

Visie voor semantische robotnavigatie in ziekenhuisgangen

Olivier VAN DEN EEDE

Promotor(en): Prof. dr. ir. Toon Goedemé

Co-promotor(en): Filip Reniers

Masterproef ingediend tot het behalen van
de graad van master of Science in de
industriële wetenschappen: Electronica-ICT
afstudeerrichting ICT

Academiejaar 2018 - 2019



©Copyright KU Leuven

Zonder voorafgaande schriftelijke toestemming van zowel de promotor(en) als de auteur(s) is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, kan u zich richten tot KU Leuven Technologicampus De Nayer, Jan De Nayerlaan 5, B-2860 Sint-Katelijne-Waver, +32 15 31 69 44 of via e-mail iiw.denayer@kuleuven.be.

Voorafgaande schriftelijke toestemming van de promotor(en) 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

Het voorwoord vul je persoonlijk in met een appreciatie of dankbetuiging aan de mensen die je hebben bijgestaan tijdens het verwezenlijken van je masterproef en je hebben gesteund tijdens je studie.

Samenvatting

De (korte) samenvatting, toegankelijk voor een breed publiek, wordt in het Nederlands geschreven en bevat **maximum 3500 tekens**. Deze samenvatting moet ook verplicht opgeladen worden in KU Lokaal.

Abstract

Het extended abstract of de wetenschappelijke samenvatting wordt in het Engels geschreven en bevat **500 tot 1.500 woorden**. Dit abstract moet **niet** in KU Loket opgeladen worden (vanwege de beperkte beschikbare ruimte daar).

Keywords: Voeg een vijftal keywords in (bv: Latex-template, thesis, ...)

Inhoudsopgave

Voorwoord	iii
Samenvatting	iv
Abstract	v
Inhoud	vi
Figurenlijst	vii
Tabellenlijst	viii
Lijst met afkortingen	ix
1 Literatuurstudie	1
1.1 Indoor navigatie & visie	1
1.2 Object detectie	1
1.2.1 Traditionele object detectie	1
1.2.2 Convolutional neural network	2
1.3 Object tracking	3
1.4 Image segmentation	3
A Uitleg over de appendices	7

Lijst van figuren

1.1	De lagen van een CNN volgens het YOLO [7] detection system.	2
1.2	Het SegNet [2] segmentatie netwerk.	4

Lijst van tabellen

Lijst van afkortingen

CNN	Convolutional Neural Network
RGB	Rood Groen Blauw
HSI	Hue Saturation Intensity
HOG	Histogram of Oriented Gradients
SVM	Support Vector Machine
SIFT	Scale-invariant feature transform
RANSAC	Random sample consensus

Hoofdstuk 1

Literatuurstudie

1.1 Indoor navigatie & visie

Op visie gebaseerde navigatie is een onderwerp dat zeer vaak onderzocht wordt.

1.2 Object detectie

Een belangrijk aspect van dit onderzoek is het detecteren van individuele objecten in het beeld van 1 enkele RGB camera. De te detecteren objecten zijn op voorhand vastgelegd, en zijn afhankelijk van de ruimte waarin de robot zich bevindt.

In de logistieke gangen van een ziekenhuis zijn er heel wat objecten te zien die we kunnen detecteren, een kleine selectie van deze objecten zijn.

- Pictogrammen
- Brandblussers
- Deurklinken

Voor deze objecten gaan we kijken naar detectie technieken uit de traditionele beeldverwerking, en naar meer *state of the art* technieken.

1.2.1 Traditionele object detectie

In openbare gebouwen zijn er heel wat pictogrammen te vinden zoals nooduitgang, hoogspanning en brandblusser. Deze pictogrammen hebben steeds een specifieke vorm, kleur en symbool. De literatuur leert ons weinig over pictogramdetectie, maar pictogrammen kunnen wel vergeleken worden met verkeersborden die bijna dezelfde kenmerken hebben. De aanpak van [4] is om 2 soorten features in een beeld te onderscheiden. In eerste instantie detecteren ze vormen op basis van kleur randen en anderzijds wordt de afbeelding omgezet naar HSI waaruit enkel de hue gebruikt wordt.

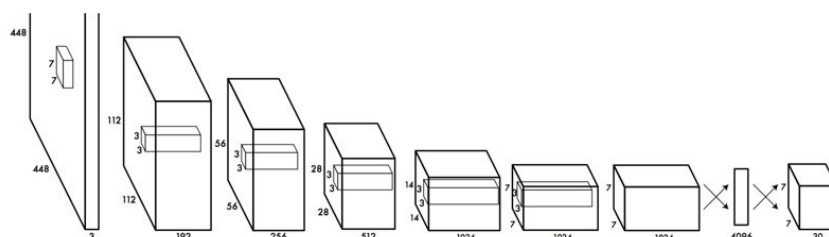
De hue is de belangrijkste component voor het onderscheiden van kleuren omdat er zo geen rekening wordt gehouden met de hoeveelheid licht en schaduwen. Een recenter onderzoek [10] bouwt voort op deze technieken, maar berekenen de Histogram of Oriented Gradients (HOG) features van het beeld. Vervolgens wordt er gebruik gemaakt van een Support Vector Machine (SVM) om te bepalen waar er zich een match bevindt.

De vorm en kleur features kunnen dan gecombineerd worden om de plaats voor een mogelijke match te vinden. Eens er een mogelijke bounding box gevonden is, kan er geprobeerd worden een template te matchen om het effectieve pictogram te achterhalen. Het grootste probleem bij de techniek van [4] is dat hun gebruikte template matching techniek niet robuust is voor schaal invariantie. Bij [10] maken ze voor de herkenningfase gebruik van SIFT[6] features en kleur informatie. Hierdoor is het probleem van schaal invariantie grotendeels opgelost. Hierbij worden de SIFT features van de kandidaat matches en de templates vergeleken, en er wordt een gemiddelde genomen van de verschillen tussen hue, saturation en value. Door middel van RANSAC en een threshold wordt er bepaald welke matches gebruikt worden. Deze techniek zou gebruikt kunnen worden voor het detecteren van pictogrammen.

1.2.2 Convolutional neural network

De laatste jaren in het domein van beeldverwerking wordt er steeds meer gegrepen naar deep learning technieken. Dit is komt omdat rekenkracht steeds beter en beter wordt, en de resultaten die bekomen worden de traditionele manieren overtreffen op verschillende vlakken. Een deep learning techniek die veel gebruikt wordt in de beeldverwerking is een CNN.

Een convolutional neural network of CNN is een supervised deep learning techniek die gebruikt kan worden om complexere beeldinterpretatie te doen. Een CNN kan bestaan uit meerdere lagen die meestal een combinatie zijn van 'convolutional-layers' en 'fully connected-layers'. Elk van deze lagen bevat een aantal neuronen met elk een eigen set van gewichten. Het doel van een CNN is om de gewichten zodanig bij te stellen zodat data die aan de eerste laag gegeven wordt een verwacht resultaat geeft aan de laatste laag. Deze laatste laag kan men de classificatielaag noemen, en geeft een representatie van wat het netwerk denkt dat er aan de input staat. In figuur 1.1 is een voorbeeld te zien van een CNN met de verschillende soorten lagen.



Figuur 1.1: De lagen van een CNN volgens het YOLO [7] detection system.

Een 'convolutional-layer' is een laag die een convolutie operatie uitvoert op zijn input, de convolutie gebeurt d.m.v een masker dat meestal voorgesteld wordt als een tensor. Door een tensormasker

te gebruiken kan de operatie uitgevoerd worden op meerdere inputdimensies tegelijkertijd, denk hierbij aan bijvoorbeeld 3 kleurkanalen.

Om uiteindelijk een classificatie te verkrijgen moet er een dimensievermindering doorgevoerd te worden, dit wordt gedaan door 'pooling layers' aan het netwerk toe te voegen na elke convolutie laag.

Een CNN kan pas gebruikt worden nadat het getraind is. Voor de training van een netwerk zijn er 2 dingen noodzakelijk, veel voorbeeld data en per voorbeeld de verwachte output (label). Bij het trainingsproces wordt alle inputdata aangelegd, en wordt er gekeken wat het netwerk aan zijn output heeft. De loss functie is een maat van hoe goed een netwerk een voorspelling kan doen van de input data, met andere woorden een vergelijking tussen de input en de output. Het doel van de training van een netwerk is het minimaliseren van deze loss functie. Dit kan gedaan worden d.m.v 'backpropagation'. Backpropagation is het steeds een klein beetje aanpassen van de gewichten in de inwendige neuronen om zo het resultaat te verbeteren en de loss functie te verkleinen. Een netwerk heeft een goede training gehad als de loss functie minimaal is.

Een voorbeeld van een CNN is het 'YOLO(You Only Look Once) detection system' [7]. Het YOLO netwerk is opgebouwd uit 24 convolutielagen en 2 fully connected layers. Dit netwerk heeft een uitgebreide training gehad op de ImageNet dataset en kan gebruikt worden om object detectie en classificatie te doen door 1 keer de input afbeelding door het netwerk te laten gaan. Door middel van een hertraining kan deze detector leren om alle objecten te detecteren en te classificeren en dus een mogelijke detector zijn voor onze toepassing.

1.3 Object tracking

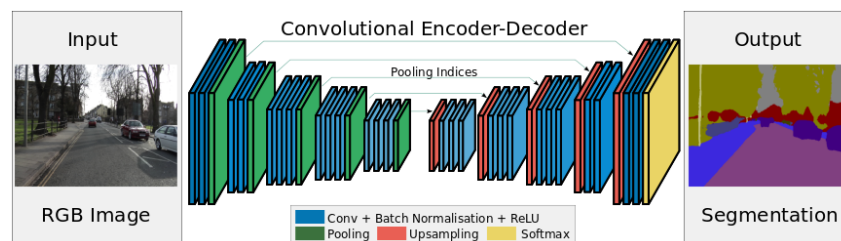
1.4 Image segmentation

Het correct segmenteren van de beelden zal een belangrijke rol spelen. Niet in elk beeld zal er een distinctief object aanwezig zijn om te detecteren. Daarom is het belangrijk om de vloer van de muren te kunnen onderscheiden. Een eenvoudige approach zou kunnen zijn om via K-means een verdeling van een beeld te doen en met een soort regressie de regio's te labelen. Volgens [11] werkt de K-means aanpak met een op textuur en kleur gebaseerde aanpak redelijk goed, maar wordt steeds de muur verbonden met het plafond omwille van kleur en textuur gelijkenissen. Hun regressie gebaseerde labeling techniek blijkt echter een slechte oplossing. Verder zoals [5] aangeeft zijn reflecties en overbelichting eigenschappen van indoor omgevingen die het moeilijk kunnen maken om een correcte segmentatie te doen. [5] stelt een techniek voor die begint met het detecteren van verticale en horizontale lijn segmenten. Dit doen ze door eerst een Canny edge detector[3] toe te passen en vervolgens een line fitting. Een zelf geleerde SVM classifier verdeelt alle lijnsegmenten in 2 categorieën namelijk horizontaal en verticaal. De vluchtlijnen van de gang worden hierbij onderverdeeld in de horizontale categorie. Alle lijnstukken krijgen een score via een reeks van operaties waarna enkel de beste lijnen bijgehouden worden. Op basis van de kleur van de vlakken tussen de lijnstukken kan een segmentatie gemaakt worden. Dit geeft een resul-

taat waarbij de vloer meestal een mooi homogeen geheel is, maar de muren worden in meerdere vlakken gesegmenteerd door eventuele kleurverschillen en objecten aan de muur.

Een andere manier om de vloer te segmenteren is voorgesteld in [8]. Zij doen een superpixel segmentatie volgens het SLIC algoritme [1], vervolgens bekijken ze de randen van de superpixels. Na observaties blijkt dat de randen van superpixels onregelmatig worden bij objectovergangen. Door het aanduiden van een paar vloerpixels kan hun algoritme superpixels aanduiden die tot de vloer behoren. Deze aanpak geeft een goede schatting van vrije ruimte op de vloer, maar is minder bruikbaar voor segmentatie van muren.

Een meer recente technologie om afbeeldingen te segmenteren is gebruik te maken van een CNN. Het netwerk voor segmentatie is verschillend van een traditioneel CNN voor bijvoorbeeld object detectie. Een voorbeeld van een segmentatienetwerk is te zien in figuur 1.2.



Figuur 1.2: Het SegNet [2] segmentatie netwerk.

Het segmentatienetwerk SegNet [2] is een combinatie van convolutielagen en pooling layers, er zijn geen fully connected layers aanwezig zoals het geval is bij een classificatie netwerk. De bedoeling van het SegNet netwerk is om als output opnieuw een afbeelding te genereren, daarom zijn de lagen opgebouwd als een zandloper, op deze manier is de output even groot als de oorspronkelijke afbeelding. Een segmentatienetwerk wordt getraind op gelijkaardige manier aan een traditioneel CNN met als verschil dat de labeling gebeurt op pixelbasis aangezien de output even groot is als de input van het systeem. De output van het systeem is een per pixel gelabelde afbeelding afhankelijk van het aantal classes waarmee het systeem getraind is.

Het SegNet netwerk is getraind op de SUN RGB-D [9] dataset. Deze dataset bevat een groot aantal indoor scenes, waarbij er onder andere segmentatie klassen zijn voor muren, vloeren en plafonds. De training is gebeurd met enkel de RGB gegevens van de dataset. Deze trainingsdata zou uiteraard nuttig kunnen zijn voor dit onderzoek.

Bibliografie

- [1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Š̂¼ssstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, Nov 2012.
- [2] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, and Senior Member. SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. pages 1–14.
- [3] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986.
- [4] Chiung Yao Fang, Sei Wang Chen, and Chiou Shann Fuh. Road-sign detection and tracking. *IEEE Transactions on Vehicular Technology*, 52(5):1329–1341, 2003.
- [5] Yinxiao Li and Stanley T. Birchfield. Image-based segmentation of indoor corridor floors for a mobile robot. *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, pages 837–843, 2010.
- [6] D G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157 vol.2, sep 1999.
- [7] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.
- [8] F. Geovani Rodríguez-Telles, L. Abril Torres-Méndez, and Edgar A. Martínez-García. A fast floor segmentation algorithm for visual-based robot navigation. *Proceedings - 2013 International Conference on Computer and Robot Vision, CRV 2013*, pages 167–173, 2013.
- [9] Shuran Song, Samuel P. Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [10] S. J. Zabihi, S. M. Zabihi, S. S. Beauchemin, and M. A. Bauer. Detection and recognition of traffic signs inside the attentional visual field of drivers. *IEEE Intelligent Vehicles Symposium, Proceedings*, (lv):583–588, 2017.

- [11] Zhong-Ju Zhang. Wall, floor, ceiling, object region identification from single image.

Bijlage A

Uitleg over de appendices

Bijlagen worden bij voorkeur enkel elektronisch ter beschikking gesteld. Indien essentieel kunnen in overleg met de promotor bijlagen in de scriptie opgenomen worden of als apart boekdeel voorzien worden.

Er wordt wel steeds een lijst met vermelding van alle bijlagen opgenomen in de scriptie. Bijlagen worden genummerd met een drukletter A, B, C,...

Voorbeelden van bijlagen:

Bijlage A: Detailtekeningen van de proefopstelling

Bijlage B: Meetgegevens (op USB)

FACULTEIT INDUSTRIËLE INGENIEURSWETENSCHAPPEN
CAMPUS DE NAYER SINT-KATELIJNE-WAVER
J. De Nayerlaan 5
2860 SINT-KATELIJNE-WAVER, België
tel. + 32 15 31 69 44
iiw.denayer@kuleuven.be
www.iw.kuleuven.be

