Pedestrian Pose Estimation Using Stereo Perception

Jorge Almeida and Vitor Santos

Abstract This paper presents an algorithm to perform pedestrian pose estimation using a stereo vision system in the Advanced Driver Assistance Systems (ADAS) context. The proposed approach isolates the pedestrian point cloud and extracts the pedestrian pose using a visibility based pedestrian 3D model. The model accurately predicts possible self occlusions and uses them as an integrated part of the detection. The algorithm creates multiple pose hypotheses that are scored and sorted using a scheme reminiscent of the Monte Carlo techniques. The technique performs a hierarchical search of the body pose from the head position to the lower limbs. In the context of road safety, it is important that the algorithm is able to perceive the pedestrian pose as quickly as possible to potentially avoid dangerous situations, the pedestrian pose will allow to better predict the pedestrian intentions. To this end, a single pedestrian model is used to detect all pertinent poses and the algorithm is able to extract the pedestrian pose based on a single stereo depth point cloud and minimal orientation information. The algorithm was tested against data captured with an industry standard motion capture system. Accurate results were obtained, the algorithm is able to correctly estimate the pedestrian pose with acceptable accuracy. The use of stereo setup allows the algorithm to be used in many varied contexts ranging from the proposed ADAS context to surveillance or even human-computer interaction.

1 Introduction

Pedestrians are one of the most vulnerable and unpredictable road users. The pedestrians ability to suddenly start motion or change direction can create a dangerous

J. Almeida(✉) · V. Santos
Department of Mechanical Engineering, University of Aveiro, Aveiro, Portugal
e-mail: {almeida.j,vitor}@ua.pt

V. Santos
IEETA, University of Aveiro, Aveiro, Portugal

L.P. Reis et al. (eds.), *Robot 2015: Second Iberian Robotics Conference*,
Advances in Intelligent Systems and Computing 417,
DOI: 10.1007/978-3-319-27146-0_38

situation in hundreds of milliseconds. In the Advanced Driver Assistance Systems (ADAS) context, the prediction of the pedestrians' intentions could potentially prevent accidents and possible injuries. For instance, the detection of the pedestrian intent to either cross a road at a crosswalk or to stop. Systems that are able to perceive pedestrian motion as soon as possible will improve safety for road users. In [1], the authors studied how humans detect the intentions of pedestrians to cross the road. The authors presented the participants videos of pedestrians crossing in natural traffic situations. The authors conclude that parameters of body language, such as legs or head movements, are indispensable for a consistent behavior prediction. Pedestrian trajectories alone are not sufficient to a correct and robust prediction. In this context, estimation of the pedestrian pose is of crucial importance to achieve a fast response system.

In this work, a technique to estimate the body pose of pedestrians is presented; with this estimation a subsequent system could potentially interpreter the poses to perform motion recognition. To achieve the proposed goal, the system must not depend on any previous manual initialization step or on a multi-frame tracking system. As such, the proposed system is able to estimate the pose from a single frame and minimal prior orientation information. The system performs a hierarchical top down geometrical search on a segmented pedestrian point cloud using an anthropomorphic constrained sampling scheme to detect body parts and limbs.

The human body pose estimation is a complex task with a large number of possible applications; robust interactive human body tracking has applications including gaming, humancomputer interaction, security, telepresence, and health care [2]. The problem is made complex due to the high dimensional search-space, frequent ambiguities between poses and high number of local minima. A great deal of research has been dedicated to detect human poses based only on monocular vision systems (survey in [3]); this is an especially ill-posed problem due to fact that many different poses present the same image projection. In this work, the use of a stereo system is proposed. This system provides dense 3D point clouds by using a state of the art stereo matching algorithm. The extra information provided by the point cloud, depth information, relieves many of the ambiguities in pose compared to the monocular system. Existing high performance depth-based systems are mostly dependent on structured light sensors [2]; theses sensors provide high precision, frame rate and definition, but are not suited to work on outdoors environments due to saturation of the sensor and range limits, therefore are not applicable in the ADAS context. The stereo setup is still the most attractive approach in the ADAS context given its low cost and low complexity, especially compared to active laser systems [4].

In section 2 the related work in markerless pose detection is presented. The proposed system is described in 3 with body parts detections in 3.4. The experimental results are presented in section 4 and final conclusions are presented in section 5.

2 Related Work

Previous work on markerless detection and tracking of a human body pose has been primarily focused in the use of intensity images, as stated above. In [3] the authors provide a survey of the different techniques used. The authors mark the distinction between model-based (generative) and model-free (discriminative) approaches, with the model-based methods using *a priori* information of the human body.

In [5], the authors propose a generic model for human detection and articulated pose estimation. The authors train detectors for anatomically defined body parts, which are then used as the likelihood in a generative model. The authors employ a flexible kinematic tree prior using pictorial structures on the configuration of body parts. In [6], the authors expand the previous work to include evidences from multiple frames. They model the temporal prior as a hierarchical Gaussian Process Latent Variable Model (hGPLVM) combined with Hidden Markov Model (HMM) to extend pedestrian tracklets. Their approach generates bottom-up evidence from 2D body models and so it constitutes a hybrid generative/discriminative approach.

The work proposed in [7] treats pose estimation as a nonlinear regression problem and proposes to estimate body poses directly from silhouette images. They employ a discriminative learning approach of body parts and embedded the algorithm in a tracking framework to facilitate disambiguation between poses. The absence of a previous model makes their technique easily adapted to different people, appearances or representations of 3D body poses.

Current monocular systems suffer from pose ambiguity problems due to the limitations of data used. These systems employ tracking architectures to solve pose ambiguity but the tracking implies the need to use multiple frames increasing the response time of these systems.

Work has also been performed using multiple monocular cameras to help with pose ambiguity. In [8], the authors propose to perform 3D human upper body pose estimation using multiple camera views. Their system creates multiple 3D pose hypotheses on a single view using a probabilistic hierarchical shape matching algorithm. These hypotheses are re-projected into other camera views and are then ranked according to their likelihood. Their system also applies a tracking mechanism integrating a motion model and observations in a maximum-likelihood approach. The need of multiple points of view severely limits the applicability of these systems.

Recently, the introduction of real-time depth cameras simplified greatly the pose estimation problem, when compared to monocular systems. The work presented in [9] makes use of a time-of-flight camera to estimate human body pose at video frame rates. The authors take a bottom-up approach to detect the body pose, starting with an interest point detector with a subsequent classification system.

Stereo has been previously applied to estimate human body pose, [10, 11]. In [12] the authors treat the pose tracking problem as a registration of two 3D point sets. The authors integrate Iterative Closest Point (ICP) with an unscented Kalman filter to yield a registration algorithm capable of tracking articulated bodies. In [13], the authors propose a system that uses stereo vision and a skin color filter. The skin color filter is used as a segmentation method to extract the point cloud belonging to the human body. The approach uses multiple models in different poses and computes an error metric to identify the correct pose. The work was performed in indoor environments and focused on upper body poses. The algorithm proposed in [14] also makes use of a variant of the ICP algorithm to match a simplified human model. The authors apply a Kalman filter based tracking architecture with a subsequent pose classification based on HMMs. All the proposed systems are either based on tracking algorithms or are not applicable in the ADAS context.

In the topic of predicting pedestrians' intentions in the ADAS context, the work by [15] presents a system that is able to predict if a pedestrian, walking towards the road curbside, will cross the road or stop. Asides from classification, the system uses dense optical flow from a stereo camera, with egomotion compensation, to obtain motion clues for the pedestrian upper torso and legs. A dimensional reduction using Principal Component Analysis (PCA) is applied to create Histogram of Orientation Motion (HOM) features. The current motion is matched to the database using Quaternion-based Rotationally Invariant Longest Common Subsequence (QRLCS) similarity metric.

On the same topic, the work by [16] presents a system that allows to detect early the intention of a pedestrian to cross a road lane. This system uses the body language as an early indicator of a crossing intent. Their system uses an infrastructure monocular vision system to extract Motion Contour Histogram of Oriented Gradients (MCHOG) feature descriptor. They apply a linear Support Vector Machine (SVM) system to identify the point when the pedestrian starts to enter the lane.

Both of these works would benefit from a more accurate and complete perception of the pedestrian motion. With additional detail the pedestrians' intentions could be inferred more accurately and also sooner. The use of stereo vision makes possible pose estimation in outdoors environments. The system is less susceptible to pose ambiguity, a serious problem in monocular systems, and performs well in outdoors environments with the desirable range. The proposed systems focus attention in the pertinent poses in ADAS context, especial attention is given to the legs pose. Previous works do not focus on this problem neither present a solution with the required characteristics; a solution that works in outdoors environments capable of, quickly and without initialization, estimate the pose of the human lower limbs during a normal walking cycle.

3 Stereo Pose Estimation

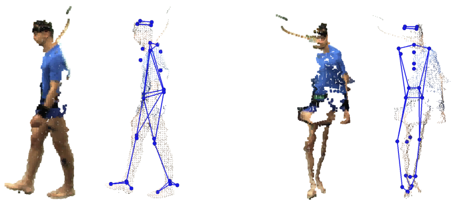
Human body poses are obtained using 3D point clouds from a stereo camera, as shown in 1. The pose estimation is performed using a method that compares the visibility of the point cloud from the stereo camera with the expected visibility from a pose hypothesis.

The visibility at each point is defined as one of three possible values: free space, occupied or occluded. A free space classification indicates that a point is visible from the camera point of view but is not occupied. A occupied point is visible from the camera and occupied by a 3D point. Finally an occluded point is a point that is not visible by the camera because there is an occupied point in front.

A dense voxel cloud is created overlapping the extracted pedestrian point cloud. A set of 3D rays interests this dense cloud, the intercepted voxels for each ray are classified according to their visibility using the pedestrian point cloud as the blocking element. After classification, this dense voxel cloud will be the base element for calculating the score of different hypotheses.

For each body part hypothesis, a set of 3D rays is used to calculate the visibility. The hypothesis score is calculated by comparing the classification of the points intercepted by the rays and the corresponding classification of the original dense voxel cloud.

When calculating the visibility of body parts hypotheses, previous detected body parts are used as blocking elements, for instance: the first detected leg will occlude the hypotheses for the second leg. This method allows to estimate the position of the occluded leg.



(a) Side view (b) Front view

Fig. 1 Example of an estimated pose. On the left the segmented pedestrian point cloud, on the right the estimated pose. The arms are not detected.

This work uses data from an industry standard motion capture system as ground truth. The motion capture system provides millimeter accurate position of a set of infrared reflective markers, visible on figure 1. To establish a direct comparison, a set of virtual markers, matching the motion capture markers, is used by the pose estimation algorithm.

3.1 Preprocessing

To extract the pedestrian point cloud three steps are applied: ground plane estimation, background subtraction and Euclidean clustering. The ground place estimation uses the RANSAC algorithm and helps to remove points near the feet. The background subtraction algorithm removes most of the points not belonging to the pedestrian. Finally, the resulting points from the two previous steps are clustered according to Euclidean distance between them and a specified threshold, the largest cluster is assumed to be the pedestrian.

This pedestrian extraction scheme works well in the dataset used, but in a more complex scenario some other state-of-the-art pedestrian detection algorithm could be used to segment the pedestrian point cloud. The developed algorithm does not require a perfect segmentation of the pedestrian from the background.

3.2 Visibility Calculation

The pose estimation algorithm here proposed assumes that a point cloud, comprised mostly of points belonging to a single pedestrian, was previously obtained. It is also assumed that the pedestrian is in an upright pose, a common assumption in the pedestrian detection context.

As stated before, ray tracing is used to calculate which voxels are either free, occupied or occluded, figure 2. The algorithm defines a set of rays using the original pedestrian cloud and the sensor position. For each ray, the intercepted voxels are classified. The end result is a dense voxel cloud in which each voxel contains the above classification, $\mathcal{V}_{\text{pedestrian}}$. This process is repeated for the pose hypotheses. Each body part pose hypothesis consists of a 3D model of the part, section 3.3, in a hypothesis pose. For each hypothesis the visibility is calculated. A score is obtained comparing the visibility of the hypothesis with the visibility of the original cloud.

In figure 2 two torso samples are presented. Each sample represents the same 3D model but in a different pose. The left hypothesis has a much larger area visible to the sensor and, as such, a much larger occluded volume. The left sample is aligned with the pedestrian, therefore the visibility will be very similar. The right sample will score a much higher value that the left sample.

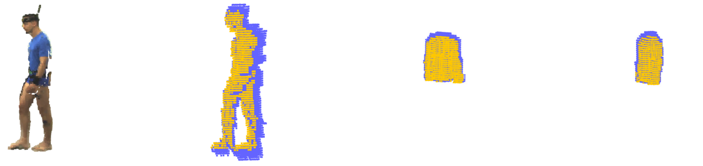

(a) Original. (b) Visibility. (c) Torso 1 (d) Torso 2

Fig. 2 Visibility dense voxel cloud representation. On the left, the original point cloud and the visibility calculated with the cloud. On the right, two different samples used to detect the torso orientation. Occupied voxels are represented as yellow squares, occluded voxels are colored blue. Empty voxels are not represented but used to score the sample.

Let $\mathcal{V} = \{v_1, \ldots, v_N\}$ represent all the voxels in the hypothesis, the score of each hypothesis Ψ_i is calculated as the sum, equation (2), of the score of every voxel, equation (1).

$$\forall v \in \mathcal{V}, s(v) = \begin{cases} P1 & \Leftarrow (v = v_{\text{pedestrian}}) \wedge (v = \text{free}) \\ P2 & \Leftarrow (v = v_{\text{pedestrian}}) \wedge (v = \text{occluded}) \\ P3 & \Leftarrow (v = v_{\text{pedestrian}}) \wedge (v = \text{occupied}) \\ P4 & \Leftarrow (v = \text{occluded}) \wedge (v_{\text{pedestrian}} = \text{occupied}) \\ P4 & \Leftarrow (v = \text{occupied}) \wedge (v_{\text{pedestrian}} = \text{occluded}) \\ 0 & \Leftarrow \text{otherwise} \end{cases} \tag{1}$$

$$\Psi_i = \frac{\sum_{n=1}^{N} s(v_n)}{N} \tag{2}$$

The different weights ($P1$, ..., $P4$) in equation (1) allow the algorithm to compensate for the different percentage of voxels with each classification.

Several performance optimizations were applied. The ray tracing can be very computationally expensive, as such, it is only performed once, for the $\mathcal{V}_{\text{pedestrian}}$ cloud. The rays and the intercepted voxels positions are reused for each pose hypotheses. The samples, after transformation, are geometrically aligned to the $\mathcal{V}_{\text{pedestrian}}$ cloud to allow the reuse of the rays. The geometric alignment of the samples also allows for a very fast indexing of the two clouds, avoiding the need for expensive nearest neighbor searches.

Ray tracing is not performed for each point in the pedestrian cloud. The rays are created starting in the sensor position and defining a square angular grid with a specific vertical and horizontal resolution, R_V and R_H respectively. The grid limits are defined from the point cloud, as to avoid unnecessary rays. The vertical and horizontal resolutions are key parameters of the algorithm. A more refined grid will account for greater detail, with the limit of the sensor own angular resolution, while a more coarse grid will correspond to lower number of rays improving computational performance.

3.3 3D Model

The proposed algorithm compares the visibility of a pose hypothesis with the visibility of the current pedestrian point cloud. To this end, a realistic geometric 3D model of a pedestrian is used. The 3D model defines the shape that will be used to calculate the visibility of each different pose hypothesis. The method is hierarchical and sequential, the first body part to be detected is the torso, followed by the head and upper legs, and finally the lower legs. As such, the 3D model was segmented into different body parts for individual use.

Let $\mathcal{P} = \{p_1, \dots, p_N\}$ represent the pedestrian point cloud with N points. The overall bounding box of $\mathcal{P}$ provides a rough approximation to the pedestrian height. The height approximation allows to estimate the size of the different body parts. The original 3D model is scaled to fit this measurement.

3.4 Detecting Body Parts

The first body part to be detected is the torso. The torso pose is extracted in three steps.

The pivot position is directly defined from the centroid position and a penetration factor. The penetration factor is used to correct the centroid in the sensor direction, placing the torso pivot inside the body and not at the surface.

The second step estimates the torso orientation θ_{torso} in the vertical direction $\hat{z}$. To this end, a set of samples is created with different orientation angles. Each sample is scored and a graphic, such as figure 3, is obtained. From this graphic, it is clearly visible that, there are two main peaks with 180° offset. The two peaks appear due to the fact that the torso shape is similar on the front and back, leading to pose ambiguity. To solve this ambiguity more information is required. In the proposed method, the θ_{torso} maximum closest to the previous estimated orientation is used.

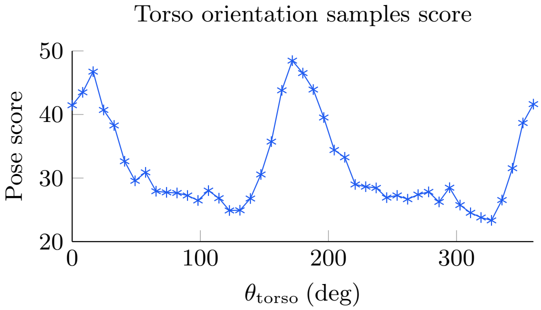


Fig. 3 Torso orientation samples score. The two peaks are created by the ambiguity between the front and back of the torso. The algorithm is able to correctly estimate the correct orientation using the peak closest to the previous orientation.

The third step estimates the torso forward inclination ϕ_{torso}, the rotation on the axis perpendicular to the vertical direction and the direction derived from the torso orientation $\hat{\phi} = \hat{z} \times \hat{\theta}$. This rotation is especially important when the pedestrian is moving quickly or running.

The head pose is estimated after the torso pose. The head pivot is directly derived from the torso pose and a set of samples is created to detect the head rotation θ_{head} in the vertical axis $\hat{z}$.

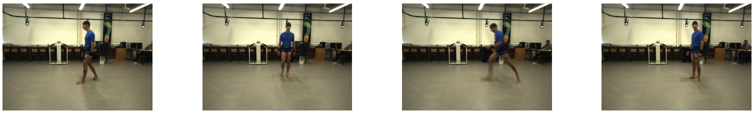
After estimation of the head pose, the legs positions are estimated. The algorithm starts by identifying which leg is more exposed to the sensor as a function of its predicted distance. This distance is based on the hip distance using the torso pose. The pose of the leg more exposed is the first to be estimated. Each leg is segmented in two parts, the upper leg and the lower leg. The upper leg comprises the distance from the hip to the knee, and the lower leg the distance from the knee to the foot.

The upper leg samples are created using two degrees of freedom, rotation on the $\hat{\phi}$ axis and rotation on the $\hat{\theta}$ axis. The upper leg pivots on the hip joint, defined by the torso pose. A set of samples is created by composed rotation of the two degrees of freedom. The samples are scored using the method described above. The lower leg samples pivots on the knee joint and rotates on the two same axes. All rotations are limited by anthropomorphic constrains.

The second leg pose is only estimated after the first. The first leg pose will influence the visibility of the second leg. The first leg will be used as an obstacle when calculating the visibility for the second leg. This method allows to estimate the position of the leg even when it is occluded. The created samples will reflect the fact that there is an obstacle in front and samples that are occluded will be correctly classified.

4 Results

The proposed algorithm was compared to a high precision industry standard motion capture system. The test trial consisted of a simulated pedestrian road crossing, figure 4. In the trial, several pedestrian trajectories were obtained. The test was composed of pedestrian trajectories parallel to the sensor, perpendicular and at an angle. The test contained trajectories where the pedestrian stopped at the simulated road entrance, and also trajectories where the pedestrian runs. The trial consisted of a total of 1588 frames, of witch 1053 were used. Frames where the pedestrian was not fully visible in the stereo camera were discarded. Also, the motion capture system was not always able to acquire all markers, in a frame, if a specific maker was not found the pose estimation marker was discarded.



(a) Walking perpendicular.

(b) Walking parallel.

(c) Running.

(d) Walk and stop.

Fig. 4 Sample images from the trial. The images present some of the several different trajectories used. The running trajectories were affected by the weak lighting conditions of the laboratory that led to some blurry images.

Quantitative results were obtained. A direct comparison was made possible by defining virtual markers analogous to the motion capture markers, on the 3D body parts. Figure 5 presents the histogram of the Euclidean distance from the motion capture markers to the pose estimation markers for the whole trial. The markers placement on the 3D body parts affects the results. Incorrect placement will appear as error on figure 5, an attempt to minimize this error was made. Table 1 presents the parameters values used in the trial.

Table 1 Parameters used in the test trial.

Parameter	Value
$P1$	10
$P2$	50
$P3$	100
$P4$	1
R_V	1.5°
R_H	0.5°

As can be observed, a large percentage, 72%, of the results are under 0.1m, and 94% of results are under 0.2m. The person's self occlusion presents some serious challenges, typically only one shoulder is visible and legs frequently occlude each other. The proposed method allows to estimate the person's orientation even with high occlusions. Given the hierarchical nature of the method, lower body parts suffer from errors in the upper parts. To account for this fact, lower body parts' samples are created with broader limits that would otherwise be necessary. Figure 6 presents the results for pose orientation. This orientation is calculated using the shoulders markers projected on the $X - Y$ plane. The figure presents a histogram of the body orientation error of the algorithm.

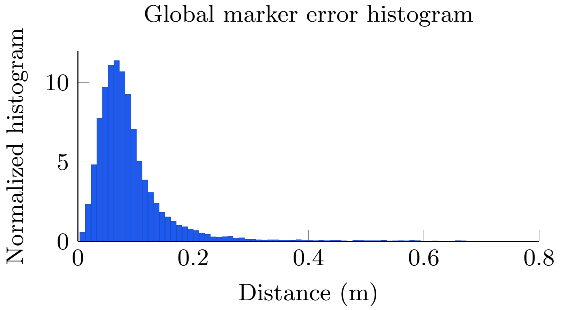


Fig. 5 Histogram of the euclidean distance between each marker of the pose estimation and the motion capture system.

The pose orientation is estimated with good accuracy. The largest errors occur when the pedestrian runs. The stereo setup used, performed poorly on low light conditions, such as the motion capture laboratory. Fast movements cause the image to become blurred due to the large exposure time. This in turn, decreased the quality of the stereo algorithm.

The stereo data used is of good quality but, nevertheless, presents some pronounced noise; the stereo noise presents the main limitation to the accuracy of the proposed approach.

The lack of a strong prior in our algorithm presents some advantages, but also disadvantages. With a good prior, the search space for each body part could be dramatically reduced, thus improving estimation accuracy. The current proposal could

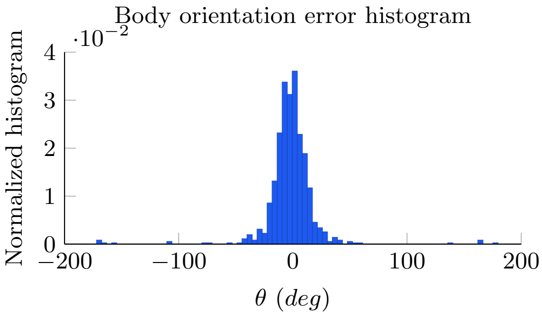


Fig. 6 Histogram of the body orientation error for the trial.

be expanded to use such a tracker. The presented algorithm, as is, could be used to initialize the tracker and also to recover from failure.

5 Conclusions

An algorithm capable of detecting human poses using stereo point clouds was presented. The algorithm is able to estimate poses using single point clouds and minimal motion orientation, used to relieve ambiguity between left and right poses. The proposed approach uses a hierarchical visibility based pose estimation algorithm. The algorithm focuses attention on the legs position, the legs motion will provide cues on the early intention of pedestrians trying to enter or cross a road.

The algorithm was tested with millimeter accurate industry motion capture data of a pedestrian simulating a possible pedestrian road crossing. Results presented show the potential of the algorithm to correctly recover poses even with noisy stereo data. The stereo setup presents some serious advantages over traditional monocular systems or even structured light systems. The point cloud data presents much less pose ambiguity than a monocular system and has the advantage of working in outdoors environments at long ranges.

Our proposed algorithm does not require any pose initialization or an elaborate pose tracking algorithm. This presents an obvious advantage by allowing the estimation of the pose of a pedestrian entering the scene without the need of a long multi-frame tracking system that would delay any conclusion. Nevertheless, the posterior application of a tracking algorithm would improve computational performance as well as performance under occlusion. The proposed algorithm could be used in the initialization step of the tracker or to recover from failure.

Future work will be focused on the implementation of a probabilistic pose tracker and finally on a system integrating the pose detection with the estimation of the pedestrians intentions in an advanced pedestrian safety system.

Acknowledgments This work was supported by the Portuguese foundation for Science and Technology (FCT) under grant SFRH/BD/73181/2010 and partially funded in the context of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014.

References

1. Schmidt, S., Färber, B.: Pedestrians at the kerb recognising the action intentions of humans. Transportation Research Part F: Traffic Psychology and Behaviour **12**(4), 300–310 (2009)
2. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. Commun. ACM **56**(1), 116–124 (2013)
3. Poppe, R.: Vision-based human motion analysis: An overview. Computer Vision and Image Understanding **108**(1–2), 4–18 (2007)
4. Geronimo, D., Lopez, A., Sappa, A., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. IEEE Trans. Pattern Anal. Machine Intell. **32**(7), 1239–1258 (2010)
5. Andriluka, M., Roth, S., Schiele, B.: Pictorial structures revisited: people detection and articulated pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, (CVPR), pp. 1014–1021 (2009)
6. Andriluka, M., Roth, S., Schiele, B.: Monocular 3D pose estimation and tracking by detection. In: IEEE Conference on Computer Vision and Pattern Recognition, (CVPR), pp. 623–630 (2010)
7. Agarwal, A., Triggs, B.: Recovering 3D human pose from monocular images. IEEE Trans. Pattern Anal. Machine Intell. **28**(1), 44–58 (2006)
8. Hofmann, M., Gavrila, D.M.: Multi-view 3D human pose estimation in complex environment. International Journal of Computer Vision **96**(1), 103–124 (2012)
9. Plagemann, C., Ganapathi, V., Koller, D., Thrun, S.: Real-time identification and localization of body parts from depth images. In: IEEE International Conference on Robotics and Automation, (ICRA), pp. 3108–3113 (2010)
10. Urtasun, R., Fua, P.: 3D human body tracking using deterministic temporal motion models. In: Pajdla, T., Matas, J.G. (eds.) Computer Vision, (ECCV). LNCS, vol. 3023, pp. 92–106. Springer, Heidelberg (2004)
11. Yang, H.D., Lee, S.W.: Reconstruction of 3D human body pose from stereo image sequences based on top-down learning. Pattern Recognition **40**(11), 3120–3131 (2007)
12. Ziegler, J., Nickel, K., Stiefelhagen, R.: Tracking of the articulated upper body on multi-view stereo image sequences. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR), vol. 1, pp. 774–781 (2006)
13. Muhlbauer, Q., Kuhnlenz, K., Buss, M.: A model-based algorithm to estimate body poses using stereo vision. In: IEEE International Symposium on Robot and Human Interactive Communication, (RO-MAN), pp. 285–290 (2008)
14. Pellegrini, S., Iocchi, L.: Human posture tracking and classification through stereo vision and 3D model matching. J. Image Video Process. 2008, 7:1–7:12, January 2008
15. Keller, C.G., Hermes, C., Gavrila, D.M.: Will the pedestrian cross? probabilistic path prediction based on learned motion features. In: Mester, R., Felsberg, M. (eds.) Pattern Recognition. LNCS, vol. 6835, pp. 386–395. Springer, Heidelberg (2011)
16. Kohler, S., Goldhammer, M., Bauer, S., Doll, K., Brunsmann, U., Dietmayer, K.: Early detection of the pedestrian's intention to cross the street. In: IEEE Conference on Intelligent Transportation Systems, ITSC, pp. 1759–1764 (2012)