Estimation of Pedestrian Walking Direction from Video

Slimane Larabi
University of Sciences and Technology
USTHB University
BP 32, El Alia, Algiers, Algeria
Email: slarabi@usthb.dz

Amina Bensebaa
University of Sciences and Technology
USTHB University
BP 32, El Alia, Algiers, Algeria
Telephone : (+213) 780 801 565
Email: abensebaa@usthb.dz

***Abstract*—In this paper we study 3D pedestrian motion direction estimation using a projective geometry based method. The idea for estimation is based on image sequence analysis, where the detection of both head and toe points will provide measurements for vanishing point estimation. Knowing that once the walk direction is fixed, these two points move along 3D parallel lines. This estimation assumes that there is a binary segmentation of the pedestrian. For each frame, the top and the bottom points of the segmentation are extracted. One line is fitted to the collection of top points and the other to the collection of the bottom points. Given these two lines, the vanishing point is estimated and thence the direction of the pedestrian. Our approach either assumes two intrinsic parameters cameras that are known or estimated at a learning stage using two previously known walking pedestrian directions. Experiments using a publicly available database show an accurate estimation of the direction of the body with less than 3 degrees error.**

Keywords*—*Pedestrian, Direction, Feet, Head, Silhouette, Vanishing point, Projective Geometry

I. Introduction

Estimating direction of pedestrians walking or running from video is a crucial task for video surveillance applications : video surveillance, pedestrian protection using driver assistance systems, traffic control systems, improving people tracking and identification systems, accident avoidance system for smart cars, etc. Many contributions have been proposed and challenges are still here, due mainly to external conditions, which degrade the image quality especially in low resolution image. In addition, even when the scene is crowded, there is need to compute an accurate value of the direction with low complexity in time and space usage.

In this paper we address the task of estimating direction of moving pedestrians. The scene is depicted by a series of images shot by a single camera. Movements of head and feet of pedestrian's silhouette determine the direction of movement of that person. Each movement - top and bottom - is defined by a line. In order to be robust against the small individual movements of the head and the feet, positions of the head and the feet, are fitted one to the top and the other to the bottom line via linear regression. These lines are parallel in 3D, according to projective geometry. Those lines would intersect in a vanishing point. The direction of the movement of the person is determined using the estimated vanishing point, the principal point coordinates and the focal length,

We begin this paper by presenting the related works. The section III is devoted to the principles of our approach. We present the conducted experiments on CASIA-A and CASIA-B dataset [1] and we discuss the obtained results in section IV. The robustness and weakness of our method and all future works are given in section V. We conclude this paper by some suggestions for improving the proposed method.

II. Related Works

The majority of current works devoted to pedestrian direction estimation, use classification methods and deliver a rough direction of the body. Experiments have been conducted on disparate datasets which made their comparison difficult.

Support Vector Machine based scheme (SVM) has served in [2] to estimate the walking direction of pedestrian from images where 90% of correct recognition is obtained for 16 directions. It has also been used in [3] to estimate the discrete probability distribution of the orientation. This approach was trained using the INRIA pedestrian dataset. The results of tests in realistic conditions, show that 49.7% fall in the same bin as a ground truth when 81.3% fall in the same or adjacent bins. Chen [4] uses sparse representation technique on each frame for body pose classification. This generating (noisy) observation on body poses which is improved by soft coupling of body pose with the movement in a particle filtering framework. For the 8 directions studied, 75% as pose accuracy is achieved. Guangzhe [5] estimates the pedestrian orientation using Adaboost as a classifier and achieve 64% accuracy for 8 orientations. In [6], D. Baltieri estimates the person orientation on single images based on appearance features. The three level HoG feature set is extracted from each people detection and they are provided as an input to an array of binary classifiers trained on a set of discrete orientations. An average of 70% is obtained for classification of 8 directions. Tao and Klette [7] proposed novel Random Decision Forests to simultaneously classify pedestrians and their directions and yield results which are comparable to those of state-of-the-art and baseline methods.

In the context of intelligent vehicles, Flohr et al.[8] present an approach for the joint probabilistic estimation of pedestrian head and body orientation. A novel approach for jointly estimating head and body orientations is proposed in [9]. Target feet positions are estimated with the multi-target tracking approach and used for head localization and cropping via a 3D head-plus-shoulder model registered through shape matching, the precision 0.87 is achieved. H. Liu and L. Ma [10] estimate the orientation based on online appearance-based classifier update. Reliable motion direction is determined as pre-estimated person orientation to update the appearance-based classifier. This approach achieves more competitive performances especially for unknown scenes. R. Raman [11] proposes an approach for direction estimation of a moving pedestrian estimation. The dimensional change-based feature is used to estimate the direction of motion ; eight discrete directions of motion are taken into account and the hidden Markov model is used for classification, 90% of accuracy is achieved on CASIA-B dataset [1].

Projective geometry can help to solve this problem. Many works have dealt with pedestrians motion for camera calibration. The horizon line is estimated by observing the human motion in different directions. Authors use the fact that the line joining the head positions of a pedestrian in two time instances is parallel to the line joining the feet (or bottom) positions to the corresponding time instances. Similarly, the line joining the head to the feet in a frame is also parallel to the line joining the head to the feet in the next frame.

Using frames of a pedestrian walking in a straight line, authors in [12], recover three orthogonal vanishing points for camera calibration method. To do this, authors assume that shoes follows periodic motion.

The published works related to vanishing points and lines are devoted to camera calibration and would either assume accurate positions of pedestrian's points, or would propose a rough detection and an approximation of the points.

In [12] left and right toes of the shoes are extracted from the pedestrian blobs when the two legs are maximally separated and used to construct two image lines. Obtaining such frames is unrealistic in real world : the presence of toes is not guaranteed, foot may be occluded or not separated from the other, and the distance between the foot is different.

Authors in [14] calculate top and bottom locations from the tracked pedestrians using the center of mass and the second order moment of the lower and the upper portion of the bounding box of the foreground region. In [15], the bottom point determined either as the end point of the major axis of the ellipse fitting the pedestrian's blob or the end of the line fitting the bounding box. These two methods [14] [15] do not locate specific point of foot and then we can't assert that the pair of lines joining the two successive top points and bottom points, are parallel in 3D.

In this paper we deal with binary images pedestrians assuming that silhouettes are noisy. Our aim is to locate the pair of parallel lines in the scene joining both the tops of heads and the feet extremities to improve the results of the state of the art in accuracy.

III. Proposed Approach

We assume that pedestrian is walking following a given direction. Image sequence is acquired and segmented. The problem of direction change is not processed here and will be in our future works.

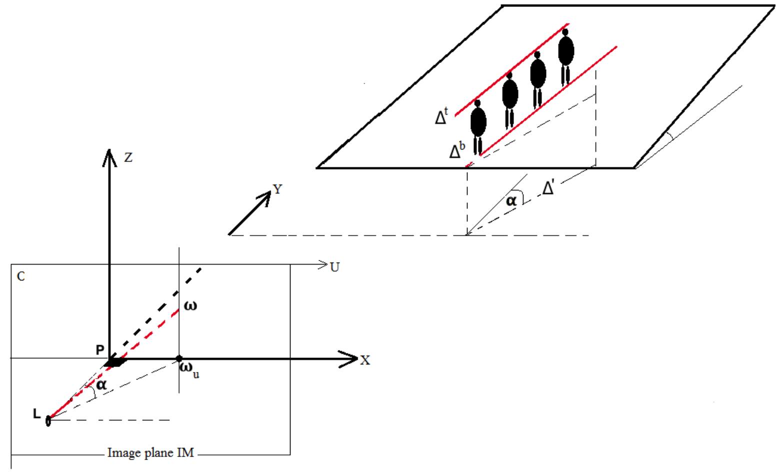

Fig. 1. Vanishing points ω associated to the direction of (Δ').

A. Walking Direction of Pedestrians from 3D to 2D

We define the pedestrian's direction as the orientation of the intersection of the plane, defined by pedestrian bodies along the motion trajectory and the plane perpendicular to the image plane as shown in figure 1 where the slope of Δ' $(\tan(\alpha))$ is the orientation of the pedestrian in $3D$.

Let $\omega(u_\omega, v_\omega)$ be the vanishing point located on the image plane associated to the direction of Δ'. The angle α may be computed using ω, the focal length f and the coordinates of the principal point $P(u_0, v_0)$ as follow:

$$\alpha = \arctan(\frac{u_\omega - u_O}{f}) \tag{1}$$

B. Determining the values of u_O, f

The values of u_O, f, needed in order to calculate the angle α, may be available either if the camera is calibrated, or by performing a learning stage using two pairs of parallel lines which have two known different directions α_1, α_2 giving two vanishing points ω_1, ω_2 in image plane. The values of u_O, f are computed as follow.

$$\tan(\alpha_1) \times f = u_{\omega_1} - u_O \ ,$$

$$\tan(\alpha_2) \times f = u_{\omega_2} - u_O$$

If α_1, α_2 are not equal to 0 and $\Pi/2$, we get from this: $f = \frac{u_{\omega_1} - u_{\omega_2}}{(\tan(\alpha_1) - \tan(\alpha_2))}$ and $u_O = u_{\omega_1} - \tan(\alpha_1) \times f$

In practical case, it is easy to perform this step by taking videos of a person moving to directions $\alpha_1 = \Pi/4$ and $\alpha_2 = -\Pi/4$ (see figure 2). We get then:

$f = \frac{u_{\omega_1} - u_{\omega_2}}{2}$, as the angle $\omega_1 L \omega_2 = \Pi/2$, f is the middle of the segment $\omega_1\omega_2$ and $u_O = u_{\omega_1} - f$.

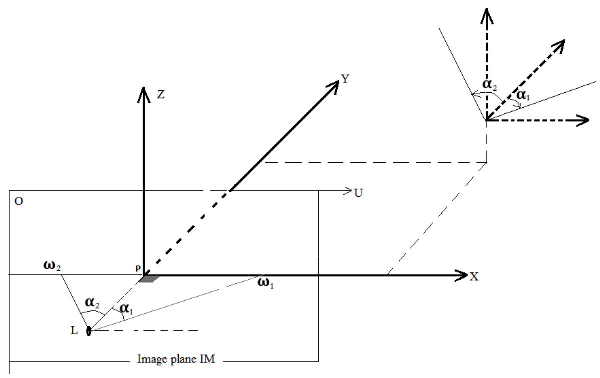

Fig. 2. Chosen directions for learning stage.

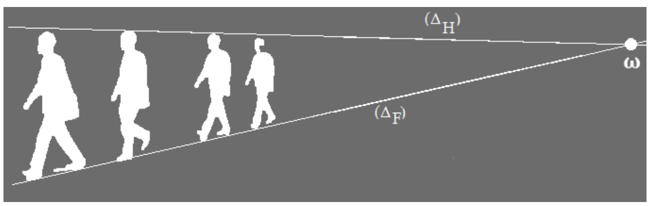

Fig. 3. Images of the parallel lines in 3D giving the vanishing point in 2D

C. *Finding pair of parallel 3D lines in the scene*

When pedestrian is walking, its images contain some geometric features which are useful for inferring information about 3D scene, especially features related to projective geometry since image formation may be seen as perspective projection.

If we look at figure 3 we can see pedestrian moves from one position to another on the same ground, the two lines defined by the limits of the heads and the limits of feet may be treated as parallel to the ground [13][14]. In the image plane, their projections (Δ_H, Δ_F) intersect in the vanishing point ω.

In theoretical case, joining the bottom points of pedestrian's images will define a line if they are corresponding to the same $3D$ point of the pedestrian's feet. This is an unrealistic situation, because feet in 3D is moving and then the bottom points are unmatched points (see figure 3). In the same case, head may change pose relatively to the body (possible inclination). Consequently in image, the two lines will be determined by minimizing the distance from the extremities (top and bottom) to the very line (The best-fitting linear regression model). The more numerous silhouettes are used in the process, the more the drawn lines in image would be accurate.

D. *Direction computation from images sequence: Algorithm and Complexity*

The first step (Learning stage) is the computation of camera parameters (f, u_0) using a sample of ground truth data. As explained is subsection III-B, two sets of frames corresponding to the motion of the pedestrian in two different directions are needed to solve two linear equations $(y = ax + b)$. In order to get best values of f, u_O, we will take more than 2 directions. More details are given in experimental results.

Once the parameters f, u_O are estimated, we compute the direction for each new set of frames using the equation 1 (Testing stage).

The Algorithm
Begin
-// $F_i, i = 0..k$ are the frames to be processed
-// focal length (f) and u-coordinate of the principal point (u_0) are computed in learning stage
For each F_i
Do

1. Locate the top p_i^t and bottom p_i^b of the pedestrian's silhouette

2. Fit via linear regression the top points p_j^t (j=0..i) to line Δ^t, and the bottom points $p_j^b (j = 0..i)$ to line Δ^b.

3. Compute the vanishing point $\{\omega\} = \Delta^t \cap \Delta^b$

EndFor

Compute the value of $\alpha = \arctan(\frac{u_\omega - u_0}{f})$.

End.

The complexity of each step $1, 2, 3$ is $O(nk) = O(n)$ it depends on the number n of pixels of outline silhouette and the k of frames, best value giving accurate estimation of the orientation is $k = 60$ (see section IV).

IV. Experimental Results

We use several datasets with different environments, several walk direction, various resolution provided by CASIA-A (binary, outdoor), CASIA-B (Binary, indoor).

Dataset CASIA-A [1] includes 20 persons. Each person has 12-image sequences, 4 sequences for each of the three directions, i.e. parallel (0), 45 degrees and 90 degrees to the image plane. The length of each sequence depends on the walker's speed and ranges from 37 to 127. The Dataset A includes 19139 images.

CASIA-B [1] dataset is composed by 124 groups of silhouettes corresponding to 13640 binary frame sequences with 11 different walk directions ranging $0°$ to $180°$ from view axis of camera. Each group contains silhouettes of 10 persons walking. For each direction, between 60 and 120 frames are available corresponding to the motion of the pedestrian. Pedestrians might carry bags, weare coat or a regular outfit or not carrying anything.

A. *Experimental Setup and Results of tests*

1) Dataset CASIA-B: For the learning stage, we used only one group (The first one). We computed the parameters (f, u_O) using frames of three known directions $18°$, $36°, 54°$. For each new frame, from images sequence, the two lines from best-fitting linear regression model, for the top points and the bottom points are estimated and vanishing point located. This computation process is repeated for all frames, and the vanishing points move depending on the data (Top and bottom points). We selected the ultimate vanishing point and evaluated the two parameters (See figure 4) and we compute the vanishing point for the directions $18°$, $36°, 54°$ as shown in figure 5. We note the noisiness of the processed data giving the non-aligned top and bottom points. Each pair of directions allow computing the f, u_O. Their average for

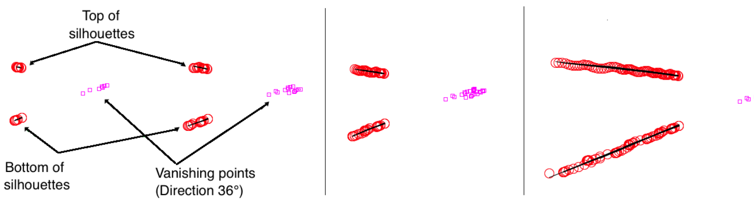

Fig. 4. The evolution of the computation process of lines joining the top and bottom of silhouettes, gives a vanishing points of a 36° direction. From left to right, the drawn lines are reevaluated for new pairs of points (Red circles) and new vanishing point are located (Magenta squares).

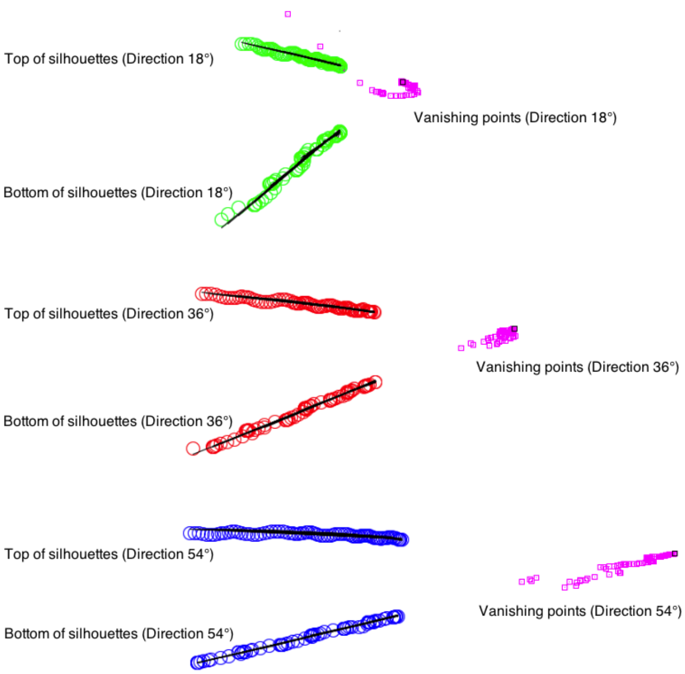

Fig. 5. Computing the vanishing points for three directions, the considered ones are drawn with black color

CASIA-B ($f = 325$pixels, $u_O = 187$pixels) are selected for the testing stage.

The 123 remaining groups have been used for test. For each set of frames, the direction is computed following the method described in subsection III-B. We computed for each set of silhouettes the vanishing point. We deduced the value of orientation angle α. Graph of figure 6 illustrates the average of the percentage error for 4 selected directions. The same percentages are obtained for other directions because they are basically the same direction (for example, $18°$ is similar to $162°$, etc). We can see that high percentage is obtained for $18°$ where (12%) corresponds to $2°$.

We studied the accuracy of computed direction related to the number of frames. Figure 7 illustrates the obtained results indicating that 60 frames are sufficient to get an accurate estimation. Figure 8(top) gives the mean errors and Standard Deviations for six directions.

2) Dataset CASIA-A: We used the same estimated values of the parameters in the learning stage for CASIA-B dataset.

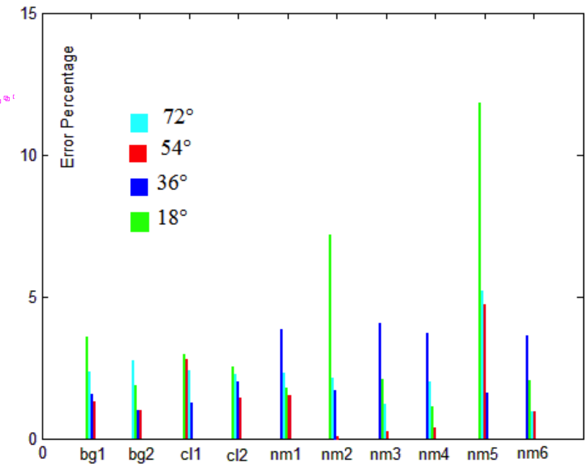

Fig. 6. Percentage error of computed directions $18°$, $36°, 54°, 72°$.

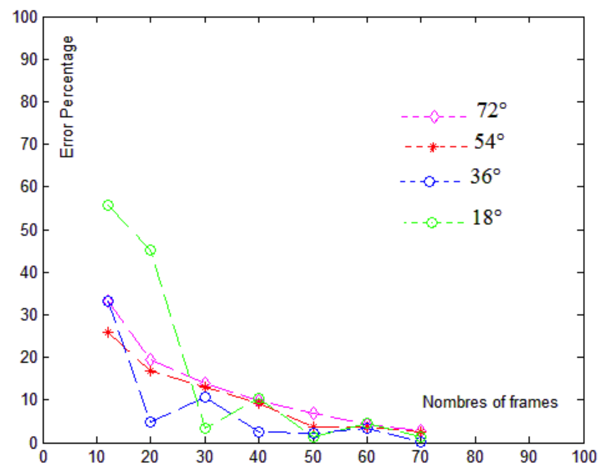

Fig. 7. Error related to frames number.

All the available data served for testing stage. The conducted test gave the results illustrated by the graph of figure 8.

We can conclude:

- The mean error and standard deviation of the obtained results are greater than of Casia-B. This is due to the result of outdoor image segmentation giving for many frames shadow regions. This as a part of silhouettes or missing region as shown by figure 9. The displacement of the top or bottom point located on the head and feet parts will disturb the linear regression fitting process and thence the intersection point will be more distant from the required vanishing point. The error and standard deviation will increase in this case.
- If we are in the case of outdoor scenes, the increasing of the error of pedestrian's orientation remains less than $6.5°$ (See figure 8).
- A higher error is obtained for the orientation $0°$ and $90°$. These two orientations are sensitive cases of our method. Indeed, for $0°$ the bottom pixels of silhouettes are unrealistic and the fitted line would produce an

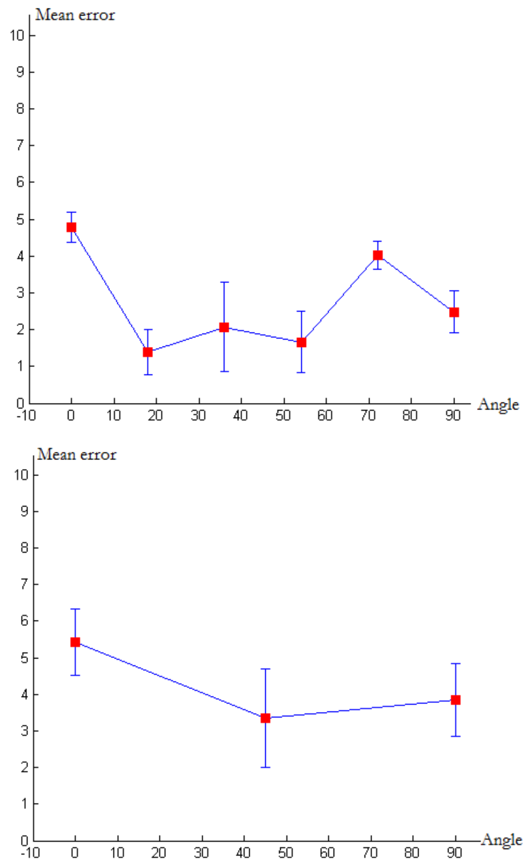

Fig. 8. Mean errors and Standard Deviations (Six directions $0°$, $18°$, $36°, 54°, 72°$, $90°$) for Casia-B dataset (Top) and for Casia-A dataset (Three directions $0°$, $45°$, $90°$)(Bottom)

Fig. 9. Sample of silhouettes taken from Casia-A and Casia-B dataset

intersection point far from the true position. For the case $90°$, the two fitted lines are theoretically horizontal and the vanishing point must be infinity. However, in practice, the intersection point may not be very far due to the fact that one or the two fitted lines aren't horizontal.

B. Evaluation Study

We compared our results with those obtained by methods that used the same experimental platform. However, these methods work on single frame of video dataset, and the published results concern the result of classification.

The average balanced accuracy of the methods conducted over CASIA Dataset B reported in [11] are: 90.87% for [11], 84.61% for [10], 72.50% for [6], 64.67% for [4] and 59.73% for [16].

These results concern the classification, not the measure of the orientation. In this case, if we translate our results to classification, knowing that all absolute errors are less that the half of the interval $(18°)$, our method achieves 100%.

The comparison in terms of absolute error of the orientation can't be made as all the results published do not give these measures.

The obtained result are accurate and surpasses the state of the art in addition to the low complexity of the computation.

V. Strongness, Weakness of the method and Future works

From conducted experiments on CASIA-B and CASIA-A dataset, and due to the obtained results, we assert that our method:

- is robust to the non-alignment of bottom points of silhouettes defining a line which is not parallel to the line joining the top of silhouettes. As found out, the bottom point is taken from one foot and sometimes from the other. The results are accurate even when the line joining these points is located between them.
- is robust for the case where the direction equal to zero, even if the top and bottom points are out of their meaning. If the pedestrian is perfectly horizontal $(90°)$ or vertical $(0°)$ to the camera the vanishing point is very noisy and is not well determined for the first frames, but after ten frames, the two fitted lines would give a correct orientation.
- is robust as far as the noisiness of the silhouettes is concerned.

In other side, we can't applied our method:

- for top view, because the features can't be extracted.
- for pedestrian changing the direction or moving on non planar ground.

Our method must be improved in order to process the change of direction and then define the frames of this change. Frames will be separated into groups such that each one will correspond to one direction. Other improvement to be investigated is to separate the bottom points of silhouettes of the same foot (Left or right) and to determine two lines in addition to the line of top points. This will decrease the error of estimated orientation.

Other interesting work concerns the orientation estimation of group of persons considering the blob instead of the silhouette.

VI. Conclusion

Starting from a video sequence of a walking pedestrian, we propose a new method allowing the computation with accuracy of the followed direction. We distinguish without ambiguity the pedestrian direction using at least 30 frames (1 second of walk) but we compute with high precision the direction from

60 frames (2 seconds of walk). The obtained results surpass all those obtained in the literature and the proposed approach can serve for real time applications.

REFERENCES

[1] S. Yu and D. Tan and T. Tan, *A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition*. ICPR, Hong Kong, China. August 2006.

[2] H. Shimizu, T. Poggio, *Direction Estimation of Pedestrian from Multiple Still Images*. IV 2004.

[3] T. Gandhi and M. M. Trivedi, *Image Based Estimation of Pedestrian Orientation for Improving Path Prediction*. IEEE Intelligent Vehicles Symposium, Eindhoven University of Technology Eindhoven, The Netherlands, 2008.

[4] C. Chen and A. Heili and J. M. Odobez, *Combined estimation of location and body pose in surveillance video*. 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp.5-10, 2011.

[5] Z. Guangzhen, T. Mrutani, S. Kajita, K. Mase, *Video based estimation of pedestrian walking direction for pedestrian protection system*. Journal of Electronics (China),Vol.29 No.1/2, March 2012.

[6] Baltieri, D. Vezzani and R. Cucchira, *People orientation recognition by mixtures of wrapped distributions on random trees*. In proc. Eur. Conf. Comput. Vis. (ECC), pp. 270-283, 2012.

[7] J. Tao and R. Klette, *Integrated Pedestrian and Direction Classification using a Random Decision Forest*. ICCV Workshop, pp. 230-237, 2013.

[8] F. Flohr, M. Dumitru-Guzu, J. F. P. Kooij and D. M. Gavrila, *Joint probabilistic pedestrian head and body orientation estimation*. IEEE Intelligent Vehicles Symposium (IV), Dearborn, Michigan, USA, 2014.

[9] E. Ricci and J. Varadarajan and R. Subramanian and S. R. Bulo and N. Ahuja and O. Lanz, *Uncovering Interactions and Interactors: Joint Estimation of Head, Body Orientation and F-formations from Surveillance Videos*. ICCV 2015.

[10] Liu, H. and Ma, L, *Online person orientation estimation based on classifier update*. In Proceeding of IEEE Int. Conf. Image Processing ICIP, Sept, pp. 1568-1572, 2015.

[11] R. Raman and P. K. Sa and B. Majhi and S. Bakshi, *Direction Estimation for Pedestrian Monitoring System in Smart Cities: An HMM Based Approach*. IEEE Access, pp. 5788-5808, DOI: 10.1109/ACCESS.2016.2608844, 2016.

[12] S. Huang and X. Ying and J. Rong and Z. Shang and H. Zha, *Camera Calibration from Periodic Motion of a Pedestrian*. CVPR 2016.

[13] F. Lv and T. Zhao, and R. Nevatia, *Self-calibration of a camera from video of a walking human*. IEEE International Conference of Pattern Recognition, 2002.

[14] I. N. Junejo, *Using Pedestrians Walking on Uneven Terrains for Camera Calibration*. Machine Vision and Applications, V 22(1), January 2011.

[15] J. Guan, F. Deboeverie, M. Slembrouck, *Extrinsic calibration of camera networks based on pedestrians*. Sensors, 16, 654; doi:10.3390/s16050654, 2016.

[16] M. Andriluka and S. Roth and B. Shiele, *Monocular 3D pose estimation and tracking by detection*. IEEE Conf. Comput. Vi. Patter Recognition. (CVPR), pp. 623-630, 2010.

[17] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, *Robust view transformation model for gait recognition*. International Conference on Image Processing, Brussels, Belgium, 2011.