

Werking van RealSense/Kinect/3D camera

Intel RealSense ZR300

In de eerste paar weken van het onderzoek heeft het onderzoeksteam de mogelijkheden onderzocht van de Intel RealSense ZR300 camera. Dit is een camera die verschillende beelden zoals videobeeld, dieptebeelden en infraroodbeelden kan opnemen. Er is voor deze camera gekozen omdat in de robot Pepper soortgelijke technologie zit die gebruik kan maken van dezelfde software als de Intel RealSense camera.

Tijdens het vooronderzoek naar de camera was één van de problemen die al snel naar voren kwam het feit dat Intel de software voor 3D camera's nog volop aan het ontwikkelen is en dat huidige versies van camera's niet meer ondersteunt en ontwikkelt worden. Zo staat op de website van Intel onder andere de volgende informatie over de SDK's en drivers.

Discontinuation Notice

The Intel® RealSense™ SDK for Windows*, the SDK components, and depth camera managers for the F200, SR300, and R200 versions will no longer be updated. You may continue to use the SDK with limited support, or use the Intel® RealSense™ Cross Platform API for camera access, and then develop on other platforms via [GitHub*](#).

For the Intel® RealSense™ SDK 2.0—our next generation SDK—support will only be available for Intel® RealSense™ cameras SR300 and D400-Series through GitHub.

Zoals zichtbaar, staat de ZR300 versie hier niet tussen. Dit komt omdat deze versie van de 3D camera's precies tussen de oudere versies en de nieuwere versies van de beschikbare camera's valt. Omdat Intel op dit moment gestopt is met het ondersteunen van software voor oudere camera types is er voor de ZR300 nooit een driver voor windows beschikbaar gekomen. Dit betekent dat de bestaande RealSense SDK voor Windows niet gebruikt kan worden.

Algoritme onderdelen SDK (voor camera's F200 en SR300)

- 3D Scanning
- Face Recognition
- Hand Tracking
- User Segmentation
- Cursor Mode

Intel RealSense Cross Platform API

Dit project is een cross platform library voor het opnemen van data met de verschillende realsense data. De library kan worden gebruikt met Windows, Linux en OS X. Deze library is gebruikt om in de eerste fases van het onderzoek beelden op te nemen met de gekochte ZR300 camera.

Kinect

Kinect is een bekende 3D camera die vooral wordt gebruikt in de game industrie met de XBOX game console. In tegenstelling tot de RealSense is de Kinect volledig ondersteunt met SDK en drivers op Windows (duh.. “Microsoft Kinect”) maar moet er voor ontwikkeling op Linux of OS X gebruikt gemaakt worden van een open source library zoals OpenKinect.

Kinect for Windows SDK key features en voordelen

- Improved body, hand and joint orientation
- Windows Store support
- Unity Pro support
- Powerful tooling
- Advanced facial tracking
- Simultaneous multi-app support

Waar het onderzoeksteam vooral in geïnteresseerd is, is de eerste feature “Improved body, hand and joint orientation”. Dit houdt in dat de SDK functionaliteit bevat waarmee het skelet van een persoon kan worden geconstrueerd door middel van de 3D beelden van Kinect.

With the ability to track as many as six people and 25 skeletal joints per person—including new joints for hand tips, thumbs, and shoulder center—and improved understanding of the soft connective tissue and body positioning, you get more anatomically correct positions for crisp interactions, more accurate avateering, and avatars that are more lifelike.

Beschrijving SDK functie skeleton - website Microsoft

Hoe werkt de skeleton functie van de kinect?

<https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/BodyPartRecognition.pdf>
<http://pages.cs.wisc.edu/~ahmad/kinect.pdf>

Note: informatie lijkt al wat ouder te zijn (2011), dus grote kans dat er ondertussen al weer wat verbeteringen zijn doorgevoerd. Zo lijkt techniek waarmee diepte berekent wordt te zijn veranderd van ‘structured light’ naar ‘time of flight’

Samengevat:

- Afleiden van een lichaamspositie is een proces wat bestaat uit twee stappen
 - Eerst wordt een depth map berekend met gebruik van structured light
 - Daarna wordt de lichaamspositie afgeleid door middel van machine learning

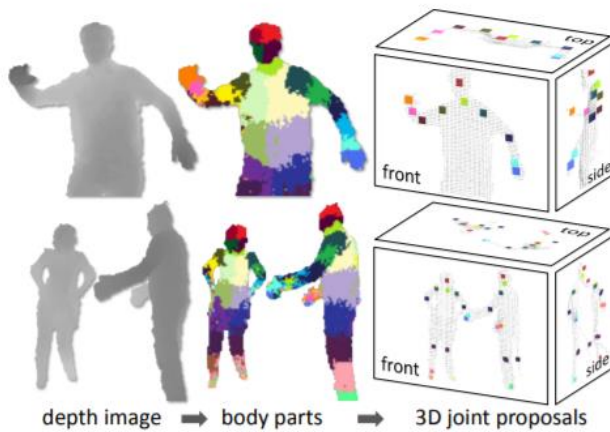
Bepalen van depth map

- Diepte map geconstrueerd door analyse van een spikkel patroon van een infrarood laserlicht (techniek ontwikkeld door PrimeSense, Microsoft heeft een licentie bij dat bedrijf)
- Deze techniek waarbij een bekend patroon wordt geanalyseerd wordt ook wel 'structured light' genoemd
- Wordt geprojecteerd op omgeving, de vervorming verteld wat over de diepte



Afleiding van lichaamsonderdelen

- De lichaamsonderdelen worden afgeleid door gebruik van een 'randomized decision forest', aangeleerd met meer dan 1 miljoen training voorbeelden.
- Deze stap start met 100.000 depth images waarvan het skeleton bekend is (opgenomen met een motion capture systeem)
- Voor elke echte afbeelding worden een aantal extra afbeelding gegenereerd door gebruik van grafische technieken
- Leer een randomized decision forest met 1 miljoen afbeeldingen, diepte afbeeldingen wordt naar lichaamsdelen gemapt



Resultaten



Microsoft denkt(of heeft al gedaan) ook na over verbeteringen:

- Weggooien van “slechte skeletons” en alleen gebruiken van “goede”.



likely



unlikely

Multiple Kinects

<https://msdn.microsoft.com/en-us/library/dn188677.aspx>

Lijkt mogelijk te zijn zolang meerdere kinects (of infrarood sources) niet in dezelfde field of view richting staan.

Waarom past de Kinect beter bij het onderzoek dan de RealSense?

Het verschil tussen de Kinect en de RealSense vat zich op dit moment samen in de volgende punten

- Kinect wordt, op dit moment, beter(professioneler) ondersteunt met software dan de RealSense. Dit kan in de toekomst eventueel veranderen.
- Kinect heeft functionaliteit in de SDK om skeletons uit te lezen, RealSense heeft dit (nog) niet

Omdat het onderzoek een half jaar duurt, is het voor het onderzoeksteam niet mogelijk om de functionaliteit voor een skeleton herkenning zelf te gaan herschrijven. Voor het onderzoek is het van belang dat een skeleton kan worden uitgelezen met een bestaand algoritme. (Die zijn wel te vinden). De volgende oplossingen zijn mogelijk het onderzoeken waard:

- Kan de Microsoft SDK gebruikt worden met de RealSense camera beelden?
- Zijn er open source algoritmes die om kunnen gaan met de beelden van de RealSense?
 - Afgelopen weken hebben we geen voortgang geboekt met het aan de praten krijgen van een mogelijke kandidaat voor het open source algoritme (OpenNI)
- Kinect kopen en de SDK daarvan gebruiken

Op dit moment heeft het onderzoeksteam de keuze gemaakt om van de Kinect gebruik te maken met de volgende redeneringen:

- Kinect + Adapter kost totaal 130 euro, dus relatief goedkoop om aan te schaffen
- In bestaande cases/onderzoeken wordt vooral gewerkt met Kinect.
- Uit dit bestaand onderzoek is gebleken dat accuraatheid van Kinect hoog genoeg is.

Bijkomende voordeel is dat door de effectieve ondersteuning vanuit Microsoft op het gebied van software, ontwikkelen met de Kinect waarschijnlijk ook behoorlijk makkelijk is.

Conclusie voor het verloop van het onderzoek:

[Wat betekent deze keuze in het licht van het verloop van het onderzoek? We gaan dus wellicht wel iets kunnen zeggen over wat er nodig is om een 'Pepper-like' robot in de toekomst te kunnen gaan gebruiken. Even bespreken samen??]

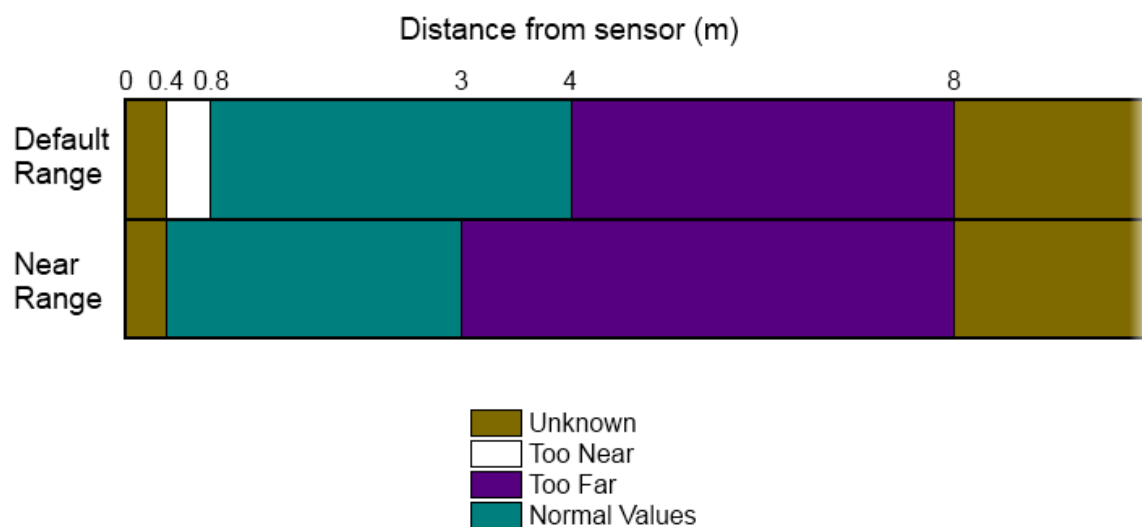
Technical Hardware Specs

Wat betreft de hardware specificaties lijkt er geen reden te zijn waarom het theoretisch onmogelijk is om vergelijkbare resultaten te halen met een RealSense camera.

Kinect

<https://msdn.microsoft.com/en-us/library/jj131033.aspx>

Kinect	Specification
Viewing angle	43° vertical by 57° horizontal field of view
Vertical tilt range	±27°
Frame rate (depth and color stream)	30 frames per second (FPS)
DepthImageFormat	The resolution is 320 x 240; 640 x 480; 80 x 60;



RealSense ZR300

<https://click.intel.com/media/ZR300-Product-Datasheet-Public-002.pdf>

Kinect	Specification
IR Laser Projector FOP	Diagonal: $80^{\circ} \pm 5\%$, Vertical: $60^{\circ} \pm 5\%$, Horizontal: $60^{\circ} \pm 5\%$
Infrared Camera FOV	Diagonal: $70^{\circ} \pm 5\%$, Vertical: $46^{\circ} \pm 5\%$, Horizontal: $59^{\circ} \pm 5\%$
Color Camera FOV	Diagonal: $75^{\circ} \pm 4\%$, Vertical: $41.5^{\circ} \pm 2\%$, Horizontal: $68^{\circ} \pm 2\%$
Fisheye Camera FOV	Diagonal: $166.5^{\circ} \pm 4\%$, Vertical: $100^{\circ} \pm 3\%$, Horizontal: $133^{\circ} \pm 3\%$
Depth Video Formats	628x468, 480x360, 320x240
Frame Rates	30, 60
Depth capture range	0.55m to 2.8m

