Estimation of Human Orientation in Images Captured with a Range Camera

Sébastien Piérard, Damien Leroy,
Jean-Frédéric Hansen, and Marc Van Droogenbroeck

INTELSIG Laboratory, Montefiore Institute, University of Liège, Belgium

Abstract. Estimating the orientation of the observed person is a crucial task for some application fields like home entertainment, man-machine interaction, or intelligent vehicles. In this paper, we discuss the usefulness of conventional cameras for estimating the orientation, present some limitations, and show that 3D information improves the estimation performance.

Technically, the orientation estimation is solved in the terms of a regression problem and supervised learning. This approach, combined to a slicing method of the 3D volume, provides mean errors as low as 9.2° or 4.3° depending on the set of considered poses. These results are consistent with those reported in the literature. However, our technique is faster and easier to implement than existing ones.

1 Introduction

The real-time interpretation of video scenes is a crucial task for a large variety of applications. As most scenes of interest contain people, analyzing their behavior is essential. It is a challenge because humans can take a wide variety of poses and appearances. In this paper, we deal with the problem of determining the orientation of persons observed laterally by a single camera (in the following, we use the term *side view*). To decrease the sensitivity to appearance, we propose to rely on geometrical information rather than on colors or textures.

There are many applications to the estimation of the orientation of the person in front of the camera: estimating the visual focus of attention for marketing strategies and effective advertisement methods [11], clothes-shopping [17], intelligent vehicles [5,6], perceptual interfaces, facilitating pose recovery [8], etc.

In most applications, it is preferable to observe the scene from a side view. Indeed it is not always possible to place a camera above the observed person. Most ceilings are not high enough to place a camera above the scene and to observe a wide area. The use of fisheye lenses raises a lot of difficulties as silhouettes then depend on the precise location of a person inside the field of view. Moreover, in the context of home entertainment applications, most of existing applications (such as games) already require to have a camera located on top or at the bottom of the screen. Therefore, in this paper, we consider a single camera that provides a side view.

J. Blanc-Talon et al. (Eds.): ACIVS 2011, LNCS 6915, pp. 519–530, 2011.

This paper explains that there is an intrinsic limitation when using a color camera. Therefore, we consider the use of a range camera, and we infer the orientation from a silhouette annotated with a depth map. We found that depth is an appropriate clue to estimate the orientation. The estimation is expressed in terms of regression and supervised learning. Our technique is fast, easy to implement, and produces results competitive with the known methods.

The paper is organized as follows. The remainder of this introduction describes related works, defines the concept of orientation, and elaborates on how we evaluate methods estimating the orientation. In Section 2, we discuss the intrinsic limitation of a color camera. Section 3 details our method based on a range camera. In Section 4, we describe our experiments and comment on the results. Finally, Section 5 concludes the paper.

1.1 Related Works

Existing methods that estimate the orientation differ in several aspects: number of cameras and viewpoints, type of the input (image or segmentation mask), and type of the output (discrete or continuous, *i.e.* classification or regression). As explained hereinbefore, it is preferable to use a side view and to rely on geometric information only. Accordingly, we limit this review of the literature to methods satisfying these requirements.

Using a single silhouette as input, Agarwal *et al.* [1] encode the silhouette with histogram-of-shape-contexts descriptors [3], and evaluate three different regression methods. They obtained a mean error of 17°, which is not accurate enough to meet the requirements of most applications. We will explain, in Section 2, that there is an intrinsic limitation when only a single side view silhouette is used.

To get more discriminant information, some authors [4,9] consider the dynamics and assume that the orientation is given by the walking direction. This direction is estimated on the basis of the temporal evolution of the foreground blob location and size. Therefore, those methods require that the person is in motion to determine his orientation. Obviously, there is no way to ensure it.

Multiple simultaneous silhouettes can also be used to improve the orientation estimation. Rybok *et al.* [14] establish that the use of several silhouettes leads to better results. They consider shape contexts to describe each silhouette separately and combine individual results with a Bayesian filter framework. Peng *et al.* [12] use two orthogonal views. The silhouettes are extracted from both views, and processed simultaneously. The decomposition of a tensor is used to learn a 1D manifold. Then, a nonlinear least square technique provides an estimate of the orientation. Gond *et al.* [8] use the 3D visual hull to recover the orientation. A voxel-based *Shape-From-Silhouettes* method is used to recover the 3D visual hull. All these methods require the use of multiple sensors, which is inconvenient.

1.2 Our Definition of the Orientation

The cornerstone observation for orientation estimation is that orientation is independent of the pose: the orientation is related to the coordinate system of the

scene, whereas the pose is specific to the body shape. To achieve this independence, it is convenient to define the orientation of a person with respect to the orientation of one body rigid part. In this paper, we use the orientation of the pelvis. The orientation $\theta = 0°$ corresponds to the person facing the camera (see Figure 1).

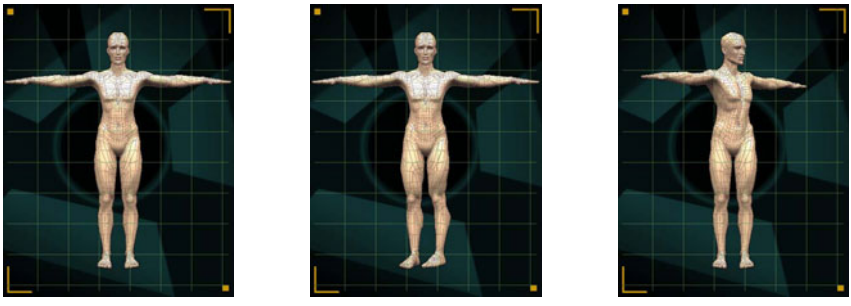

Fig. 1. Defining the orientation of a person requires the choice of a body part. In this paper, we use the orientation of the pelvis. This figure depicts three examples of configurations corresponding to an orientation $\theta = 0°$. Note that the position of the feet, arms, and head are not taken into account to define the orientation.

According to our definition, evaluating the orientation of the pelvis is sufficient to estimate the orientation of the observed person. But, evaluating the orientation of the pelvis is not a trivial task, even if a range camera is used. As a matter of fact, one would have first to locate the pelvis in the image, and then to estimate its orientation from a small number of pixels. Therefore, we need to get information from more body parts. Unfortunately, it is still an open question to determine the set of body parts that can be used as clues for the orientation. Therefore, we simply use the whole silhouette.

1.3 Criterion Used for Evaluation

The criterion we use to evaluate our method is the mean error on the angle. The error $\Delta\theta$ is defined as the smallest rotation between the true orientation θ and the estimated orientation $\hat{\theta}$. The angle $\Delta\theta$ is insensitive to the rotation direction and, therefore, $\Delta\theta \in [0°, 180°]$. $\Delta\theta$ is the angle between the two vectors $(\cos\theta, \sin\theta)$ and $(\cos\hat{\theta}, \sin\hat{\theta})$. If $\bullet$ denotes the scalar product,

$$\cos\Delta\theta = (\cos\theta, \sin\theta) \bullet (\cos\hat{\theta}, \sin\hat{\theta}) = cos\left(\hat{\theta} - \theta\right). \tag{1}$$

Since the error is comprised within $\Delta\theta \in [0°, 180°]$, it is computed as

$$\Delta\theta = \cos^{-1}\left(\cos\left(\hat{\theta} - \theta\right)\right). \tag{2}$$

In this paper, we present results as the mean error made by the orientation estimator.

2 The Intrinsic Limitation of a Color Camera

With a color camera, it would be mandatory to decrease the sensitivity to appearance by relying on shapes instead of colors or textures. The purpose of background subtraction algorithms is precisely to detect silhouettes by determining objects in motion. Edge detection techniques are less reliable as the detected edges relate to colors and textures. Therefore, the most reliable (which does not mean that it is robust!) information that can be used with a color camera is the silhouette. Most authors share this point of view. But, there is not enough information in a silhouette to recover the orientation of the observed person, as illustrated in Figure 2.

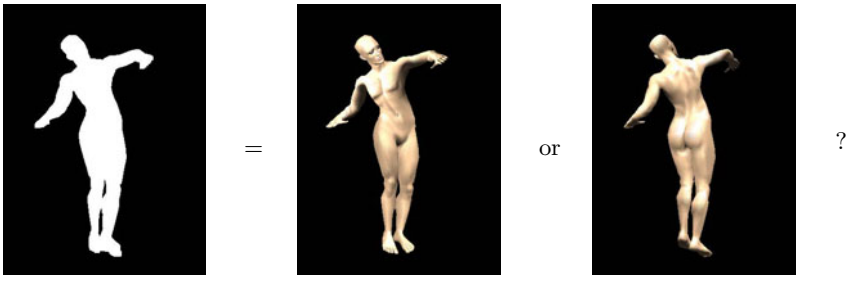

Fig. 2. There is not enough information in a sole silhouette to recover the orientation. With a side view, the intrinsic limitation is the possible confusion between θ and $180° - \theta$. The two configurations depicted on the right hand side correspond to $\theta = 30°$ and $\theta = 150°$, with two mirror poses.

If several authors wrongly believe that a 180° ambiguity is inherent [12], Figure 2 clearly establishes that the intrinsic limitation is not to confuse the orientations θ and $\theta + 180°$, but to confuse θ and $180° - \theta$. In mathematical words, it is only possible to estimate $\sin\theta$, not $\cos\theta$. This can be proved if one assumes that (1) the rotation axis of the observed person is parallel to the image plane (that is, we have a side view) and the projection is orthographic (the demonstration of this property is beyond the scope of this paper). In other words, the perspective effects should be negligible.

Given the intrinsic limitation previously mentioned, the best that an estimator can do is to choose randomly between $\hat{\theta} = \theta$ and $\hat{\theta} = 180° - \theta$. Assuming that all orientations are equally likely, the expected mean error is 45° for the optimal estimator. However, in practice, the mean error is lower than 45°, but still substantial. As shown in our experimental results (see Section 4.1), this is because the perspective effects become significant (and thus a source of information) when the observed person is getting closer to the camera. For example, it is easier to distinguish the direction of the feet when the camera is close enough to the observed person, because they are viewed from above. Indeed, when the

person stands at 3 meters from the camera (which is a typical distance for home entertainment applications), a vertical opening angle of 37° is required. Clearly, the assumption of a near orthographic projection is invalid.

From the above discussion, it follows that there are apparently four ways to address the problem of estimating the orientation.

1. If we could assume that the angle is always comprised in the $[-90°, 90°]$ interval, then using a sole silhouette would permit to recover the orientation. But this assumption is not possible in most of the applications.
2. If we could place two cameras to get orthogonal views, it would be possible to estimate $\sin(\theta)$ and $\cos(\theta)$ independently from those two views, and to recover the orientation during a simple post-processing step. However, the use of two cameras is a huge constraint because is not convenient.
3. The use of the perspective effects to overcome the intrinsic limitation could be considered. However, perspective effects most often consist in small details of silhouettes, and it does not seem a good idea to rely on small details because noise could ruin them in a real application.
4. Another possible solution to the underdetermination is to use a range camera. It is a reliable way to get more geometric information with a single sensor. This is the approach followed in this paper. It should be noted that some manufacturers have recently developed cheap range cameras for general public (see *Microsoft*'s *kinect*).

3 Our Method

3.1 Data

We found it inappropriate and intractable to use real data for training the orientation estimator. Hand-labeling silhouettes with the orientation ground-truth is an error prone procedure (it is difficult to obtain an uncertainty less than 15° [17]). An alternative is to use motion capture to get the ground-truth. However, it is easy to forget a whole set of interesting poses, leading to insufficiently diversified databases. Moreover, using a motion capture system (and thus sequences) has the drawback to statistically link the orientation with the pose. Therefore, our experiments are based on synthetic data instead of real data.

In order to produce synthetic data, we used the avatar provided with the open source software *MakeHuman* [2,15] (version 0.9). The virtual camera (a pinhole camera without any lens distortion) looks towards the avatar, and is placed approximately at the pelvis height. For each shooting, a realistic pose is chosen [13], and the orientation is drawn randomly. We created two different sets of $20,000$ human silhouettes annotated with depth: one set with a high pose variability and the other one with silhouettes closer to the ones of a walker. They correspond to the sets $\mathcal{B}$ and $\mathcal{C}$ of [13] and are shown in Figure 3. Each of these sets is divided into two parts: a learning set and a test set.

Fig. 3. Human synthetic silhouettes annotated with depth, with a weakly constrained set of poses (upper row) and a strongly constrained set of poses (lower row)

3.2 Silhouette Description

In order to use machine learning algorithms, silhouettes have to be summarized in a fixed amount of information called *attributes*, which can be numerical or symbolic. Therefore, we need to describe the silhouettes annotated with depth.

We want the descriptors to be insensitive to uniform scaling, to translations, and to small rotations (this guarantees that the results remain identical even if the camera used is slightly tilted). The most common way to achieve this insensitivity consists in applying a normalization in a pre-processing step: input silhouettes are translated, rescaled, and rotated before computing their attributes. The normalization runs as follows. We use the centroid for translation, a size measure (such as the square root of the silhouette area) for scaling, and the direction of the first principal component (PCA) for rotation. Note that the depth information is not considered here.

Once the pre-processing step has been applied, we still have to describe a silhouette annotated with depth. Most shape descriptors proposed in the literature are only applicable to binary silhouettes. Consequently, we extract a set of binary silhouettes from the annotated silhouette. Each silhouette is further described separately, and all the descriptors are put together. The binary silhouettes (hereafter named *slices*) are obtained by thresholding the depth map. Let S be the number of slices, $s \in \{1, 2, \ldots, S\}$ the slice index, and $th(s)$ the threshold for the slice s. There are several ways to obtain S slices with thresholds. In this paper, we compare the results obtained with slicing methods based on the surface, on moments, and on the extrema. Note that all our slicing methods produce ordered silhouettes, that is, if $s \leq r$ are two indices, then the s slice is included in the r slice.

Slicing based on surface. Slicing based on the surface produces silhouettes whose surface increases linearly. $th(s)$ is such that the ratio between the area of the s slice and the binary silhouette is approximately equal to s/S. Figure 4 presents the result for 5 slices.

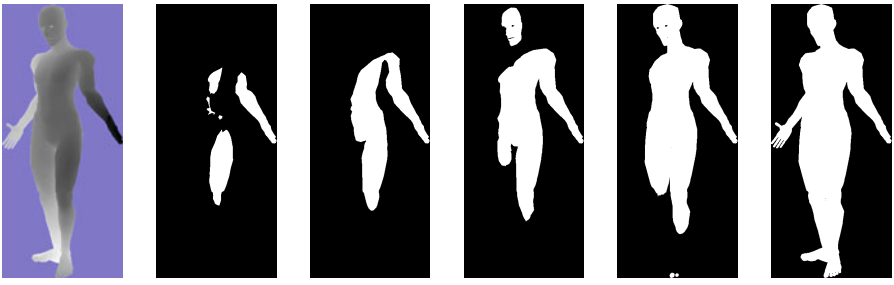

Fig. 4. Binary silhouettes produced by a surface-based slicing with 5 slices

Slicing based on moments. Let μ denote the mean of depth, and σ the standard deviation of depth. $th(s) = \mu - 2\sigma + 4\sigma\frac{s-1}{S-1}$.

Slicing based on extrema. Let m and M denote, respectively, the minimum and the maximum values of depth. $th(s) = m + \frac{s}{S}(M - m)$.

It should be noted that the number of slices has to be kept small. Indeed, increasing the number of slices involves a higher computational cost, and a larger set of attributes which may be difficult to manage for regression algorithms.

After the slicing process, each slice is described separately, and all the descriptors are put together. A wide variety of shape descriptors have been proposed for several decades [10,16]. In a preliminary work, we have derived the orientation on the basis of a single binary silhouette (in the $[-90°, 90°]$ interval). It can be proved that the descriptor should distinguish between mirrored images. Therefore, the descriptors insensitive to similarity transformations are not suited for our needs. Several shape descriptors were evaluated, and we have found that the use of the Radon transform or the use of one shape context offer the best performances. These descriptors are briefly described hereafter.

Descriptors based on the Radon transform. We have used a subset of the values calculated by a Radon transform as attributes. Radon transform consists in integrating the silhouettes over straight lines. We have used 4 line directions, and 100 line positions for a given direction.

Descriptors based on the shape context. Shape contexts have been introduced by Belongie *et al.* [3] as a mean to describe a pixel by the location of the surrounding contours. A shape context is a log-polar histogram. In our implementation, we have a single shape context centered on the gravity center (of the binary silhouette) which is populated by all external and internal contours. We have used a shape context with 5 radial bins and sectors of 30 degrees (as in [3]).

3.3 Regression Method

The machine learning method selected for regression is the *ExtRaTrees* [7]. It is a fast method, which does not require to optimize parameters (we do not have to setup a kernel, nor to define a distance), and that intrinsically avoids overfitting.

In practice, it is not possible to estimate the orientation directly. Indeed, it is possible to find two silhouettes annotated with depth such that (1) they are almost identical, and (2) their orientations are $\theta \simeq -180°$ and $\theta \simeq 180°$. Therefore, the function that maps the silhouette annotated with depth to the orientation presents discontinuities. In general, discontinuities are a problem for regression methods. For example, the *ExtRaTrees* use an averaging that leads to erroneous values at the discontinuity. Our workaround to maintain continuity consists in the computation of two regressions: one regression estimates $\sin\theta$ and the other one estimates $\cos\theta$. The estimate $\hat{\theta}$ is derived from:

$$\hat{\theta} = \tan^{-1}\left(\widehat{\sin\theta}, \widehat{\cos\theta}\right) \tag{3}$$

The same approach was also followed by Agarwal *et al.* [1]. Note that $\widehat{\sin\theta}^2 + \widehat{\cos\theta}^2$ is not guaranteed to be equal to 1.

4 Experiments

Prior to the orientation estimation in images captured with a range camera, we analyzed the impact of perspective effects. We have avoided the range normalization problem by slicing the 3D volume of a human isolated from the background. But perspective effects cannot be ignored.

4.1 The Impact of Perspective Effects

In our first experiment, we ignore the depth information, and try to estimate the orientation from the binary silhouette. The purpose of this experiment is twofold. Firstly, we want to highlight the impact of perspective effects in the orientation estimation. Secondly, we want to show that perspective effects cannot (alone) lead to satisfactory results. Figure 5 shows that the mean error on the orientation depends on the distance between the camera and the observed person. When the camera moves away from the avatar, the vertical opening angle is decreased to keep the silhouettes at approximately the same size. For example, the opening angle is 50° at 3 meters, and 8° at 20 meters. The larger the distance from the camera to the observed person is, the more the projective model approximates an orthographic projection.

Clearly, the perspective plays a significant role. Despite these effects, binary silhouettes do not suffice to achieve an acceptable mean error, even for the set of strongly constrained poses (13° at 3 meters). Our second experiment shows that better results are obtained from the 3D information. However, the fact that the results depend on the distance between the person and the camera leads to the following question: which distance(s) should we choose to learn the model? We did not find the answer to this question. On the one hand, we would like to fill the learning database with samples corresponding to the operating conditions (typically a distance of about 3 meters). On the other hand, we want to avoid learning details that are not visible in practice because of noise, and therefore to use a large distance in order to reduce perspective effects.

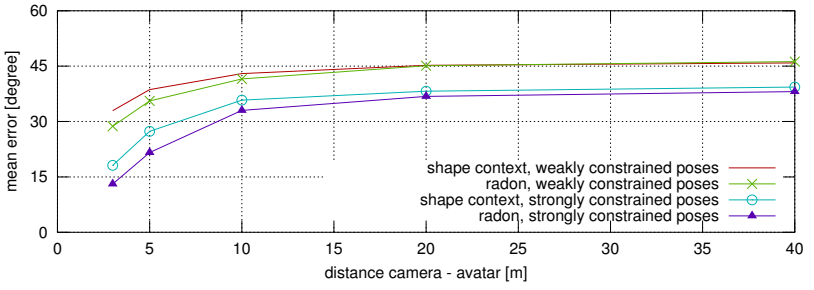

Fig. 5. The mean error on the orientation estimation depending on the distance between the camera and the observed person

4.2 The Role Played by 3D in Orientation Estimation

In our second experiment, we evaluate the performance that can be reached from 3D information. The virtual camera is placed at a distance of 3 meters from the avatar. We report the results obtained with the descriptors that have been previously mentioned. The mean error results are provided in Tables 1 and 2 for, respectively, the sets of strongly and weakly constrained poses. The following conclusions can be drawn:

1. The mean error is lower for a strongly set of poses than for a weakly set of poses. The diversity of the poses in the learning set has a negative impact on the results.
2. Although increasing the number of slices always improves the performance, the number of slices only affects the performance slightly when it exceeds 2 or 3. Thus there is no need to have a high resolution for the distance values in the depth map.
3. The surface-based slicing method systematically outperforms the two other slicing strategies.
4. We are able to obtain mean errors as low as 9.2° or 4.3° on a 360° range of orientations depending on the set of poses considered. So the role played by 3D in orientation estimation is much more important than the one played by perspective effects. Moreover, these results demonstrate that several viewpoints (as used in [8] and [14]) are useless for the orientation estimation.

It is difficult (if not impossible) to compare our results with those reported for techniques based on a classification method (such as the one proposed by Rybok *et al.* [14]) instead of a regression mechanism. Therefore, we limit our comparison to results expressed in terms of an error angle. However, one should keep in mind that a perfect comparison is impossible because the set of poses used has never been reported by previous authors. Agarwal *et al.* [1] obtained a mean error of 17° with a single viewpoint. Gond *et al.* [8] obtained a mean error of 7.57° using several points of view. Peng *et al.* reported 9.56° when two orthogonal

Table 1. Mean errors obtained with a strongly constrained set of poses

Radon:	surface-based slicing	moments-based slicing	extrema-based slicing
1 slice	13.0°	—	13.1°
2 slices	6.4°	10.6°	8.6°
3 slices	6.1°	6.4°	7.9°
4 slices	5.8°	6.3°	7.6°
5 slices	5.7°	6.1°	7.3°

shape context:	surface-based slicing	moments-based slicing	extrema-based slicing
1 slice	18.2°	—	18.1°
2 slices	4.8°	12.2°	8.3°
3 slices	4.5°	4.9°	7.3°
4 slices	4.4°	5.1°	6.8°
5 slices	4.3°	4.7°	6.7°

Table 2. Mean errors obtained with a weakly constrained set of poses

Radon:	surface-based slicing	moments-based slicing	extrema-based slicing
1 slice	28.9°	—	28.8°
2 slices	11.1°	24.4°	19.4°
3 slices	9.9°	11.7°	15.5°
4 slices	9.4°	11.5°	14.0°
5 slices	9.2°	10.7°	13.2°

shape context:	surface-based slicing	moments-based slicing	extrema-based slicing
1 slice	32.8°	—	32.6°
2 slices	11.1°	28.1°	23.7°
3 slices	10.0°	12.4°	18.8°
4 slices	9.5°	12.5°	16.4°
5 slices	9.2°	11.3°	15.2°

views are used. All these results were obtained with synthetic data, and can thus be compared to our results. The results reported by Gond *et al.* and Peng *et al.* are of the same order of magnitude as ours, but our method is much faster and simpler to implement. In contrast with existing techniques, we do not need complex operations such as camera calibration, shape from silhouettes, tensor decomposition, or manifold learning.

4.3 Observations for a Practical Application in Real Time

We applied our method to a real application driven by a *kinect*. A simple background subtraction method has been used to extract the silhouettes annotated with depth. We didn't filter the depth signal, nor the slices. The estimated

orientation has been applied in real time to an avatar, and projected on a screen in front of the user. A light temporal filtering has been applied to the orientation signal to avoid oscillations of the avatar. This allowed a qualitative assessment. The models have been learned from synthetic data without noise. It appears that the model learned with the descriptor based on the Radon transform is efficient, and that it outperforms the model learned with the descriptor based on the shape context. Without any kind of filtering, it seems thus that region-based descriptors are preferable to limit the impact of noise.

5 Conclusions

Estimating the orientation of the observed person is a crucial task for a large variety of applications including home entertainment, man-machine interaction, and intelligent vehicles. In most applications, only a single side view of the scene is available. To be insensitive to appearance (color, texture, ...), we rely only on geometric information (the silhouette and a depth map) to determine the orientation of a person.

Under these conditions, we explain that the intrinsic limitation with a color camera is to confuse the orientations θ and $180° - \theta$. When the camera is close enough from the observed person, the perspective effects bring a valuable information which helps to overcome this limitation. But, despite perspective effects, performances remain disappointing in terms of the mean error of on the estimated angle. Therefore, we consider the use of a range camera and provide evidence that 3D information is appropriate for orientation estimation.

We address the orientation estimation in terms of regression and supervised learning with the *ExtRaTrees* method and show that mean errors as low as 9.2° or 4.3° can be achieved, depending on the set of poses considered. These results are consistent with those reported in the literature. However, our technique has many advantages. It requires only one point of view (and therefore a single sensor), it is fast and easy to implement.

Acknowledgments. S. Piérard has a grant funded by the FRIA, Belgium.

References

1. Agarwal, A., Triggs, B.: Recovering 3D human pose from monocular images. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(1), 44–58 (2006)
2. Bastioni, M., Re, S., Misra, S.: Ideas and methods for modeling 3D human figures: the principal algorithms used by MakeHuman and their implementation in a new approach to parametric modeling. In: Proceedings of the 1st Bangalore Annual Compute Conference, pp. 10.1–10.6. ACM, Bangalore (2008)
3. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4), 509–522 (2002)

4. Bhanu, B., Han, J.: Model-based human recognition: 2D and 3D gait. In: Human Recognition at a Distance in Video. Advances in Pattern Recognition, ch.5, pp. 65–94. Springer, Heidelberg (2011)
5. Enzweiler, M., Gavrila, D.: Integrated pedestrian classification and orientation estimation. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA, pp. 982–989 (June 2010)
6. Gandhi, T., Trivedi, M.: Image based estimation of pedestrian orientation for improving path prediction. In: IEEE Intelligent Vehicles Symposium, Eindhoven, The Netherlands (June 2008)
7. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Machine Learning 63(1), 3–42 (2006)
8. Gond, L., Sayd, P., Chateau, T., Dhome, M.: A 3D shape descriptor for human pose recovery. In: Perales, F.J., Fisher, R.B. (eds.) AMDO 2008. LNCS, vol. 5098, pp. 370–379. Springer, Heidelberg (2008)
9. Lee, M., Nevatia, R.: Body part detection for human pose estimation and tracking. In: IEEE Workshop on Motion and Video Computing (WMVC), Austin, USA (February 2007)
10. Loncaric, S.: A survey of shape analysis techniques. Pattern Recognition 31(8), 983–1001 (1998)
11. Ozturk, O., Yamasaki, T., Aizawa, K.: Tracking of humans and estimation of body/head orientation from top-view single camera for visual focus of attention analysis. In: International Conference on Computer Vision (ICCV), Kyoto, Japan, pp. 1020–1027 (2009)
12. Peng, B., Qian, G.: Binocular dance pose recognition and body orientation estimation via multilinear analysis. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, USA (June 2008)
13. Piérard, S., Van Droogenbroeck, M.: A technique for building databases of annotated and realistic human silhouettes based on an avatar. In: Workshop on Circuits, Systems and Signal Processing (ProRISC), Veldhoven, The Netherlands, pp. 243–246 (November 2009)
14. Rybok, L., Voit, M., Ekenel, H., Stiefelhagen, R.: Multi-view based estimation of human upper-body orientation. In: IEEE International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, pp. 1558–1561 (August 2010)
15. The MakeHuman team: The MakeHuman (2007), `http://www.makehuman.org`
16. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37(1), 1–19 (2004)
17. Zhang, W., Matsumoto, T., Liu, J., Chu, M., Begole, B.: An intelligent fitting room using multi-camera perception. In: International Conference on Intelligent User Interfaces (IUI), pp. 60–69. ACM, Gran Canaria (2008)