Inferring Heading Direction from Silhouettes

Amina Bensebaa, Slimane Larabi and Neil M. Robertson

Abstract Due to the absence of features that may be extracted from face, heading direction estimation for low resolution images is a difficult task. For such images, estimating heading direction requires to taking into account all information that may be inferred from human body in image, particularly its silhouette. We propose in this paper a set of geometric features extracted from shape shoulders-head, feet and knees shapes which jointly allow the estimation of body direction. Other features extracted from head-shoulders are proposed for the estimation of heading direction based on body direction. The constraint of camera position related to proposed features is discussed and results of experiments conducted are presented.

1 Introduction

Heading direction estimation is one of challenging tasks for computer vision researchers especially in case of low resolution images. In case of high and medium resolution images, many approaches has been proposed to solve this problem. A survey may be found in [11]. All of these approaches try to find the most discriminate set of facial features which permit to estimate the pose. The objective to reach for any proposed technique is to verify a set of criteria such as: Accuracy, Monocular, Autonomous, Multi-person, Identity and Lighting invariant, Resolution independent, Full range of head motion and Real time [11].

Face extraction in low-resolution images is an important task in the process of heading direction estimation. Few works have been devoted for this purpose and

A. Bensebaa (✉) · S. Larabi
Computer Science Department, USTHB University, BP 32 El Alia, Algiers, Algeria
e-mail: bensebaa.amina@gmail.com

S. Larabi
e-mail: slarabi@usthb.dz

N. M. Robertson
Edinburgh Research Partnership in Engineering and Mathematics, Heriot-Watt University, Edinburgh, EH14 4AS, UK
e-mail: n.m.robertson@hw.ac.uk

all present difficulties for detecting faces when the resolution of images decreases [18]: Labeled training examples of head images are used to train various types of classifiers such as support vector machines, neural networks, nearest neighbor and tree based classifiers [3, 4, 13]. The disadvantage of these methods is the requirement of all combinations of lighting conditions and skin/hair colour variations in order to estimate an accurate classification.

Contextual features has been used in addition to visual ones in order to improve the quality of heading direction estimation [1, 8, 9]. Using multiple views camera, Voit et al. [17] estimate head pose for low resolution image by appearance-based method. The head size varies around 20×25 and the obtained results are satisfactory due to the use of multiple cameras. Additional contextual information: multiple calibrated camera and a specific scene allows estimating of absolute coarse head pose for wide-angle overhead cameras by integrating 3D head position [16].

Head-shoulders shape has been studied and many methods have been proposed for the purpose of human detection in images using wavelet decomposition technique and support vector machine [14] or background subtraction algorithm [12]. In other side, Head-shoulders shape has been used for human tracking and head pose estimation. In [12], the direction of head movements is detected and tracked throughout video frames. Templates are captured for a specific position of the camera (mounted sufficiently high above to provide a top-view of the scene) and do not use all positions of the head pose. Shape context is used but this descriptor is sensitive to the locations of pixels of the shape outline.

Another important feature that may contribute for heading-direction estimation is the legs shape. However, the use of detectors on the lower parts of the body has been introduced in many works for human body pose calculation and human action recognition [15]. Legs shape has been also used for human segmentation. Lin et al. [10] modeled the parts of the body particularly the legs in order to detect and segment human. The proposed approach is based on the matching of part-template tree images hierarchically proposed and used initially in [6, 7].

The problem or heading for low-resolution images without adding contextual information requires yet more contributions in order to deal with complex scenes where human are relatively far from the camera. The performance of proposed methods are principally limited because they are based on extracted features from the head which are very dependent on camera placement and the chosen texture and skin color models depend on the resolution of the head in the image and therefore doesn't work for lower resolution.

In this paper, we investigate what can be done from shoulders-head and legs shapes for heading direction estimation in case of low-resolution images. Firstly, a set of features are extracted from shoulders-head and legs shapes and used for inferring body direction. In the next, heading direction is estimated using body direction and features extracted from head-shoulders shape. Section 2 covers the theoretical aspects of body and heading direction estimation based on features extracted from shoulders-head and legs shapes. Experiments are conducted to validate our approach and obtained results are presented in Sect. 3.

Fig. 1 Some shapes of legs for which it is easy to infer body direction

Fig. 2 Shapes of legs with inflected knee

2 Basic Principle of the Method

Assuming that silhouettes of humans are extracted from images of low resolution, our aim is to estimate body and heading directions. Geometric features are extracted from silhouette due to the absence of other features that may be extracted from the face for such images. We will focus in this paper on the parts head, shoulders, knees and feet shapes which may be considered as a good features to achieve this task. Body direction is firstly estimated using features extracted from head and shoulders, knees and feet shapes. Secondly, heading direction is inferred from estimated body direction and features of head and shoulders shape.

2.1 Features Extraction from Silhouette

A shape leg is a part of human silhouette which plays a dominant role in the process of inferring body direction from image. Indeed, our visual system is able to infer body direction seeing only the outline shape legs (see Fig. 1). We propose three determinant cues of shapes legs and head-shoulders that allow inferring body direction when they are extracted from outline shape. These features cannot be computed for a fixed top down camera because head-shoulders are confused with body silhouette.

The **first one** is the inflections of the knees. When a leg is well separated from the other and the knee is inflected, a coarse body direction can be inferred without ambiguity. Figure 2a illustrates an example of shape legs where feet are cut. Our visual system can easily give an estimate of body direction because the feet have limited possibilities of poses due to the geometry of one leg (high inflexion). Figure 2b illustrates the correct poses and the directions can be inferred using the feet shapes, however Fig. 2c shows impossible situation. The directions of the lines joining inflexion points of the same leg are used to infer the body direction.

The **second one** is the direction of shape foot. Indeed, our visual system encounters difficulties by looking at legs shapes without feet and cannot estimate body direction for many configurations even if the body is moving and legs are well separated but without inflexion of knees. For example, seeing to the outlines of Fig 3a, without
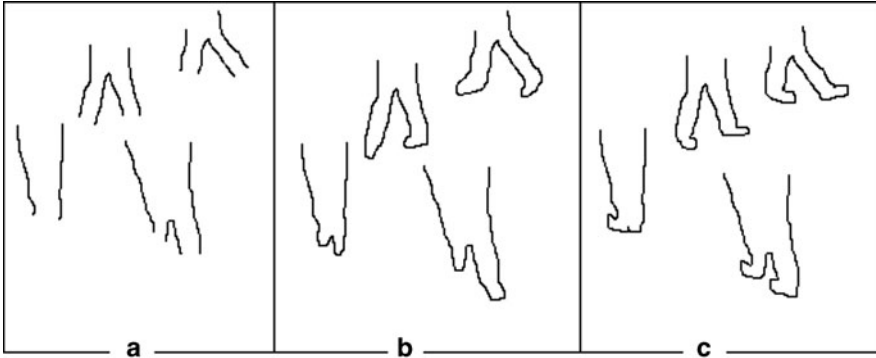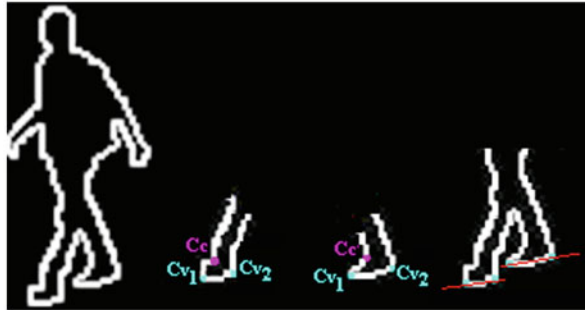
Fig. 3 Ambiguity in body direction estimation in case of missed shape feet

Fig. 4 Steps of body
direction estimation based on
foot directions

feet we cannot recognize to what direction body is moving. This ambiguity is clear seeing at the original shapes (see Fig. 3b) and at new shapes obtained drawing feet (see Fig. 3c). The base lines of the feet are good features because they indicate the body direction. Their use is explained in Sect. 2.2.

The **third feature** concerns the variation of silhouette's width along the shape head-shoulders and the length of each shoulder. The ratio of the width of the upper part (head) and the lower part (shoulders) with the varying of the shoulders length are related to the angle of rotation. We noticed that there's an opposite relationship between the ratio and the orientation angle.

2.2   Inferring Body Direction

Body Direction Estimation Using Feet's Features: This task consists to split the lower human shape into separated legs, separated lower legs or grouped legs (The two first cases include the case where the knee of one leg is inflected). We associate to each foot a **base line** defined by two extremities of the foot located between the heel and the toes. The outline of lower part is processed in order to determine the baseline of the feet located between the heel and the toes. Firstly, high convexities points Cv_1 and Cv_2 characterizing the outline foot are located (see Fig. 4). Secondly,

Fig. 5 Body orientation from feet (In *red color* the feet orientations of foot and in *blue* the body orientation)
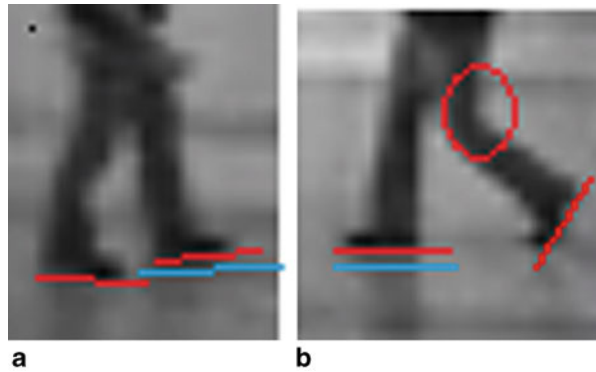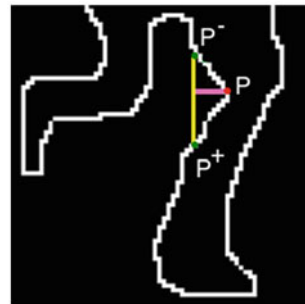

Fig. 6 Location of inflection points on outline legs

the last point of interest Cc representing a high concavity on this outline is located, such as the distances $CcCv_2$ is minimal. The convex point that represents toes, will be the closest point to the concave point of the feet outline, the other convex point will obviously correspond to the heel. Thus the base line joins the two convexities of the foot and the orientation of feet corresponds to the vector carried by the feet base line.

Applying the 2D quasi-invariant, the angle between the two vectors measured in 3D-space varies slowly in the image as viewpoint varies [2]. As in the scene the disposition of foot vectors is restricted by the human physic constraints, it will be the same case in image plane; the body direction is inferred as the average of foot directions. Once the base lines of feet are extracted, body orientation is computed as the resultant vector of the two orientations (see Fig. 5a). When one foot is not put on the ground, which correspond to a high inflection of the knee, the resultant vector will have the direction of the base line of the other foot (see Fig. 5b).

Body Direction Estimation Using Knee's Features: Extraction of inflection points consists to find the best concave or convex pixels of the lower part of the silhouette using the Chetverikov's algorithm [5]. Among the selected points of inflection p, p^* which is the farthest to the line binding p^- and p^+ is chosen. The position of p^-, p^+ to p^* is a parameter (see Fig. 6).

Many types of knees inflexion may be located (see Fig. 7). The direction of the body follows the direction of the inflected knee considered as the direction of the
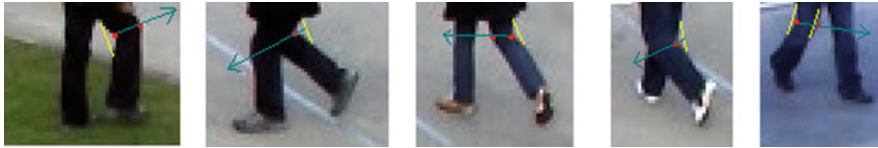
Fig. 7 Some cases of knee inflexion and the inferred direction of them

Fig. 8 The pixel *p* is the farthest from the line *L*

Table 1 Body direction inferred from head-shoulders features

Body direction	Ratio R_w
[0°, 15°]	≥ 1.82
[15°, 30°]	[1.70, 1.81]
[30°, 45°]	[1.61, 1.69]
[45°, 60°]	[1.51, 1.60]
[60°, 75°]	[1.36, 1.5]
[75°, 90°]	[1.4, 1.5]

line joining the concave point to the convex one. Only the direction left towards right and inversely will be considered.

Body Direction Estimation Using Head-Shoulders Features: Applying the algorithm of D. Chetverikov [5], the two concave points (left and right) delineating the head and the two convex points (left and right) extremities of shoulders are located. Head is separated by locating the pixel having the minimum angle among the selected point candidates. The two convex pixels are located based on high curvature. Each pixel is characterized by the fact that it is the farthest from the line (L) connecting the beginning of the shoulder and the end pixel of the head-shoulders outline (see Fig. 8).

When human is in the centre of field view of the camera, the average of computed ratios R_w (ratio of the widths of head and shoulders) estimated are given by Table 1 and the Fig. 9 illustrates an example corresponding to the rotation of a person towards the left using the ratio R_w of head-shoulders.
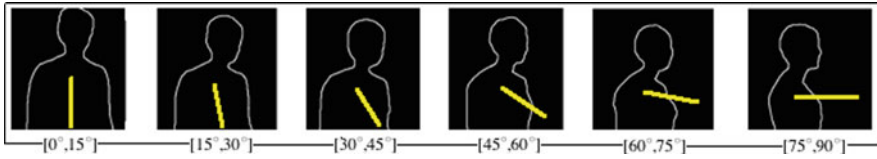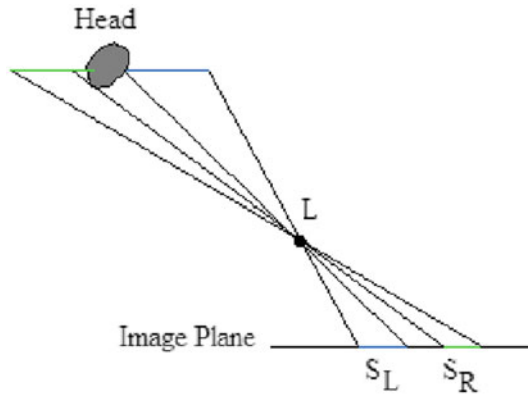
[0°,15°] — [15°,30°] — [30°,45°] — [45°,60°] — [60°,75°] — [75°,90°]

Fig. 9 Estimating body direction using the ratio R_w

Fig. 10 Case of occlusion of shoulder by head

2.3 Inferring Head Direction from Shoulders-Head Shape

We assume now that body direction is estimated based on the three features proposed above (head-shoulders, knee inflexion and feet). In order to estimate the heading direction, we will base our approach on two features extracted from head-shoulders outline.

Features Extraction The **first feature** concerns the lengths of shoulders S_L and S_R on shape head-shoulders. In some cases, the end of the neck is not visible on one side due to head occlusion. In this case, it will be replaced by the point of high curvature on head-shoulders outline.

The lengths of shoulders are important cues for both head and body directions estimation and the difference between lengths of S_L and S_R arises from one of the following configurations:

- Depending on the camera and body positions, the head can occlude a part of one shoulder and then decreases the shoulder length. For example, when the camera is on top at the right or at the left of the person (see Fig. 10).
- When human body is rotating, one of shoulders becomes less visible. This occurs for example when the camera is on top even if the person is in front of the camera. In this case, length of one shoulder decreases until that the two sides of the shape head-shoulders do not correspond to shoulders.

Consequently, when the direction of body and head is in front to the camera, the lengths $L(S_L)$, $L(S_R)$ of shoulders are identical. Otherwise, when the head is rotating or when body is at the lateral side of the camera, this equality is not verified because

Fig. 11 Intersection of shoulders in case where **a** body and head are in front, **b** body and head rotating
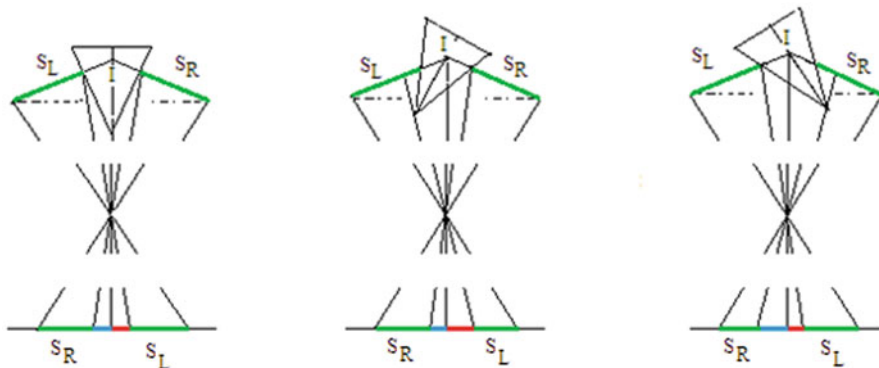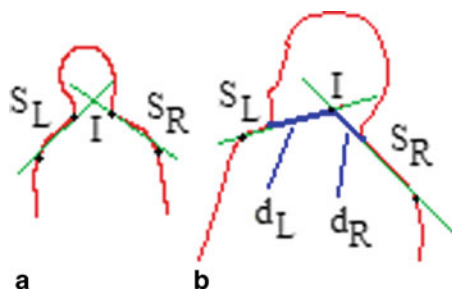

Fig. 12 Different poses of head where d_R, d_L are illustrated with *blue* and *red color* in case of human is in the *center* of the field of view

in both cases the head occludes a part of one shoulder (see Fig. 10). We proved geometrically that without occlusion by head, the lengths of one shoulder decreases when body is rotating.

The **second feature** which completes the first one, concerns the occluded parts of shoulders that permit to estimate head rotation. Let I be the intersection point of the lines joining extremities of shoulders S_L and S_R (see Fig. 11). When body and head are in front to the camera, the distances d_L and d_R from I to shoulders are identical in the scene and in image plane. However, when head or body are rotating, these distances are different in image because a part of shoulder is occluded by head and thus in image the distance d_L or d_R includes the occluded segment of the shoulder and a part of the neck. The distances d_L, d_R will be used to infer the heading direction.

Coarse Estimation of Head Direction Heading direction is estimated assuming that in previous steps, the body orientation, the difference ΔL between the lengths of shoulders (S_L) and (S_R) and the difference Δd between the distances d_L and d_R are computed. We distinguish three cases: body is in the center, at the left, or at right of the view field. For the two first cases, We give in Table 2 the results obtained of heading direction applying a geometric reasoning depending on the values of ΔL and δd and body direction. The third case is symmetrical to the second one. Figure 12 illustrates the variation of ΔL and δd in case where human in the center of the field of view of the camera.

Table 2 Heading direction inferred in cases where body is in front and at the left

Body in the center	$\delta d = 0$	$\delta d > 0$	$\delta d < 0$
$\Delta L = 0$	Head in Front	Not possible	Not possible
$\Delta L < 0$	Not possible	Rotation to left	Not possible
$\Delta L > 0$	Not possible	Not possible	Rotation to right
Body at the left			
$\Delta L = 0$	Low Rotation to right	Not possible	Not possible
$\Delta L < 0$	Not possible	Head in front or rotating to left	Not possible
$\Delta L > 0$	Not possible	Not possible	Hight rotation to right

2.4 Study of the Camera Position Constraint

As we are interested in this work to images of low resolution which means a far field of view, the camera may be:

- Fixed at the top and far from the scene. In this case, none from the features: head, shoulders, legs and feet can't be located using the blob representing human.
- Fixed so as its optical axis is oblique or horizontal towards the scene. in this case, whatever the position of the camera relatively to human in the scene: in front or at the lateral position, its head-shoulders, legs and feet are viewed. Consequently, the availability of the proposed features depends only on the pose, which means that inflexion of knees or feet base lines may be missed, what is required is the presence of the head-shoulders outline.

3 Results

We applied our method on PETS data set. Firstly silhouettes are extracted and body direction is firstly computed. In the next, heading direction is estimated. We used all features extracted from head-shoulders, feet and knees outlines.

Figure 13 illustrates some poses, extracted silhouettes and computed body directions. Body direction is computed using the ratio R_w having respectively the values
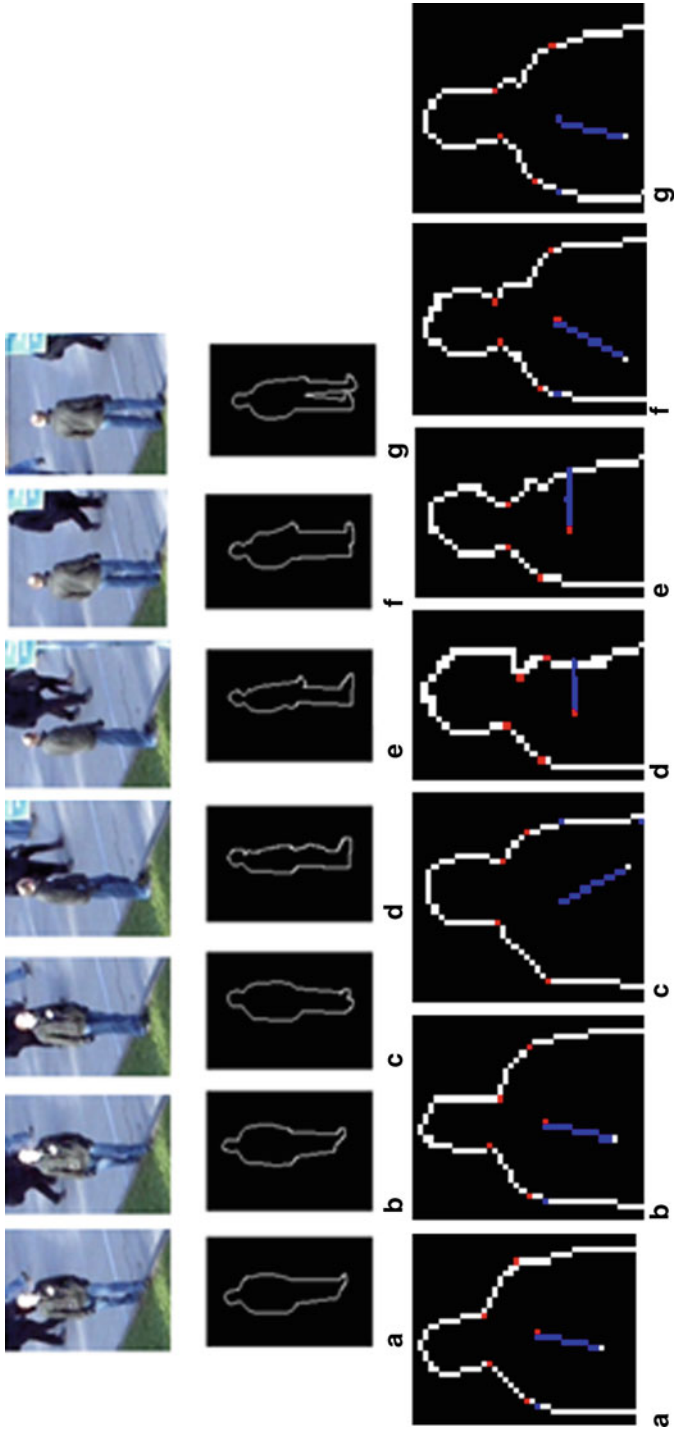
Fig. 13 Some poses and extracted silhouettes and the computed body directions based on R_w values

2.6, 2.89, 2.25, 1.33, 1.36, 2.27, 2.09 giving the directions: $[0°, 15°], [0°, 15°],$ $[15°, 30°], [75°, 90°], [75°, 90°], [15°, 30°], [0°, 15°]$. As the computed body direction for the two last poses $(f), (g)$ are done using only the first feature which cannot differentiate if the body is in front or of back with regard to the camera.

The orientation of feet, when are located in the image, eliminates the ambiguity (in front or of back). Figure 14 illustrates some body poses which combine only features of head-shoulders and feet (knees inflexions are not visible).

The combination of features used for body direction depends on what can be extracted in image. The features extracted from feet and knees are more strong than those extracted from head-shoulders which just allows us to calculate the direction. Figure 15 illustrates the results obtained when inflexion of knees are used in addition of the ratio R_w.

Heading direction estimation is based on estimated body direction and the values of d_L, d_R computed using head-shoulders outline. We can see in Figure 16 the use of all presented features for estimating heading direction. Figure 17 summarizes this combination of features and shows that a good estimation is made even if the images are of low resolution.

4   Conclusion

We proposed in this paper a method for heading direction for images based on geometric features which can be extracted from silhouette even if images are of low resolution. Body direction is inferred from features extracted from outlines of knees and feet and head-shoulders. This direction is used in addition to features extracted from outlines of head-shoulders for estimating heading direction. The proposed method has been applied on real images and achieves good estimation of heading direction. Also, the features extracted are independent from camera pose, except the top view where head-shoulders, knees and feet cannot be located on human shape.
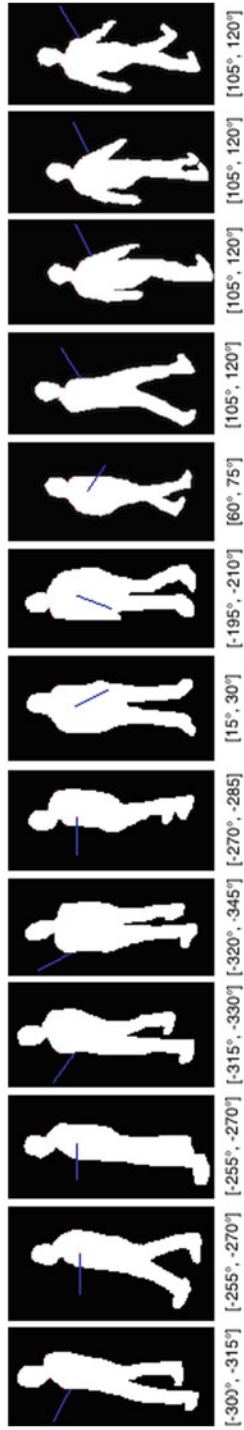
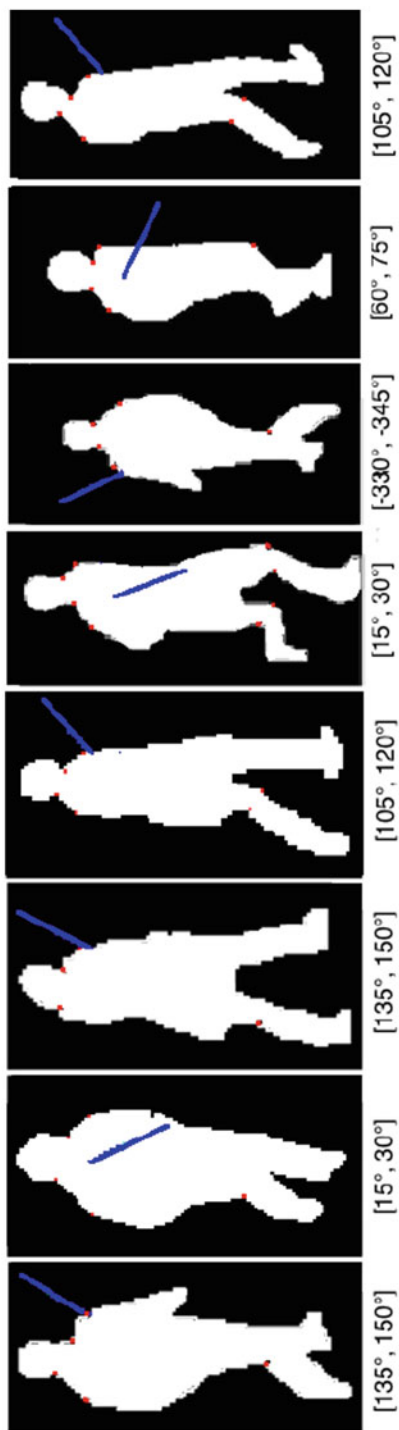**Fig. 14** Body orientation using the features: feet and R_w ratio

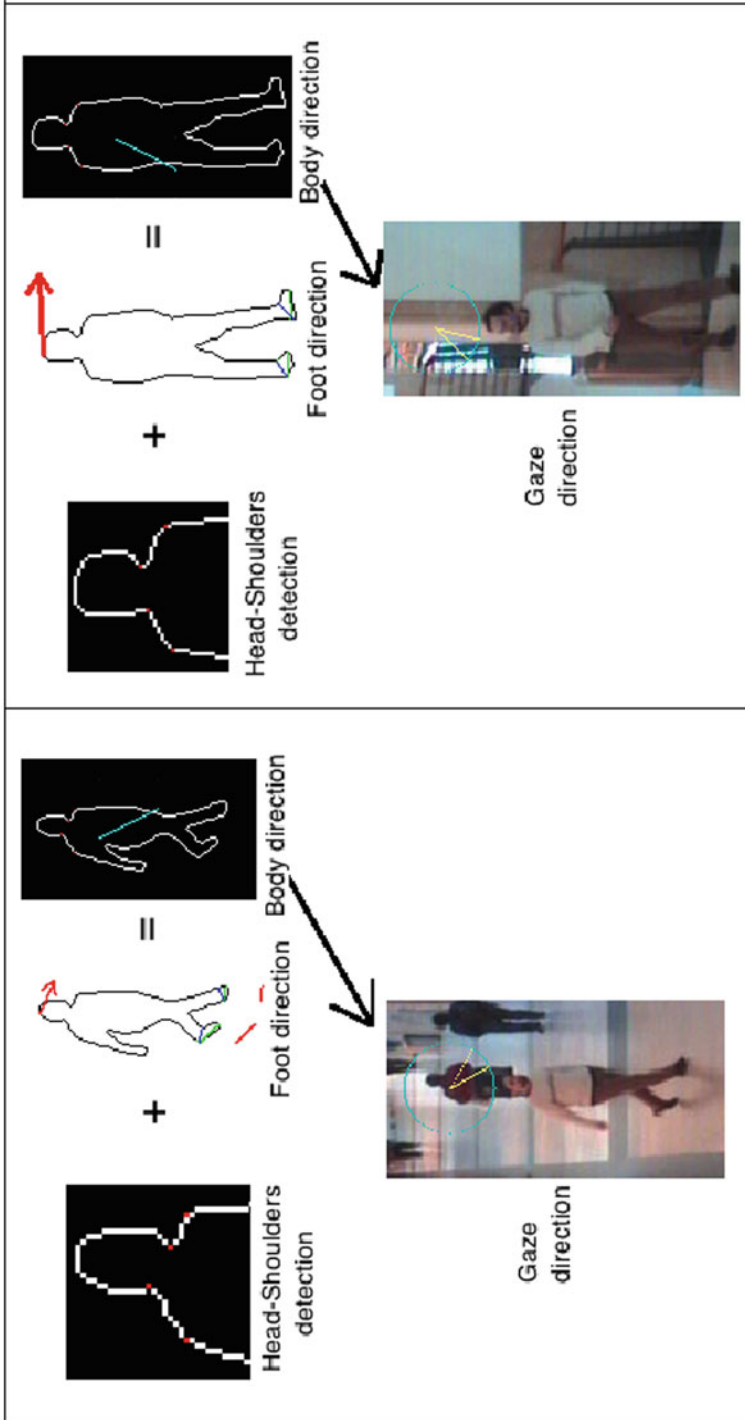**Fig. 15** Body orientation using the features: knee inflexion and R_w ratio

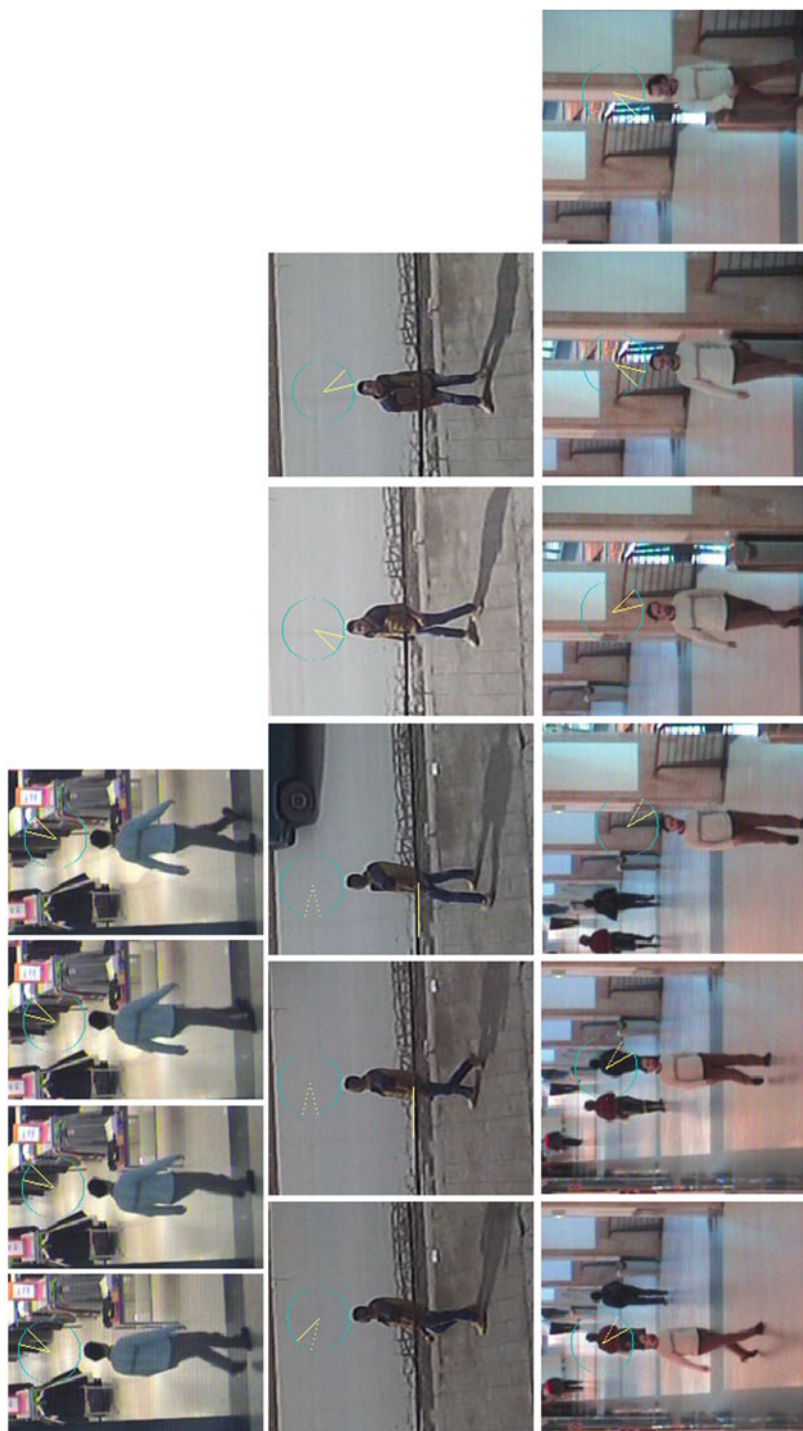**Fig. 16** Step of heading direction estimation

Fig. 17 Heading and body directions from combined features

References

1. Ba SO, Odobez JM (2011) Multiperson visual focus of attention from head pose and meeting contextual cues. IEEE Trans Pattern Anal Mach Intell 33(1):101–116
2. Binford TO, Levitt TS (1993) Quasi-invariants: theory and exploitation. In: Proceedings of DARPA Image Understanding Workshop, pp 819–829
3. Benfold B, Reid I (2008) Colour invariant head pose classification in low resolution video. In: Proceedings of the 19th British Machine Vision Conference
4. Benfold B, Reid I (2011) Unsupervised learning of a scene-specific coarse gaze estimator. In: Proceedings of the International Conference on Computer Vision (ICCV), pp 2344–2351
5. Chetverikov D (2003) A simple and efficient algorithm for detection of high curvature points in planar curves. In: Computer analysis of images and patterns, 10th international conference, CAIP 2003, Groningen, the Netherlands, August 2003, pp 25–27
6. Gavrila DM (1999, Jan) The visual analysis of human movement: a survey. Comput Vis Image Underst 73(1):8298
7. Gavrila DM (2007) A bayesian, exemplar-based approach to hierarchical shape matching. IEEE Trans Pattern Anal Mach Intell 29(8):1408–1421
8. Lanz O, Brunelli R (2008) Joint Bayesian tracking of head location and pose from low-resolution video. In: Multimodal technologies for perception of humans, pp 287–296
9. Launila A, Sullivan J (2010) Contextual features for head pose estimation in football games. In: International conference on pattern recognition (ICPR 2010), Turkey, pp 340–343
10. Lin Z, Davis LS (2010) Shape-based human detection and segmentation via hierarchical part-template matching. IEEE Trans Pattern Anal Mach Intell 32(4):604–618
11. Murphy-Chutorian E, Trivedi MM (2009, April) Head pose estimation in computer vision: a survey. Pattern Anal Mach Intell, IEEE Trans 31(4):607–626
12. Ozturk O, Yamasaki T, Aizawa K (2009) Tracking of humans and estimation of body/head orientation from top-view single camera for visual focus of attention analysis. In: IEEE 12th international conference on computer vision workshops (ICCV workshops), pp 1020–1027
13. Robertson NM, Reid ID (2006) A general method for human activity recognition in video. Comput Vis Image Underst 104(2–3):232–248
14. Sun Y, Wang Y, He Y, Hua Y (2005) Head-and-shoulder detection in varying pose. In: Advances in natural computation, first international conference, ICNC, Changsha, China, pp 12–20
15. Singh VK, Nevatia R, Huang C (2010) Efficient inference with multiple heterogeneous part detectors for human pose estimation. In: Computer vision ECCV 2010, pp 314–327
16. Tian YL, Brown L, Connell C, Sharat P, Arun H, Senior A, Bolle R (2003) Absolute head pose estimation from overhead wide-angle cameras. In: IEEE international workshop on analysis and modeling of faces and gestures, AMFG 2003, pp 92–99
17. Voit M, Nickel K, Stiefelhagen R (2006) A Bayesian approach for multi-view head pose estimation. In: IEEE international conference on multisensor fusion and integration for intelligent systems, pp 31–34
18. Zheng J, Ramirez GA, Fuentes O (2010) Face detection in low-resolution color images. In: Proceedings of the 7th international conference on image analysis and recognition, ICIAR'10, Portugal, pp 454–463