

Augmented Reality Bordspellen

Wouter Franken

Thesis voorgedragen tot het behalen
van de graad van Master of Science
in de ingenieurswetenschappen:
computerwetenschappen,
hoofdspecialisatie Mens-machine
communicatie

Promotor:
Prof. dr. ir. Philip Dutre

Assessor:
TODO

Begeleider:
Ir. J. Baert

© Copyright KU Leuven

Zonder voorafgaande schriftelijke toestemming van zowel de promotor als de auteur is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, wend u tot het Departement Computerwetenschappen, Celestijnenlaan 200A bus 2402, B-3001 Heverlee, +32-16-327700 of via e-mail info@cs.kuleuven.be.

Voorafgaande schriftelijke toestemming van de promotor is eveneens vereist voor het aanwenden van de in deze masterproef beschreven (originele) methoden, producten, schakelingen en programma's voor industrieel of commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

Voorwoord

DANKWOORD

Wouter Franken

Inhoudsopgave

Voorwoord	i
Samenvatting	iv
Lijst van figuren en tabellen	v
Lijst van afkortingen en symbolen	vii
1 Inleiding	1
1.1 Lorem ipsum 4–5	1
1.2 Lorem ipsum 6–7	1
2 Het eerste hoofdstuk	3
2.1 Eerste onderwerp in dit hoofdstuk	3
2.2 Tweede onderwerp in dit hoofdstuk	3
2.3 Besluit van dit hoofdstuk	4
3 Naïeve legoblokdetectie	5
3.1 Evaluatiemethode en implementatie	5
3.2 Blokdetectie op basis van simpele geometrische informatie	6
3.3 Blokdetectie met behulp van een 2D grid	14
3.4 Besluit van dit hoofdstuk	18
4 Legoblokdetectie op basis van CAD modellen	19
4.1 Begrippen en algoritme	19
4.2 Features	22
4.3 Evaluatie features	24
4.4 Classificatie methodes	28
4.5 Evaluatie classifier methodes	29
4.6 Discussie / Besluit	34
5 Het laatste hoofdstuk	11
5.1 Eerste onderwerp in dit hoofdstuk	11
5.2 Tweede onderwerp in dit hoofdstuk	11
5.3 Besluit van dit hoofdstuk	11
5 Besluit	11
A De eerste bijlage	15
A.1 Meer lorem	15
A.2 Lorem 51	16

INHOUDSOPGAVE

B De laatste bijlage	17
B.1 Lorem 20-24	17
B.2 Lorem 25-27	18
Bibliografie	19

Samenvatting

ABSTRACT

Lijst van figuren en tabellen

Lijst van figuren

3.1	Simpele (links) en complexe scene (rechts) voor performantie evaluatie van naïeve blokdetectie.	5
3.2	Twee voorbeelden die aangeven hoe de pose vanuit een contour gevonden kan worden.	9
3.3	Resultaat van het eerste naïeve algoritme.	10
3.4	Resultaat van het eerste naïeve algoritme bij veel schaduw.	11
3.5	Resultaat van het eerste naïeve algoritme bij rode tint in het beeld.	11
3.6	Gemiddelde performantie van het eerste naïeve algoritme.	12
3.7	Ruis in threshold boor morphologische operaties.	15
3.8	Fout van dit algoritme in de vorm van schaduw.	15
3.9	Resultaat van het tweede naïeve algoritme.	17
3.10	Voorbeeld van het Tree Adaptive A* algoritme (uit [HSKM11]).	17
4.1	Pose van het CAD model van een 2x2 legoblokje.	20
4.2	Uitgebreide Haar-like featureset [LM02].	22
4.3	Een 9x9 MB-LBP operator: elk blok bestaat uit 9 pixels [LZL ⁺ 07].	23
4.4	De 20 afbeeldingen die als testset werden gebruikt.	26
4.5	Feature performantie tijdens detectie.	27
4.6	Feature robuustheid.	28
4.7	Robuustheid classifier methodes.	32
4.8	Classifier methodes performantie tijdens detectie.	32
4.9	Impact schaalfactor op performantie en robuustheid van HOG feature.	33
4.10	Een voorbeeld van een legoblokconstructie waarin het groen omlijnde vlak het enige is wat we van die legoblok zien.	34
4.11	De vergroting van een rand in een muur van legoblokken.	35

Lijst van tabellen

4.1	Parameters gebruikt tijdens training en detectie van features met cascade classificatie als classificatie methode.	25
4.2	Feature training tijdsduur.	28

LIJST VAN FIGUREN EN TABELLEN

4.3 Parameters gebruikt tijdens training en detectie van features met SVM als classificatie methode.	30
4.4 Classifier methodes training tijdsduur.	31

Lijst van afkortingen en symbolen

Afkortingen

Symbolen

Hoofdstuk 3

Naïeve legoblokdetectie

In dit hoofdstuk bespreken we twee naïeve algoritmes om legoblokken te detecteren. Beide algoritmes zullen via een thresholding operatie de blokken lokaliseren. Vervolgens zal het eerste algoritme de geometrie van een legoblok gebruiken om te bepalen wat de pose is van deze eerste legoblok. Het tweede algoritme zal een 2 dimensionaal grid bepalen waar de blokken zich bevinden.

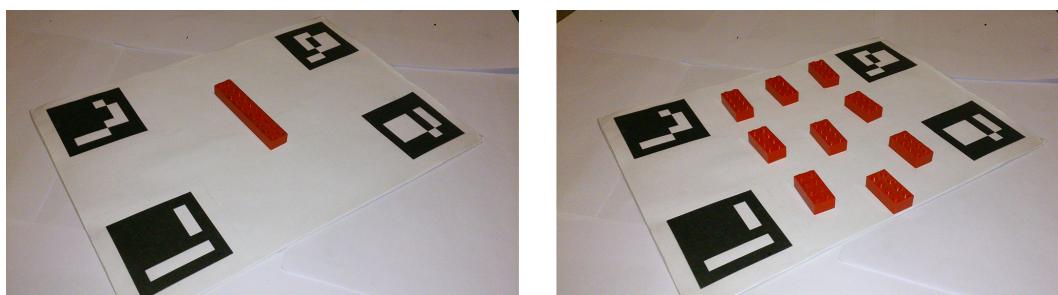
TODO: OVERZICHT VAN HET HOOFDSTUK

3.1 Evaluatiemethode en implementatie

In deze sectie bespreken we eerst kort op welke manier de performantie van beide algoritmes worden geëvalueerd. Vervolgens wordt kort overlopen welke machine en welke bibliotheken er zijn gebruikt voor de implementaties van de algoritmes.

3.1.1 Evaluatiemethode

De performantie van beide algoritmes werden geëvalueerd door de tijd op te meten van het volledige lemmingenspel zowel wanneer een eenvoudige scene werd gebouwd als wanneer een complexe scene werd gebouwd. Figuur 3.1 toont beide scenes die hiervoor zijn gebruikt. Merk op dat dit scenes zijn die simpel en complex zijn voor



FIGUUR 3.1: Simpele (links) en complexe scene (rechts) voor performantie evaluatie van naïeve blokdetectie.

3. NAÏVE LEGOLOKDETECTIE

deze algoritmes, maw. het zijn scenes die beide algoritmes kunnen detecteren en respectievelijk weinig / veel legoblokken bevatten.

De opmeting gebeurde even lang als een lemming nodig heeft om van begin tot eindpunt te geraken. Aangezien de snelheid gedefinieerd is als 0.5 cm/s en de afstand minimum 23 cm is, duurt de opmeting dus minimum 11.5 seconden. Vervolgens wordt per opgemeten gedeelte het gemiddelde genomen om geen grote uitscheters te hebben.

3.1.2 Implementatie

De volledige implementatie en experimenten in dit hoofdstuk zijn uitgevoerd op een Sony Xperia Z C6602 met een Qualcomm Quad-core 1.5 GHz Krait processor, Adreno 320 GPU, 2GB RAM en Android 4.4.4 KitKat. Alles werd geprogrammeerd in Java 7 met gebruik van de Cardboard VR SDK van Google voor stereo rendering voor de Google Cardboard en de OpenCV 2.4.9 bibliotheek voor implementaties van computer vision algoritmes.

3.2 Blokdetectie op basis van simpele geometrische informatie

Dit eerste algoritme verloopt in twee delen om uiteindelijk de locatie en pose van een legoblok te bepalen: vinden van de contour van de legoblok en bepalen van de pose van de legoblok uit de contour. Eerst bespreken we deze twee onderdelen in een theoretische sectie, vervolgens wordt uitgelegd hoe het algoritme werd geïmplementeerd en ten slotte bespreken we de resultaten, performantie en voor-en nadelen van deze methode.

3.2.1 Begrippen en veronderstellingen

Begrippen

Thresholding is een operatie die vaak in beeldverwerking wordt gebruikt om twee delen in een afbeelding te scheiden van elkaar (vaak voorgrond en achtergrond). Het resultaat is een binaire afbeelding waarin de achtergrond meestal wordt aangeduid met zwart en de voorgrond met wit.

YUV, **RGB** en **HSV** zijn drie verschillende kleurenruimtes die elk op een andere manier kleuren definiëren. Bij YUV gebeurt dit door de helderheid (Y) van de kleur te scheiden van twee kleurcomponenten (U en V). In RGB zijn er enkel drie kleurcomponenten waarbij de helderheid dus in deze componenten verwerkt zit. HSV scheidt tint (H), verzadiging (S) en helderheid (V) van elkaar. Omdat HSV drie componenten scheidt die een duidelijk verschillende invloed hebben op de uiteindelijke kleur wordt deze kleurenruimte vaak gebruikt voor kleurthresholding, het is immers eenvoudiger om kleuren van elkaar te scheiden zonder de verzadiging of de helderheid te beïnvloeden.

Pinhole model is het model dat OpenCV gebruikt om 3D punten van een scene te transformeren naar de afbeelding via een perspectieftransformatie. In de algoritmes uit dit hoofdstuk wordt het gebruikt om 2D punten om te zetten naar 3D. Dit is de formule van het pinhole model:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Hierbij is (u, v) het 2D punt, (X, Y, Z) het 3D punt, de eerste matrix in het rechterlid de camera intrinsics matrix (bekomen door camera calibratie), de tweede matrix in het rechterlid is de camera extrinsics matrix (afgeleid uit de ligging van de markers) en s , tenslotte, is een schaalfactor. Wanneer de twee cameramatrices bekend zijn kunnen we een 3D punt omzetten in 2D en vice versa. Een belangrijke opmerking hierbij is dat, bij omzetting van 2D naar 3D, wat informatie te kort is (het aantal dimensies verhoogt immers). Deze extra informatie kan worden gegeven in twee vormen: ofwel kennen we de uiteindelijke Z coördinaat, ofwel kennen we de uiteindelijke X en Y coördinaten.

Veronderstellingen

In dit algoritme worden enkele veronderstellingen gemaakt, indien hier niet aan wordt gehouden zijn de resultaten onbepaald. In deze sectie bespreken we kort welke veronderstellingen zijn gemaakt en waarom.

- De achtergrond bestaat uit wit en zwart, belangrijk is vooral dat ze geen rode tint bevat. Deze restrictie werd opgelegd omdat we in het algoritme kleurthresholding gebruiken om de rode blokken te scheiden van de achtergrond.
- Omdat het algoritme gebruik maakt van kleurthresholding is het ook aangeraden om zo weinig mogelijk schaduw te hebben. Anders kan de contour van de legoblok moeilijk gedetecteerd worden waardoor legoblokken mogelijks slecht worden gedetecteerd. In de resultaten (sectie 3.2.3) tonen we dat dit werkelijk een probleem is.
- De blokken die worden gebruikt zijn steeds balkvormig, geen enkele blok of constructie van blokken mag een hoek bevatten. Dit is belangrijk omdat het algoritme ervan uit gaat dat de contour van een te detecteren legoblok bestaat uit maximaal zes hoekpunten. Dit impliceert ook dat legoblokken elkaar niet mogen aanraken vanuit het standpunt van de camera gezien (tenzij ze in elkaars verlengde liggen).
- Zoals hierboven reeds aangehaald gaat het algoritme ervan uit dat de contour van een blok maximaal zes hoekpunten bevat. Dit impliceert dat het algoritme beter werkt wanneer we legoblokken vanuit perspectief zien want dan bestaat de omtrek van een blok uit exact zes hoekpunten.

3. NAÏEVE LEGOBLOKDETECTIE

3.2.2 Algoritme

Bepalen van de contour

Eerst wordt de nieuwe frame van het YUV formaat omgezet naar het HSV formaat. Dit is een erg belangrijke stap vanwege twee redenen: ten eerste kan OpenCV niet werken met het YUV formaat en ten tweede is het HSV formaat (zoals hierboven reeds aangehaald) veruit het eenvoudigste formaat om te gebruiken bij thresholding operaties.

Op de HSV frame wordt vervolgens kleurthresholding toegepast om de legoblokken te scheiden van de wit-zwarte achtergrond. In dit eerste algoritme werd enkel gewerkt met de kleur rood. Experimenteel werd ondervonden dat om rood te scheiden van een wit-zwarte achtergrond de HSV waarden in het volgende interval moeten liggen:

$$160 < H < 180; 153 < S < 255; 30 < V < 255$$

TODO: IS CANNY EDGE DETECTIE WEL NODIG, AANGEZIEN WE AL EEN THRESHOLD HEBBEN?

Contouren worden dan bepaald met behulp van de `findContours` methode van OpenCV. Deze methode gebruikt het algoritme dat wordt beschreven in [S+85], het wordt hier niet besproken omdat dit out of scope van deze thesis is.

In de volgende stap wordt deze contour benadert met behulp van de `approxPolyDP` methode van OpenCV. Deze methode benadert deze contours met een polygon. Deze stap is nodig omdat een contour uit een enorme hoeveelheid punten bestaat, na deze benadering blijven de belangrijkste punten over: de hoekpunten van de legoblok. Deze methode bevat een parameter ϵ die aangeeft wat de maximale afwijking mag zijn van de benadering ten opzichte van de oorspronkelijke contour. Om ervoor te zorgen dat we een zo goed mogelijke benadering verkrijgen in alle gevallen, maken we ϵ steeds groter tot we net een maximum van zes hoekpunten hebben bereikt. Deze adaptieve techniek behoudt de kwaliteit van de contour maar in ruil daarvoor moeten we de veronderstelling maken dat de omtrek van de contour van de threshold bestaat uit maximaal zes hoeken.

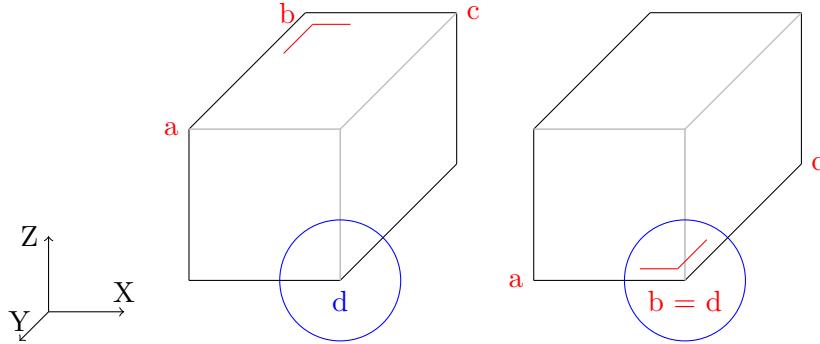
Nu kan het zijn dat er contours zijn met een te kleine oppervlakte door ruis in de kleurthreshold. Om deze te verwijderen bekijken we de oppervlakte van alle contours en worden contours verwijderd met een oppervlakte die 100x kleiner is dan de oppervlakte van de grootste contour.

Het bepalen en verfijnen van de contour is afgelopen, nu moet bepaald worden wat de pose is van de legoblok. Dat gebeurt in het tweede deel van het algoritme.

Bepalen van de pose

Nu de contour is gevonden moet hieruit worden achterhaald wat de pose van de legoblok is, hiervoor bepalen we eerst in welk vlak van de legoblok elk hoekpunt van de contour ligt: onder- of bovenvlak. Dit geeft ons de Z coördinaat van al deze hoekpunten waardoor ze kunnen worden omgezet van 2D naar 3D via het pinhole model. Bij uitbreiding zijn dan alle hoekpunten van de legoblok in 3D gekend en dus kennen we ook de pose van de legoblok.

3.2. Blokdetectie op basis van simpele geometrische informatie



FIGUUR 3.2: Twee voorbeelden die aangeven hoe de pose vanuit een contour gevonden kan worden.

Eerst leggen we uit hoe kan achterhaald worden in welk vlak (boven of onder) elk hoekpunt van de contour ligt. Deze redenering wordt verduidelijkt met twee voorbeelden in figuur 3.2, waarin legoblokjes worden getoond met hun contourlijnen aangeduid in zwart.

Uit de geometrische informatie van een legoblok is geweten dat minstens drie opeenvolgende hoekpunten in hetzelfde Z-vlak liggen. Om te bepalen welke drie hoekpunten dit zijn zetten we alle 2D contourpunten om naar 3D (geprojecteerd op vlak $Z = 0$) en bepalen we welke drie opeenvolgende 3D hoekpunten (a , b en c) een hoek vormen die het dichtst bij 90 graden ligt. Van deze hoeken zijn we dan zeker dat ze in hetzelfde vlak liggen. In het eerste voorbeeld van figuur 3.2 ligt de hoek tussen drie hoekpunten in het bovenvlak dichtst bij 0, in het tweede voorbeeld is dit de hoek in het ondervlak. Een ander geval is onmogelijk omdat, na projectie op het ondervlak, hoeken tussen punten in een verschillend Z -vlak sterker zullen afwijken van 90 graden.

Vervolgens bepalen we het hoekpunt (d) dat in 3D (geprojecteerd op vlak $Z = 0$) zich het dichtst bij de camera bevindt, hiervan kan met zekerheid gezegd worden dat het werkelijk in het vlak $Z = 0$ ligt. Dit hoekpunt is in de voorbeelden aangeduid met een blauwe cirkel. Nu kan met zekerheid gezegd worden of a , b en c in het onder- of bovenvlak van de legoblok liggen:

- Als $d == a$ OF $d == b$ OF $d == c$, dan liggen a , b en c in het ondervlak;
- Anders liggen a , b en c in het bovenvlak.

In het eerste voorbeeld van figuur 3.2 liggen de hoekpunten a , b en c dus in het bovenvlak aangezien hoekpunt d niet gelijk is aan één van deze hoeken. In het tweede voorbeeld is dit wel het geval en dus liggen de drie punten in het ondervlak. Dit geeft ons (via het pinhole model) de 3D posities van alle hoekpunten van de contour en bij uitbreiding van de volledige legoblok.

Ten slotte kunnen nog de afmetingen en precieze positie van de blok worden bepaald. In plaats deze informatie uit slechts één frame te halen, wordt het achter-

3. NAÏEVE LEGOLOKDETECTIE

haald via een voting systeem over meerdere frames. Dit helpt om fouten in de vorige stappen van het algoritme te verkleinen. Het voting systeem werkt als volgt:

De blok wordt in verschillende frames gedetecteerd en wanneer alle hoekpunten dichter dan 0.8 cm bij elkaar liggen wordt dit gezien als dezelfde blok. In dat geval kan gestemd worden op de grootte en positie van de blok door het gemiddelde te nemen van de groottes en posities van alle blokken die zo dicht bij elkaar liggen. De waarde 0.8 cm is niet toevallig gekozen: het is exact de helft van de breedte van een 2x2 legoblok, zo dicht kunnen legoblokken dus nooit bij elkaar liggen.

Bovendien kan een legoblok nog in een verschillende state verkeren om te bepalen wanneer de blok een deel van het spel wordt (*actief*) en wanneer de blok er geen deel meer van uitmaakt (*inactief*): Wanneer een blok minstens in drie frames werd gedetecteerd wordt hij actief, maar indien hij minstens drie frames achter elkaar niet werd gedetecteerd wordt hij inactief en wanneer hij na vijf frames achter elkaar niet is gedetecteerd wordt hij zelfs verwijderd. Dit mechanisme maakt het algoritme flexibeler om blokken te kunnen toevoegen of verwijderen uit het spel wanneer de speler dat wil.

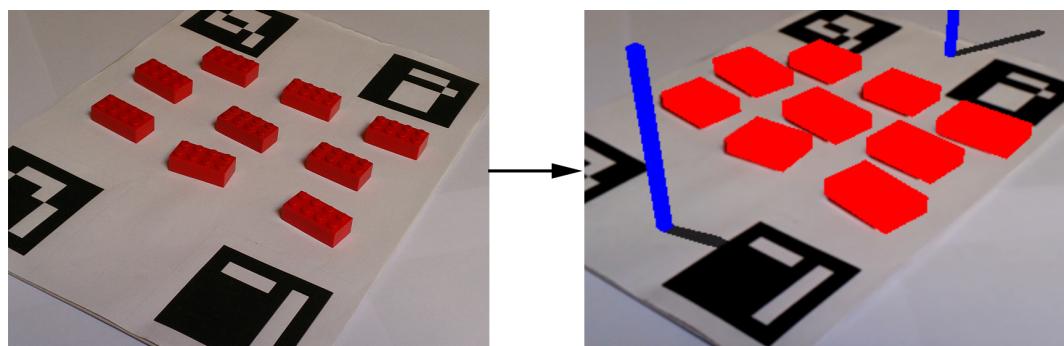
3.2.3 Resultaten en performantie

Resultaten

Figuur 3.3 toont het resultaat van het algoritme. De linkerafbeelding toont de legoblokken en de rechteraafbeelding wat de detectie van legoblokken oplevert.

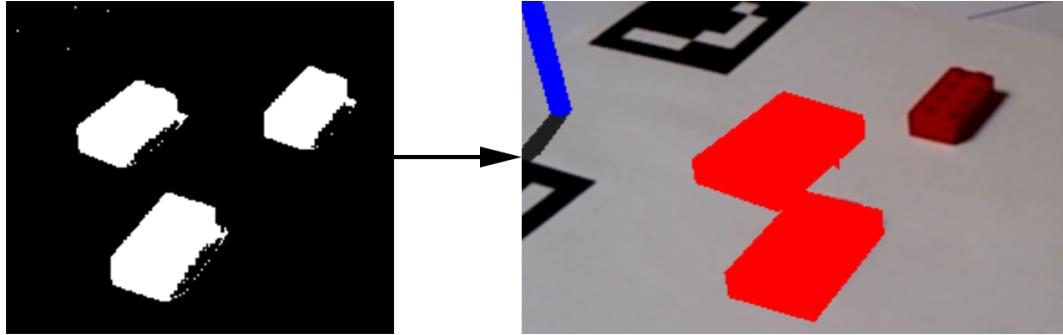
Merk op dat de rendering van de gedetecteerde legoblokken een kleine afwijking bevat van de werkelijke legoblokken. Dit wordt veroorzaakt door een trage performantie van het algoritme, waardoor de rendering in feite van enkele frames geleden.

Verder valt ook op te merken dat de virtuele legoblokken iets groter zijn dan de werkelijke legoblokken. Dit was nodig omdat de virtuele wereld in feite bestaat uit een discreet grid. Om er zeker van te zijn dat de lemmings niet gedeeltelijk in een legoblok wandelen worden ze iets groter gemaakt dan hun werkelijke grootte.



FIGUUR 3.3: Resultaat van het eerste naïeve algoritme.

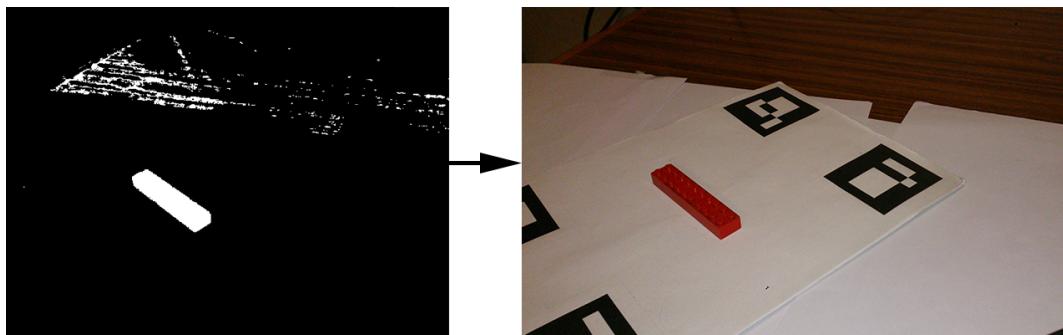
3.2. Blokdetectie op basis van simpele geometrische informatie



FIGUUR 3.4: Resultaat van het eerste naïeve algoritme bij veel schaduw.

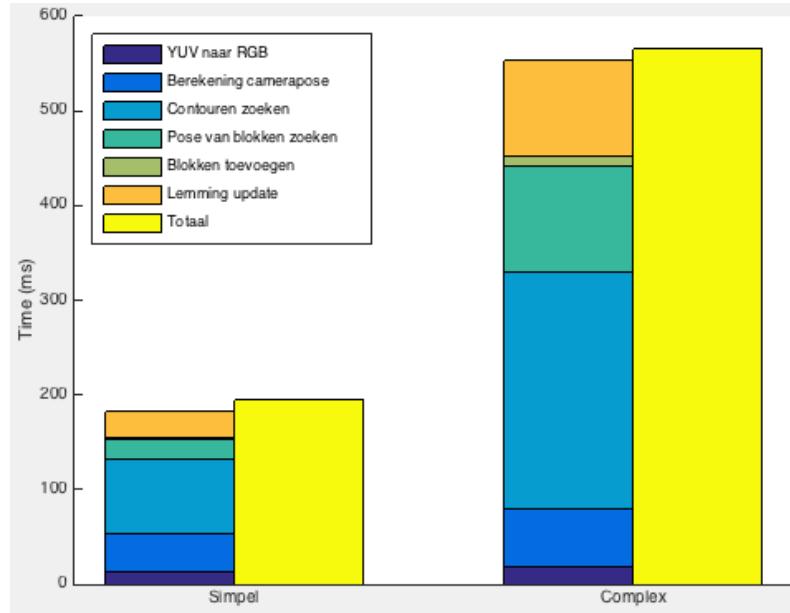
In figuur 3.4 wordt aangetoond dat schaduw een negatief effect heeft op het resultaat van het algoritme. We zien in de threshold dat er wat artefacten op de plaats van de schaduw aanwezig zijn, dit resulteert uiteindelijk in legoblokken die moeilijk worden gevonden.

Zoals aangegeven in sectie 3.2.1 is een rode tint in het beeld is absoluut uit den boze, elke kleur in het beeld dat een beetje in de grenzen van de threshold valt vormt een probleem. Dit tonen we aan in figuur 3.5, waarin een bruine achtergrond in beeld komt. We zien duidelijk dat veel ruis in de threshold voorkomt. Het gevolg hiervan is vooral dat het algoritme enorm veel trager is dan normaal: soms duren alle berekeningen samen wel tot 5 seconden. Dit komt omdat ruis in de threshold het algoritme doet denken dat op die plaatsen ook legoblokken staan en daardoor zal het al die contouren ook willen verwerken wat veel tijd vraagt. Dit is natuurlijk onaanvaardbaar voor een spel met AR. Bovendien gebeurde het zelden wel eens dat de camera zelfs zijn belichting aanpaste (bijvoorbeeld wanneer plots grote hoeveelheden donkere kleuren in beeld komen), hierdoor werd de threshold voor de legoblok zelf ook slechter wat ertoe leidde dat de blok helemaal niet meer gedetecteerd werd.



FIGUUR 3.5: Resultaat van het eerste naïeve algoritme bij rode tint in het beeld.

3. NAÏVE LEGOBLOKDETECTIE



FIGUUR 3.6: Gemiddelde performantie van het eerste naïeve algoritme.

Performantie

Figuur 3.6 toont de performantie van het eerste naïeve algoritme voor de simpele scene (links) en complexe scene (rechts). De gele blok stelt telkens de totale tijd die het algoritme nodig heeft, de gekleurde blokken geven de onderverdelingen aan van deze tijd:

- YUV omzetten naar BGR;
- Camerapose berekenen (via markers);
- Threshold berekenen en contouren van legoblokken zoeken (eerste deel van dit algoritme);
- Pose van de blokken zoeken (tweede deel van dit algoritme);
- Blokken toevoegen aan de virtuele wereld, mergen van blokken en voten voor grootte en positie van blokken.
- Lemmings updaten (o.a. pad van de lemming updaten).

Merk op dat de som van de gekleurde blokken niet exact gelijk is aan de grootte van de gele blok. Dit komt omdat enkel de belangrijkste onderdelen uit het algoritme zijn opgenomen in deze grafiek, enkele implementatie specifieke details zijn weggelaten omdat ze erg weinig tijd innemen en onbelangrijk zijn.

Uit de figuur kunnen we vooral opmaken dat het gedeelte van het zoeken van de contouren en zoeken van de pose erg veel verschillen tussen de complexe en simpele

3.2. Blokdetectie op basis van simpele geometrische informatie

scene. Dit is normaal aangezien het aantal legoblokken hoger ligt en dus moet dit naïeve algoritme veer iteraties doen in een complexe scene.

We merken ook op dat het updaten van de lemmingen een stuk langer duurt, dit is ook aannemelijk omdat bij de complexe scene het pad van een lemming van begin tot eindpunt een stukje langer is dan bij de simpele scene. Aangezien dit pad ook steeds opnieuw wordt berekend (om *on-the-fly* toevoegen van blokken toe te kunnen staan) is dit verschil significant.

Als we even kijken naar de totale tijd de volledige berekening inneemt, dan zien we toch dat dit algoritme erg traag is. Dit komt neer op een 5-tal fps voor de simpele scene en 1.8 fps voor de complexe scene, wat natuurlijk erg traag is voor een AR spel.

3.2.4 Voor- en nadelen

Deze naïeve implementatie heeft enkele voordelen:

- Ze laat toe dat legoblokken *on-the-fly* worden toegevoegd aan de virtuele wereld.

Aan deze methode zijn echter heel wat nadelen verbonden:

- Legoblokken mogen niet in een hoek naast elkaar staan omdat dan de veronderstelling dat een contour maximaal zes hoeken bevat niet meer geldig is.
- Enkel rode legoblokken kunnen worden gedetecteerd. Dit nadeel is echter eenvoudig weg te werken zoals wordt besproken in sectie 3.3.
- Legoblokken mogen niet in meerdere niveau's gebouwd worden. Muren zouden in principe wel kunnen (ze hebben immers maximaal zes hoekpunten) maar dan moet op één of andere manier de grootte van deze muur bepaald worden, dit wordt uitgebreider behandeld in hoofdstuk 5. Andere constructies in de hoogte kunnen echter niet vanwege dezelfde reden als het eerste nadeel.
- Het algoritme kan niet goed om met veranderingen qua belichting. Dit komt omdat we puur thresholden op basis van kleur maar deze waarden zijn sterk afhankelijk van welke belichting we gebruiken. In hoofdstuk 5 komt een calibratiemethode aan bod die dit probleem verhelpt.
- Zoals aangegeven in de resultaten geeft een rode tint in het beeld een erg slecht resultaat, vooral qua performantie.
- Performantie is erg laag.

3.3 Blokdetectie met behulp van een 2D grid

Dit algoritme is een sterke vereenvoudiging van het vorige. In plaats van werkelijk de pose en positie van de legoblok te bepalen, gebruiken we enkel de threshold. Eerst behandelen we de aanpassingen ten opzichte van het vorige algoritme en vervolgens bespreken we de gevolgen die deze aanpassingen hebben.

3.3.1 Begrippen

Deze sectie behandelt enkele nieuwe begrippen die in het algoritme worden gebruikt.

Morphologische operaties zijn operaties die de geometrie van vormen in een binaire afbeelding kunnen wijzigen. De twee basis operaties zijn *dilate* en *erode*. In deze operaties komt er op neer dat we de convolutie nemen van de afbeelding met een kernel, die eender welke vorm of grootte kan hebben. Hierbij wordt de kernel over de afbeelding geschoven en dan wordt de pixel in het ankerpunt van de kernel vervangen door een maximum of minimum van alle pixels die binnen de kernel vallen. Bij een *dilate* operatie is dit het maximum en bij een *erode* operatie is dit het minimum. Bij een *dilate* en *erode* operatie wordt een zwart vlak dus respectievelijk kleiner en groter.

Naast de basis morphologische operaties bestaan ook nog de veel gebruikte *open* en *close* operatie. *Open* is in feite een *dilate* operatie toepassen op een afbeelding die eerder een *erode* operatie onderging. De *close* operatie is het omgekeerde. De *open* operatie wordt vaak gebruikt om kleine witte vlekjes te verwijderen, terwijl de *close* operatie wordt gebruikt om kleine zwarte vlekjes te verwijderen.

3.3.2 Algoritme

Het eerste deel van het vorige algoritme wordt sterk ingekort aangezien we enkel een threshold nodig hebben. Maar er zijn ook een tweetal toevoegingen om tegemoet te komen aan de nadelen van het vorige algoritme:

Ten eerste berekenen we opnieuw een kleurenthreshold maar deze keer niet enkel voor de kleur rood, ook voor geel en blauw. Op die manier willen we aantonen dat het algoritme zonder problemen met meerdere kleuren overweg kan. Dit zijn kleurengrenzen die werden gebruikt om in een HSV afbeelding te thresholden zodat het resultaat een afbeelding is waarin de legoblokken wit zijn en de achtergrond zwart:

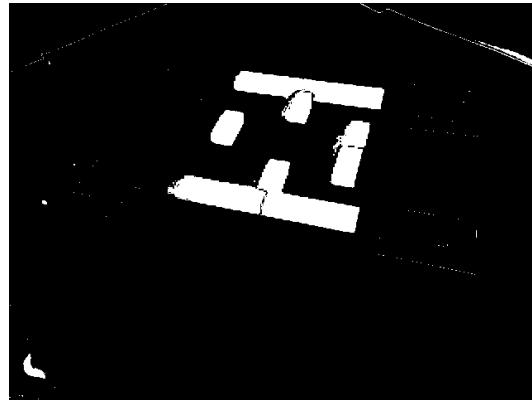
$$\text{Rood} : 160 < H < 180; 153 < S < 255; 30 < V < 255$$

$$\text{Geel} : 135 < H < 160; 147 < S < 255; 30 < V < 255$$

$$\text{Blauw} : 0 < H < 112; 42 < S < 255; 13 < V < 255$$

Ten tweede, omdat we nu thresholden op meerdere kleuren kan het gebeuren dat er wat ruis optreedt in onze threshold. Deze ruis vertoont zich in hoopjes witte of zwarte lijntjes op plaatsen waar het niet hoort (zie figuur 3.7). Om dit te vermijden worden de morphologische operaties *open* en *close* toegepast op de threshold.

3.3. Blokdetectie met behulp van een 2D grid



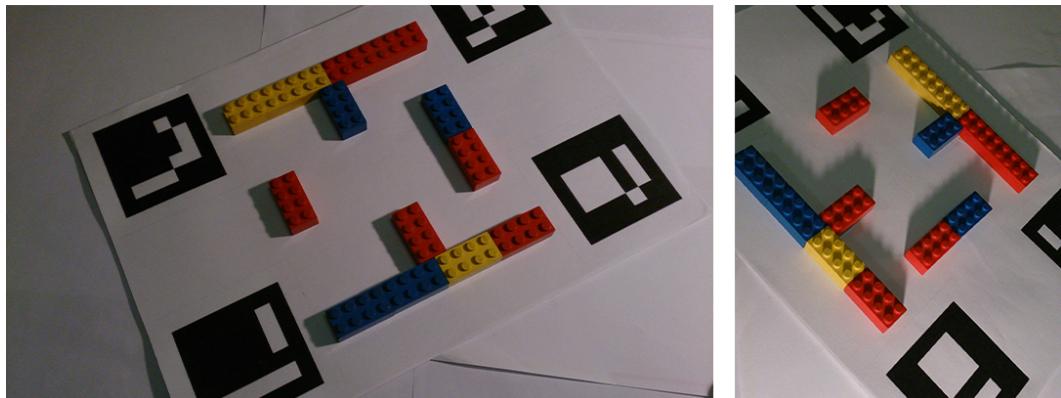
FIGUUR 3.7: Ruis in threshold boor morphologische operaties.

Het tweede deel van het vorige algoritme ('het bepalen van de pose') valt volledig weg, logisch want we hebben geen contour bepaald. In plaats daarvan wordt het grid van onze virtuele wereld naar 2D omgezet (via het pinhole model). Dit 2 dimensionaal grid kunnen we vervolgens overlappen met de eerder berekende threshold en wanneer een gridnode zich in een witte regio van de threshold bevindt beschouwen we deze gridnode als 'bezet door een legoblok'.

3.3.3 Gevolgen

Dit algoritme is een grove versimpeling van het eerste algoritme aangezien de positie en grootte van een legoblok niet meer expliciet wordt bepaald. Nu vragen we ons af welke consequenties dat heeft.

Ten eerste wordt een fout gemaakt in het aantal gridnodes die bezet worden door een legoblokje. Deze fout wordt gemaakt omdat een onbezette deel schuilt achter de legoblok en dus niet zichtbaar is in een 2D camera frame. Sterker nog: we



FIGUUR 3.8: Fout van dit algoritme in de vorm van schaduw.

3. NAÏVE LEGOLOKDETECTIE

kunnen precies definiëren dat die fout even groot is als de schaduw op het grondvlak wanneer een lichtbron vanuit de camera schijnt in dezelfde richting als de camera kijkt. Afbeelding 3.8 toont de fout in de vorm van schaduw als de camera op de plaats van de lichtbron stond. We zien dat de fout relatief klein is wanneer de camera hoog boven het grondvlak staat (links) en groter wordt wanneer de camera dichter naar het grondvlak toe beweegt (rechts).

Ten tweede heeft dit algoritme ook een aantal voordelen in die zin dat enkele nadelen uit het vorige algoritme werden weggewerkt:

- Legoblokken in meerdere kleuren kunnen worden gebruikt, het gevolg is wel dat morphologische operaties de threshold moeten oppoetsen om er kleine foutjes uit te halen. Dit werd hoogstwaarschijnlijk veroorzaakt door de belichting die niet altijd constant is, zodat de thresholdwaarden niet altijd even nauwkeurig zijn. Merk op dat dit even goed in het eerste algoritme had kunnen worden geïmplementeerd.
- Legoblokken kunnen nu ook in een hoek tegen elkaar staan, wat bij het eerste algoritme niet was toegestaan. Een nadeel is echter wel dat meerdere niveau's in dit algoritme helemaal uitgesloten zijn, terwijl dit in het vorige algoritme, althans in theorie en met restricties, wel mocht. Dit komt omdat, als we de constructie hoger maken, dan de fout die we in dit algoritme maken steeds groter wordt. Er is immers een nog groter gedeelte van het grondvlak, waar geen legoblokken op staan, verborgen achter de legoblokconstructie.

3.3.4 Resultaten en performantie

Resultaten

Voor dit algoritme is het niet mogelijk om een rendering te maken van de virtuele legoblok (zoals bij het eerste algoritme in sectie 3.3). Dit komt omdat de legoblokken in principe nooit volledig worden berekend: enkel een threshold wordt gebruikt waar vervolgens een grid over wordt gelegd.

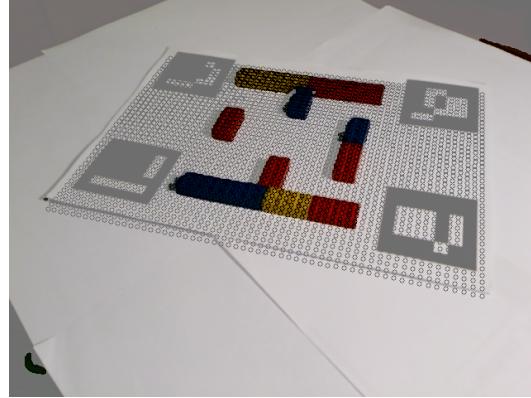
Figuur 3.9 toont het resultaat door het zwarte deel van de threshold deels transparant te maken. We kunnen duidelijk zien dat de legoblokken goed werden gedetecteerd via kleurthresholding. Ook werd over het grondvlak het grid gelegd, hierdoor is het duidelijk welke nodes niet toegankelijk zijn voor lemmings.

Performantie

TODO: BESPREKING PERFORMANTIE

Omdat de performantie van dit algoritme erg laag is, aangezien we elke keer opnieuw een volledig pad moeten berekenen en dus ook veel gridnodes moeten omzetten van 3D naar 2D, gebruiken we in plaats van A* om paden te zoeken Tree Adaptive A*. In deze uitbreiding zullen grote delen van reeds berekende paden herbruikbaar zijn waardoor al deze gridnodes niet opnieuw van 3D naar 2D moeten worden omgezet.

3.3. Blokdetectie met behulp van een 2D grid



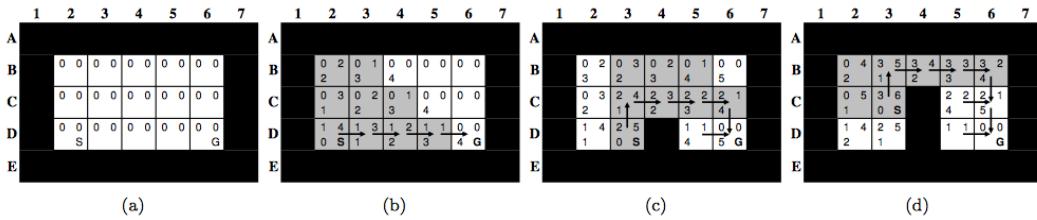
FIGUUR 3.9: Resultaat van het tweede naïeve algoritme.

3.3.5 Uitbreiding: Tree Adaptive A*

In deze sectie leggen we uit hoe Tree Adaptive A* (Tree-AA*) werkt en in welke mate het de performantie heeft verhoogd. Dit algoritme is in dieper detail uitgelegd in [HSKM11].

Algoritme

Tree Adaptive A* is een algoritme dat een uitbreiding is op Path Adaptive A*, wat op zijn beurt een uitbreiding is op Adaptive A*. Het algoritme gaat uit van de *freespace assumptie*, waarin een agent zich door een onbekend terrein een pad moet banen. Met een onbekend terrein bedoelen we dat hij de obstakels niet weet zijn, hij kan enkel obstakels om zich heen voelen. Hierdoor zal de agent elke keer hij een onbekend obstakel voelt het pad met optimale kost moeten aanpassen. Bij elke aanpassing mag de agent wel gebruik maken van de suffixen van vorige minimale kost paden (behalve natuurlijk de stukjes paden die door het obstakel gingen). De vorige paden die herbruikbaar zijn worden steeds opgeslagen in een datastructuur die de *herbruikbare boom* wordt genoemd (het is namelijk een boom met zijn root in de doelcel). Zo moet de agent telkens slechts een deel van het pad aanpassen wat het zoeken van een pad een heel stuk efficiënter maakt. De grootste kost zit dan in de initiële zoektocht van een pad maar aangezien de meeste obstakels dan nog niet



FIGUUR 3.10: Voorbeeld van het Tree Adaptive A* algoritme (uit [HSKM11]).

3. NAÏEVE LEGOBLÖKDETECTIE

door de agent kunnen worden gezien zal dit best meevalen.

We zetten onze uitleg nog kracht bij met het voorbeeld uit [HSKM11] dat getoond wordt in figuur 3.10. Alle grijze tegels zijn geëxpandeerde nodes en de zwarte tegels zijn de obstakels waar de agent weet van heeft. Figuur 3.10a toont het beginsituatie, figuur 3.10b toont de situatie waarin gezocht wordt naar een pad met minimale kost tot de doelcel. Nog geen obstakels zijn gevonden omdat deze zich niet bevinden rondom de huidige cel van de agent. Het zoeken stopt wanneer men net cel D6 wil gaan expanderen, het resulterende pad is dan D1, D2, D3, D4, D5 en D6. De agent gaat dan naar cel D3 en voelt van daaruit een obstakel in cel D4. Dan wordt de het vorige minimale kost pad deels verwijderd. Hierdoor blijft het stukje tussen D5 en D6 over in de *herbruikbare boom* omdat dit nog steeds deel kan uitmaken van een minimaal kost pad (we weten immers niet of op plaats D5 ook een obstakel zou liggen). De agent vindt vervolgens een nieuw pad dat ook aan de boom wordt toegevoegd. In de vierde stap wordt een obstakel ontdekt in cel C4 waardoor het stuk pad D3, C3, C4, C5 verdwijnt uit de boom terwijl het stukje C5, C6, D6 in de boom blijft. Bij het vinden van het nieuwe pad wordt het stukje C6, D6 herbruikt. Voor een meer gedetailleerde uitleg over het algoritme: zie paper [HSKM11].

Het grote voordeel aan dit algoritme voor een AR spel is dat deze *herbruikbare boom* kan gedeeld worden tussen lemmingen. Dus als de wereld niet veranderd moeten weinig stukjes pad herberekend worden.

Performantie

3.4 Besluit van dit hoofdstuk

Hoofdstuk 4

Legoblokdetectie op basis van CAD modellen

In het voorgaande hoofdstuk werd een naïeve methode aangebracht om legoblokken te detecteren. Deze methode worstelde echter met verschillende nadelen waardoor het (bijna) niet mogelijk was om constructies van legoblokken te detecteren die bestaan uit meerdere niveau's. Om dit soort constructies wel te kunnen detecteren moeten we op zoek naar een meer generiek algoritme dat op basis van alle geometrische informatie een legoblok kan detecteren (zodanig dat de blok kan gevonden worden in eender welke legoconstructie). Daarom wordt in dit hoofdstuk een algoritme behandeld dat op basis van features geëxtraheerd uit CAD modellen legoblokken kan detecteren in een videoframe.

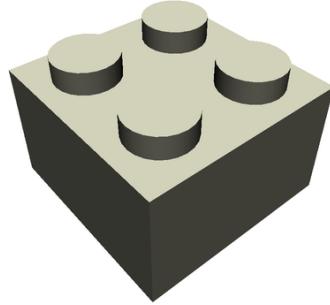
Eerst leggen we enkele begrippen uit en schetsen we kort het algoritme in sectie 4.1. Vervolgens bespreken we verschillende soorten features voor dit algoritme in sectie 4.2 en vergelijken we ze met elkaar in sectie 4.3.1. Ook enkele classificatie methodes voor dit algoritme komen aan bod en worden vergeleken respectievelijk in secties 4.4 en 4.5.1. Tenslotte geven we in sectie 4.6 aan waarom deze methode niet kan worden gebruikt in een AR spel om legoblokken te detecteren.

4.1 Begrippen en algoritme

4.1.1 Begrippen

Computer Aided Design (CAD) modellen zijn collecties van punten die de geometrie van een 3 dimensionaal object beschrijven, meestal worden ze gebruikt om objecten in een virtuele 3D wereld voor te stellen. Afbeelding 4.1 toont een afbeelding van het CAD model van een legoblok dat later in dit hoofdstuk nog wordt gebruikt.

Window sliding is een methode waarbij een rechthoek (het window) geschoven wordt over een afbeelding. Elke keer de rechthoek zich op een nieuwe plaats op de afbeelding bevindt wordt meestal een andere afbeelding vergeleken met het gedeelte van de afbeelding in het window. Deze methode wordt onder andere gebruikt bij object detectie om een object te zoeken in een afbeelding. In dit geval wordt dan,



FIGUUR 4.1: Pose van het CAD model van een 2x2 legoblokje.

steeds wanneer de rechthoek verschuift, bekeken of het te detecteren object zich in dit window bevindt. Omdat er tijdens window sliding echter geen rekening wordt gehouden met de schaal van het object doen we dit meermaals en wordt de schaal van het window telkens aangepast.

Een **classifier** is een functie die bij object detectie gebruikt wordt om tijdens window sliding te beslissen of het window het object bevat. In ons geval is de taak van de classifier dus om tijdens het window sliding te beslissen of het window een legoblok bevat.

AdaBoost is een algoritme dat vaak wordt gebruikt in object detectie algoritmes om classifiers sterker te maken tijdens de training. Het idee is dat enkel die zwakke classifiers (die op basis van 1 sample classificeren) worden gekozen die de kleinste trainingsfout maken en deze te combineren tot één sterke classifier [FS95]. Hier wordt het bij de Haar-like features (zie sectie 4.2.1) ook gebruikt om features te selecteren.

Een **integral image** is een afbeelding die gegenereerd wordt van een oorspronkelijke afbeelding door elke pixel vervangen met de som van alle pixels die in de linkerboven rechthoek liggen ten opzichte van de pixel [VJ01]. Dus indien de pixel in de linkerbovenhoek pixel $(0, 0)$ is, dan geldt voor elke pixel in de integral image:

$$p(x, y) = \sum_{x' \leq x, y' \leq y} p(x', y')$$

4.1.2 Algoritme

Een CAD model bevat alle geometrische informatie over een object (het moet dat object immers in 3D voorstellen) en dus kan dit model gebruikt worden om een object in een afbeelding te detecteren. Het grote voordeel is dan dat aangezien dit model alle geometrische informatie bevat het object in om het even welke pose kan worden gedetecteerd. Dit is de methode die werd aangebracht in [AME⁺14].

Om CAD modellen te kunnen gebruiken voor object detectie moet eerst nuttig informatie uit deze modellen worden gehaald. In een afbeelding is echter slechts een deel van het object te zien (de afbeelding is namelijk 2D), daarom is het beter eerst afbeeldingen te genereren van de verschillende poses van het CAD model (figuur 4.1 toont een voorbeeld).

Deze poses bevatten niet het exacte object dat we zoeken in de testafbeelding (het heeft misschien een andere kleur, andere belichting, andere achtergrond, etc.). Daarom moet nog informatie (in vorm van een classifier) geëxtraheerd worden uit deze afbeeldingen dat we vervolgens kunnen gebruiken voor de detectie. Dit gebeurt tijdens een trainingsfase.

Training

De training bestaat uit het berekenen van een classifier uit een aantal parameters en een hoop positieve en negatieve samples. We hebben ervoor gezorgd dat een deel van de negatieve samples bestaan uit het grondvlak zonder legoblokjes. Hiermee zijn de negatieve samples representatief voor de gebieden waarvan we willen dat ze niet als legoblok worden aanzien (de achtergrond).

De positieve samples zijn opgebouwd uit een set positieve afbeeldingen. Deze set bestaat enerzijds uit de virtueel gegenereerde poses van een 2x2 legoblok en anderzijds uit foto's van zulke legoblok, beide op een witte achtergrond (we gebruiken tijdens detectie immers ook een witte achtergrond). De poses werden gegenereerd van slechts een kwart van de legoblok omdat deze symmetrisch is. De foto's zijn nuttig omdat ze helpen de invloed van belichting op de classifier te minimaliseren aangezien ze werden getrokken in een realistische belichting.

Tijdens de training wordt een classifier berekend die zo goed mogelijk de positieve samples scheidt van de negatieve samples. Om de verschillende samples met elkaar te kunnen vergelijken worden eerst features berekend van alle samples die we vervolgens kunnen vergelijken met elkaar. Aangezien dit proces afhangt van welke classifier methode wordt gekozen, bespreken we dit uitgebreider in sectie 4.4.

Detectie

Na de training kan met behulp van de gegenereerde classifier en window sliding de legoblok gedetecteerd worden in een testafbeelding. Hiervoor wordt voor elke window de feature vector berekend die vervolgens aan de classifier wordt gegeven. Deze beslist dan of dit window het object bevat of niet.

Omdat de keuze van classifier methode en feature type erg belangrijk is voor het welslagen van dit algoritme worden de verschillende feature types en classifier methodes met elkaar vergeleken in de rest van dit hoofdstuk. De experimenten gebeuren allemaal op een desktop en niet op de smartphone waarop de uiteindelijke legobrick detectie moet worden geïmplementeerd. Dit is vooral omdat een desktop eenvoudiger is om op te experimenteren en omdat het voldoende is om het beste feature type en de beste classificatie methode te bepalen.

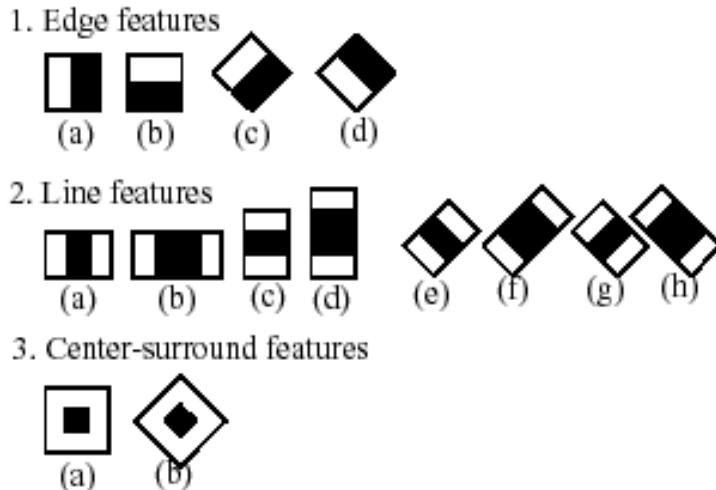
4.2 Features

In deze sectie worden drie verschillende soorten features besproken en met elkaar vergeleken: Haar-like [VJ01], Multiscale Block Local Binary Patterns [LZL⁺07] (MB-LBP) en Histogram of Oriented Gradients [DT05] (HOG). Tenslotte wordt ook een parts-based feature besproken (gebaseerd op HOG), dat gebruikt werd in [AME⁺14], en waarom deze niet is opgenomen in de evaluatie [FGMR10].

4.2.1 Haar-like

Bij de berekening van een Haar-like feature vector wordt het verschil genomen van de som van pixels tussen verschillende rechthoekige regio's. Oorspronkelijk werden vier soorten features gebruikt: twee die elk bestonden uit twee rechthoekige gebieden, één soort met drie rechthoekige gebieden en één soort met vier rechthoekige gebieden [VJ01]. Later werden extra features toegevoegd tot de uiteindelijke featureset 14 soorten features bedraagt [LM02]. De uitgebreide featureset is te zien in figuur 4.2: het verschil wordt bepaald door de som te nemen van pixels in de witte regio's en af te trekken van de som van pixels in de zwarte regio's. Alle soorten features zijn eenvoudig uit te rekenen aan de hand van een *integral image* (of een afgeleide daarvan [LM02]).

Om goede features te vinden wordt eerst een enorme hoeveelheid aan features aangemaakt. Aangezien dit aantal features een heel stuk hoger is dan het aantal pixels in het window moet een efficiënte methode gebruikt worden om de juiste features te selecteren. Hiervoor wordt AdaBoost gebruikt, een methode die origineel wordt gebruikt om een classifier sterker te maken. Aangezien AdaBoost het feature selecteert dat de kleinste fout maakt op de classificatie van positieve en negatieve samples, impliceert dit ook dat elke feature een threshold bevat die aangeeft wanneer



FIGUUR 4.2: Uitgebreide Haar-like featureset [LM02].

deze feature een detectiewindow als juist of onjuist classificeert. De uiteindelijke feature vector van het volledige detectiewindow is de set van de geselecteerde features.

4.2.2 MB-LBP

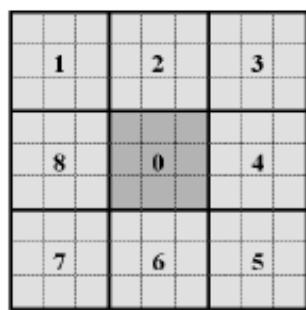
Een MB-LBP feature is een uitbreiding op de originele LBP feature waarin elke pixel werd vergeleken met zijn acht buren. Deze acht buren werden steeds in dezelfde volgorde overlopen en telkens een buur een grotere intensiteit had dan de pixel zelf werd een '1' genoteerd, in de andere gevallen een '0'. Zo wordt voor elke pixel een getal verkregen van acht bits dat de gradiënt van de pixel voorstelt in de verschillende richtingen [OPH96].

Hier wordt de MB-LBP feature gebruikt in plaats van de oorspronkelijke LBP omdat deze meer robuust is. In de MB-LBP variant worden, in plaats van een pixel met zijn buren te vergelijken, blokken van pixels met hun buren vergeleken (zie figuur 4.3). Hierdoor is de feature meer robuust en encodeert het bovendien, naast microstructuren, ook macrostructuren in een afbeelding [LZL⁺07].

Om de feature vector van een volledig window te berekenen wordt het window in cellen opgesplitst en vervolgens histogrammen berekend van de MB-LBP features binnen een cel. De histogrammen van alle cellen worden ten slotte aaneengeschakeld om de feature vector te bekomen.

4.2.3 HOG

Bij een HOG feature worden eerst gradiënten van alle pixels in een window berekend. Vervolgens wordt het window onderverdeeld in cellen en wordt voor elke cel een histogram van de gradiënten gemaakt. Om problemen met belichting en contrast te voorkomen worden cellen gecombineerd tot blokken waarover deze histogrammen worden genormaliseerd. Ten slotte worden alle histogrammen aaneengeschakeld (net als bij MB-LBP) om een feature vector van het window te bekomen [DT05].



FIGUUR 4.3: Een 9x9 MB-LBP operator: elk blok bestaat uit 9 pixels [LZL⁺07].

4.2.4 Deformed parts-based

Het deformed parts-based feature bestaat uit twee onderdelen: een HOG feature van de volledige window (root filter) en aantal kleinere HOG features van een subwindow (part filters). Deze verschillende part filters worden gedefinieerd door een anker punt ten opzichte van de plaats van de root filter en een functie die alle mogelijke plaatsen van deze part filter beschrijft relatief ten opzichte van het anker punt. De score van de hypothese dat een window het te detecteren object bevat wordt berekend door de sommatie te nemen van de verschillen (voor elke filter) tussen de score van een filter (HOG feature) en een kost voor de vervorming van de filter. Dit model maakt het HOG feature een stuk meer flexibel voor vervormingen of andere poses van een object [FGMR10].

Een groot nadeel van deze methode is dat ze nogal wat rekenwerk vraagt om uit te voeren en dus ongeschikt is voor gebruik in real time. Dit probleem is eerder al aangehaald in andere state-of-the-art papers die de performantie hebben kunnen verhogen. Hoewel bijvoorbeeld in paper [YLWL14] wordt aangegeven dat 40 FPS haalbaar is, is dit slechts na parallelisatie (over zes threads) en bovendien op een desktop computer. De desktop die in deze recente paper werd gebruikt heeft een 2.66GHz Hexacore Intel X5650 CPU. Dit is een heel stuk krachtiger dan de smartphone (met een 1GHz Quadcore Qualcomm Snapdragon S4 Pro APQ8064) die wij uiteindelijk willen gebruiken om het volledige algoritme op te implementeren. Het zou dus erg moeilijk zijn en bovendien veel tijd vragen om dit optimaal te implementeren op de smartphone. Dit verklaart waarom dit feature type voor ons geen interessante piste was en waarom deze niet is opgenomen in onze feature evaluatie.

4.3 Evaluatie features

Deze sectie behandelt de evaluatie van de feature types. Eerst wordt kort besproken hoe de evaluatie verloopt en vervolgens worden alle soorten features vergeleken op vlak van robuustheid en performantie.

4.3.1 Evaluatiemethode

De evaluatie van de verschillende features gebeurt door eerst een classifier te trainen voor alle features en vervolgens de geleerde classifiers te gebruiken voor detectie.

Training

Bij de training werden de features steeds getraind tot een cascade classifier (zie sectie 4.4.1), dit om geen invloed op de resultaten te krijgen door een andere keuze in classificatie methode. Om de training uit te voeren werd gebruik gemaakt van het programma `opencv_traincascade`, dat deel is van de OpenCv bibliotheek. Er werden ongeveer even veel positieve als negatieve samples gebruikt: de positieve samples begon met een set van 128 afbeeldingen die via transformaties op een

Parameter	Waarde
Training	
# classif. in cascade	20
Wind. size	64x64
Min. hit rate	0.999
Max. false alarm rate	0.5
Mode (Haar-like)	ALL
Detectie	
Schaalfactor	1.1
Min. # buren	15
Min. grootte	10x10 (px)
Max. grootte	100x100 (px)

TABEL 4.1: Parameters gebruikt tijdens training en detectie van features met cascade classificatie als classificatie methode.

willekeurig gekozen negatieve achtergrond (uit de negatieve samples) werd vergroot tot 3000 samples. Deze transformaties helpen dus om een groter aantal positieve samples te bekomen maar ook om het effect van belichting en rotaties van legoblokken kleiner te maken.

Omdat de training erg lang kan duren werd deze maar eenmaal uitgevoerd per soort feature om toch een grote indicatie te geven van hoe lang zulke training duurt. De parameters die tijdens de training werden gebruikt zijn aangegeven in tabel 4.1.

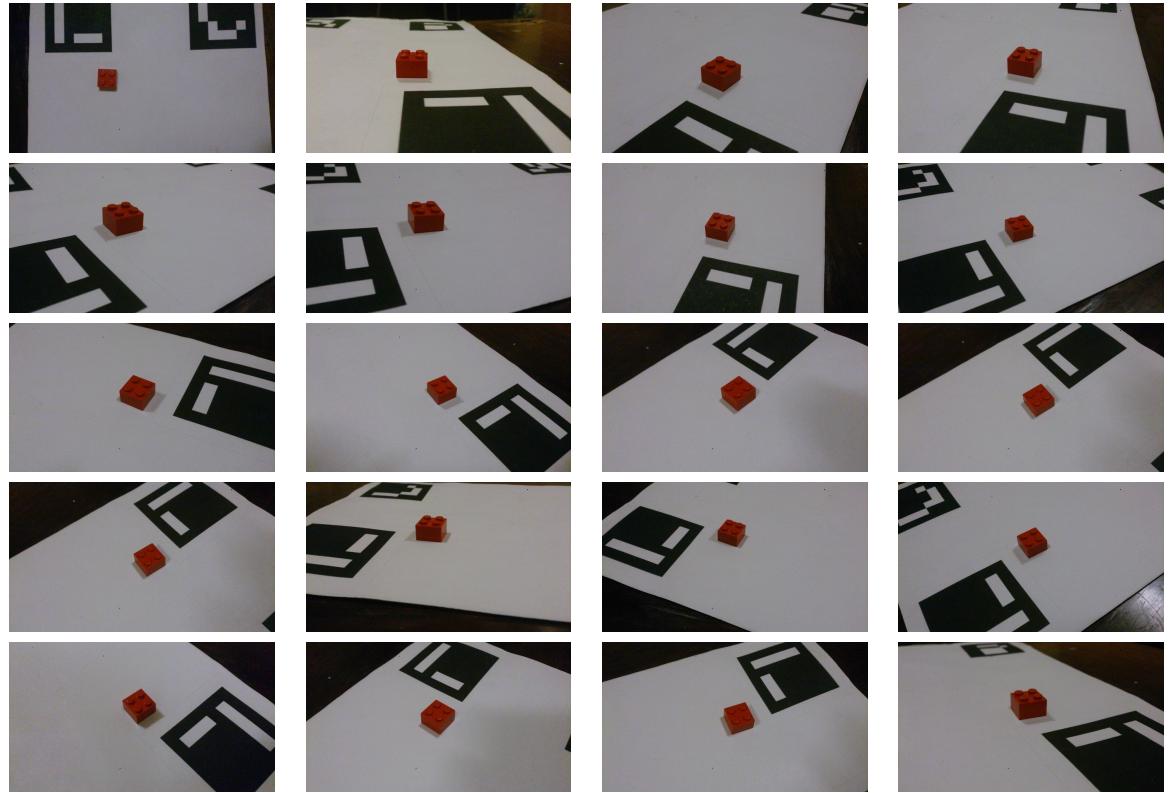
De training werd uitgevoerd op een machine met een 2.4 GHz CPU met 12 cores en 47GB RAM. Dit is niet dezelfde machine als gebruikt wordt tijdens de detectie omdat die machine veel te traag is om de training op uit te voeren. Bovendien is het niet nuttig om performantie van de training en detectie met elkaar te vergelijken omdat deze waarden ver uit elkaar liggen.

Detectie

De detectie gebeurt op 20 afbeeldingen van een legoblok in verschillende poses. Hierbij werden 3 standaardposes gekozen: bovenaanzicht, zijaanzicht en een aanzicht waarin de legoblok vanuit een hoekpunt wordt gezien. De rest van de poses zijn allemaal willekeurig gekozen. Deze afbeeldingen zijn te zien in figuur 4.4: de drie afbeeldingen in links in de eerste rij zijn de drie specifiek gekozen aanzichten, de rest zijn de willekeurige. Deze keuze van poses zal een idee geven over hoe robuust / performant de features zijn tegen wijzigingen in pose en welke poses robuuster / performanter zijn dan andere. De parameters die gebruikt werden tijdens de detectie zijn ook aangegeven in tabel 4.1.

Bij het vergelijken van features zal naar performantie en robuustheid worden gekeken. Voor de experimenten die focussen op performantie werd de detectie 10x uitgevoerd op alle 20 testafbeeldingen en per afbeelding het gemiddelde genomen.

4. LEGOLOKDETECTIE OP BASIS VAN CAD MODELLEN



FIGUUR 4.4: De 20 afbeeldingen die als testset werden gebruikt.

De experimenten werden uitgevoerd op een **machine** met een Intel Core 2 Duo 2,4 GHz processor, een NVIDIA GeForce 320M 256 MB grafische kaart, 8 Gb RAM en Mac OS X 10.10.2. Alle implementaties zijn geschreven in C++ en maken gebruik van OpenCV 2.4.9.

4.3.2 Vergelijking

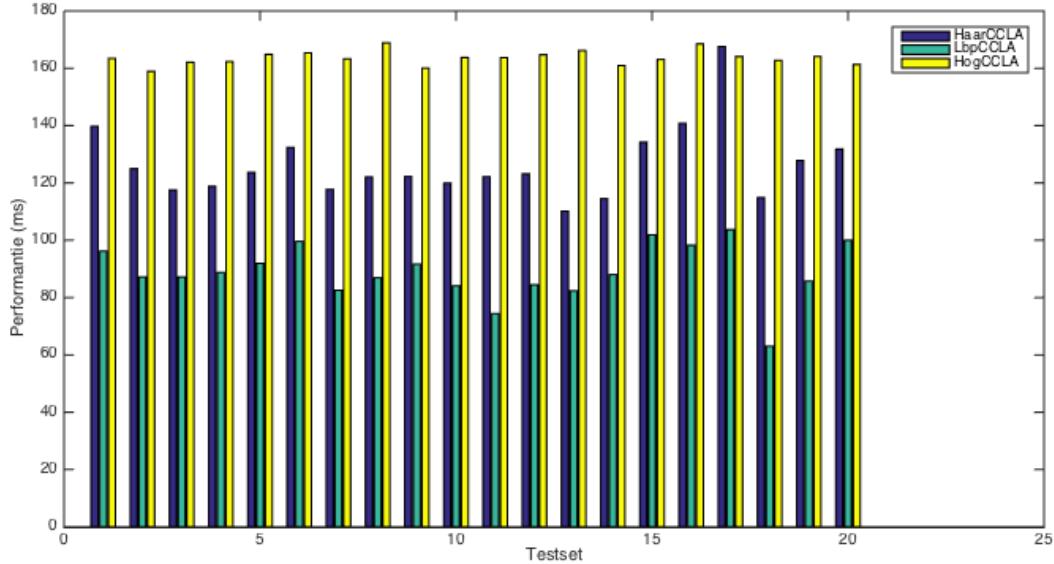
In deze sectie worden de eerste drie besproken feature types met elkaar vergeleken. Ze worden vergeleken qua performantie en robuustheid.

Performantie

De vergelijking van performantie tussen verschillende soorten features bestaat uit twee onderdelen: enerzijds de snelheid bij training van de classifier en anderzijds de snelheid bij detectie.

De duur van de training wordt aangegeven in tabel 4.2. Het is belangrijk op te merken dat tijdens de HOG training slechts 1 core wordt gebruikt terwijl de andere soorten alle CPU cores gebruikten, dit is inherent aan de implementatie van OpenCV. We kunnen er wel uit afleiden dat HOG waarschijnlijk snelle kan worden getraind dan MB-LBP indien alle cores werden gebruikt. Het opmerkelijkste is het groot

4.3. Evaluatie features



FIGUUR 4.5: Feature performantie tijdens detectie.

verschil is tussen de tijd nodig om een Haar-like classifier te trainen en tijd nodig om een LBP of HOG classifier te trainen. Dit valt als volgt te verklaren: bij het trainen van een Haar-like classifier wordt eerst voor elke window een enorme hoeveelheid aan features gegenereerd die vervolgens worden uitgeselecteerd met AdaBoost [FS95]. Deze hoeveelheid aan features is vele malen groter dan het aantal pixels in het window, in tegenstelling tot MB-LBP en HOG waarbij voor elk window enkel een aantal histogrammen worden opgesteld.

De performantie tijdens detectie van de 20 testafbeeldingen wordt weergegeven in figuur 4.5. Het is duidelijk dat MB-LBP de snelste is, dit komt omdat bij MB-LBP integerwaarden worden vergeleken terwijl bij de rest floating point getallen worden vergeleken. Haar-like is bovendien sneller dan HOG, wat valt te verklaren door het feit dat aan Haar-like een intensieve training vooraf gaat waarna specifieke features zijn geselecteerd die eenvoudig te berekenen zijn via een *integral image*. Bij HOG moet tijdens de detectie voor elk window een volledig histogram worden berekend wat rekenintensiever is dan het uitrekenen een beperkte set van eenvoudig te berekenen features.

Feature	Training	# cores
Haar-like	> 4 dagen	12
MB-LBP	$\pm 2\text{h}$	12
HOG	$\pm 3\text{h}$	1

TABEL 4.2: Feature training tijdsduur.

Robuustheid

Uit de resultaten van figuur 4.6 kan worden afgeleid dat detectie met behulp van HOG de beste resultaten geeft: geen enkele afbeelding uit de testset was vals positief en op 4 afbeeldingen uit de testset na werd de legoblok overal correct gevonden. Hoewel met MB-LBP alle legoblokken werden gevonden, vertoont het heel wat meer valse positieven. Haar-like ligt tussen beide in: het detecteert één legoblok meer als HOG maar heeft wel twee valse positieven, terwijl HOG er geen enkele heeft.

4.3.3 Besluit

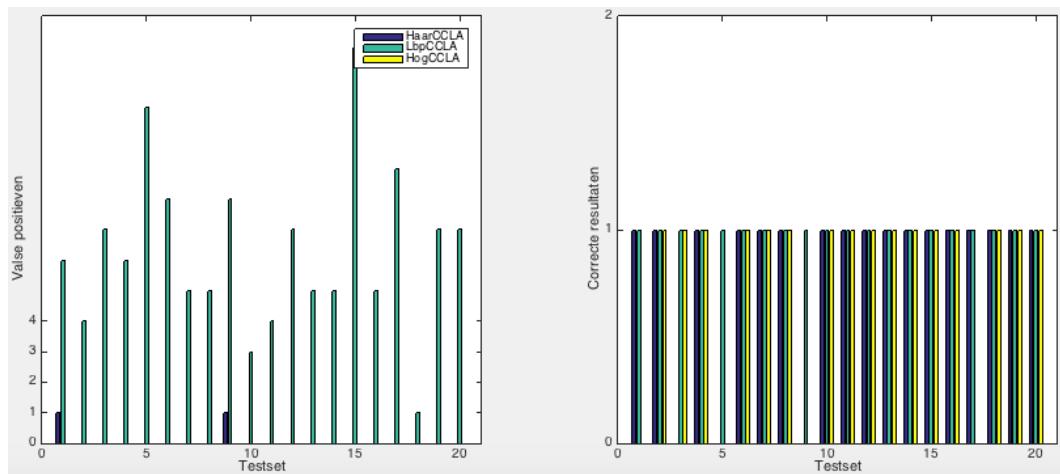
We kunnen besluiten dat HOG het meest robuuste feature type is maar dat MB-LBP een stuk sneller. Ook is belangrijk dat hoewel Haar-like features redelijk robuust en performant zijn, er erg veel tijd in kruipt om ze trainen.

4.4 Classificatie methodes

Deze sectie behandelt de classificatie methodes. De taak van de classificatie methodes is om op basis van berekende features een legoblok te detecteren. De volgende classificatie methodes besproken: Cascade Classifiers [VJ01] (CCLA) en Support Vector Machines [Joa99] (SVM).

4.4.1 CCLA

Bij een **CCLA** wordt een classifier getraind door een aantal feature classifiers te combineren. Elk van deze classifiers bestaat uit een aantal features op basis waarvan elke subwindow van een frame wordt geclasseerd. Zulke cascade classifier werkt door eerst elke subwindow van de frame te laten evalueren door de eerste classifier, indien de eerste classifier de subwindow aanvaardt triggert dit een evaluatie door de



FIGUUR 4.6: Feature robuustheid.

volgende classifier in de cascade. Enkel wanneer een window door alle classifiers is aanvaard, kan er vanuit worden gegaan dat dit window het te detecteren object bevat. Door een cascade van classifiers te gebruiken moet elke classifier van de cascade niet al te sterk zijn waardoor negatieve windows al sneller kunnen worden afgevoerd, wat de detectie des te sneller maakt.

4.4.2 SVM

Een **SVM** classificeert op een andere manier: elk subwindow wordt voorgesteld als een p-dimensionale vector in een p-dimensionale ruimte. Met behulp van (p-1)-dimensionale hyperplane worden de subwindows gesplitst in twee groepen: de ene groep wordt aanvaard, de andere niet. De *hit threshold* parameter bepaalt hoe ver een vector van de hyperplane moet liggen om te worden aanvaard of afgekeurd. Deze lineaire classifier was het oorspronkelijke algoritme [Vap63], later werd ook een methode ontwikkelt voor non-lineaire classificatie [BGV92].

4.5 Evaluatie classifier methodes

Deze sectie behandelt de evaluatie van de classifier methodes. Eerst wordt kort besproken hoe de evaluatie verloopt, vervolgens worden beide methodes (CCLA en SVM) vergeleken op vlak van robuustheid en performantie. Ten slotte wordt de impact bekeken van de *schaalfactor* parameter op robuustheid en performantie voor zowel CCLA als SVM.

4.5.1 Evaluatiemethode

De evaluatie van de verschillende classifier methodes gebeurt door elke soort classifier te trainen voor één bepaalde feature en vervolgens de geleerde classifiers te gebruiken voor detectie.

Training

Bij de vergelijking tussen de classificatie methodes werd steeds voor de HOG features (zie sectie 4.2.3) gekozen om geen invloed op de resultaten te krijgen door een andere keuze in features. Om de training uit te voeren werd gebruik gemaakt van het programma `opencv_traincascade` (van de OpenCv bibliotheek) om een CCLAclassifier te trainen en van de SVMLight bibliotheek [Joa99] om een SVM classifier te trainen. Als positieve samples werd de oorspronkelijke set van 128 afbeeldingen gebruikt en als negatieve samples werden dezelfde afbeeldingen gebruikt als bij de features (maar ze werden wel opgesplitst in tiles van 64x64 omdat de implementatie van SVMLight dat vereist).

Opnieuw werd training maar eenmaal uitgevoerd per classifier methode om een grofke indicatie te geven van hoe lang zulke training duurt. De parameters die tijdens de training werden gebruikt zijn aangegeven in tabellen 4.1 en 4.3 voor CCLA en SVM respectievelijk.

4. LEGOLOKDETECTIE OP BASIS VAN CAD MODELLEN

Parameter	Waarde
Training	
Padding (HOG)	0x0 (px)
Wind. stride (HOG)	8x8 (px)
Wind. size (HOG)	64x64 (px)
Detectie	
Schaalfactor	1.1
Hit threshold	0.4
Padding (HOG)	0x0 (px)
Wind. stride (HOG)	8x8 (px)
Wind. size (HOG)	64x64 (px)

TABEL 4.3: Parameters gebruikt tijdens training en detectie van features met SVM als classificatie methode.

De training werd opnieuw uitgevoerd op een machine met een 2.4 GHz CPU met 12 cores en 47GB RAM, vanwege dezelfde reden als bij de feature evaluatie.

Detectie

De detectie gebeurt, net als bij features, op 20 afbeeldingen van een legoblok in verschillende poses. Deze afbeeldingen zijn te zien in figuur 4.4: de drie afbeeldingen in rechts in de eerste rij zijn de drie specifiek gekozen aanzichten, de rest zijn de willekeurige. De parameters die gebruikt werden tijdens de detectie zijn ook aangegeven in tabellen 4.1 en 4.3.

Bij het vergelijken van classifier methodes zal naar performantie, robuustheid en de impact van de schaalfactor op de performantie en robuustheid worden gekeken. Voor de experimenten die focussen op performantie werd de detectie 10x uitgevoerd op alle 20 testafbeeldingen en per afbeelding het gemiddelde genomen. In het geval van experimenten met de schaalfactor wordt bij de performantie het gemiddelde genomen over alle testafbeeldingen.

De experimenten werden, net als bij de features, uitgevoerd op een machine met een Intel Core 2 Duo 2,4 GHz processor, een NVIDIA GeForce 320M 256 MB grafische kaart, 8 Gb RAM en Mac OS X 10.10.2. Alle implementaties zijn geschreven in C++ en maken gebruik van OpenCV 2.4.9 (zowel training als detectie).

4.5.2 Evaluatie

Robuustheid

Figuur 4.7 toont de robuustheid van de twee classifier methodes. Merk op dat de grafiek met valse positieven is weggelaten omdat deze in beide gevallen overal nul was. We merken op dat de robuustheid van de SVM hoger ligt dan die van de CCLA. De oorzaak hiervan dat SVM alle features gebruikt om die lineair te classificeren,

terwijl CCLA slechts een subset van alle features gebruikt om een zo klein mogelijke fout te maken op de classificatie.

Performantie

De vergelijking van performantie tussen verschillende soorten classificatie methodes bestaat uit twee onderdelen: enerzijds de snelheid bij training van de classifier en anderzijds de snelheid bij detectie.

De duur van de training wordt aangegeven in tabel 4.2. Dit geeft aan dat een SVM classifier een heel stuk sneller kan getraind worden dan een CCLA classifier. Dit komt omdat de SVM classifier features moet uitrekenen en een hyperplane bepalen tussen de features. CCLA, daarentegen, moet een hele cascade van classifiers opbouwen waarbij telkens opnieuw goede features moeten worden gekozen, dit vraagt heel wat meer rekenwerk.

De performantie tijdens detectie van de 20 testafbeeldingen wordt weergegeven in figuur 4.5. Dat de detectie langer duurt bij een SVM is logisch aangezien het een lineaire classifier is war vergeleken kan worden met één enkele classifier uit een cascade. Om een voldoende goede detectie te behalen moet een SVM dus erg correct kunnen classificeren, dit vergt meer tijd aangezien een cascade juist is ontwikkeld om efficiënter te zijn dan een enkele classifier.

Schaalfactor

De schaalfactor is een parameter die bepaalt hoe sterk de schaal van het window wordt gewijzigd tussen een minimum en maximum grootte. Dus wanneer deze kleiner wordt zullen normaal gezien meer iteraties nodig zijn, terwijl een nauwkeurigere detectie plaatsvindt.

Om de impact van de schaalfactor te bepalen, wordt de performantie en robuustheid van de HOG feature opgemeten wanneer de schaalfactor 1.01, 1.05, 1.10, 1.15, ... , 4 bedraagt. Uit figuur 4.9 is het duidelijk dat aan de verwachtingen voldaan is: de performantie (rechts) is hoger (het aantal ms daalt) bij een hogere schaalfactor, terwijl het aantal correcte detecties (links) daalt. Het aantal valse positieven werd ook gemeten maar deze was voor beide classifiers bijna altijd 0.

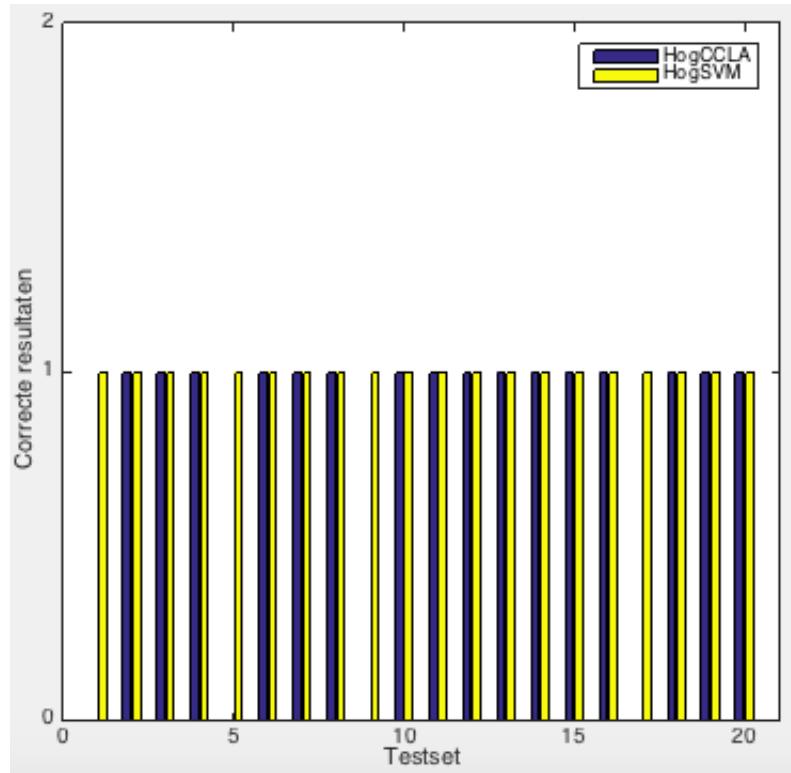
Bij de grafiek die het **aantal correcte detecties** toont zien we dat in het begin SVM robuuster is dan CCLA, daarna wisselt het wat af.

Dat het aantal correcte poses in het begin hoger ligt bij SVM dan bij CCLA terwijl zijn performantie over het algemeen hoger ligt kan op dezelfde manier verklaard worden als bij de performantie is geredeneerd. Vanwege het lineaire karakter van een

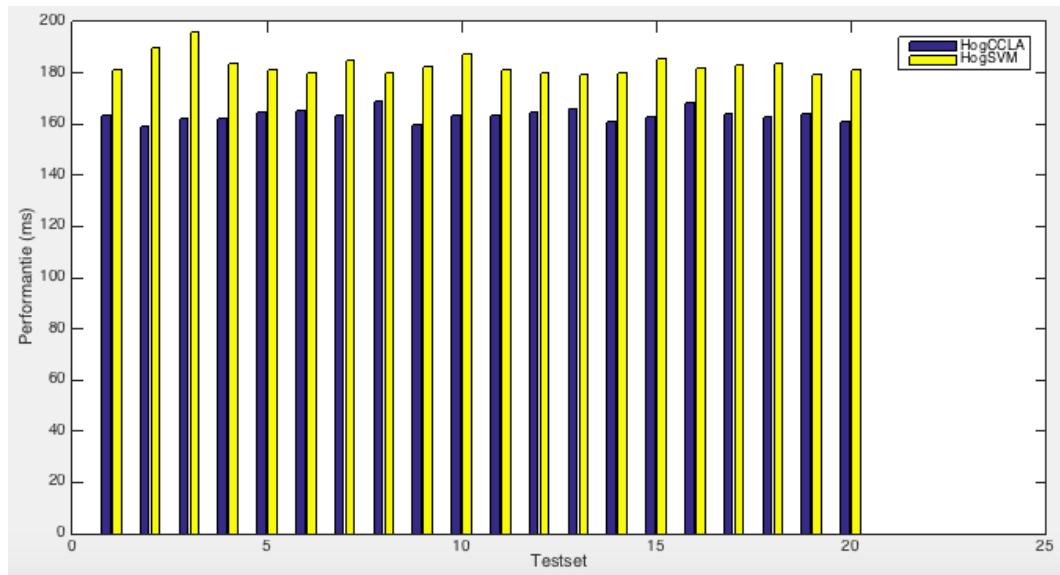
Feature	Training	# cores
CCLA	± 3h	1
SVM	TODO	1

TABEL 4.4: Classifier methodes training tijdsduur.

4. LEGOBLOKDETECTIE OP BASIS VAN CAD MODELLEN



FIGUUR 4.7: Robuustheid classifier methodes.



FIGUUR 4.8: Classifier methodes performantie tijdens detectie.

SVM is het dan ook logisch dat wanneer een hogere robuustheid wordt gehaald (in het begin) de performantie lager ligt.

Het wisselend karakter van de eerste grafiek ligt aan de aard van de parameter: indien de schaalfactor toevallig een waarde heeft waardoor het window even groot wordt als een bepaalde legoblok zal deze worden gedetecteerd. Merk op dat deze eerste grafiek slechts tot 1.65 wordt getoond, dit komt omdat het aantal correcte detecties hierna toch steeds 0 is. Dat het reeds bij 1.65 altijd 0 is, komt omdat wanneer de schaalfactor te hoog wordt slechts enkele verschillende groottes als window worden gebruikt. Indien de groottes van deze windows niet toevallig dicht genoeg liggen bij de grootte van de legoblok zal deze niet gedetecteerd worden.

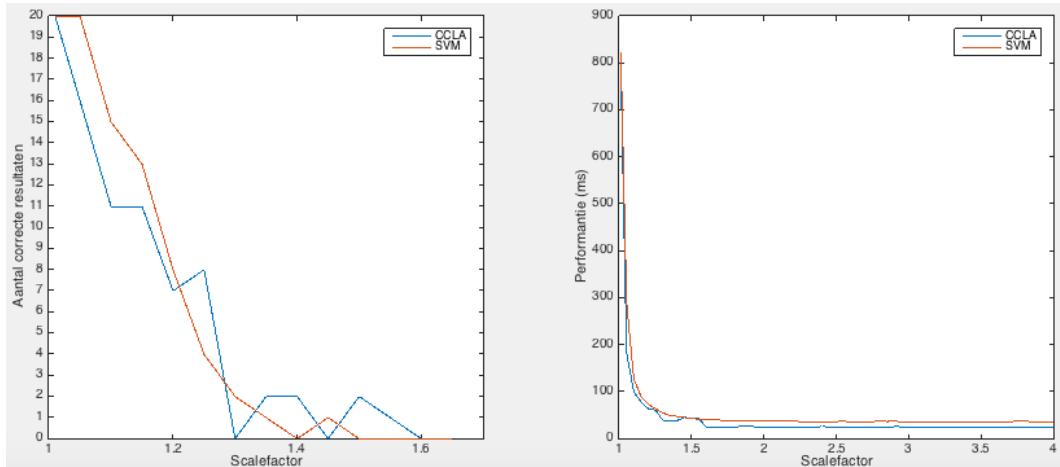
Opmerkelijk is de vorm van de grafiek die de **performantie** toont, het geeft aan dat beide classifier methodes een performantie hebben van de orde $O(1/(s - 1))$ met s de schaalfactor. Merk op dat de orde $1/(s - 1)$ is en niet $1/s$ omdat de verticale asymptoot moet liggen op $s = 1$ want als de schaalfactor 1 is duurt de detectie oneindig lang, het window wordt dan immers niet groter.

4.5.3 Besluit

In deze evaluatie werd aangetoond dat SVM een meer robuuste methode is dan CCLA, maar dat CCLA wel de meer performante is. Vervolgens werd aangetoond dat een hogere schaalfactor de performantie verhoogt maar de robuustheid verlaagt. Tenslotte werd duidelijk dat de performantie van beide methodes van orde $O(1/(s-1))$ is met s de schaalfactor.

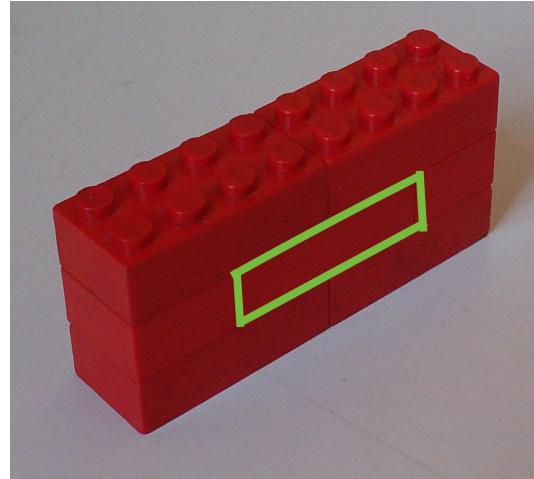
4.6 Discussie / Besluit

In dit hoofdstuk werd onderzocht welke feature types en classifier methodes beter werken voor detectie van legoblokken. Zo is aangetoond dat HOG de meer robuuste



FIGUUR 4.9: Impact schaalfactor op performantie en robuustheid van HOG feature.

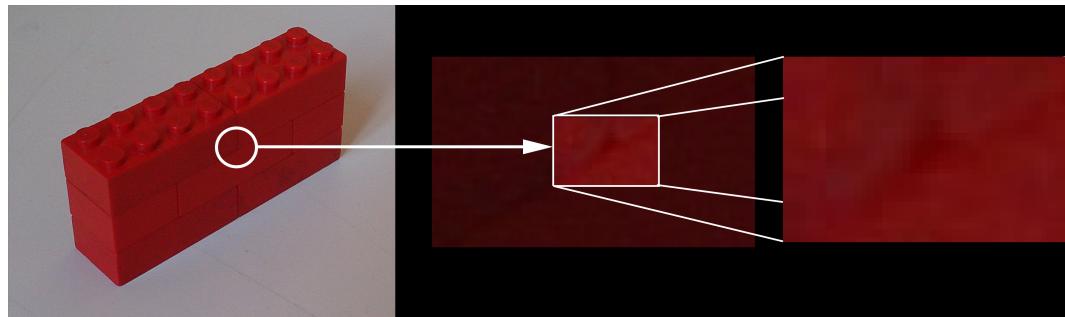
4. LEGOLOKDETECTIE OP BASIS VAN CAD MODELLEN



FIGUUR 4.10: Een voorbeeld van een legoblokconstructie waarin het groen omlijnde vlak het enige is wat we van die legoblok zien.

feature type is en dat MB-LBP de meest performante van de drie is. Verder werd aangegeven dat een vierde type feature (op basis van deformed parts-based modellen) te traag is om nuttig te gebruiken in een AR applicatie. Ten slotte toonde de vergelijking tussen classifier methodes aan dat detectie met een SVM classifier vaak trager is maar tegelijk ook meer robuust, dat een hogere schaalfactor leidt tot hogere performantie maar lagere robuustheid en dat de performantie van de classifier methodes van orde $1/(s-1)$ is met s de schaalfactor. De methode die in dit hoofdstuk werd onderzocht is echter niet nuttig om te gebruiken in een AR spel om legoblokken te detecteren en hier is waarom:

Hoewel het wel mogelijk is om een enkele legoblok te detecteren is het erg moeilijk om in een constructie van legoblokken de blokken afzonderlijk te detecteren. Dit resulteerde telkens in erg slechte resultaten, zelfs wanneer het beste type feature en de beste classificatie methode werd gekozen. Dit wordt hoofdzakelijk veroorzaakt door de enorme hoeveelheid aan occlusie die een legoblok in een constructie grotendeels



FIGUUR 4.11: De vergroting van een rand in een muur van legoblokken.

verbergt. Hierdoor is slechts een erg beperkt deel van de legoblok (in sommige gevallen slechts één vlak, zie figuur 4.10) zichtbaar wat het erg moeilijk maakt om met features een legoblok te detecteren. Zelfs indien op voorhand een optimale pose zou worden gekozen kan voor sommige legoblokken enkel één vlak zichtbaar zijn in de afbeelding.

Men zou dan kunnen zeggen dat op basis van de randen de verschillende legoblokken in de constructie kunnen worden gedetecteerd. Dit is echter niet waar omdat deze randen erg zwak zijn wanneer legoblokken met eenzelfde kleur naast elkaar staan. Indien we dit probleem van naderbij zouden bekijken bekomen we een resultaat zoals in figuur 4.11. In deze figuur is een deel van de afbeelding dat legoblokranden bevat tweemaal uitvergroot. Deze rand neemt slechts twee à drie pixels in beslag en verandert bovendien erg weinig van kleur. Dit maakt het enorm moeilijk, zo niet onmogelijk om de verschillende legoblokken op basis van hun randen te detecteren.

Deze redenering leidt tot een volgend resultaat: het is erg moeilijk, zo niet onmogelijk, om in een legoconstructie alle legoblokken apart te detecteren. Daarom moeten we proberen om, in plaats van de blokken individueel te proberen detecteren, ze als groter geheel te detecteren. Zulke legoconstructie kunnen we zien als een flexibele legoblok die eender welke hoogte, diepte en breedte kan hebben. Dat is exact waarvoor het deformed parts-based feature type gebruikt wordt maar zoals aangehaald in sectie 4.2.4 is deze methode te traag om te gebruiken in een mobiele AR applicatie. Dat is de reden waarom in de volgende sectie een sneller alternatief wordt voorgesteld dat gebruik maakt van het idee dat lego constructies moeten gedetecteerd worden in plaats van individuele legoblokken. Hierbij wordt echter wel afgestapt van het idee om CAD modellen te gebruiken.

Bijlagen

Bibliografie

- [AME⁺14] Mathieu Aubry, Daniel Maturana, Alexei A Efros, Bryan C Russell, and Josef Sivic. Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3762–3769. IEEE, 2014.
- [BGV92] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.
- [DT05] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [FGMR10] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.
- [FS95] Yoav Freund and Robert E Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.
- [HSKM11] Carlos Hernández, Xiaoxun Sun, Sven Koenig, and Pedro Meseguer. Tree adaptive a*. In *The 10th International Conference on Autonomous Agents and Multiagent Systems- Volume 1*, pages 123–130. International Foundation for Autonomous Agents and Multiagent Systems, 2011.
- [Joa99] Thorsten Joachims. Svmlight: Support vector machine. *SVM-Light Support Vector Machine <http://svmlight.joachims.org/>, University of Dortmund*, 19(4), 1999.
- [LM02] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–900. IEEE, 2002.

BIBLIOGRAFIE

- [LZL⁺07] Shengcai Liao, Xiangxin Zhu, Zhen Lei, Lun Zhang, and Stan Z Li. Learning multi-scale block local binary patterns for face recognition. In *Advances in Biometrics*, pages 828–837. Springer, 2007.
- [OPH96] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.
- [S⁺85] Satoshi Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.
- [Vap63] Vladimir Vapnik. Pattern recognition using generalized portrait method. *Automation and remote control*, 24:774–780, 1963.
- [VJ01] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.
- [YLWL14] Junjie Yan, Zhen Lei, Longyin Wen, and Stan Z Li. The fastest deformable part model for object detection. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2497–2504. IEEE, 2014.

Fiche masterproef

Student: Wouter Franken

Titel: Augmented Reality Bordspellen

Engelse titel: Augmented Reality Boardgames

UDC: 681.3

Korte inhoud:

Hier komt een heel bondig abstract van hooguit 500 woorden. LATEX commando's mogen hier gebruikt worden. Blanco lijnen (of het commando) zijn wel niet toegelaten!

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Thesis voorgelegd tot het behalen van de graad van Master of Science in de ingenieurswetenschappen: computerwetenschappen, hoofdspecialisatie Mens-machine communicatie

Promotor: Prof. dr. ir. Philip Dutre

Assessor: TODO

Begeleider: Ir. J. Baert