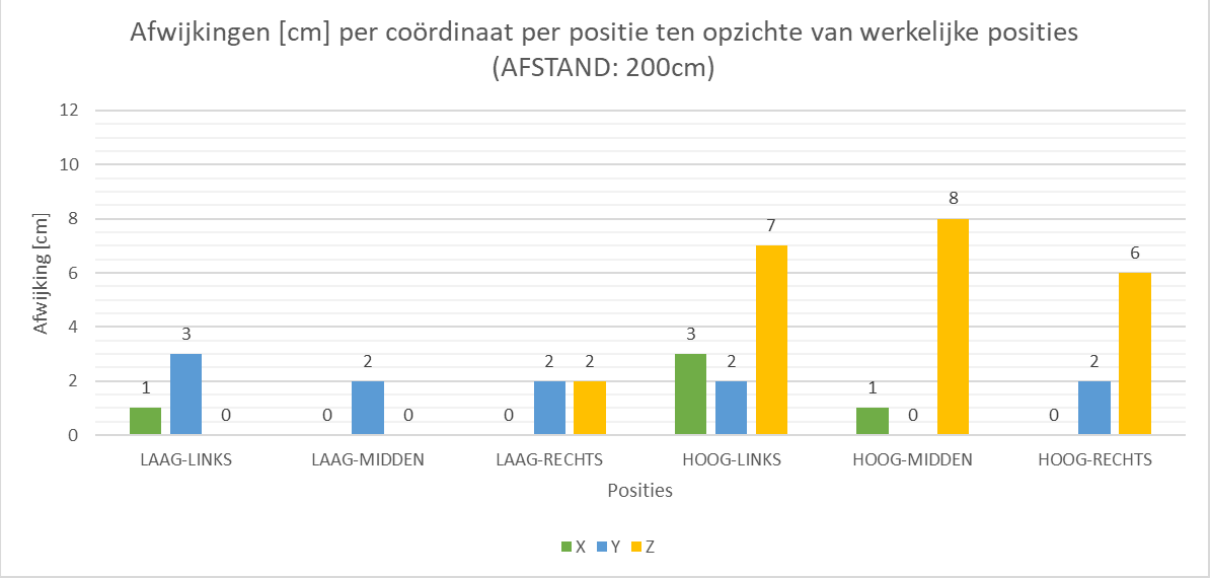
4.3.2 Resultaten diepte (Z): 2 meter

Tabel 4 Resultaten (diepte (Z) 2 meter)

AFSTAND			LAAG-LINKS				LAAG-MIDDEN				LAAG-RECHTS		
200			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	100	27	200	WERK.	0	27	200	WERK.	-100	27	200
		METING	101	24	200	METING	0	25	200	METING	-100	25	198
					200				200				198
		AFW [cm]	1	3	0	AFW [cm]	0	2	0	AFW [cm]	0	2	2
		AFW [%]	1,0	11,1	0,0	AFW [%]	##########	7,4	0,0	AFW [%]	0,0	7,4	1,0
			HOOG-LINKS				HOOG-MIDDEN				HOOG-RECHTS		
			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	100	-80	200	WERK.	0	-80	200	WERK.	-100	-80	200
		METING	103	-78	207	METING	-1	-80	208	METING	-100	-78	206
					207				208				206
		AFW [cm]	3	2	7	AFW [cm]	1	0	8	AFW [cm]	0	2	6
		AFW [%]	3,0	2,5	3,5	AFW [%]	##########	0,0	4,0	AFW [%]	0,0	2,5	3,0



Figuur 83 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 200cm)



Figuur 84 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 200cm)

Bovenstaande Tabel 4, Figuur 83 en Figuur 84 geven de resultaten van de testen, uitgevoerd bij een diepte (Z) van 2 meter, weer. Hierbij zetten de reeds eerder besproken afwijking in Y-positie van de onderste helm en in Z-positie van de bovenste helm zich voort. Wat procentuele afwijkingen en afwijkingen in centimeter betreft valt het wel op dat het om kleinere afwijkingen gaat. Het is hierbij wel belangrijk om te vermelden dat het om andere positioneringen gaat. Zo werd de bovenste helm meer naar boven geplaatst en werd de gehele testopstelling van de helmen ook verder in horizontale richting naar links en rechts verplaatst. De precieze afstanden van positioneringen zijn terug te vinden in de tabellen in combinatie met de afwijkingen. De precieze waarden van de afwijkingen in procenten en in centimeters zijn ook gevisualiseerd en weergegeven op beide grafieken.

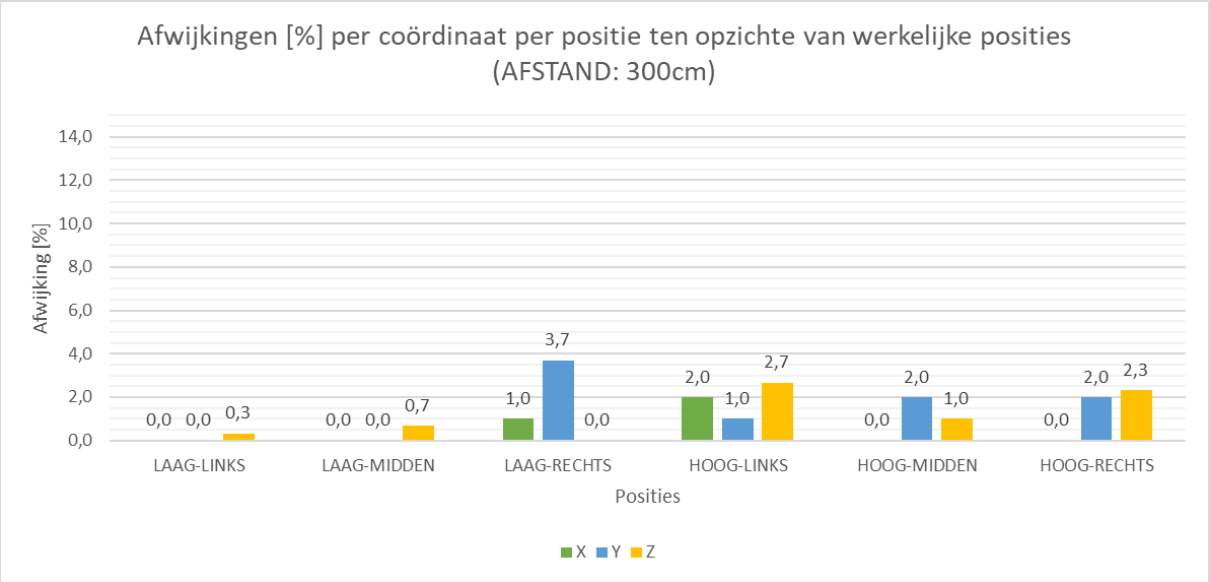
Naast de 2 eerder besproken afwijkingen zijn er over het algemeen ook enkele andere afwijkingen merkbaar. Het gaat hier om verschillen van maximaal 3 cm. Deze zouden te wijten kunnen zijn aan onnauwkeurigheden binnen de positionering van de stereovisie cameraopstelling en/of testopstelling (helmen), maar ook door eventuele fouten binnen het controleren en verifiëren van de werkelijke afstanden. Afwijkingen ten gevolge van het 3D-positiebepalingssysteem op basis van stereovisie zouden natuurlijk ook altijd mogelijk kunnen zijn. Exacte oorzaken vallen op dit moment moeilijk te controleren en/of concluderen.

Verdere evaluatie van meerdere resultaten zal een beter zicht bieden op de hier besproken afwijkingen. Deze zouden zich namelijk kunnen voortzetten binnen andere verplaatsingen.

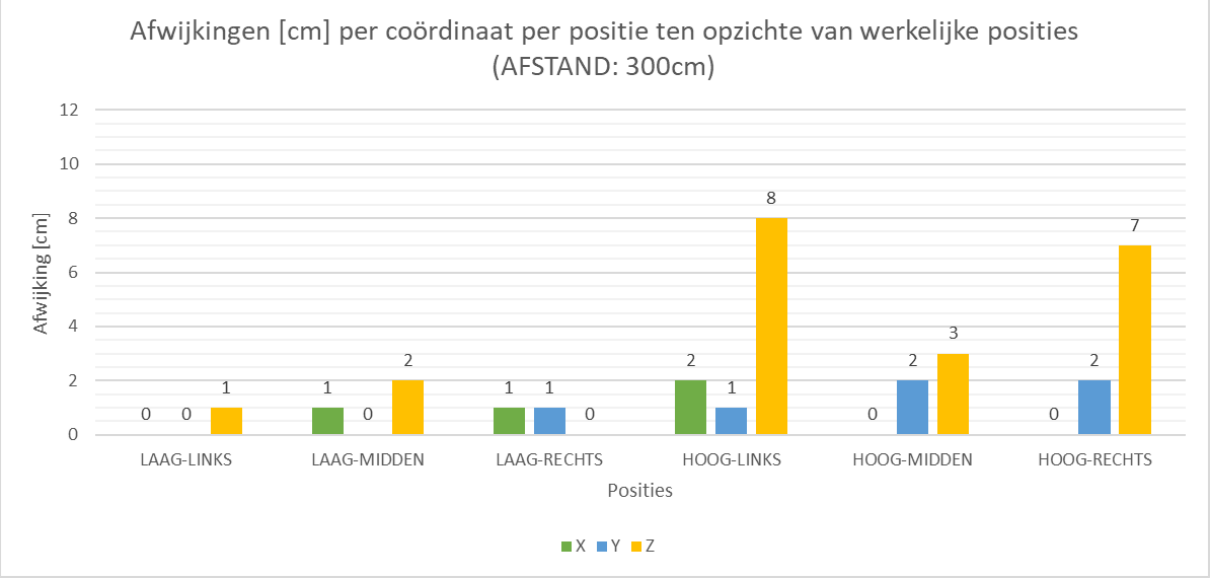
4.3.3 Resultaten diepte (Z): 3 meter

Tabel 5 Resultaten (diepte (Z) 3 meter)

AFSTAND			LAAG-LINKS				LAAG-MIDDEN				LAAG-RECHTS		
300			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	100	27	300	WERK.	0	27	300	WERK.	-100	27	300
		METING	100	27	299	METING	1	27	302	METING	-101	28	300
					299				302				300
		AFW [cm]	0	0	1	AFW [cm]	1	0	2	AFW [cm]	1	1	0
		AFW [%]	0,0	0,0	0,3	AFW [%]	##########	0,0	0,7	AFW [%]	1,0	3,7	0,0
			HOOG-LINKS				HOOG-MIDDEN				HOOG-RECHTS		
			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	100	-100	300	WERK.	0	-100	300	WERK.	-100	-100	300
		METING	98	-99	308	METING	0	-98	303	METING	-100	-98	307
					308				303				307
		AFW [cm]	2	1	8	AFW [cm]	0	2	3	AFW [cm]	0	2	7
		AFW [%]	2,0	1,0	2,7	AFW [%]	##########	2,0	1,0	AFW [%]	0,0	2,0	2,3



Figuur 85 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 300cm)



Figuur 86 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 300cm)

Bovenstaande Tabel 5, Figuur 85 en Figuur 86 geven de resultaten van de testen, uitgevoerd bij een diepte (Z) van 3 meter, weer. Uit deze resultaten valt af te lijnden dat de procentuele afwijkingen over het algemeen vrij laag zijn. Er is hier een maximale afwijking van 3,7% zichtbaar voor de Y-positie van de onderste helm bij de verplaatsing van 1 meter naar rechts. Deze afwijking bedraagt in centimeters slechts 1 cm.

Bij het observeren van de andere resultaten in centimeters valt er over het algemeen te zien dat de afwijkingen vrij laag zijn. Gemiddeld zo een 3 cm. Deze zouden te wijten kunnen zijn aan onnauwkeurigheden binnen de positionering van de stereovisie cameraopstelling en/of testopstelling (helmen), maar ook door eventuele fouten binnen het controleren en verifiëren van de werkelijke afstanden. Afwijkingen ten gevolge van het 3D-positiebepalingssysteem op basis van stereovisie zouden natuurlijk ook altijd mogelijk kunnen zijn. Er zijn bij deze resultaten wel 2 extreme uitschieters merkbaar.

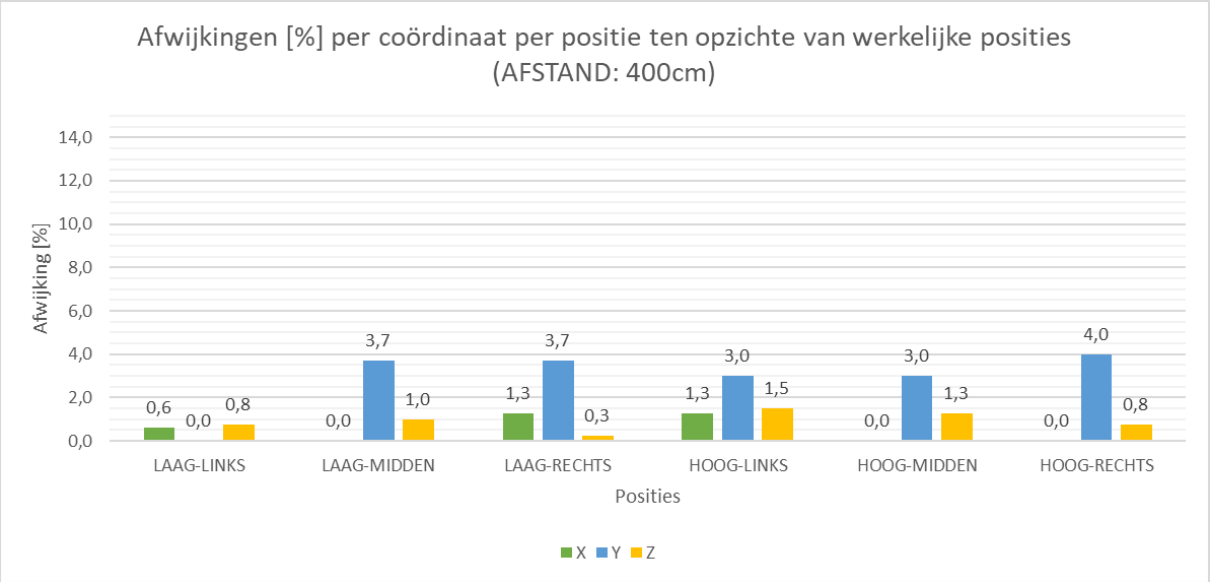
Wat extreme uitschieters binnen de resultaten omtrent afwijkingen in centimeter betreft, gaat het over de Z-posities van de bovenste helm bij de verplaatsingen naar links en rechts, maar ook lichtjes (3cm) in het midden. Deze trend was ook al eerder zichtbaar. Ook al ging het hier om andere posities. De afwijking voor de middelste positie is wel sterk gedaald tot slechts 3 cm.

De reeds eerder besproken afwijking van de onderste helm in Y-positie is op deze afstand voor elke positie zo goed als niet meer merkbaar. Slechts een kleine afwijking van 1 cm bij de verplaatsing naar rechts over 1 meter is merkbaar. Deze afwijking is hier in principe verwaarloosbaar.

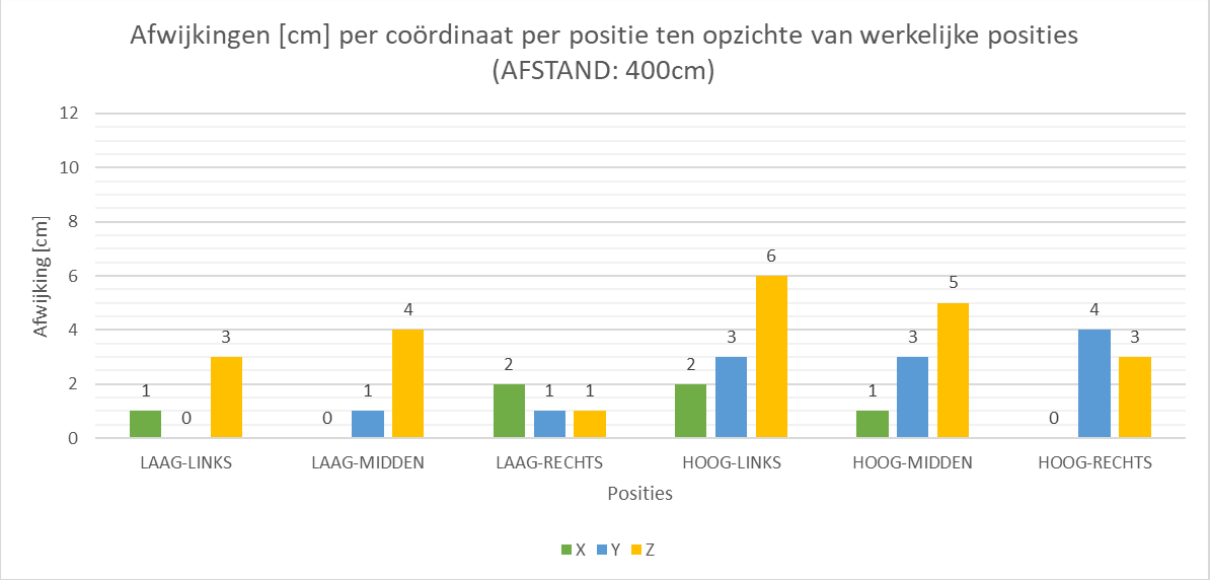
4.3.4 Resultaten diepte (Z): 4 meter

Tabel 6 Resultaten (diepte (Z) 4 meter)

AFSTAND			LAAG-LINKS				LAAG-MIDDEN				LAAG-RECHTS		
400			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	160	27	400	WERK.	0	27	400	WERK.	-160	27	400
		METING	161	27	403	METING	0	28	404	METING	-158	28	401
					403				404				401
		AFW [cm]	1	0	3	AFW [cm]	0	1	4	AFW [cm]	2	1	1
		AFW [%]	0,6	0,0	0,8	AFW [%]	##########	3,7	1,0	AFW [%]	1,3	3,7	0,3
			HOOG-LINKS				HOOG-MIDDEN				HOOG-RECHTS		
			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	160	-100	400	WERK.	0	-100	400	WERK.	-160	-100	400
		METING	158	-97	406	METING	-1	-97	405	METING	-160	-96	403
					406				405				403
		AFW [cm]	2	3	6	AFW [cm]	1	3	5	AFW [cm]	0	4	3
		AFW [%]	1,3	3,0	1,5	AFW [%]	##########	3,0	1,3	AFW [%]	0,0	4,0	0,8



Figuur 87 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 400cm)



Figuur 88 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 400cm)

Bovenstaande Tabel 6, Figuur 87 en Figuur 88 geven de resultaten van de testen, uitgevoerd bij een diepte (Z) van 4 meter, weer. Uit deze resultaten valt af te lijden dat de procentuele resultaten overal vrij laag zijn. Er is hier een maximale afwijking van 4% zichtbaar voor de Y-positie van de bovenste helm bij de verplaatsing van 1,6 meter naar rechts. Deze afwijking bedraagt in centimeters slechts 4 cm.

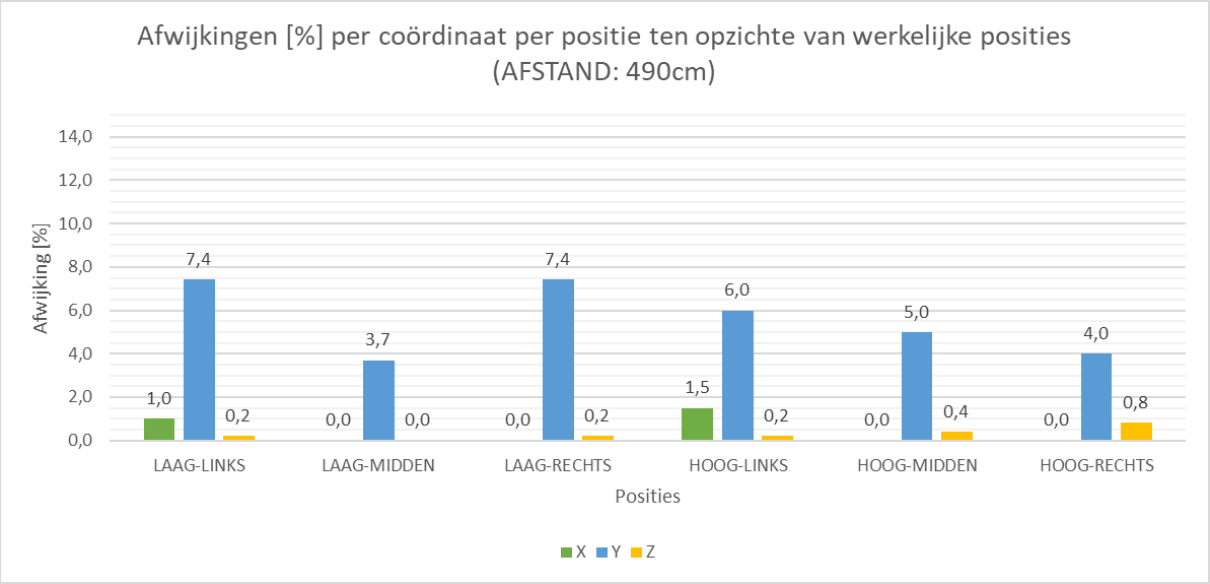
Bij het observeren van de afwijkingen in centimeters, zijn er over het algemeen overal wel wat lichte afwijkingen merkbaar. Gemiddelde gaat het hier over zo een 3 centimeter. De afwijkingen van de Y-posities van de bovenste helm zijn ten opzichte van de andere afwijkingen aan de hoge kant. Deze zouden te wijten kunnen zijn aan onnauwkeurigheden binnen de positionering van de stereovisie cameraopstelling en/of testopstelling (helmen), maar ook door eventuele fouten binnen het controleren en verifiëren van de werkelijke afstanden. Afwijkingen ten gevolge van het 3D-positiebepalingssysteem op basis van stereovisie zouden natuurlijk ook altijd mogelijk kunnen zijn.

Over het algemeen valt het op dat de afwijkingen in centimeter voor de Z-posities vrij hoog zijn. Deze variëren van slechts 1 centimeter, voor de laagste helm bij een horizontale verplaatsing over 1,6 meter naar rechts, tot een afwijking van 6 centimeter, voor de bovenste helm bij een horizontale verplaatsing van 1,6 meter naar links. De waarden, wat afwijking in Z-positie voor de bovenste helm betreft, waren bij de eerdere diepteposities ook altijd al aan de hoge kant. Het gaat hier telkens wel over andere posities en verplaatsingen. Bij deze diepte van 4 meter, zijn er ook lichte verhogingen tot 4 centimeter merkbaar bij de laagste helm. Deze afwijkingen zouden eventueel te maken kunnen hebben met een incorrecte werkelijke diepte afstand. Maar onnauwkeurigheden wat positionering van de stereovisie cameraopstelling betreft en eventuele foute 3D-positiebepalingen zijn natuurlijk ook nooit uitsluitbaar.

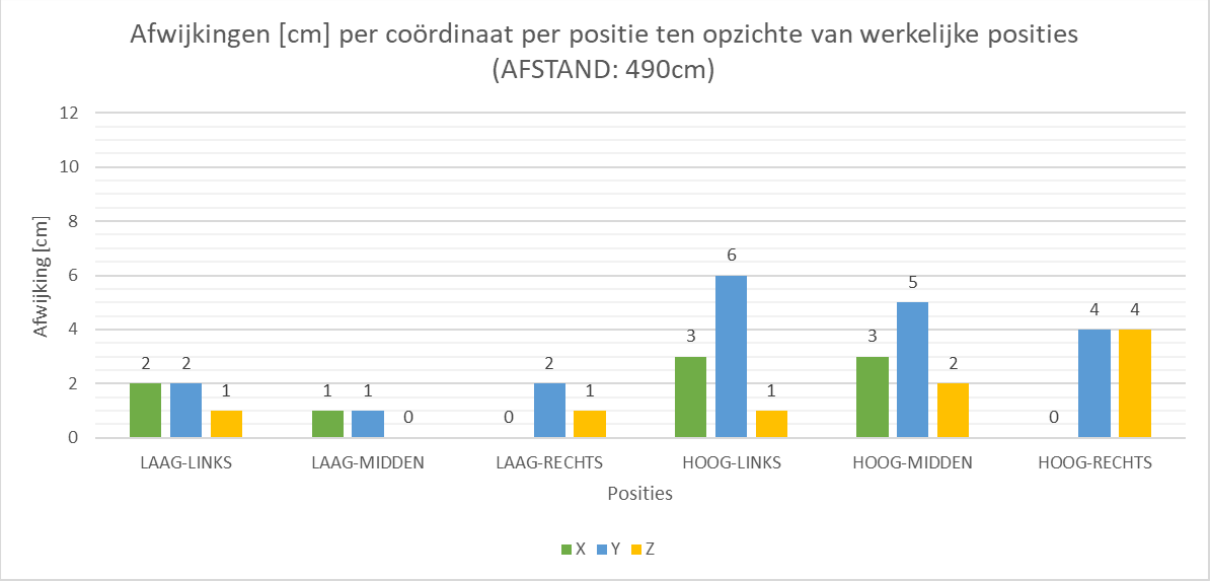
4.3.5 Resultaten diepte (Z): 4,9 meter

Tabel 7 Resultaten (diepte (Z) 4,9 meter)

AFSTAND			LAAG-LINKS				LAAG-MIDDEN				LAAG-RECHTS		
490			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	200	27	490	WERK.	0	27	490	WERK.	-200	27	490
		METING	198	29	491	METING	1	28	490	METING	-200	29	489
					492				490				
		AFW [cm]	2	2	1	AFW [cm]	1	1	0	AFW [cm]	0	2	1
		AFW [%]	1,0	7,4	0,2	AFW [%]	##########	3,7	0,0	AFW [%]	0,0	7,4	0,2
			HOOG-LINKS				HOOG-MIDDEN				HOOG-RECHTS		
			X	Y	Z		X	Y	Z		X	Y	Z
		WERK.	200	-100	490	WERK.	0	-100	490	WERK.	-200	-100	490
		METING	197	-94	489	METING	-3	-95	492	METING	-200	-96	494
					489				492				494
		AFW [cm]	3	6	1	AFW [cm]	3	5	2	AFW [cm]	0	4	4
		AFW [%]	1,5	6,0	0,2	AFW [%]	##########	5,0	0,4	AFW [%]	0,0	4,0	0,8



Figuur 89 Grafische voorstelling afwijkingen [%] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 490cm)



Figuur 90 Grafische voorstelling afwijkingen [cm] per coördinaat per positie ten opzichte van werkelijke posities (AFSTAND: 490cm)

Bovenstaande Tabel 7, Figuur 89 en Figuur 90 geven de resultaten van de testen, uitgevoerd bij een diepte (Z) van 4,9 meter, weer. Zoals reeds eerder vermeld was een afstand van 5 meter niet haalbaar, gezien limitaties van de gebruikte testruimte.

Bij het observeren van de procentuele afwijkingen valt er over het algemeen te stellen dat deze overal vrij laag (rond 1%) liggen. Behalve voor alle Y-waarden. Hier variëren de afwijkingen van 3,7%, bij de onderste helm in het midden (X = 0), tot 7,4%, bij de lage helmen bij een horizontale verplaatsing van 2 meter naar links en 2 meter naar rechts.

Bij het observeren van de afwijkingen in centimeters, valt er te zien dat de lage helm over het algemeen op elke positie zeer nauwkeurige resultaten geeft. De maximale afwijkingen hier bedragen slechts 2 centimeter. Bij de hoge helm zijn er vooral afwijkingen merkbaar wat Y-positie betreft, zoals de procentuele afwijkingen ook al bevestigden. Het gaat hier om afwijkingen in Y-positie van 4 tot 6 centimeter. Ook valt er te zien dat er voor de linker- en middenpositie van de helm, wat horizontale verplaatsing betreft, een afwijking in X-positie zichtbaar is. Het gaat hier beide over 3 centimeter, wat slechts een kleine afwijking is. Als laatste valt er ook nog te zien dat de hoge helm enkele kleine afwijkingen heeft, wat diepte Z betreft. Bij de linker positie (horizontale verplaatsing van 2 meter naar links ten opzichte van het midden) bedraagt deze afwijking slechts 1 centimeter. In het midden (X = 0 (werkelijk)) bedraagt de afwijking 4 centimeter. Ten slotte bedraagt de afwijking voor de rechter positie (horizontale verplaatsing van 2 meter naar rechts ten opzichte van het midden) 4 centimeter.

Deze afwijkingen zouden te wijten kunnen zijn aan onnauwkeurigheden binnen de positionering van de stereovisie cameraopstelling en/of testopstelling (helmen), maar ook door eventuele fouten binnen het controleren en verifiëren van de werkelijke afstanden. Afwijkingen ten gevolge van het 3D-positiebepalingssysteem op basis van stereovisie zouden natuurlijk ook altijd mogelijk kunnen zijn.

4.3.6 Algemene bespreking resultaten

In het algemeen bevestigen de resultaten een voldoende nauwkeurige werking van het 3D-positiebepalingssysteem op basis van stereovisie. Er zijn natuurlijk wel enkele afwijkingen merkbaar binnen de resultaten. De maximale afwijkingen in centimeter worden weergegeven op onderstaande Tabel 8. Er valt hier te zien dat de maximale afwijking slechts 10 cm bedraagt. Dit bij de Z-coördinaat. Deze waarde werd behaald bij de bovenste helm in het midden bij een afstand (diepte 'Z') van 1 meter ten opzichte van de stereovisie cameraopstelling. De andere maximale waardes bedragen 6 cm voor de Y-coördinaat en 3 cm voor de X-coördinaat. Beperkte en aanvaardbare afwijkingen dus. Wat de alternatieve methode betreft waren er geen verschillen in resultaten waarneembaar.

Tabel 8 Maximale afwijkingen in centimeter per coördinaat, binnen de verwerkte resultaten, ten opzichte van de werkelijk coördinaten (posititie)

Maximale Afwijkingen	Waarde (cm)
X	3
Y	6
Z	10

Over het algemeen geeft het 3D-positiebepalingssysteem op basis van stereovisie de grootste afwijkingen bij korte afstanden (1 en 2 meter ten opzichte van de stereovisie cameraopstelling). Bij verdere afstanden (3, 4 en 4,9 meter) zijn er meerdere afwijkingen merkbaar, maar allemaal van beperktere grootteorde. Zeker in vergelijking met hun werkelijke afstanden en de afstand waarop de testopstelling (helmen) zich bevond ten opzichte van de stereovisie cameraopstelling (diepte 'Z').

Deze afwijkingen zouden te wijten kunnen zijn aan onnauwkeurigheden binnen de positionering van de stereovisie cameraopstelling en/of testopstelling (helmen), maar ook door eventuele fouten bij het controleren en verifiëren van de werkelijke afstanden. Afwijkingen ten gevolge van het 3D-positiebepalingssysteem op basis van stereovisie zouden natuurlijk ook altijd mogelijk kunnen zijn.

Aperam wenst voor het gebruik van het systeem, ter gebruik binnen de digitalisering van het veiligheidskegel veiligheidshulpsysteem, een maximale afwijking van 50 cm ten opzichte van werkelijke posities te bekomen. Op basis van de uitgevoerde testen en bekomen resultaten werd deze eis behaald. De maximaal bekomen afwijking bedroeg slechts 10 centimeter. Dit is slechts 20% van de maximale waarde van de afwijking die behaald mocht worden. Het systeem zou volgens deze vereiste dus gebruikt kunnen worden binnen de digitalisering van het veiligheidshulpmiddel.

4.4 Alternatieve testen

Naast de eerder besproken testen werden er ook nog enkele extra alternatieve testen uitgevoerd. Dit ter verdere evaluatie van de werking van het systeem. Deze testen, met bijhorende resultaten van het 3D-positiebepalingssysteem op basis van stereovisie, kunnen jammer genoeg niet gebruikt worden ter officiële verificatie van het systeem. Dit gezien de werkelijke 3D-posities niet bepaald konden worden en er dus geen geverifieerd vergelijkingsmateriaal is ter bevestiging van de bepaalde 3D-posities aan de hand van het ontwikkelde systeem. De resultaten van deze testen werden echter wel op manuele en subjectieve wijze geëvalueerd. Vandaar worden deze toch nog vermeld binnen deze masterthesis.

4.4.1 Video rechte lijn

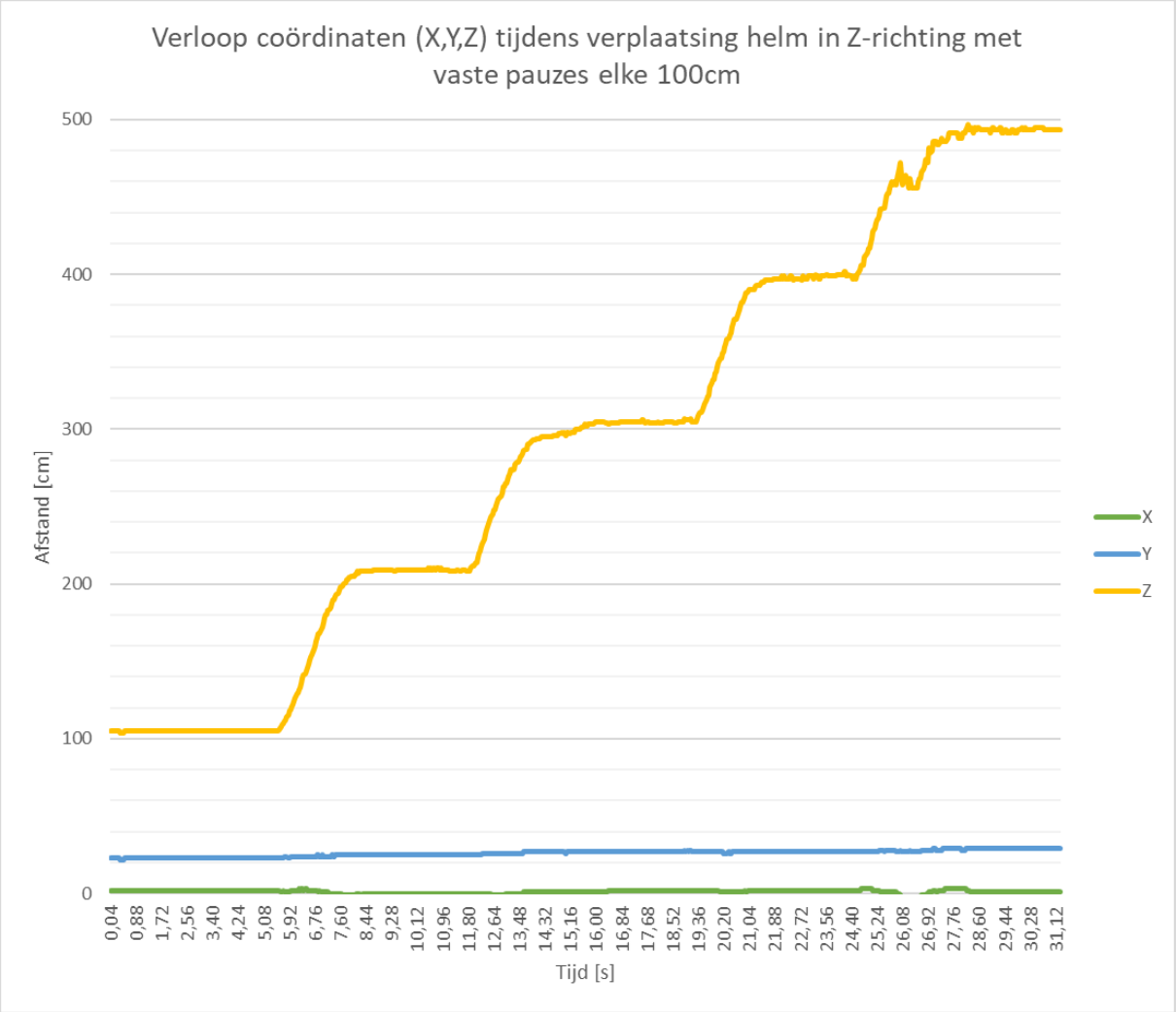
Binnen de eerste alternatieve test werd er gebruik gemaakt van een (synchrone) stereo video-opname. Deze video-opname werd als input van het 3D-positiebepalingssysteem op basis van stereovisie gebruikt.

De opname zelf werd gemaakt aan de hand van het stereovisie camerasysteem. Er werd hiervoor een apart stukje code ontwikkeld. Binnen deze code werd er aan de hand van de vervormde, gerectificeerde en getransformeerde beelden een opname van beide camera streams gemaakt en weggeschreven naar 2 videobestanden. 1 met het linker beeld en 1 met het rechter beeld.

Op de stereo video wordt er gebruik gemaakt van 1 veiligheidshelm, namelijk de onderste helm van de testopstelling. De bovenste werd verwijderd. Bij het begin van de video bevindt de testopstelling met de helm zich op een afstand van 1 meter van de stereovisie cameraopstelling. Vervolgens werd de gehele testopstelling over 1 rechte lijn (X=0) achteruit verplaatst. Dit met als doel enkel een afwijking in diepte ‘Z’ te hebben. Het verplaatsen gebeurde zo nauwkeurig mogelijk, maar kon natuurlijk niet perfect in 1 rechte lijn gebeuren. Bij elke vaste verplaatsing van 1 meter in diepte, gebaseerd op de eerder aangebrachte markeringen op de grond met tape, werd er gepauzeerd, om na enkele seconden weer verder te bewegen naar een volgende positie, 1 meter verder. Dit tot de maximale afstand van 4,9 meter in diepte ‘Z’ bereikt werd.

De resultaten van het 3D-positiebepalingssysteem, met de stereo video als input, werden weggeschreven naar een Excel-bestand en van hieruit verder verwerkt en gevisualiseerd. Voor de visualisatie van de resultaten werd er gekozen voor een lijndiagram. Dit gezien het verloop van elke coördinaat (X, Y, Z) in functie van de tijd hierop duidelijk weergegeven kan worden. De bekomen grafiek met resultaten is te zien op Figuur 91.

Hierop is duidelijk het verloop van elke coördinaat in tijd zichtbaar. Deze komen overeen met de werkelijke posities en verplaatsingen binnen de video, maar kunnen zoals reeds eerder vermeld niet geverifieerd worden. Wat de Y-coördinaat betreft is er een stijging merkbaar. Deze stijging zou te wijten kunnen zijn aan onnauwkeurigheden binnen de camera positionering. In dit geval zouden de camera’s niet waterpas staan, maar onder een lichte hoek naar onder staan gericht.



Figuur 91 Verloop coördinaten (X, Y, Z) tijdens verplaatsing helm in Z-richting met vaste pauzes elke 100cm

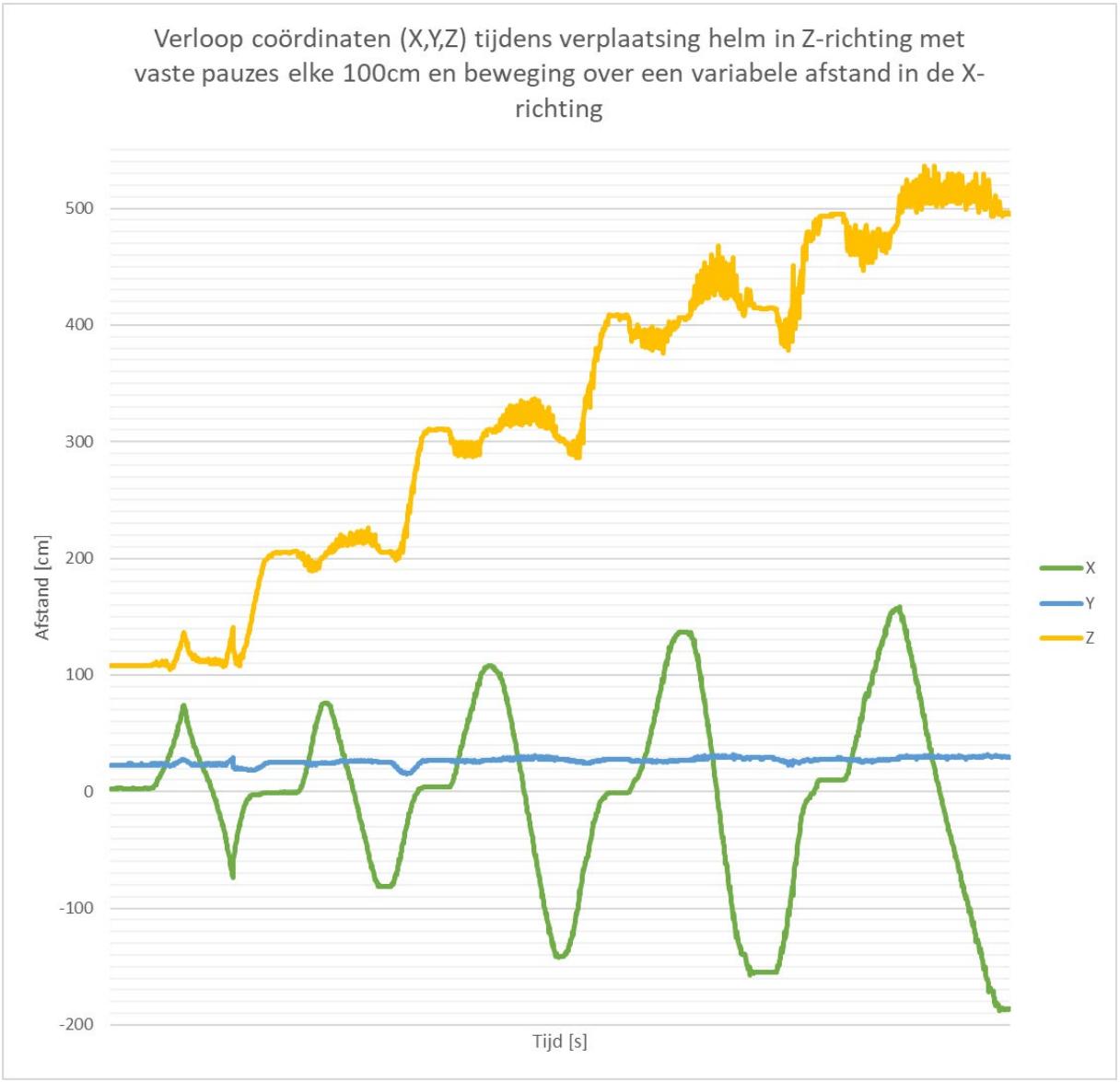
4.4.2 Video beweging

Binnen een 2de alternatieve test werd er ook gebruik gemaakt van een stereo video-opname. Ook deze werd op dezelfde manier als de vorige, zoals besproken binnen deel 4.4.1, opgenomen. In plaats van enkel te bewegen op een rechte lijn (X=0) in diepte (Z) met vaste pauzes elke meter, werd er nu gekozen voor meer random verplaatsingen. Naast een verplaatsing in de diepte (Z-positie) over telkens een vaste afstand met pauzes elke meter, werd de testopstelling met 1 helm nu ook telkens nog eens in horizontale richting bewogen. Een extra variatie in X-positie dus. Ook deze verplaatsingen werden zo nauwkeurig mogelijk uitgevoerd op basis van de aangebrachte markeringen en vierkant patroon van de tegels op de grond. Natuurlijk kon deze verplaatsing niet zeer nauwkeurig uitgevoerd worden. De testopstelling bewoog soms vrij stroef of bewoog redelijk veel.

Ook deze resultaten werden weggeschreven naar een Excel-bestand en van hieruit verder verwerkt en gevisualiseerd. Ook hier in de vorm van een lijndiagram in functie van de tijd. Het verloop van iedere coördinaat (X, Y, Z) binnen deze bewegingen is te zien op Figuur 92.

Het verloop van de coördinaten, gevisualiseerd op het lijndiagram, komt overeen met de bewegingen van de helm op de video. Het verloop van de Z-coördinaat ziet er vrij onrustig uit, wat ook het geval was binnen de video. Dit gezien de stroefheid en snelheid van de verplaatsing.

Ondanks dat de bekomen resultaten overeenkomen met wat er op de videbeelden gebeurt, kan ook dit resultaat niet gebruikt worden ter officiële verificatie van het 3D-positiebepalingssysteem op basis van stereovisie. Dit gezien er geen vergelijking gemaakt kan worden met de werkelijke posities met bijhorende 3D-coördinaten.



Figuur 92 Verloop coördinaten (X, Y, Z) tijdens verplaatsing helm in Z-richting met vaste pauzes elke 100 cm en bewegingen over een variabele afstand in de X-richting

4.4.3 Extreme posities

Binnen de laatste alternatieve test werd er gebruik gemaakt van de live toepassing van het 3D-posititbepalingssysteem op basis van stereovisie. Er werden hier enkele extreme posities uitgetest. Het gaat hier van de beeldgrenzen van de zone van overlap tussen beide beelden, tot uiterste limieten van de gebruikte testruimte. Met uiterste limieten wordt er verwezen naar de uiterste afstanden en verste posities ten opzichte van de stereovisie cameraopstelling.

Eén veiligheidshelm werd telkens manueel door een persoon op zo een positie vastgehouden. De resultaten van het 3D-positiebepalingssysteem werden vervolgens manueel en subjectief beoordeeld. Dit ter controle van de realiteit ervan. Binnen deze test ging het eerder om evalueren van de reactie van het systeem op deze 'extreme' posities (Kan de positie kloppen?,

Realistische waarden? , etc.). Resultaten werden dus ook niet bijgehouden. Dit gezien een betrouwbare verificatie van de afstanden zeer lastig was. De bekomen, subjectief beoordeelde, resultaten kwamen overeen met de werkelijke posities. Alle bekomen resultaten gaven een realistische weergave van de positie. Voor het beoordelen werd er voornamelijk gebruik gemaakt van gekende afstanden uit eerdere testen. Zo ook de markeringen op de grond.

Bovenstaande test kan uiteraard niet gebruikt worden ter verificatie van de werking van het 3D-positiebepalingsssyteem op basis van stereovisie. Deze werd eerder gebruikt ter manuele beoordeling van het systeem binnen extreme situaties. Dit voornamelijk ook ter evaluatie van de ontwikkeling.

5 Conclusie

Binnen deze masterthesis werd er een 3D-positiebepalingssysteem op basis van stereovisie ontwikkeld. Dit met evaluatie en verificatie van de werking ervan aan de hand van veiligheidshelmdetectie. Deze masterthesis kan opgesplitst worden in een aantal fasen. Allereerst een onderzoeksfase, die wordt vervolgd door de effectieve ontwikkeling van het 3D-positiebepalingssysteem op basis van stereovisie, met als laatst de evaluatie en verificatie van de werking van het 3D-positiebepalingssysteem op basis van enkele testen.

Binnen de onderzoeksfase van deze masterthesis werd er op basis van literatuuronderzoek uitgevoerd om de theorie rondom 3D-positiebepaling op basis van stereovisie te begrijpen en er inzichten in te verkrijgen. Hierin werd er ook onderzoek gedaan naar bepaalde keuzes die gemaakt dienden worden binnen de ontwikkeling van het systeem. De uitwerking hiervan valt te zien binnen hoofdstuk 2, de literatuurstudie.

Na de onderzoeksfase volgde de ontwikkelingsfase. Binnen de fase werd het effectieve 3D-positiebepalingssysteem op basis van stereovisie ontwikkeld. Er dienden heel wat keuzes en afwegingen gemaakt worden, die allemaal binnen hoofdstuk 3, de ontwikkeling, van deze scriptie tot in detail besproken en uitgelegd worden. Ook de werking van het systeem, met bijhorende coördinatenstelsels en tussenstappen, wordt binnen deze fase zorgvuldig besproken.

Binnen een laatste fase van de masterthesis werd het ontwikkelde 3D-positiebepalingssysteem op basis van stereovisie, onderworpen aan enkele testen. Dit ter evaluatie en verificatie van de werking ervan. Voor het uitvoeren van deze testen werd er gebruik gemaakt van het ontwikkelde YOLOv5s-helmdetectiemodel, dat ook binnen de ontwikkeling besproken wordt. Resultaten van deze testen tonen een nauwkeurige werking van het systeem aan. De maximale afwijkingen van elke coördinaat, bepaald door het 3D-positiebepalingssysteem ten opzichte van de werkelijke coördinaten van enkele 3D-posities, bedragen respectievelijk 3cm, 6cm en 10cm. Dit wat X-, Y-, en Z-positie betreft. Deze afwijkingen zouden te wijten kunnen zijn aan onnauwkeurigheden binnen de positionering van de stereovisie cameraopstelling en/of testopstelling (helmen), maar ook door eventuele fouten binnen het controleren en verifiëren van de werkelijke 3D-posities. Afwijkingen ten gevolge van het 3D-positiebepalingssysteem op basis van stereovisie zouden natuurlijk ook altijd mogelijk kunnen zijn. Andere testen, waarin het verloop van de coördinaten van de 3D-postie van een helm in beweging, in functie van de tijd beoordeeld werd, geven ook goede resultaten weer. Alle testen, inclusief de opstelling en onzekerheden, worden in meer detail besproken en uitgelegd binnen hoofdstuk 4, ‘Verificatie ontwikkeld 3D-positiebepalingssysteem’, van deze masterthesis.

Over het algemeen kan er geconcludeerd worden dat het mogelijk is om 3D-posities van één of meerdere objecten te bepalen aan de hand van 2 camera’s. Het gebruik van 2 camera’s, met een verschil in perspectief, stelt ons in staat aan de hand van verschillende technieken dieptes en afstanden binnen de 3D-ruimte te bepalen. Dit meer specifiek in het gebied van overlap tussen beide vervormde, gerectificeerde en getransformeerde camerabeelden.

Gezien de bekomen resultaten en behaalde vereisten van Aperam voor het 3D-positiebepalingssysteem, kan het ontwikkelde 3D-positiebepalingssysteem op basis van stereovisie binnen deze masterproef, gebruikt worden binnen de verdere digitalisering van het veiligheidshulpmiddel voor transport aan de hand van een rolbrug.

Toekomstperspectief

Naar de toekomst toe zouden er verdere stappen ondernomen kunnen worden tot verdere evaluatie, verificatie en optimalisatie van de werking van het 3D-positiebepalingssysteem op basis van stereovisie. Enkele mogelijke stappen worden aangehaald binnen onderstaande opsomming. Bij de toepassing van het systeem binnen een specifieke applicatie zouden er optimalisaties van het 3D-positiebepalingssysteem in functie van de applicatie toegepast kunnen worden.

- Evaluatie van gemaakte keuzes binnen de ontwikkeling van het systeem.
- Vergelijking en evaluatie van invloeden van verschillende keuzes binnen de ontwikkeling van het systeem.
- Verder evaluatie en verificatie van de werking van het ontwikkelde systeem op basis van meerdere en andere testen.
- Algemene verdere optimalisatie van de werking van het systeem.

6 Bibliografie

[1] "Aperam Genk," 2019. https://www.aperamgenk.be/ (accessed May 15, 2022).

[2] A. Zaarane *et al.*, "Distance measurement system for autonomous vehicles using stereo camera," *Array*, vol. 5, no. December 2019, p. 100016, Mar. 2020, doi: 10.1016/j.array.2020.100016.

[3] K. Karthika, S. Adarsh, and K. I. Ramachandran, "Distance Estimation of Preceding Vehicle Based on Mono Vision Camera and Artificial Neural Networks," *2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020*, Jul. 2020, doi: 10.1109/ICCCNT49239.2020.9225406.

[4] Z. Xu, Z. Zhou, and L. Wang, "Research on monocular vision distance measurement algorithm based on reference target," *2019 IEEE 4th Int. Conf. Signal Image Process. ICSIP 2019*, pp. 192–197, Jul. 2019, doi: 10.1109/SIPROCESS.2019.8868888.

[5] M. Domnguez-Morales, A. Jimnez-Fernndez, R. Paz-Vicente, A. Linares-Barranco, and G. Jimnez-Moreno, "Stereo Matching: From the Basis to Neuromorphic Engineering," *Curr. Adv. Stereo Vis.*, Jul. 2012, doi: 10.5772/45901.

[6] J. Mrovlje and D. Vran, "Distance measuring based on stereoscopic pictures," *9th Int. PhD Work. Syst. Control Young Gener. Viewp.*, vol. 2, no. October, pp. 1–6, 2008.

[7] Ashraf Anwar Fahmy, "Stereo Vision Based Depth Estimation Algorithm In Uncalibrated Rectification," *Int. J. Video&Image Process. Netw. Secur. IJVIPNS-IJENS Vol13 No02* , Apr. 2013.

[8] Y. Zhang *et al.*, "Adaptive Unimodal Cost Volume Filtering for Deep Stereo Matching," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 12926–12934, Sep. 2019, doi: 10.48550/arxiv.1909.03751.

[9] A. Beato, "Understanding Comfortable Stereography," 2011. https://www.affonsobeato.com/Affonso_Beato/Understanding_Comfortable_Stereography.html (accessed Apr. 07, 2022).

[10] M. Elgendy, *Deep Learning for Vision Systems*.

[11] Y. Dawood, K. Ruhana Ku-Mahamud, and E. Kamioka, "Distance Measurement for Self-Driving Cars Using Stereo Camera Oracle View project Ant Colony Optimization View project," no. October, 2017, [Online]. Available: https://www.researchgate.net/publication/320336266

[12] T. S. Hsu and T. C. Wang, "An improvement stereo vision images processing for object distance measurement," *Int. J. Autom. Smart Technol.*, vol. 5, no. 2, pp. 85–90, 2015, doi: 10.5875/AUSMT.V5I2.460.

[13] "Object Detection | Papers With Code." https://paperswithcode.com/task/object-detection (accessed Apr. 11, 2022).

[14] B. Dickson, "An introduction to object detection with deep learning – TechTalks," *TechTalks*, Jun. 21, 2021. https://bdtechtalks.com/2021/06/21/object-detection-deep-learning/ (accessed Apr. 11, 2022).

[15] B. Dickson, "What is the difference between supervised and unsupervised machine learning? – TechTalks," Feb. 10, 2020. https://bdtechtalks.com/2020/02/10/unsupervised-learning-vs-supervised-learning/ (accessed Apr. 11, 2022).

[16] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, Jul. 2018, doi: 10.1109/ICCUBEA.2018.8697857.

[17] "Ultimate Guide to Object Detection Using Deep Learning [2022] | upGrad blog." https://www.upgrad.com/blog/ultimate-guide-to-object-detection-using-deep-learning/ (accessed Apr. 13, 2022).

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, Nov. 2013, doi: 10.48550/arxiv.1311.2524.
[19] R. Girshick, “Fast R-CNN,” 2015. doi: 10.1109/ICCV.2015.169.
[20] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2015, doi: 10.48550/arxiv.1506.01497.
[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 779–788, Jun. 2015, doi: 10.48550/arxiv.1506.02640.
[22] “YOLOv5 Documentation.” https://docs.ultralytics.com/#yolov5 (accessed Jun. 01, 2022).
[23] M. A. Mahammed, A. I. Melhum, and F. A. Kochery, “Object Distance Measurement by Stereo VISION,” *2013 Int. J. Sci. Appl. Inf. Technol.*, vol. 2, no. 2, pp. 5–8, 2013.
[24] “Internet Protocol (IP) Cameras – How do They Work & What are the Benefits?,” *Safesite Facilities*. https://www.safesitefacilities.co.uk/knowledge-base/internet-protocal-cameras-how-do-they-work
[25] “DS-2CD2643G2-IZS | Pro Series (All) | Hikvision.” https://www.hikvision.com/en/products/IP-Products/Network-Cameras/Pro-Series-EasyIP-/ds-2cd2643g2-izs/ (accessed May 25, 2022).
[26] “CISCO CBS2508PPD: Switch, 8-Port, Gigabit Ethernet, PoE+ at reichelt elektronik.” https://www.reichelt.com/de/en/switch-8-port-gigabit-ethernet-poe--cisco-cbs2508ppd-p319638.html?r=1 (accessed May 25, 2022).
[27] “OpenCV: Camera Calibration and 3D Reconstruction.” https://docs.opencv.org/3.4/d9/d0c/group__calib3d.html#ga617b1685d4059c6040827800e72ad2b6 (accessed May 25, 2022).
[28] OpenCV, “About OpenCV.” https://opencv.org/about/
[29] “hjp: doc: RFC 2326: Real Time Streaming Protocol (RTSP).” https://www.hjp.at/doc/rfc/rfc2326.html (accessed May 28, 2022).
[30] A. Noda, S. Tabata, M. Ishikawa, and Y. Yamakawa, “Synchronized High-Speed Vision Sensor Network for Expansion of Field of View,” *Sensors 2018, Vol. 18, Page 1276*, vol. 18, no. 4, p. 1276, Apr. 2018, doi: 10.3390/S18041276.
[31] “What is the role of frames in making a video slow? - Quora.” https://www.quora.com/What-is-the-role-of-frames-in-making-a-video-slow (accessed May 28, 2022).
[32] “What is the difference between hd, full hd, ultra hd, 2K, 4k, 8k, 10K resolutions? | DSPLAY.” https://dsplay.tv/site/en/suporte/qual-a-diferenca-entre-resolucoes-hd-full-hd-ultra-hd-2k-4k-8k-10k/ (accessed May 29, 2022).
[33] Adorama, “Understanding Focal Length and How to Use It,” 2021. https://www.adorama.com/alc/understanding-focal-length-and-how-to-use-it/
[34] H. Kim, S. H. Sim, and B. F. Spencer, “Automated concrete crack evaluation using stereo vision with two different focal lengths,” *Autom. Constr.*, vol. 135, p. 104136, Mar. 2022, doi: 10.1016/J.AUTCON.2022.104136.
[35] E. Ali Yasin, “The Depth I: Stereo Calibration and Rectification,” 2020. https://python.plainenglish.io/the-depth-i-stereo-calibration-and-rectification-24da7b0fb1e0
[36] A. Y. Eser, “OpenCV Camera Calibration,” *Medium*, Mar. 01, 2020. https://aliyasineser.medium.com/opencv-camera-calibration-e9a48bdd1844 (accessed May 30, 2022).
[37] OpenCV, “Geometric Image Transformations.” https://docs.opencv.org/3.4/da/d54/group__imgproc__transform.html

[38] F. Zhou, H. Zhao, and Z. Nie, “Safety Helmet Detection Based on YOLOv5,” *Proc. 2021 IEEE Int. Conf. Power Electron. Comput. Appl. ICPECA 2021*, pp. 6–11, Jan. 2021, doi: 10.1109/ICPECA51329.2021.9362711.
[39] PyTorch, “PyTorch.” https://pytorch.org/
[40] Roboflow, “Roboflow - Overview.” https://docs.roboflow.com/
[41] Google, “Google Colaboratory - Frequently Asked Questions.” https://research.google.com/colaboratory/faq.html#:~:text=Colaboratory%2C or “Colab” for,learning%2C data analysis and education.
[42] “GitHub - ultralytics/yolov5: YOLOv5 ? in PyTorch > ONNX > CoreML > TFLite.” https://github.com/ultralytics/yolov5 (accessed Jun. 02, 2022).
[43] “What Is CUDA | NVIDIA Official Blog.” https://blogs.nvidia.com/blog/2012/09/10/what-is-cuda-2/ (accessed Jun. 02, 2022).
[44] “ComputerVision/yoloCustomObjectDetection.py at master · niconielsen32/ComputerVision · GitHub.” https://github.com/niconielsen32/ComputerVision/blob/master/yoloCustomObjectDetection.py (accessed Jun. 02, 2022).
[45] “Camera model: intrinsic parameters – Hoàng-Ân Lê.” https://lhoangan.github.io/camera-params/ (accessed Jun. 03, 2022).

Bijlage A

Specificaties Hikvision DS-2CD2643G2-IZS IP-camera [25]

Camera	
Image Sensor	1/3" Progressive Scan CMOS
Min. Illumination	Color: 0.005 Lux @ (F1.6, AGC ON), 0 Lux with IR on
Shutter Speed	1/3 s to 1/100,000 s
Slow Shutter	Yes
P/N	P/N
Wide Dynamic Range	120 dB
Day & Night	ICR Cut
Angle Adjustment	Pan: 0° to 360°, tilt: 0° to 90°, rotate: 0° to 360°
Lens	
Focal Length & FOV	2.8 to 12 mm, horizontal FOV 95.8° to 29.2°, vertical FOV 50.6° to 16.4°, diagonal FOV 114.6° to 33.4°
Aperture	F1.6
Lens Mount	Ø14
DORI	
DORI	2.8 to 12 mm: D: 64 m to 200 m, O: 25.4 m to 79.4 m, R: 12.8 m to 40 m, I: 6.4 m to 20 m
Illuminator	
IR Range	Up to 60 m
IR Wavelength	850 nm
Video	
Max. Resolution	2688 × 1520
Main Stream	50 Hz: 25 fps (2688 × 1520, 1920 × 1080, 1280 × 720) 60 Hz: 30 fps (2688 × 1520, 1920 × 1080, 1280 × 720)
Sub-Stream	50 Hz: 25 fps (640 × 480, 640 × 360) 60 Hz: 30 fps (640 × 480, 640 × 360)
Third Stream	50 Hz: 10 fps (1920 × 1080, 1280 × 720, 640 × 480, 640 × 360) 60 Hz: 10 fps (1920 × 1080, 1280 × 720, 640 × 480, 640 × 360)
Video Compression	Main stream: H.265/H.264/H.264+/H.265+ Sub-stream: H.265/H.264/MJPEG Third stream: H.265/H.264
Video Bit Rate	32 Kbps to 8 Mbps
H.264 Type	Baseline Profile/Main Profile/High Profile
H.265 Type	Main Profile
H.264+	Main Stream supports
H.265+	Main Stream supports
Bit Rate Control	CBR/VBR
Scalable Video Coding (SVC)	Yes
Region of Interest (ROI)	1 fixed region for main stream and sub-stream
Audio	
Environment Noise Filtering	Yes
Audio Sampling Rate	8 kHz/16 kHz/32 kHz/44.1 kHz/48 kHz
Audio Compression	G.711ulaw/G.711alaw/G.722.1/G.726/MP2L2/PCM/MP3/AAC

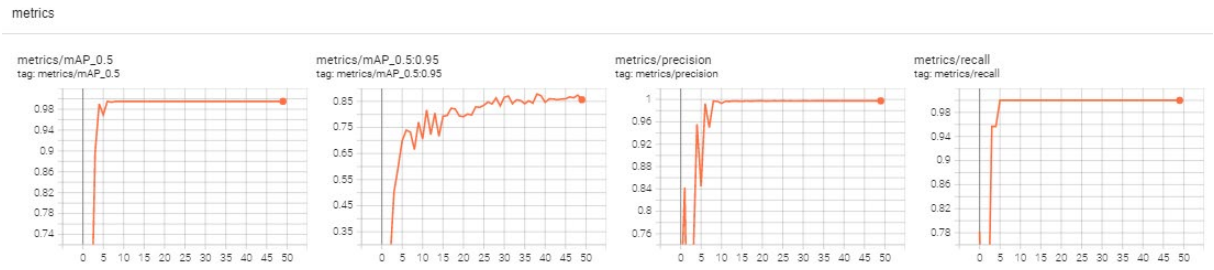
Audio Bit Rate	64Kbps(G.711ulaw/G.711alaw)/16Kbps(G.722.1)/16Kbps(G.726)/16Kbps-64Kbps(AAC)/32-192Kbps(MP2L2)/8-320Kbps(MP3)
Network	
Simultaneous Live View	Up to 6 channels
API	Open Network Video Interface (Profile S, Profile G), ISAPI, SDK
Protocols	TCP/IP, ICMP, HTTP, HTTPS, FTP, DHCP, DNS, DDNS, RTP, RTSP, NTP, UPnP, SMTP, IGMP, 802.1X, QoS, IPv6, UDP, Bonjour, SSL/TLS, PPPoE
User/Host	Up to 32 users. 3 user levels: administrator, operator and user
Security	Password protection, complicated password, HTTPS encryption, IP address filter, Security Audit Log, basic and digest authentication for HTTP/HTTPS, TLS 1.1/1.2, WSSE and digest authentication for Open Network Video Interface
Network Storage	MicroSD/SDHC/SDXC card (256 GB) local storage, and NAS(NFS, SMB/CIFS), auto network replenishment (ANR)
Client	iVMS-4200, Hik-Connect, Hik-Central
Web Browser	Plug-in required live view: IE 10, IE 11 Plug-in free live view : Chrome 57.0+, Firefox 52.0+ Local service: Chrome 57.0+, Firefox 52.0+
Image	
Day/Night Switch	Day, Night, Auto, Schedule
Image Enhancement	BLC, HLC, 3D DNR
Image Parameters Switch	Yes
Image Settings	Rotate, mirror, privacy masks, saturation, brightness, contrast, sharpness, gain, white balance adjustable by client software or web browser
Interface	
Audio	1 input (line in), two-core terminal block, max. input amplitude: 3.3 vpp, input impedance: 4.7 KΩ, interface type: non-equilibrium; 1 output (line out), two-core terminal block, max. output amplitude: 3.3 vpp, output impedance: 100 Ω, interface type: non-equilibrium
Alarm	1 input, 1 output (max. 12 VDC, 30 mA)
On-Board Storage	Built-in micro SD, up to 256 GB
Reset Key	Yes
Ethernet Interface	1 RJ45 10M/100M self-adaptive Ethernet port
Event	
Basic Event	Motion detection (human and vehicle targets classification), video tampering alarm, exception
Face Detection	Yes
Deep Learning Function	
Perimeter Protection	Line crossing detection, intrusion detection Supports human and vehicle targets classification
General	
Linkage Method	Upload to FTP/memory card/NAS, notify surveillance center, trigger record, trigger capture, Email

Language	33 languages English, Russian, Estonian, Bulgarian, Hungarian, Greek, German, Italian, Czech, Slovak, French, Polish, Dutch, Portuguese, Spanish, Romanian, Danish, Swedish, Norwegian, Finnish, Croatian, Slovenian, Serbian, Turkish, Korean, Traditional Chinese, Thai, Vietnamese, Japanese, Latvian, Lithuanian, Portuguese (Brazil), Ukrainian
General Function	Anti-flicker, heartbeat, password reset via e-mail, pixel counter
Software Reset	Yes
Storage Conditions	-30 °C to 60 °C (-22 °F to 140 °F). Humidity 95% or less (non-condensing)
Startup and Operating Conditions	-30 °C to 60 °C (-22 °F to 140 °F). Humidity 95% or less (non-condensing)
Power Supply	12 VDC ± 25% PoE: 802.3at, Class 4
Power Consumption and Current	12 VDC, 1.08 A, max. 13 W PoE: (802.3at, 42.5 V to 57 V), 0.36 A to 0.27 A, max. 15 W
Power Interface	Ø5.5 mm coaxial power plug
Material	Aluminum alloy body
Dimension	308.5 × 97.9 × 93 mm (12.2" × 3.9" × 3.7")
Package Dimension	385 × 190 × 180 mm (15.2" × 7.5" × 7.1")
Weight	Approx. 1390 g (3.1 lb.)
With Package Weight	Approx. 1961 g (4.3 lb.)
Approval	
EMC	FCC (47 CFR Part 15, Subpart B); CE-EMC (EN 55032: 2015, EN 61000-3-2: 2014, EN 61000-3-3: 2013, EN 50130-4: 2011 +A1: 2014); RCM (AS/NZS CISPR 32: 2015); IC (ICES-003: Issue 6, 2016); KC (KN 32: 2015, KN 35: 2015)
Safety	UL (UL 60950-1); CB (IEC 60950-1:2005 + Am 1:2009 + Am 2:2013); CE-LVD (EN 60950-1:2005 + Am 1:2009 + Am 2:2013); BIS (IS 13252(Part 1):2010+A1:2013+A2:2015)
Environment	CE-RoHS (2011/65/EU); WEEE (2012/19/EU); Reach (Regulation (EC) No 1907/2006)
Protection	IP67 (IEC 60529-2013), IK10 (IEC 62262: 2002)

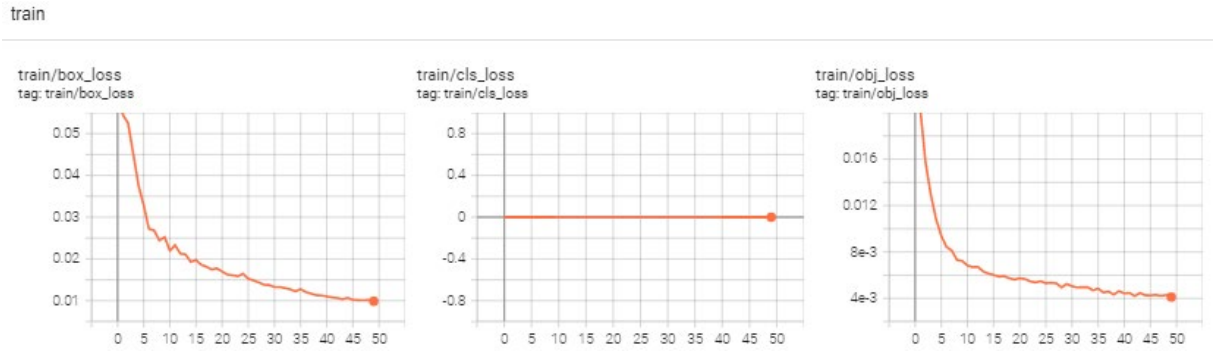
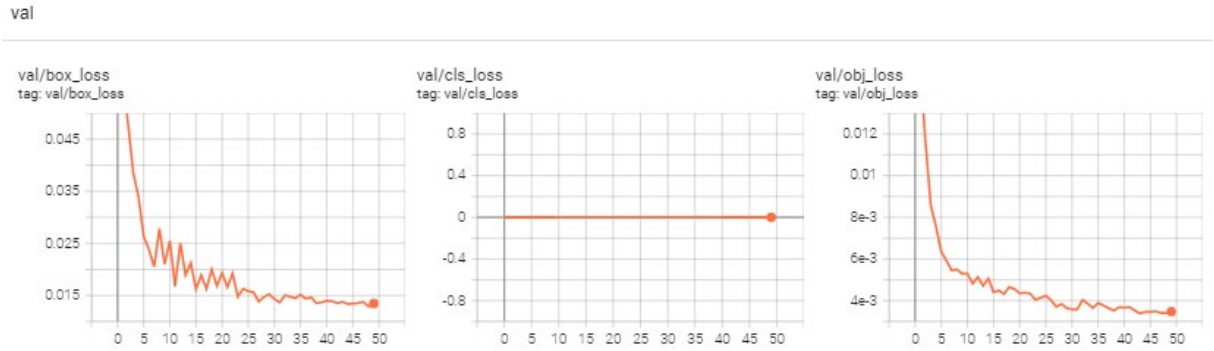
Bijlage B

Visualisatie trainingsresultaten YOLOv5s-helmdetectiemodel aan de hand van TensorBoard

Metrics:



Train:



Validatie: