TEKST presentatie

Human pose estimation is het proces waarbij met de positie van pesonen schat en dit is nog een redelijk nieuw concept en waarschijnlijk nog niet gekend onder de meesten van jullie. Daarom vinden we het belangrijk om deze techniek te vergelijken met een meer gekende techniek: 3D motion capture, waarvoor men een speciaal pak aantrekt om de bewegingen vast te leggen.

De grootste voordeel van motion capture is de precisie, het bepaald de positie exact.

De nadelen zijn dan de prijs, die kan oplopen tot meer dan 100.000 euro, dit komt door de nood aan vele camera’s en dure software. en dat alles in dezelfde studio moet gefilmd worden waar de camera’s zijn geïnstalleerd.

De nadelen van van deze techniek zijn dan direct ook de voordelen van Human pose estimation, het is gratis. Iedereen die wil kan er gebruik van maken, maar dan moet je wel in staat zijn om het te installeren en dit is niet zo vanzelfsprekend. Je kan het ook gebruiken op gelijk welke foto of video, een simpele camera, bijvoorbeeld die van je gsm en een laptop is al genoeg om de positie te schatten van één of meerdere personen. . Ik moet er wel bij vertellen dat we enkel 2D data krijgen als maar 1 camera wordt gebruikt. Voor 3D data zijn wel degelijk 3 camera’s nodig.

Het grootste voordeel van motion capture is ook het grootste nadeel van human pose estimation, namelijk dat deze techniek slecht de positie schat, daarom moet je opletten voor welke toepassing je deze techniek gebruikt.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Maar hoe werkt Openpose nu eigenlijk?

Over het algemeen zijn er 2 manieren Van human pose estimation. Ten eerste heb je de top-down aanpak. Hierbij maak je eerst gebruik van een personendetector en schat je dan voor elke persoon apart hun pose.

Deze methode heeft enkele minpunten, namelijk dat de looptijd evenredig is met het aantal personen en als een persoon in de eerste stap niet herkend is, is er geen mogelijkheid om de fout toch nog goed te maken.

Een bottom-up aanpak daarentegen heeft deze problemen niet. Deze methode bepaalt alle belangrijke knooppunten en probeert dan de punten op de juiste manier aan elkaar te linken zodat het de poses van verschillende mensen vormt.

OpenPose is een voorbeeld van zo’n bottom-up aanpak en is een van de meest efficiënte neurale netwerken voor positiebepaling. OpenPose was ook de eerste open-source software voor realtime lichaams-, voet-, hand- en gezichtsdetectie.

Hier zie je de architectuur van het algoritme, men noemt dit een “twobranch multistage CNN”, heel kort betekent dit het volgende: CNN staat voor convolutioneel neuraal netwerk, “two-branch” wijst op het feit dat het CNN twee takken heeft en multistage wil zeggen dat het netwerk meerdere keren op elkaar is gestapeld. De twee takken worden meerdere keren na elkaar doorlopen om zo het resultaat steeds te verbeteren.

Concreet, het systeem neemt een foto als input en geeft als output voor elke persoon op de foto de 2D locaties van zijn lichaamsdelen. Eerst wordt een eerste voorspelling gedaan van de set van 2D vectorvelden (L). Dit zijn part affinity fields (PAF’s), zij leggen de graad van associatie tussen de lichaamsdelen vast. Dan worden aan de hand van de foto en deze part affinity fields ook een schatting gedaan van keypoints van de lichaamsdelen, deze werden betrouwbaarheidskaarten genoemd en worden voorgesteld door heatmaps. Dit zijn respectievelijk de eerste en de tweede tak van het CNN.

Deze eerste voorspelling wordt dan opnieuw aan het netwerk gevoerd en verbeterd. Na zes keer is het resultaat goed genoeg en wordt het samengevoegd om de uiteindelijke output te geven. Omdat het netwerk meerdere keren doorlopen wordt, noemt men dit een multistage CNN.

Er zijn heel wat factoren die invloed hebben op de output van openpose en die ervoor kunnen zorgen dat de uitkomst niet betrouwbaar is.

Een eerste is de invloed van resolutie. Om dit te onderzoeken, zijn we gestart met een foto van 1500 op 1500 pixels als referentie. We hebben deze foto dan herschaald naar foto’s van lagere en hogereresoluties.

Bij foto’s met een lage resolutie zagen we als snel een spreiding van 7°. Als we dan keken naar de afwijking van de voorgestelde verhoging/ verlaging van het zadel was er een spreiding van meer dan 5mm op het voorstel. Dit is een veel te grote foutenmarge voor een bikefitting waar juist die enkele millimeters bepalend kunnen zijn.

OpenPose maakt slechts een schatting van de positie OpenPose kiest niet altijd de keypoints juist op de gewrichten. We hebben online een foto van een professionele bikefitting gevonden en ons programma daarop laten lopen. Hierop duidelijk te zien dat een afwijking van de keypoints een grote invloed heeft op de aanpassingen die ons algoritme voorstelt. Zo ligt het juiste knooppunt van de heup verder naar achter, waardoor de hoek van de knie in het echt scherper is. Hierdoor kom je vaak een andere waarde uit voor een hoek dan de werkelijke waarde. Dit vormt een probleem aangezien we de optimale zadelhoogte baseren op werkelijke hoeken.

Een ander probleem is de omzetting van pixels naar centimeter of millimeter. Zoals Isaac al zei vragen we hiervoor nog de lengte van het dijbeen. We berekenen die lengte dan ook in OpenPose.

Zo krijg je telkens een verhouding van centimeter per pixel. En aan de hand van die waarde kan je dan uitendelijk de output van ons programma in centimeter uitdrukken.

Het grote probleem is nu dat je nooit op dezelfde manier de lengte van je dijbeen zal meten zoals OpenPose die berekent. Hierdoor heb je vanaf de start al te maken met meetfouten en zal je dus nooit een betrouwbare output krijgen.

Bij het bepalen van de schouder hoek hebben we dan te maken met een ander probleem. De ideale hoek is ongeveer 90°. Maar openpose ziet de rug als een rechte lijn (toon de figuur), terwijl de rug in het echt redelijk gebogen is. Daardoor kan het zijn dat de hoek in werkelijkheid wel 90° is, maar dat openpose een compleet andere waarde geeft, in dit geval zal het steeds een kleinere hoek zijn.

Bij bikefitting gaat heet vaak om die laatste millimeters. We hebben echter gemerkt dat OpenPose hiervoor geen goede oplossing biedt. Er zijn veel te veel factoren die een invloed hebben op het uiteindelijke resultaat waardoor het niet meer betrouwbaar is. Waarschijnlijk zal je een beter en sneller resultaat zal hebben indien je de zadelhoogte op het gevoel aanpast en in gedachte houdt dat je knie nog lichtjes gebogen moet zijn als je voet beneden staat.

In de vorige toepassingen namen we aan dat de foto vanuit een goede positie werd genomen, voor deze toepassing wilden we een concreet antwoord vind op de vraag: "Wat is het effect van de camerapositie?” Bij het nemen van een foto of video die je wil analyseren zijn er heel wat variabelen: hoe ver je staat, hoe hoog de camera staat en vanuit welke hoek t.o.v. de persoon de foto genomen is. Daarom is het belangrijk dat we testen wat de invloed is van deze variabelen op de uitkomst van ons programma. We hebben dit op een systematische manier getest, zo hoopten we duidelijke conclusies te kunnen trekken uit de data.

Om dit te testen hebben we foto’s getrokken vanop 1 meter, 1,5 meter, waarbij we eenmaal centraal, eenmaal 50cm naar links en eenmaal 50cm rechts gingen en vanop 2 meter afstand. Op elke positie werden 5 foto’s genomen, 40cm .60, 80, 100 en 120cm boven de grand.

Doordat niet elke squat gelijk gelijk was tussen de reeksen en dat de kanteling van de camera niet steeds hetzelfde was kunnen we geen conclusie trekken uit onze data.

We zagen dat de hoeken en afstanden die we hebben gemeten heel willekeurig veranderden. Aangezien openpose in 2D werkt en we hoeken meten van iets in 3D hebben we te maken met een projectie. Daarom is het logisch dat je de foto of video best zo loodrecht mogelijk op het vlak van de hoek, dan zal er het minst distorsie zijn. Ook de persoon staat best volledig op de foto voor het beste resultaat