

Live actieherkenning met de Kinect sensor in Python

Bert De Saffel

Student number: 01614222

Supervisors: Prof. dr. ir. Peter Veelaert, Prof. dr. ir. Wilfried Philips

Counsellors: ing. Sanne Roegiers, ing. Dimitri Van Cauwelaert

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Information Engineering Technology

Academic year 2018-2019

Hoofdstuk 1

Gerelateerde werken

- Bron [1] gaat eerder over hoe het skelet bepaalt wordt
 - voorstel van een methode om op een accurate manier de 3D posities van de joints te bepalen, vanuit slechts één dieptebeeld, zonder temporale informatie
 - Het bepalen van lichaamsdelen is invariant van pose, lichaamsbouw, kleren, etc...
 - Kan runnen aan 200 fps
 - Wordt effectief gebruikt in de Kinect software (onderzoeksteam is van Microsoft)
 - Een dieptebeeld wordt gesegmenteerd in verschillende lichaamsdelen, aangegeven door een kleur, op basis van een kansfunctie; Elke pixel van het lichaam wordt apart behandeld en gekleurd. Een verzameling van dezelfde kleuren wordt een joint
 - Aangezien tijdsaspect weg is, is er enkel interesse in de statische poses van een frame. Verschillen van pose in twee opeenvolgende frames is minuscule zodat die genegeerd worden
- Bron [2]
 - ✓ Bevat bruikbare datasets van skelet-, diepte- en kleurenbeelden
 - Ook hier praten ze over de vaak voorkomende uitdagingen: Intra-en interklasse variaties, de omgeving en de grootte van de verzameling van acties die er eigenlijk bestaan.
 - Hier tonen ze ook weer het nut van de Kinect sensor aan, en gebruiken de Kinect
 - Ze geven een nieuw algoritme om menselijke actieherkenning uit te voeren vanuit een dieptebeeld, een view-invariante representatie van poses en het systeem werkt real-time.
 - ! Het real-time component bevat drie zaken:
 - * Het verkrijgen van de 3D locaties van de joints → via bron [1]
 - * Het berekenen van HOJ3D (histogram)
 - * Classificatie
 - Histogram gebaseerde representatie van 3D poses (HOJ3D genoemd) = partitie van 3D ruimte in n "bins", gebruik maken van een bolcoördinatensysteem. Selectie van 12 joints die een compacte representatie van het skelet weergeven. (hand en pols, voet en enkel worden gecombineerd).
 - Het centrum van deze 3D ruimte is de heup joint. Er is ook een vector α , parallel met de grond, door de heup (van links naar rechts), en een vector θ loodrecht op de grond en door het centrum. Deze 3D ruimte (figuur 3b) wordt opgesplitst in n partities.
Voor θ : [0, 15], [15, 45], [45, 75], [75, 105], [105, 135], [135, 180]. (7 bins)

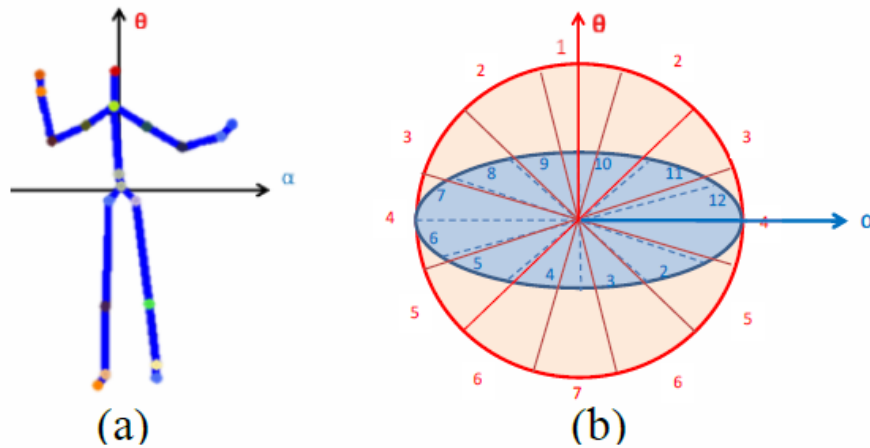


Figure 3: (a) Reference coordinates of HOJ3D. (b) Modified spherical coordinate system for joint location binning.

Voor α : 30 graden voor elke bin, dus 12 bins.

in totaal $7 * 12 = 84$ bins

Via deze bolcoördinaten kan elke 3D joint gelokaliseerd worden in een unieke bin

- De 3 joints die gebruikt worden om het bolcoördinatenstelsel te oriënteren staan uiteraard vast. De overige 9 joints worden onderverdeeld in één van de 84 bins.
- Om de representatie robust te maken, wordt één enkele joint over verschillende, naburige bins verdeeld (8 burens), op basis van gewichtsfunctie:

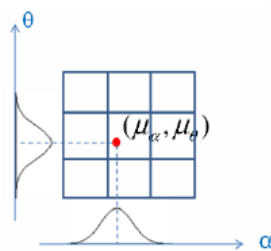


Figure 4: Voting using a Gaussian weight function.

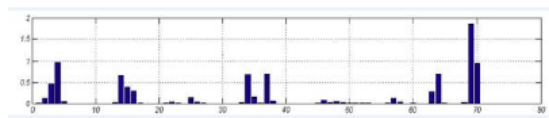


Figure 5: Example of the HOJ3D of a posture.

- Linear discriminant analysis (LDA) wordt toegepast om dominante features eruit deze histogram te halen.
- Ze stellen voor om kleurenbeelden te combineren met dieptebeelden om algoritmen te ontdekken met beter herkenning

- Ze beweren sneller te zijn dan bron [3]
- Bron [3] (pre-kinect era)
 - Actieherkenning met behulp van reeksen van dieptebeelden
 - Gaan ervan uit dat efficiënte tracking van skeletbeelden nog niet mogelijk is. (is gepubliceerd zelfde jaar dat Kinect beschikbaar was, 2010)
 - Hun oplossing is dus niet gebaseerd op het tracken van de skeletbeelden
- Bron [4]
 - real time activity recognition systeem, om routines te herkennen. Ze stellen zelf voor om het in een smart home te gebruiken.
 - Systeem is gebaseerd op twee categorieën:
 - * Informatie verzamelen met dieptebeelden
 - * Feature extraction gebaseerd op joint informatie en trainen van activiteiten.
- Bron [5]
 - Probleem: output van de actiecategorie EN de start en eind tijd van de actie.
 - Ze beweren dat actieherkenning reeds goed opgelost is, maar niet actiedetectie. Hun definities zijn:
 - * Actieherkenning: De effectieve actieherkenning indien het systeem weet wanneer hij moet herkennen
 - * Actiedetectie: een langdurige video, waarbij de start en stop van een actie niet gedefinieerd zijn = untrimmed video (videos waarbij er meerdere acties op hetzelfde moment kunnen voorkomen, alsook een irrelevante achtergrond). **sluit heel goed aan op onze masterproef**
 - Uitdaging in bestaande oplossingen: groot aantal onvolledige actiefragmenten. Voorbeelden:
 - * Bron [6]:
 - maakt gebruik van untrimmed classificatie:
 - * Bron [7]:
 - Spreekt over de onzekerheid van het voorkomen van een actie en de moeilijkheid van het gebruik van de continue informatie
 - Pyramid of Score Distribution Feature (PSDF) om informatie op meerdere resoluties op te vangen
 - PSDF in combinatie met Recurrent Neural networks bieden performantiewinst in untrimmed videos.
 - Onbekende parameters: actielabel, actieuitvoering, actiepositie, actielengte
 - Oplossing? Per frame een verzameling van actielabels toekennen, gebruik makend van huidige frame actie-informatie en inter-frame consistentie.

Bibliografie

- [1] J. Shotton, A. Fitzgibbon, A. Blake, A. Kipman, M. Finocchio, B. Moore, and T. Sharp, “Real-time human pose recognition in parts from a single depth image.” IEEE, June 2011, best Paper Award. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/real-time-human-pose-recognition-in-parts-from-a-single-depth-image/>
- [2] L. Xia, C. Chen, and J. Aggarwal, “View invariant human action recognition using histograms of 3d joints,” in Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, 2012.
- [3] W. Li, Z. Zhang, and Z. Liu, “Action recognition based on a bag of 3d points.” IEEE, June 2010, pp. 9–14. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/action-recognition-based-bag-3d-points/>
- [4] S. K. Ahmad Jalal and D.-S. Kim, “Detecting complex 3d human motions with body model low-rank representation for real-time smart activity monitoring system,” March 2018.
- [5] L. W. Z. W. X. T. Yue Zhao, Yuanjun Xiong and D. Lin, “Temporal action detection with structured segment networks,” 2017.
- [6] G. Singh and F. Cuzzolin, “Untrimmed video classification for activity detection: submission to activitynet challenge,” arXiv preprint arXiv:1607.01979, 2016.
- [7] J. Yuan, B. Ni, X. Yang, and A. Kassim, “Temporal action localization with pyramid of score distribution features,” 06 2016, pp. 3093–3102.