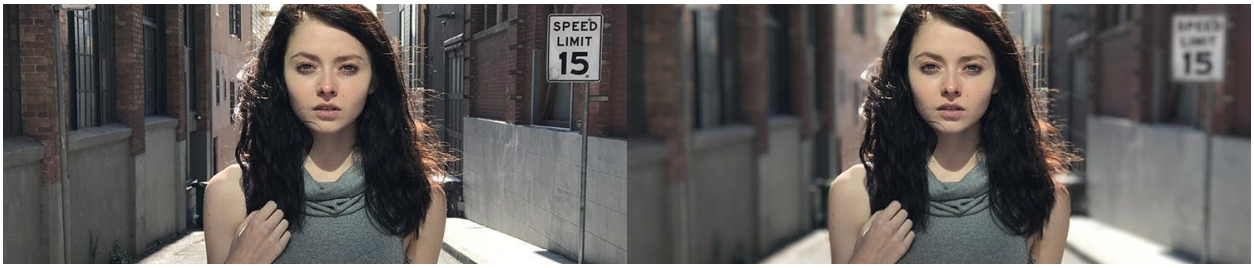
SYNTHETISCHE SCHERPTEDIEPTE, HOE WERKT DAT PRECIES?

Video Technieken 3

Tom Gunst

Scherptediepte is de afstand tussen de dichtstbijzijnde en verste punten ten opzichte van de camera die acceptabel scherp worden afgebeeld. Dit wordt beïnvloed door de grote van de sensor, brandpuntsafstand, scherpstelafstand en de diafragma opening waarbij een grote opening een ondiepe scherptediepte geeft en een klein diafragma een diepe scherptediepte veroorzaakt. Een ondiepe scherptediepte is een belangrijke esthetische eigenschap binnen de portretfotografie omdat het zijn onderwerp mooi kan isoleren van een anders vaak rommelige, afleidende achtergrond. Dit is echter moeilijk te bekomen met smartphone camera's vanwege zijn kleine sensor, vast diafragma en groothoeklens. Toch lanceerde Apple in 2016 "portrait mode" Voor de iphone 7 plus. Met deze functie kunnen nu ook amateurs foto's trekken met hun smartphone waarbij het scherptediepte effect van professionele camera's wordt nagebootst. Dit heet synthetische scherptediepte. In deze paper onderzoek ik hoe dit precies in zijn werk gaat.



Aangezien een smartphone enkel bezit over digitale pixel data van een tweedimensionaal plaatje heeft een computer niet zomaar de capaciteit objecten te herkennen nog de ruimtelijke indeling ervan te bepalen. Toch is het belangrijk dat de smartphone een driedimensionale voorstelling van de gefotografeerde scène bezit om te weten hoeveel bepaalde pixels geblurred moeten worden en welke niet. Dit doen ze aan de hand van "depth maps". Deze depth maps, ofwel "disparity maps" worden door middel van "computer stereo vision" achterhaald. Dit is het digitaal equivalent van het biologische principe "stereopsis", waarmee mensen en andere dieren diepte kunnen waarnemen.

Stereopsis (Binoculaire dispariteit)

Omdat onze ogen op een korte laterale afstand van elkaar staan, ziet elk oog vanuit een iets andere hoek. Dit resulteert in de projectie van twee licht verschillende beelden op het netvlies (Binoculaire dispariteit). De beelden worden naar de hersenen verstuurd en verwerkt in de visuele cortex. Aangezien het verschil van de twee beelden groter zal zijn bij dichter gelegen objecten dan bij verder gelegen objecten, kunnen we diepte waarnemen. Een gelijkaardig concept wordt gebruikt door smartphones en andere computers om diepte te achterhalen. Dit heet Computer stereo vision.

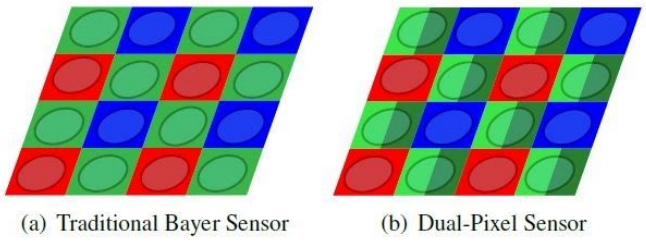
Dual Camera techniek

Dit is wellicht de meest voor de hand liggende manier. Het linker en rechter beeld worden met elkaar vergeleken door een stereo algoritme. Hierbij gaat men op zoek naar corresponderende pixels in beide beelden. Deze geven een verschuiving, waarbij pixels met een grotere verschuiving ten opzichte van elkaar, zich dichter bij de camera bevinden dan verder gelegen punten. Vervolgens kan men door middel van triangulatie (techniek voor het berekenen van lengte) de afstand van de gefotografeerde objecten tot de camera berekenen.

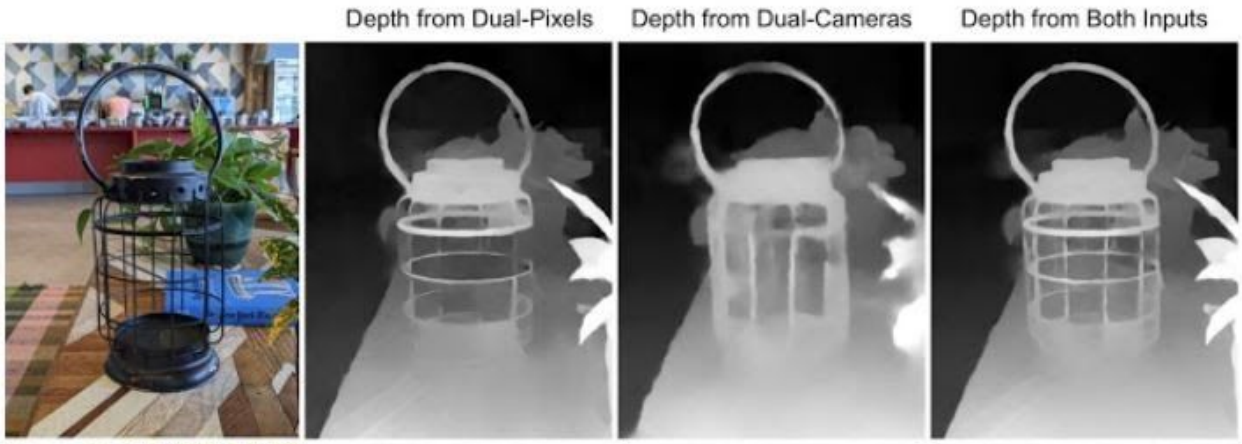
Dual pixel techniek

Diezelfde techniek kan ook worden toegepast bij smartphones die maar over één enkele camera beschikken. Zij maken gebruik van het welbekende dual pixel autofocus systeem. Dit houd in dat elke pixel op de sensor wordt onderverdeeld in een linker en rechter helft.

Zo ontstaat er net als bij de dual camera techniek een kleine verschuiving tussen twee beelden waarmee men opnieuw diepte kan berekenen. Het verschil is dat pixels veel dichter op elkaar staan dan twee aparte camera's, deze dual pixel manier resulteert in een goede meting van fijne details. Anderzijds is deze techniek minder accuraat voor het meten van verre objecten, daarom is het zo dat deze dual pixel methode vaak wordt gecombineerd met de dual camera techniek voor meer accurate depth maps.



(a) Traditional Bayer Sensor (b) Dual-Pixel Sensor



Structure from motion techniek

In plaats van één enkele foto te maken, beweeg je de camera in een opwaartse beweging om een reeks frames vast te leggen. Dit introduceert een parallax. Door middel van een computeralgoritme genaamd Structure-from-motion (SfM) kan men een 3D-model van de wereld trianguleren, waarbij de afstand tot elk punt in de scène wordt geschat. Hoewel deze techniek mogelijk is met elke soort camera en dus een goedkoop alternatief voor de dual pixel of dual camera techniek kan zijn, is deze techniek minder gebruiksvriendelijk aangezien de kwaliteit van het resultaat afhangt van de snelheid en stabiliteit van de beweging.

Time of flight camera

Range image of a human face captured with a time-of-flight camera (artist's depiction)

Deze camera's maken geen gebruik van stereopsis voor het berekenen van diepte. Een time of flight camera of structured-light direct depth sensor is een camera die infrarood flitsen uit stuurt. Door de timing van de weerkaatste lichtstralen te meten, kan men diepte achterhalen. Hoewel dit in mijn ogen de meest voor de hand liggende techniek leek te zijn, blijken dergelijke systemen duur en niet goed te werken in openlucht.

Semantische segmentatie

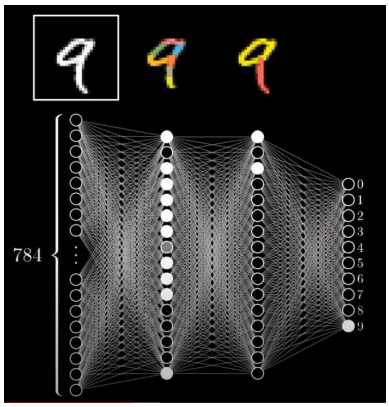
Deze derde techniek wordt vooral gebruikt bij front facing cameras. De selfie camera is namelijk meestal fixed-focus en heeft daardoor geen nood aan dual pixel autofocus. Hierdoor kan men dus geen diepte achterhalen. Aangezien deze camera in 99% van de gevallen toch gebruikt wordt om selfies te nemen is het meestal voldoende om een mask te maken van de gefotografeerde persoon, en een uniforme blur toe te voegen aan de achtergrond. Het nadeel hiervan is dat het geen blurvariatie of voorgrondblur kan toevoegen.

(a) A segmentation mask obtained from the front-facing camera.

Deze techniek gebruikt semantische segmentatie en getrainde “neural networks”, die mensen en hun accessoires kunnen segmenteren en identificeren. Segmentatie betekent dat een afbeelding wordt onderverdeeld in groepjes pixels die bij elkaar horen. In dit geval is dit meestal een persoon en zijn achtergrond. Segmentatie gebeurt op verschillende manieren. Ik geef vervolgens drie technieken om een idee te geven van hoe segmentatie in zijn werk gaat.

> **Randdetectie**: Segmenten van een afbeelding onderscheiden zich meestal door een abrupte overgang in intensiteit. Deze randen duiden de scheidingslijnen tussen de verschillende segmenten aan.
>
> **Histogram gebaseerd**: Segmenten van hetzelfde type hebben ongeveer dezelfde kleur en intensiteit. Door deze techniek te combineren met randdetectie krijg je betere resultaten dan met randdetectie alleen.
>
> **Groeiende zones**: Deze methode tracht segmenten te vinden door te vertrekken vanuit startpunten waaruit zones gevormd worden door omliggende pixels te vergelijken en aan de zone toe te voegen indien ze gelijkwaardig zijn.

Vervolgens is het aan een neuraal netwerk, die in dit geval getraind zijn in het herkennen van mensen, om te bepalen welke segmenten bij elkaar horen (bijvoorbeeld gezicht + haar + t-shirt + armen en een kop koffie in de hand)



Neurale netwerken zijn vrij complex. In het kort komt het er op neer dat deze aan de hand van typerende kenmerken, segmenten kunnen classificeren, filteren en ten slotte identificeren. Eens het netwerk weet welke pixels deel uit maken van bijvoorbeeld een gezicht, arm of kapsel, voegt het getrainde netwerk deze samen tot een geheel zijnde de gefotografeerde persoon. Voor een duidelijke en meer gedetailleerde uitleg omtrent neurale netwerken zou ik aanraden dit filmpje te bekijken: [But what is a Neural Network? | Deep learning, chapter 1]

Machine learning to predict depth

Ter toevoeging op voorgaande traditionele stereo algoritmes voor het berekenen van diepte, gebruiken sommige applicaties ook nog machine learning voor het verder verbeteren van de diepte schatting. Stereo vision lijdt namelijk aan “the aperture problem” wat wilt zeggen dat men ten gevolge van de laterale positie van dual camera’s of dual pixels, geen paralax kan detecteren bij horizontale lijnen in de achtergrond.

Andere clues voor het schatten van diepte zijn bijvoorbeeld, dat punten ver gelegen van het focuspunt, zelfs met een smartphonecamera, iets minder scherp zullen zijn. Hieruit kan men een ruwe schatting van diepte maken. Daarnaast kunnen we als mensen ook accuraat schatten hoe ver alledaagse objecten in beeld liggen omdat we hun grote kennen en die objecten dus kleiner afgebeeld zullen zijn dankzij perspectief. Een computer kan dit ook toepassen met behulp van machine learning en een convolutional neural network, die de objecten kunnen identificeren. Vervolgens worden van deze nu bekende objecten, de pixels geteld en kan men ongeveer schatten hoe ver bijvoorbeeld een auto zich in de achtergrond bevindt.

Bokehvorming

Bokeh is de manier waarop een lens onscherpe lichtpunten weergeeft. Een opvallende eigenschap is dat het highlights in de onscherpe achtergrond omtovert tot heldere schijven. De vorm van echte optische bokeh wordt voornamelijk bepaald door de vorm van het diafragma. Zo zal een zes-messig diafragma, zeshoeken creëren in de onscherpte. Bij synthetische onscherpte kiest men voor de ideale cirkelvormige bokeh.

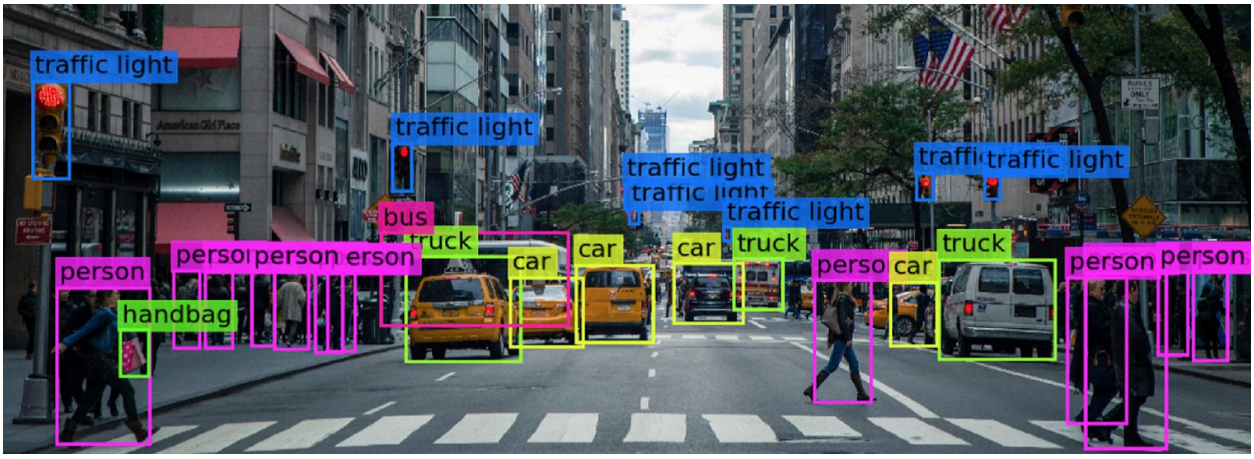
Deze genereert men door over elke pixel een doorzichtige cirkelvormige schijf te plaatsen, waarvan de grootte afhankelijk is van zijn diepte in beeld.

SLR Bokeh | Old Portrait Mode Bokeh | New Portrait Mode Bokeh

Conclusie

Aan de basis van synthetische scherptediepte licht steeds Computer stereo visie, wat gebaseerd is op biologische stereopsis. Afhankelijk van de complexiteit van de aanwezige camera of camera's, zal men verschillende technieken kunnen toepassen voor het creëren van synthetische scherptediepte. Hoewel elke methode op zichzelf synthetische scherptediepte kan verkrijgen kan men in ideale omstandigheden het beste resultaat krijgen met dual camera's die beschikken over dual pixel autofocus waarop men ter toevoeging segmentatie en machine learning kan loslaten.

Synthetische scherptediepte is maar één van de vele technologische toepassingen waarbij computervisie wordt gebruikt. Het ziet er naar uit dat computervisie naarmate de tijd, ervaring en onderzoek met stroomversnelling zal evolueren. Ik ben dan ook heel benieuwd naar de mogelijkheden die deze technologie in de toekomst zal bieden.



Bronnen

3Blue1Brown. (2017). But what is a Neural Network? | Deep learning, chapter 1 [Videobestand]. Geraadpleegd van https://www.youtube.com/watch?v=aircAruvnKk

busam, B., Hog, M., McDonagh, S., & Slabaugh, G. (2019). SteReFo: Efficient Image Refocusing with Stereo Vision. CVF, 1–10. Geraadpleegd van https://arxiv.org/abs/1909.13395v1

Garg, R., Wadhwa, N., Ansari, S., & Barron, J. (2019). Learning Single Camera Depth Estimation using Dual-Pixels. arXiv.org, 1–19. Geraadpleegd van https://arxiv.org/abs/1909.13395v1

Improvements to Portrait Mode on the Google Pixel 4 and Pixel 4 XL. (2019, 16 december). Geraadpleegd van https://ai.googleblog.com/2019/12/improvements-to-portrait-mode-on-google.html

KDCloudy. (2017). Portrait Mode on Smartphones: How Does It Work? [Videobestand]. Geraadpleegd van https://www.youtube.com/watch?v=nzb6E8A1yy0

Learning to Predict Depth on the Pixel 3 Phones. (2018, 29 november). Geraadpleegd van https://ai.googleblog.com/2018/11/learning-to-predict-depth-on-pixel-3.html

Lens Blur in the new Google Camera app. (2014, 16 april). Geraadpleegd van https://ai.googleblog.com/2014/04/lens-blur-in-new-google-camera-app.html

Marques Brownlee. (2017). Portrait Mode: Explained! [Videobestand]. Geraadpleegd van https://www.youtube.com/watch?v=xhybjeRciYg&t=1s

Portrait mode on the Pixel 2 and Pixel 2 XL smartphones. (2017, 17 oktober). Geraadpleegd van https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html

Wadhwa, N., Garg, R., Jacobs, D. E., Feldman, B. E., Kanazawa, N., Carroll, R., … Levoy, M. (2018). Synthetic depth-of-field with a single-camera mobile phone. ACM Transactions on Graphics, 37(4), 1–13. https://doi.org/10.1145/3197517.3201329

Wikipedia contributors. (2020a, 21 februari). Triangulation. Geraadpleegd van https://en.wikipedia.org/wiki/Triangulation

Wikipedia contributors. (2020b, 5 maart). Binocular disparity. Geraadpleegd van https://en.wikipedia.org/wiki/Binocular_disparity

Wikipedia contributors. (2020c, 14 mei). Parallax. Geraadpleegd van https://en.wikipedia.org/wiki/Parallax

Wikipedia contributors. (2020d, 7 juni). Image segmentation. Geraadpleegd van https://en.wikipedia.org/wiki/Image_segmentation