




Direction estimation of moving pedestrian groups for intelligent vehicles

Amina Bensebaa¹ · Slimane Larabi¹ 

Published online: 2 May 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

In this paper, we consider direction estimation of pedestrian group for video surveillance and intelligent vehicles applications. A theoretical study of the position of vanishing points in image plane associated with all directions in the scene leads to the definition of the notion of directional areas. Image plane is divided along the x-axis into a set of bounded areas; each one is associated with a specific direction. The pedestrian direction is inferred directly depending on the belonging area of the vanishing points computed from video sequence. Top and bottom points of walking pedestrian define two parallel lines in 3D. The vanishing point is estimated from video sequence and from the direction of the pedestrian. The obtained results demonstrate the efficacy and robustness of the proposed method and confirm the improvement with respect to state-of-the-art approaches.

Keywords Directional areas · Pedestrian group · Intelligent vehicles · Video surveillance · Motion analysis · Silhouette

1 Introduction

Estimating direction of walking pedestrian group from video is useful for video surveillance, pedestrian protection, intelligent vehicles, traffic control systems, people tracking, identification systems and accident avoidance.

Our aims in this paper are to estimate the direction of walking pedestrian groups whatever the quality of image segmentation. In order to ensure a fast estimation, we studied the relationship between directions in the scene and vanishing points in image plane. Directional areas are defined as bounded regions along the x-axis in image plane; each one is associated with a specific direction. Once the vanishing point is computed from video sequence, the pedestrian direction is inferred directly depending on its belonging area.

Considered pair of parallel lines in the scene is defined by joining the top points of the head and the bottom points of the feet of walking pedestrian. In image plane, the top and bottom of pedestrian's silhouette are fitted to a top and a bottom line via linear regression and the vanishing point is calculated as their intersection.

We begin this paper by presenting the related works. Section 3 is devoted to the principles of our approach. In Sect. 4, we present the proposed algorithm and its complexity. Processing pedestrian group and the conducted experiments on available datasets are presented in the two last sections. We conclude this paper by some suggestions for improving the proposed method.

2 Related works

The majority of works devoted to pedestrian direction estimation use classification methods and delivered a rough direction of the body. Support vector machine (SVM)-based scheme has served in [1] to estimate the walking direction of pedestrian from images where 90% of correct recognition is obtained for 16 directions. It has also used in [2] to estimate the discrete probability distribution of the orientation. This approach was trained with the INRIA pedestrian dataset, and results of tests in realistic conditions show that 49.7% fall in the same bin as ground truth and 81.3% fall in the same or adjacent bins. Chen [3] use sparse representation technique on each frame for body pose classification, 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 considered,

✉ Slimane Larabi
slarabi@usthb.dz

¹ Computer Science Department, USTHB University, Bab Ezzouar, Algeria

75% as pose accuracy is achieved. Guangzhe [4] estimate the pedestrian orientation using AdaBoost as a classifier and achieve 64% accuracy for 8 orientations. In [5], D. Baltieri estimate person orientation on single images based on appearance features. The three-level HoG feature set is extracted from each people detection and provided as 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 [6] proposed novel random decision forests to simultaneously classify pedestrians and their directions and yield results comparable to those of state-of-the-art and baseline methods. In [7], authors present an approach for the joint probabilistic estimation of pedestrian head and body orientation in the context of intelligent vehicles. Quantitative experiments showed that joint tracking of head and body decreases the angular mean error for head/body orientation by $13^\circ/10^\circ$ compared to single-frame estimation and further by $3^\circ/2^\circ$ compared to independent tracking. A novel approach for jointly estimating head, body orientations is proposed in [8]. 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. Liu and Ma [9] estimate the orientation based on online appearance-based classifier update. Reliable motion direction is determined acting as pre-estimated person orientation to update the appearance-based classifier. This approach achieves more competitive performances, especially for unknown scenes. Raman [10] propose 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 considered; the hidden Markov model is used for classification; and 90% of accuracy is achieved on Casia-B dataset [11].

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 exploit the fact that the line joining the head positions of a pedestrian at two time instances is parallel to the line joining the feet (or bottom) positions at these same corresponding time instances. Similarly, the line joining the head to the feet in a frame is parallel to the line joining the head to feet 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 follow periodic motion.

The published works related to vanishing points and lines are devoted to camera calibration and then either assume accurate positions of pedestrian's points, or propose a rough detection and then 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 separating the foot is different.

Authors in [13] 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 portions of the bounding box of the foreground region. In [14], 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 line fitting the bounding box. These two methods [13,14] do not locate specific point of foot, and then we cannot assert that the pair of lines joining, respectively, two successive top points and bottom points are parallel in 3D. In addition, these two works served only for camera calibration.

Recently, deep-learning approach proposed by Raza et al. [15] outperforms existing classifiers and achieves accuracy of 0.92 for full-body orientation estimation using only eight orientation classes. In [16], a novel framework is proposed for jointly estimating head, body orientations of targets and geometric formations involving interacting targets of conversational groups from social scenes. However, the direction of the group is not considered, but human's direction is estimated individually.

The purpose of our system is to provide robust estimates direction of walking pedestrian groups whatever the segmentation results and image resolution. Our contributions are multiples:

- The notion of directional areas in image plane is introduced and served to infer directly the direction and the trajectory of pedestrian group.
- Robustness of the method to shadow and to low-resolution images.
- Fast and accurate direction estimation. If the direction is delivered with a delay of 20 frames, we surpass the results of the state of the art for direction classification

3 Proposed approach

3.1 Persons detection

3.1.1 Background estimation

We assume that background is not available and may change during time. To extract the foreground, background is estimated using the n first frames. Let $I_i(x, y)$ be the luminance value of the pixel $p(x, y)$ in the frames F_i . Luminance of $p(x, y)$ of the background B noted $B(x, y)$ is calculated as the average of $I_i(x, y)$, $i = 1..n$. Background is estimated

periodically in order to tackle weather change for outdoor scenes and lighting change for indoor scenes. Once background is estimated, a chosen threshold serves to extract the foreground.

3.1.2 Shadow elimination

To eliminate shadow, we use the method of Cucchiara et al. [17]. Let $F_H^k(p)$, $F_S^k(p)$ and $F_V^k(p)$ be the values in HSV color system of the pixel p of the frame k . The pixel p of the foreground is a shadow if $M^k(p)$ defined by Eq. (1) is equal to 1.

$$M^k(p) = \begin{cases} 1 & \text{if } \alpha \leq \frac{F_V^k(p)}{B_V^k(p)} \leq \beta \\ & \wedge |F_S^k(p) - B_S^k(p)| \leq \tau^S \\ & \wedge |F_H^k(p) - B_H^k(p)| \leq \tau^H \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The α parameter defines the strongness of the light source, whereas β is set to avoid detection of pixels where background changed slightly by noise as shadows. The choice for the threshold parameters τ^H and τ^S is set experimentally. A preliminary sensitivity analysis for α , β , τ^H and τ^S is reported in [17]. When shadow is dark such as frames of MGP01 dataset [18], this method fails and shadow is considered as a part of foreground.

Once blobs are extracted corresponding to one person or group of persons, they are tracked along the frames sequence.

3.2 Basic principle of direction estimation

Assuming that image formation may be considered as perspective projection of the 3D scene, images of walking pedestrian contain geometric properties useful for inferring information about 3D scene, especially those related to projective geometry. When pedestrian moves on planar ground, the two lines joining, respectively, the tops of the head and the bottoms of the feet may be considered as parallel to the ground [13,19]. In the image plane, their projections intersect at the vanishing point. Let $\omega(u_\omega, v_\omega)$ be the located vanishing point on the image plane associated with the pedestrian direction, and let L be the projection center of the camera (see Fig. 1). We consider the pedestrian’s direction as the value of the orientation angle α of $L\omega$, computed using u_ω , the focal length f and u_O (the u -coordinates of the principal point O) as follows:

$$\alpha = \arctan\left(\frac{u_\omega - u_O}{f}\right) \quad (2)$$

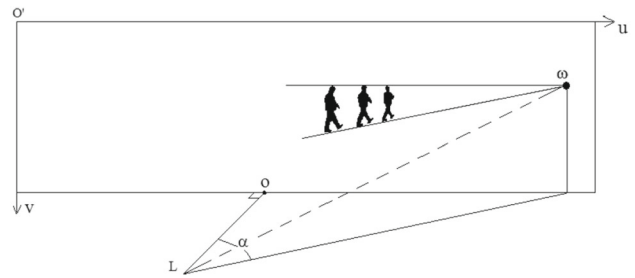


Fig. 1 Vanishing point ω associated with the direction of walking. $(O'uv)$ is the image plane coordinate frame centered on top left and O is the principal point

For a group of pedestrians located as a blob in image, we will consider that lines joining the top and bottom points of the blob are parallel in 3D. During walking, the top points and the bottom points may appertain to different pedestrians of the same group. When this occurs, the two lines in 3D are not parallel, the deviation of their images will produce uncertainty in estimated direction.

The values of u_O and f , needed for calculating α angle, may be available either by camera calibration, or by performing a leaning stage. With at least two pairs of parallel lines having two known different directions α_1, α_2 , two vanishing points ω_1, ω_2 are then computed and the values of u_O, f are derived as indicated by Eqs. 3, 4.

$$\tan(\alpha_1) \times f = u_{\omega_1} - u_O \quad (3)$$

$$\tan(\alpha_2) \times f = u_{\omega_2} - u_O \quad (4)$$

If $\alpha_1 \neq 0$ and $\alpha_2 \neq 0$, the values of f, u_O are estimated by solving the system of linear Eqs. (3) and (4). Otherwise, if $\alpha_1 = 0$, then $u_{\omega_1} = u_O$, and it is necessary to have $\alpha_2 \neq 0$ required to compute f .

In theoretical case, joining the top points and bottom points of pedestrian’s images will define two lines if they correspond to same 3D points of pedestrian’s head and feet. Due to the uncertainty, in image, the two lines will be determined by minimizing the distance from the extremities (top, bottom) to the considered line (the best-fitting linear regression model). Note that the more are numerous the used silhouettes in this process, the more is the accuracy of fitted line in image.

3.3 Direction estimation of a linear motion

The first step is the computation of camera parameters (f, u_O) as explained in Sect. 3.2. In order to get accurate values of f, u_O , we will take more than two directions. More details are given in validation section. Once the parameters f, u_O are estimated, algorithm 1 delivers the pedestrian’s direction for each frame of the video sequence starting at the second frame (delay = 1 frame).

Algorithm 1 Direction estimation for a linear motion

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 previously.

$-i = 0$ # the order number of the first frame F_0 .

- Locate the top and bottom points p_0^h, p_0^f of the pedestrian's silhouette for the frame F_0

For each $F_i, (i = 1..k)$

Do

1. Locate the top point p_i^h and bottom point p_i^f of the pedestrian's silhouette on F_i .

2. Fit via linear regression the top points $p_k^h (k=0..i)$ to new line δ^h

3. Fit via linear regression the bottom points $p_k^f (k=0..i)$ to new line δ^f

4. Compute the vanishing point $\{\omega\} = \delta^h \cap \delta^f$

5. Compute the value of $\alpha = \arctan((u_\omega - u_0)/f)$
 # α is the estimated direction for the frame F_i

EndFor

End.

For the video sequence of direction 18° taken from Casia-B dataset [11], we show in Fig. 2 a sample of extracted top and bottom points. Due to head occlusion and error of segmentation, the top, bottom points are not aligned. Consequently, the fitted lines produce vanishing points in different areas far from the correct position.

3.4 Directional areas in image plane

Let n be the number of considered directions. Without loss of generality, we assume that $n = 11$ (case of Casia-B dataset [11]). The directions $d_i, i = 0..10$ are defined by the intervals of orientation angles (see Fig. 3): $[-9^\circ + 9^\circ[, [+9^\circ, 27^\circ[, [27^\circ 45^\circ[, [45^\circ 63^\circ[, [63^\circ 81^\circ[, [81^\circ 99^\circ[, [99^\circ 117^\circ[, [117^\circ 135^\circ[, [135^\circ 153^\circ[, [153^\circ 171^\circ[, [171^\circ 189^\circ[.$

The direction $d_i, i = 0..10$ is correlated to u_ω (u -coordinates of the vanishing point ω) located on image plane. Indeed, from Eq. 2, $\tan(\alpha) \times f + u_0 = u_\omega$. Figure 4 illustrates the u -coordinates of vanishing points associated with some

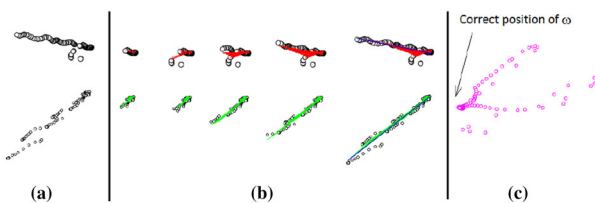


Fig. 2 **a** Located top and bottom points of silhouettes (black circles) for video sequence of Casia-B dataset [11]. Silhouettes are moving from right to left. **b** Evolution of the fitted lines to top (red color) and to bottom points (green color). **c** Computed vanishing points (magenta circles) by applying Algorithm 1

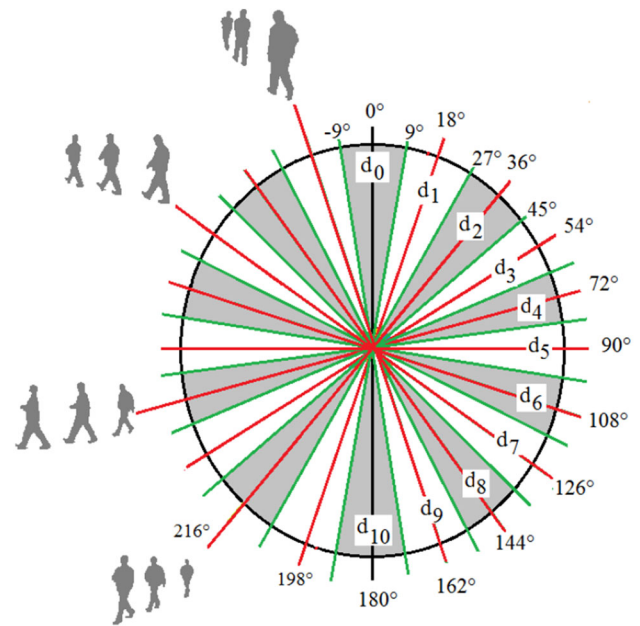


Fig. 3 The 11 directions considered for Casia-B dataset [11]. The pedestrian walks from the center to outside of the circle

directions using images taken from Casia-B dataset video sequence. Note that the more the value of α is large, the more the u_ω increases. When α surpasses 90° , u_ω will start from $-\infty$ and reaches the principal point O when $\alpha = 180^\circ$. If we translate the u -axis origin from O' (left top of image) to the principal point O , $u_\omega - u_0$ will be noted u and the direction $d = \alpha$ is given by Eq. 5.

$$d = \arctan\left(\frac{u}{f}\right) \tag{5}$$

Graph of Fig. 5 shows the variation of the orientation angle according to u -coordinate. Note that this graph is intrinsic to the camera because it depends on the parameters f, u_0 , and it indicates also the circularity of the direction d_i . Once the values of f, u_0 are known, u -coordinates of the bounds of the interval associated with d_i are computed and define the area of that direction as shown in Fig. 6. Note that d_0 is associated with central area ($u = 0$) and the more the angle is high, the more the area associated is large. In order to determine in what direction the pedestrian is walking, it is sufficient to find in what area the vanishing point belongs.

3.4.1 Properties of the directional areas

- The directional areas are delimited in image plane depending on the intrinsic parameters of the camera (f, u_0) and on interval bounds of all assumed directions in the scene.

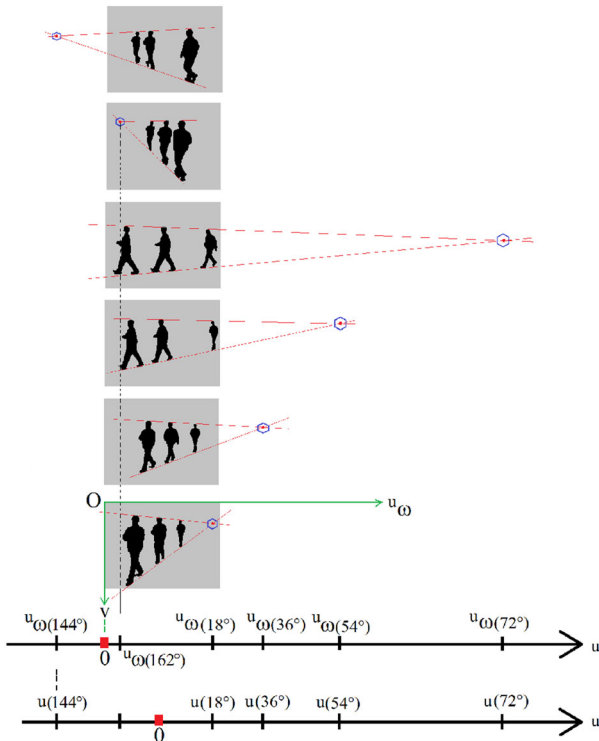


Fig. 4 The abscise of vanishing points of some directions for Casia-B dataset [11]. In bottom, the translated origin of the u-axis

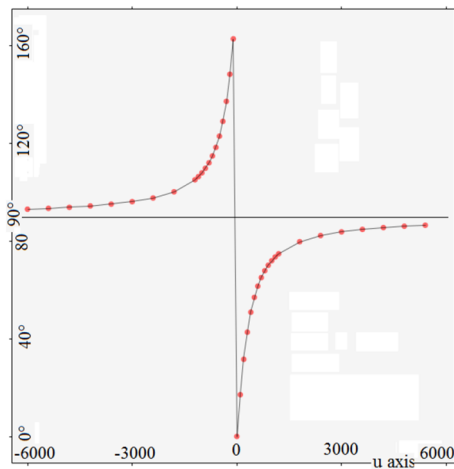


Fig. 5 Variation of orientation angle with regard to u

- While vanishing points are in the same area, the pedestrian is walking in the same direction.
- If vanishing points move from one area to the neighboring area, the pedestrian is changing the direction increasing or decreasing the orientation angle.
- Pedestrian’s trajectory is inferred directly from the trajectory of vanishing points ω on directional areas. As example, circular direction is associated with a trajectory of ω moving from one area until the most right (or

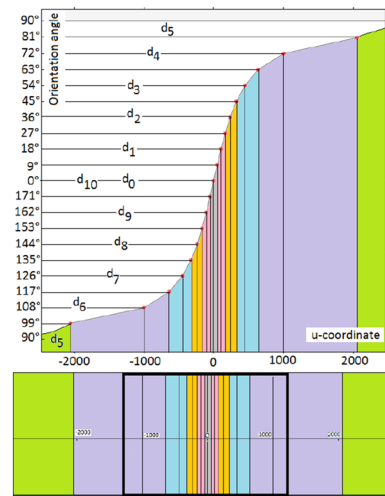


Fig. 6 (Top) The u-coordinates of the 11 directions for Casia-B dataset. (Bottom) The directional areas associated in the image plane. In bold is drawn a rectangle delimiting the physical image plane

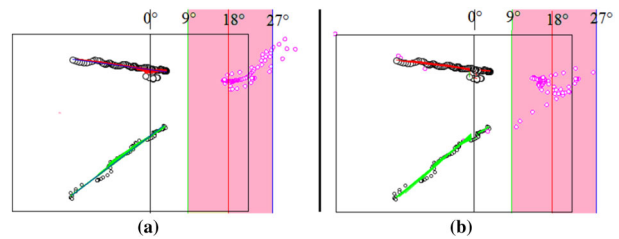


Fig. 7 **a** Vanishing points computed by applying algorithm 1, **(b)** and algorithm 2. Note that vanishing points that are outside of the directional area in **(a)** (in magenta color) are due to remoteness of the top points from the fitted line

left) area and again appears from the most left (or right) area and moves toward the starting area.

4 Direction estimation for a nonlinear motion

The intersection point (vanishing point) is sensitive to occlusion, segmentation error, direction change and non-planar ground. Figure 7a shows an example where head of pedestrians is occluded in the frames 46–55, the top points in this case are taken from shoulders. The fitted line to top points of frames 1–45 is relatively far from the new top points. In this case, we consider only the previous frames such that the slope of fitted line including the new top points is approximately stable. Otherwise, the current frame will be considered as the first for the fitting process. This will allow to get good estimation in case of missing parts of silhouette, direction change and non-planar ground because the top and bottom points define two parallel lines. Figure 7b illustrates how the fitted lines are corrected giving accurate position of vanish-

ing points, appertaining two the directional area. Algorithm 2 summarizes the steps of this computation.

Algorithm 2 Direction estimation for a nonlinear motion

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 previously
 - $i_{start} = 0$ # the order number of the starting frame
 - *start_again*:
 { - $i = i_{start}$
 - Locate the top points p_i^h, p_{i+1}^h and bottom points p_i^f, p_{i+1}^f of the pedestrian's silhouette for the frames F_i, F_{i+1}
 - Fit via linear regression the top points to line δ^h , and the bottom points to line δ^f .
 - Compute the vanishing point $\{\omega\} = \delta^h \cap \delta^f$
 - Compute the direction $\alpha = \arctan((u_\omega - u_0)/f)$.
 # α is the estimated direction for the frames F_i, F_{i+1}
 - $i=i+2$
 }

For each $F_j, (j = i..k)$

Do

1. Locate the top points p_j^h and bottom points p_j^f of the pedestrian's silhouette
2. **If** (p_j^h is near from δ^h)
Then Fit via linear regression the top points $p_k^h (k = i_{start}..j)$ to new line δ^h
Else $i_{start} = j$, Goto *start_again*
3. **If** (p_j^f is near from δ^f)
Then Fit via linear regression the bottom points $p_k^f (k = i_{start}..j)$ to new line δ^f
Else $i_{start} = j$, Goto *start_again*
4. Compute the vanishing point $\{\omega\} = \delta^h \cap \delta^f$
5. Compute the value of $\alpha = \arctan((u_\omega - u_0)/f)$.
 # α is the estimated direction for the frame F_j

EndFor

End.

Complexity of the algorithm 2.

We assume that the silhouette is encompassed by a rectangle ($w \times h$). The top (respectively, bottom) pixel is located as the middle of the first (resp. last) line of silhouette pixels in this rectangle. In the worst case, $w+w$ pixels are processed. As w may be equal to (n) (the width of the source image), and knowing that the complexity of fitting regression lines is linear, then the complexity of the algorithm 2 is $O(n)$.

5 Direction estimation for pedestrian groups

For intermediate crowded scenes, groups of pedestrians are often present in the field of view of intelligent vehicles and camera surveillance. In image, blobs are extracted instead of single silhouette. Despite the fact that top and bottom extracted points from these blobs may be associated with



Fig. 8 Silhouettes of pedestrian groups from MGP01 dataset [18]. Note that shadow is not removed

different pedestrians of the group or may be associated with the shadow as shown in Fig. 8, their motion in the scene defines two lines. We considered that lines are parallel in the scene and then we estimated the walking direction under this assumption. The uncertainty in direction estimation with the presence of shadow and displacement of pedestrians inside of the group is evaluated in experimental section.

6 Evaluation

We begin this section by presenting the used datasets. Next, we give the results of foreground extraction and estimated camera's parameters. The obtained results for direction estimation for single and groups of pedestrians are then presented and discussed.

6.1 Used datasets

We used the following datasets characterized by different environments, several walk direction and various resolutions.

Dataset Casia-A [11] includes 19139 images for three directions ($0^\circ, 45^\circ$ and 90°) taken in outdoor scenes.

Casia-B [11] dataset is composed by 124 groups of silhouettes corresponding to 13640 binary frame sequences with 11 different walk directions ($0^\circ, 18^\circ, 36^\circ, 54^\circ, 72^\circ, 90^\circ, 108^\circ, 126^\circ, 144^\circ, 162^\circ, 180^\circ$) taken in indoor scenes.

MGP01 dataset: (Moving Group of Pedestrians) [18]: Videos are taken in our campus using smartphone camera. This dataset is characterized by the presence of shadows due to the high contrast scene and constitutes in addition to a challenge for shadow removing. The ground-truth data are computed from the ground geometry of the scene. There are in the scene many groups of pedestrians walking in different directions at different resolutions.

EPFL Campus dataset [20,21]: Video sequences called Campus were shot outside on the university campus with 3 DV cameras. Up to four people are simultaneously walking in front of them. As the ground truth related to body direction is not available, we hand-estimated it based on the ground geometry.



Fig. 9 Extracted silhouettes for MGP01 [18] (top), Campus dataset [20,21] (bottom). Note that shadow is still present and some parts (head) are missing for some silhouettes of MGP01 dataset

6.2 Background estimation and silhouettes extraction

Applying the method proposed in Sect. 3.1, background is first estimated using 20 frames. Figure 9 shows sample of extracted silhouettes of MGP01 and Campus datasets. Despite the fact that the foreground is noisy with missing/additional parts and the presence of shadows, we did not apply any further processing in order to improve it. This allows to test the robustness of our method behavior to noisy data.

6.3 Camera’s parameters estimation

Casia-A and Casia-B datasets

We computed the parameters (f, u_0) using three video sequences in which pedestrian is walking in directions 18° , 36° and 54° . For each new frame from video sequence, the two lines of best-fitting linear regression model for, respectively, the top points and bottom points are estimated, and the vanishing point is located. This computation process is repeated for all frames of the same video sequence, and the vanishing points move depending on the data (top and bottom points). We selected the ultimate vanishing point for each direction and evaluated the two parameters (see Fig. 10). We note that the noisiness of the processed data generates non-aligned top and bottom points. Each pair of directions allows computing the values of f, u_0 . Their averages for Casia-B ($f = 325$ pixels, $u_0 = 187$ pixels) are selected for the testing stage for Casia-A and Casia-B datasets. We notice that frames with occluded head have been eliminated from used video sequences in this process. The 123 remaining video sequence of Casia-B and all video sequences of Casia-A served for direction estimation.

MGP01 dataset

We used a prior knowledge related to the ground’s scene. Indeed, there are parallel lines on the ground located as boundaries between lozenge slabs. Two pairs of parallel lines with known directions ($45^\circ, -45^\circ$) are used to solve the

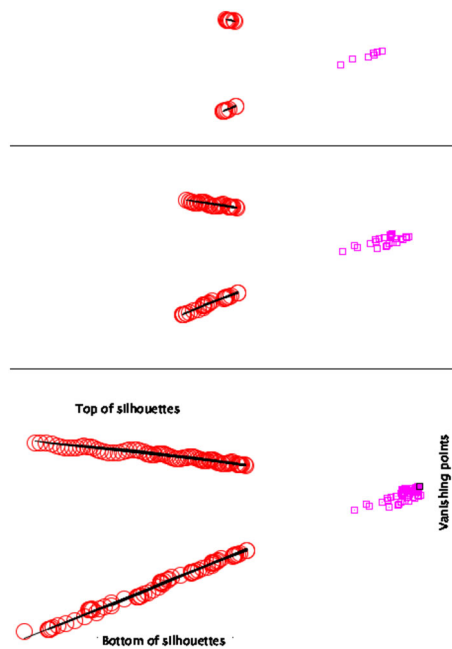


Fig. 10 Evolution of vanishing points and computed lines joining the top and bottom of silhouettes for the direction 36° . From top to bottom, the drawn lines are reevaluated for new pairs of points (red circles) and new vanishing point is located (magenta squares). The considered point is drawn with black color

linear system of Eqs. 3, 4. Using the two vanishing points located on image, we obtained ($f = 2141, u_0 = 964$).

Campus dataset

The same process has been applied for Campus. The computed and used values of (f, u_0) are (206.4, 181).

6.4 Accuracy of estimated direction according to the delay

Estimated direction depends on the number of silhouette’s extremities (top, bottom) used in the fitting process. The more the number of points is larger, the more the fitted lines are accurate. We studied using Casia-A and Casia-B datasets the accuracy of estimated direction according to the delay. Figure 11 shows that increasing the number of pairs of points allows the convergence of vanishing point inside the directional area (colored area). The results indicate that after 10 frames, the direction is well recognized with accuracy.

Estimated direction is more accurate for Casia-B than Casia-A dataset due to image segmentation giving for many frames of Casia-A dataset shadow or missing regions. Displacements of the top or bottom point located on the head and feet parts disturb the linear regression fitting process, and consequently, the intersection point is outside of the directional area. The error and standard deviation increase in this case.

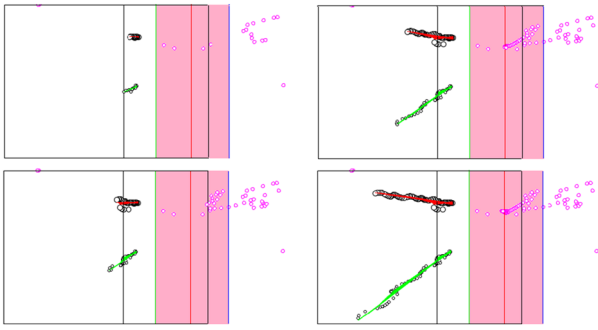


Fig. 11 Convergence of vanishing points toward the directional area when pairs of (top, bottom) points for a video sequence of directions 18° from Casia-B are numerous

Table 1 Accuracy(%) comparison of direction estimation results over Casia-B dataset with [10]

| D_i | Our (0) | Our (20) | Our (40) | [10] |
|-------|---------|----------|----------|-------|
| d_0 | 74.78 | 83.97 | 93.34 | 93.05 |
| d_1 | 91.99 | 98.09 | 100 | 90.99 |
| d_2 | 95.44 | 98.73 | 100 | 87.97 |
| d_3 | 76.35 | 85.35 | 94.33 | 89.75 |
| d_4 | 79.71 | 92.17 | 98.1 | 92.60 |
| BA | 83.65 | 91.66 | 97.15 | 90.87 |

Our(n) indicates our method using a delay of n frames, (BA) indicates balanced accuracy (average)

6.5 Direction estimation: results and comparison

6.5.1 Datasets Casia-B and Casia-A

Due to the linear regression technique used for fitting top points and bottom points, the error of estimation is very large for the first frames as seen in the previous subsection. As the interval between the 11 directions is 18° , then if the error is less than 9° the direction estimated is considered well classified.

In order to compare our method and [10], we merged the 11 different walk directions of Casia-B Dataset into five discrete directions:

$$D_1 = \{0^\circ, 18^\circ\}, D_2 = \{36^\circ, 54^\circ\}, D_3 = \{72^\circ, 90^\circ, 108^\circ\}, \\ D_4 = \{126^\circ, 144^\circ\}, D_5 = \{162^\circ, 180^\circ\}.$$

We give in Table 1 the accuracy(%) computed for each direction of Casia-B dataset and the highest scores in the state-of-the-art methods [10]. In Table 2, we compare the average accuracy with those obtained by the state-of-the-art methods. Without delay, the average rate (balanced accuracy) is 83.65% which is less than the rates obtained by [9,10]. However, with a delay of 20 frames, we surpass all works of the state of the art. When the delay is 40 frames, we reach the rate 97.15%.

Table 2 Average accuracy (%) comparison of direction estimation results over Casia-B dataset with state-of-the-art methods

| Our (0) | Our (20) | Our (40) | [10] | [9] | [5] | [3] |
|---------|----------|----------|-------|-------|-------|-------|
| 83.65 | 91.66 | 97.15 | 90.87 | 84.61 | 72.50 | 64.67 |

Table 3 Accuracy (%) comparison of direction estimation results over Casia-A dataset with [10]

| D_i | Our (0) | Our (20) | Our (40) | [10] |
|-------|---------|----------|----------|-------|
| d_1 | 79.73 | 84.37 | 94.44 | 92.5 |
| d_2 | 93.22 | 98.001 | 99.91 | 92.5 |
| d_3 | 91.51 | 97.34 | 100 | 98.75 |
| BA | 88.15 | 93.23 | 98.11 | 94.58 |

Table 4 Average accuracy (%) comparison of direction estimation results over Casia-A dataset with state-of-the-art methods

| Our (0) | Our (20) | Our (40) | [10] | [9] | [5] | [3] |
|---------|----------|----------|-------|-------|-------|-------|
| 88.15 | 93.23 | 98.11 | 94.58 | 90.32 | 77.67 | 66.42 |

We give in Table 3 the accuracy rate for each estimated direction of Casia-A dataset. The results obtained in [10] are also reported. The average accuracy is compared with the state-of-the-art methods and given in Table 4. With a delay of 30 frames, we surpass the published works and we reach the rate 98.11% with a delay of 40 frames.

6.5.2 Campus dataset

In the first step, silhouettes are extracted applying our proposed method. Once extracted, silhouettes are tracked and their top and bottom points are located. The fitting lines are then computed, and the direction is derived. The error in direction depends on the segmentation errors and on the number of used silhouettes. Note that the more silhouettes are used, the more the error decreases. Our method takes into account the direction change as described in algorithm 2. For direction change, new fitted lines are considered because the new top and bottom points are far from the previous fitted lines. Figure 12 shows an example where the error on computed direction decreases for each processed frame. The different tests indicate that mean error μ , σ and standard deviation depend on the delay. With a delay of 30 frames, the direction is estimated with accuracy, $(\mu, \sigma) = (4.7, 4.4)$. For delay of 0, 10, 20, the (μ, σ) is equal, respectively, to $(14.9, 18.2)$, $(10.9, 13.2)$, $(7.7, 9.4)$.

6.5.3 MGP01 dataset

The same experiments have been conducted on MGP01 dataset. Note that this dataset contains groups of moving

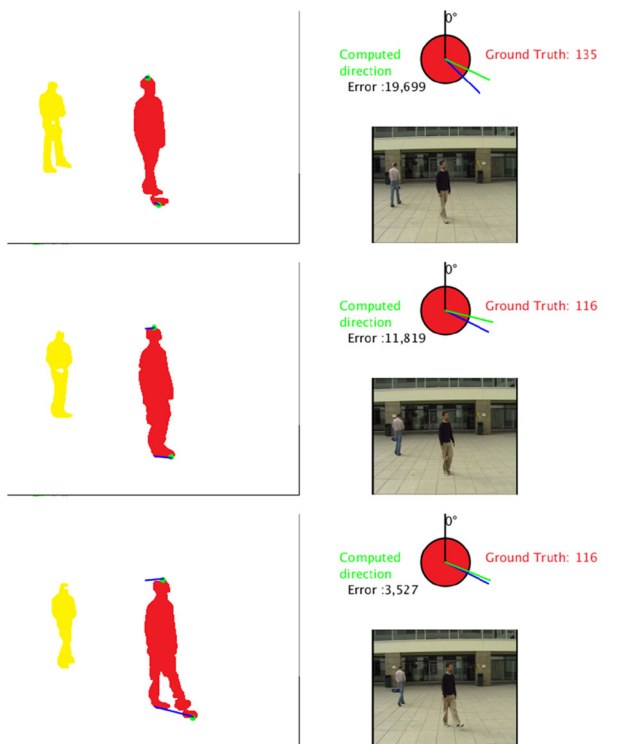


Fig. 12 Sample of direction estimation for Campus dataset

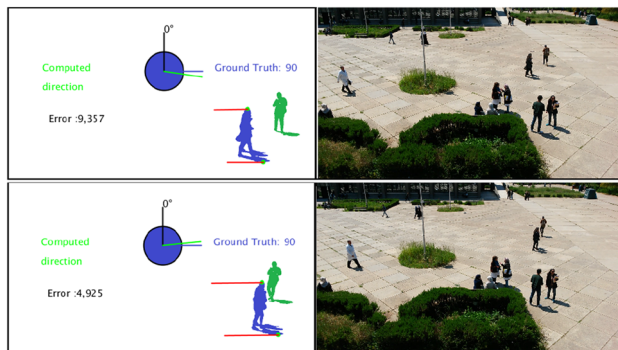


Fig. 13 Sample of direction estimation of moving group of pedestrians for MGP01 dataset

pedestrians. Figure 13 shows despite the effect that shadow is present, our method delivers correct direction for pedestrian group whatever their orientation. Direction is estimated with accuracy starting from 30 frames, $(\mu, \sigma) = (5.58, 2.53)$. However, for single pedestrian, for a delay of 30 frames, $(\mu, \sigma) = (7.67, 3.9)$. Note that pedestrian group in this dataset moves as a block because pedestrians have the same positioning inside the group. As the shadow follows a linear trajectory, the computed vanishing point for each frame is inside of the directional area.

6.6 Discussion

Accurate results are obtained for Campus dataset relatively to other ones because the frame segmentation is relatively more correct. Our method competes the state-of-the-art methods and provides for moving pedestrians (single or group) an accurate estimation of the direction and achieves a high rate accuracy for direction classification starting from the 20th frame . In other side, we did not compare our results on MGP01 dataset because we did not find work devoted to direction estimation of walking pedestrian group, which is crucial for intelligent vehicles. In addition to our proposed dataset, there is a need to put available datasets made in urban scenes from vehicle with the presence of walking pedestrian groups.

7 Conclusion

We mapped in this work the walking direction of pedestrians from 3D to 2D by introducing the notion of directional areas in image plane. Saying in what direction is walking the pedestrian is now synonymous to say to what directional area the pedestrian’s silhouette is going. Compared to the state of the art, our method is not sensitive to noise, fast and delivers an accurate estimation of the direction. As future works, we are interested to apply our method for direction estimation of rigid objects (vehicles, bicycle).

Compliance with ethical standards

Conflict of interest The authors Amina Bensebaa and Slimane Larabi declare no competing financial interests.

Funding This research was not funded.

References

1. Shimizu, H., Poggio, T.: Direction estimation of pedestrian from multiple still images. In: IEEE, Intelligent Vehicles Symposium, Parma, Italy, 14–17 June 2004
2. Gandhi, T., Trivedi, M.M.: Image based estimation of pedestrian orientation for improving path prediction. In: IEEE Intelligent Vehicles Symposium, Eindhoven University of Technology, Eindhoven, The Netherlands, 4–6 June 2008
3. Chen, C., Heili, A., Odobez, J.M.: Combined estimation of location and body pose in surveillance video. In: 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 5–10 (2011)
4. Guangzhen, Z., Mrutani, T., Kajita, S., Mase, K.: Video based estimation of pedestrian walking direction for pedestrian protection system. J. Electron. **29**(1/2), 72–81 (2012)
5. Baltieri, D., Vezzani, R., Cucchira, R.: People orientation recognition by mixtures of wrapped distributions on random trees.

- In: Proceedings of European Conference on Computer Vision (ECCV), pp. 270–283 (2012)
6. Tao, J., Klette, R.: Integrated pedestrian and direction classification using a random decision forest. In: Conference: IEEE International Conference on Computer Vision (ICCV), Sydney, pp. 230–237 (2013)
 7. Flohr, F., Dumitru-Guzu, M., Kooij, J.F.P., Gavrila, D.M.: Joint probabilistic pedestrian head and body orientation estimation. In: IEEE Intelligent Vehicles Symposium (IV), Dearborn, Michigan, USA, 8–11 June 2014
 8. Ricci, E., Varadarajan, J., Subramanian, R., Buló, S.R., Ahuja, N., Lanz, O.: Uncovering interactions and interactors: joint estimation of head, body orientation and F-formations from surveillance videos. In: IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 Dec 2015
 9. Liu, H., Ma, L.: Online person orientation estimation based on classifier update. In: Proceeding of IEEE International Conference on Image Processing ICIP, pp. 1568–1572 (2015)
 10. Raman, R., Sa, P.K., Majhi, B., Bakshi, S.: Direction estimation for pedestrian monitoring system in smart cities: an HMM based approach. *IEEE Access* **4**, 5788–5808 (2016)
 11. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: ICPR, Hong Kong, China (2006)
 12. Huang, S., Ying, X., Rong, J., Shang, Z., Zha, H.: Camera calibration from periodic motion of a pedestrian. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA (2016)
 13. Junejo, I.N.: Using pedestrians walking on uneven terrains for camera calibration. *Mach. Vis. Appl.* **22**(1), 137–144 (2011)
 14. Guan, J., Deboeverie, F., Slembrouck, M.: Extrinsic calibration of camera networks based on pedestrians. *Sensors* **16**, 654 (2016). <https://doi.org/10.3390/s16050654>
 15. Raza, M., Chen, Z., Rehman, S., Wang, P., Bao, P.: Appearance based pedestrians head pose and body orientation estimation using deep learning. *Neurocomputing* **272**, 647–659 (2018)
 16. Varadarajan, J., Subramanian, R., Rota Buló, S., Ahuja, N., Lanz, O., Ricci, E.: Joint estimation of human pose and conversational groups from social scenes. *Int. J. Comput. Vis.* **126**(2–4), 410–429 (2018)
 17. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with HSV color information. In: IEEE Intelligent Transportation Systems Conference Proceedings, Oakland, 25–29 Aug 2001
 18. MGP01, <http://perso.usthb.dz/~slarabi/MGP01/index.html> (2017)
 19. Lv, F., Zhao, T., Nevatia, R.: Self-calibration of a camera from video of a walking human. In: IEEE International Conference of Pattern Recognition, pp. 281–304 (2002)
 20. Berclaz, J., Fleuret, F., Turetken, E., Fua, P.: Multiple object tracking using K-shortest paths optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(9), 1806–1819 (2011)
 21. Fleuret, F., Berclaz, J., Lengagne, R., Fua, P.: Multi-camera people tracking with a probabilistic occupancy map. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 267–282 (2008)



Amina Bensebaa received the master degree in Intelligent Informatics Systems from the University of Sciences and Technology Houari Boumediene, Algeria, in 2012. She is also a PhD student at the University of Sciences and Technology Houari Boumediene, Algeria, under Professor Larabi's supervision. Her current research interests concern image and video analysis, content-based image retrieval and pattern recognition.



Slimane Larabi received his PhD in Computer Science from the National Institute Polytechnic of Toulouse, France, 1991. In January 1992, he joined the Computer Science Department of US-THB University in Algeria, where he is currently Professor. He leads research in Computer Vision Group of the Laboratory of Artificial Intelligence Research. His work spans a range of topics in vision including: image description, human action recognition, head and body pose estimation and video analysis. He also proposed and leads the Master of Visual Computing in the same university and teaches several courses: data visualization, game design, multimedia systems, artificial intelligence and computer vision. He conducted many projects in different areas, especially in computer vision, augmented reality, data visualization.