Fish-Eye Camera Video Processing and Trajectory Estimation Using 3D Human Models

Konstantina Kottari, Kostas Delibasis, Vassilis Plagianakos, Ilias Maglogiannis

Fish-eye Camera Video Processing and Trajectory Estimation Using 3D Human Models

K. Kottari[1], K.K. Delibasis[1],V. Plagianakos[1], I. Maglogiannis[2]

[1] University of Thessaly, Dept. of Computer Science and Biomedical Informatics, Lamia, Greece
[2]University of Piraeus, Dept. of Digital Systems, Piraeus, Greece
kottarikonstantina@gmail.com, kdelibasis@yahoo.com, vpp@dib.uth.gr; imaglo@unipi.gr

Abstract. Video processing and analysis applications are part of Artificial Intelligence. Frequently, silhouettes in video frames lack depth information, especially in case of a single camera. In this work, we utilize a three-dimensional human body model, combined with a calibrated fish-eye camera, to obtain three-dimensional (3D) clues. More specifically, a generic 3D human model in various poses is derived from a novel mathematical formalization of a well-known class of geometric primitives, namely the generalized cylinders, which exhibit advantages over the existing parametric definitions. The use of the fish-eye camera allows the generation of rendered silhouettes, using these 3D models. Moreover, we present a very efficient algorithm for matching that 3D model with a real human figure in order to recognize the posture of a monitored person. Firstly, the silhouette is segmented in each frame and the calculation of the real human position is calculated. Subsequently, an optimization process adjusts the parameters of the 3D human model in an attempt to match the pose (position and orientation relatively to the camera) of real human. The experimental results are promising, since the pose, the trajectory and the orientation of the human can be accurately estimated.

Keywords: fish-eye camera video processing, three-dimensional human modelling, posture recognition, minimization, generalized cylinders, and elliptical intersections.

1 Introduction

The field of automated human activity recognition utilizing fixed cameras of indoor environments has gained significant interest during the last years. It finds a variety of applications in diverse areas, such as assistive environments, smart homes, support for the elderly or the chronic ill, surveillance and security, traffic control, industrial processes, etc.

This work focuses on fish-eye camera video processing for pose estimation of sitting or standing/walking humans. Therefore, human silhouette segmentation of the video sequence is a prerequisite. Recognizing a human pattern is often possible via volume intersection [1] or a voxel-based approach [2,3]. Stereometry based models have also

been constructed through calibrated camera pairs. Using triangulation, the depths of the points are calculated. This approach has been taken into account by Pl*ä*nkers and Fua [4] and Haritaoglu et al. in [5]. Stereo vision is also used by Jojic et al. [6], with the optional aid of projected light patterns. The proposed algorithm is based on a parametric three-dimensional (3D) human model with limited degrees of freedom so that it allows efficient manipulation for standing/walking and sitting postures. Our aim is to estimate human position, trajectory and standing/sitting state, which would be useful towards human behavior recognition.

The first step in applications dealing with human activity recognition from video is the foreground segmentation. Most video segmentation algorithms are based on background subtraction. The background has to be modelled, since it may change due to a number of reasons, including: motion of background objects, changes in light conditions, or video compression artifacts. In this work, we employ the forward and inverse camera model that was proposed in [7]. We follow a "top-down" approach that matches the model rendered through the calibrated fish-eye camera, with the segmented frame of the video. Then, an optimization algorithm is utilised to find the model parameters and determine human orientation and pose. The rest of the paper is structured as follows: Section 2 discusses the technical details of the proposed algorithms, Section 3 presents some initial results, while Section 4 concludes the paper.

2 Proposed Methodology

2.1 Generalized Cylinders

For the generation of the human model, we utilized the concept of generalized cylinders (GC), as proposed in [8]. More specifically, let C_1 be a piecewise smooth curve defined in a Cartesian coordinate system *OXYZ*, as:

$$r_1(t) = (x(t), y(t), z(t)), \mathrm{t} \in [\mathrm{a}, \mathrm{b}] \subset \mathrm{R} \tag{1}$$

and C_2 be a planar curve defined in an orthogonal local Cartesian system *OXY*. Letus now consider the surface *S* that is generated by moving the curve C_2 along C_1, so that its plane is perpendicular to the tangent vector of C_1, and the origin of *OXY* belongs to C_1.If we express the planar curve C_2 in polar coordinates $r_2 = r_2(u), u \in [0, 2\pi]$ and introduce a scale factor *s*(*t*)and a rotation factor $\phi(t)$ along the tangent vector of *C*1 as function of position along C_1, then the surface equation of the GC becomes:

$$\begin{aligned} x(t,u) = x(t) &+ \frac{s(t)y'(t)r(u+\phi(t))\cos(u+\phi(t))}{P_2(t)} \\ &+ \frac{s(t)x'(t)z'(t)r(u+\phi(t))\sin(u+\phi(t))}{P_1(t)P_2(t)} \end{aligned} \tag{2}$$

$$y(t,u) = y(t) - \frac{s(t)x'(t)r(u+\phi(t))\cos(u+\phi(t))}{P_2(t)} + \frac{s(t)y'(t)z'(t)r(u+\phi(t))\sin(u+\phi(t))}{P_1(t)P_2(t)} \quad (3)$$

$$z(t,u) = z(t) - \frac{s(t)(x'(t))^2 + (y'(t))^2 r(u+\phi(t))\sin(u+\phi(t))}{P_1(t)P_2(t)} \quad (4)$$

where, $(t,u) \in [a,b] \times [0,2\pi]$, $P_1(t) = \sqrt{(x'(t))^2 + (y'(t))^2 + (z'(t))^2}$ and $P_2(t) = \sqrt{x'(t))^2 + y'(t))^2}$ .The complete proof is given in [8].

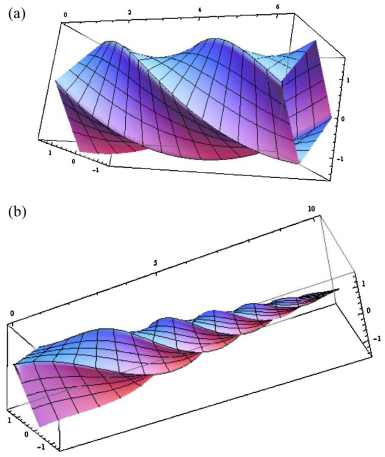


Fig. 1. Two examples of surfaces derived from equation (2) – (4) from [8]

2.2 3D Model Construction

In this work, a free triangulated model of a standing human (Fig.2) is utilised, defined by the Cartesian coordinates of approximately 27,000 vertices [9]. Since we are interested in simulating the rendering of the human model through the fish-eye camera in real time, we discard the triangle information of the model and we treat it as a cloud of points.

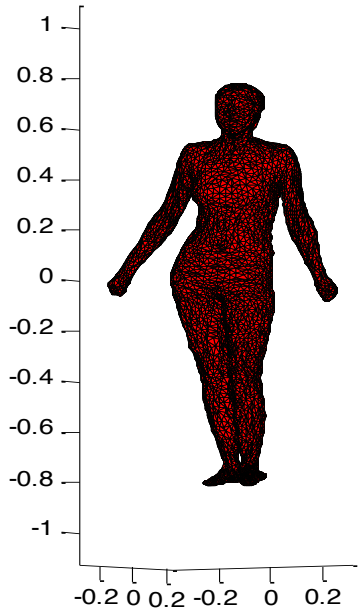
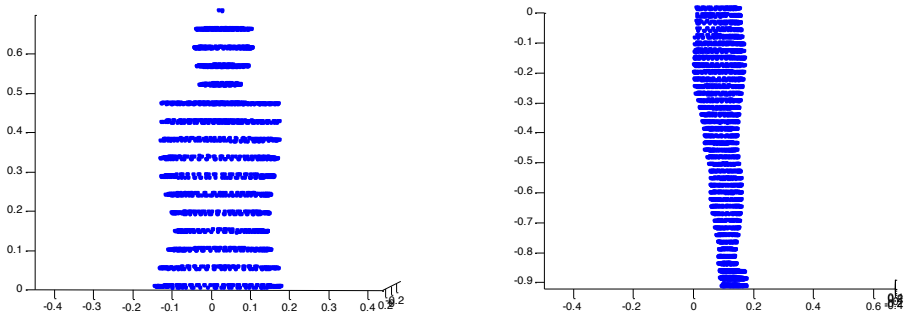


Fig. 2. Triangulatedmodel of a standing human [9].

Therefore, we compute the intersections of the model in a number of horizontal planes, in distance of two centimetres along the *Z* axis (feet – head direction) as shown in Fig.3 (a). The same process is repeated, along hands and legs – see Fig. 3(b).



(a) (b)

Fig. 3. Elliptical intersections of torso and leg.

Each intersection is estimated for approximating an ellipse with its semi-axes a_{semi}, b_{semi} parallel to *X* and *Y* axis of coordinate system, as shown in Fig.4.

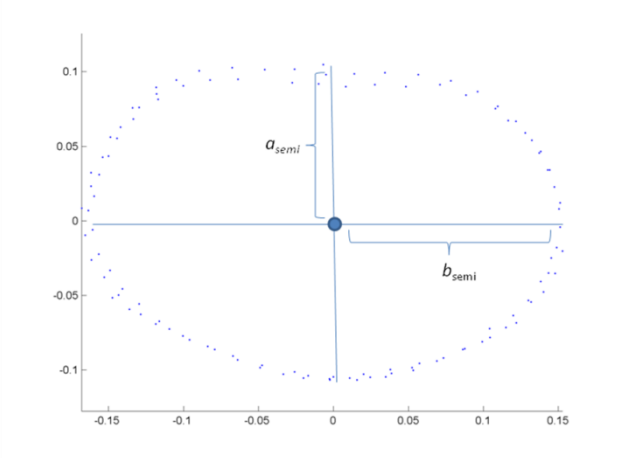


Fig. 4. Elliptical intersection of the human torsowith the*XY* plane.

The parametrical equation of GC (4) can be simplified by using a piecewise straight line as curve C_1, each segment of which is defined by vector (a_0, b_0, c_0) and by assigning these ellipses as planar closed curveC_2as follows:

$$x = a_0 t + \frac{b_0 r \cos(u)}{P_2} + \frac{a_0 c_0 r \sin(u)}{P_2}$$

$$y = b_0 t - \frac{a_0 r \cos(u)}{P_2} + \frac{b_0 c_0 r \sin(u)}{P_2} \tag{5}$$

$$z = c_0 t - \frac{(a_0{}^2 + b_0{}^2) r \sin(u)}{P_2}$$

where $r = \dfrac{a_{semi} b_{semi}}{\sqrt{(b_{semi}\cos(u))^2 + (a_{semi}\sin(u))^2}}$ and a_0, b_0, and c_0 are determined by the direction of the leading axis. In Fig.5 we see the result of elliptical patterning of model intersections, through the insertion of ellipses to the GC Eq. (5). Note that a torsional inconstancy at the knees section is being observed. This phenomenon has been explained in [8, section 4] and does not affect the optimization process that matches the 3D model to the segmented human silhouette.

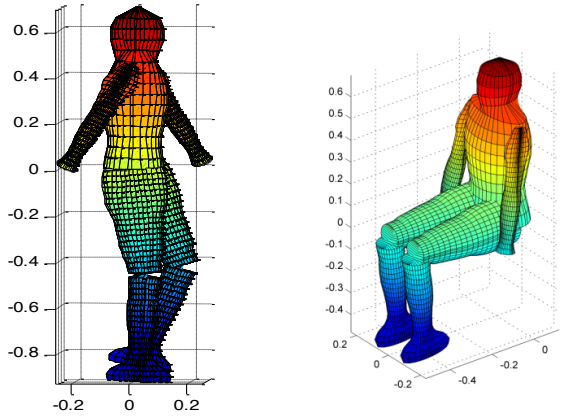


Fig. 5.3D standing (left) and sitting human (right).

The estimation of human posture (sitting or standing) is based on the construction of the human model. Having the 3D standing human model constructed as described above, its transformation to match the sitting position can be easily performed by changing the model parameter (angles) at waist and knees. The result of that transformation is shown in Fig. 5 right, while Fig.6 depicts both models as they are utilised by the video-processing algorithm.

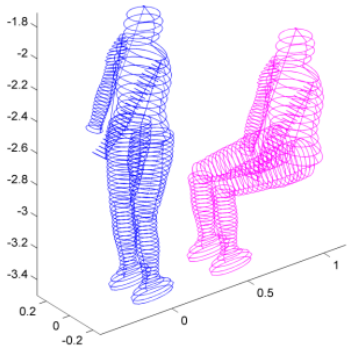


Fig. 6. 3D Human silhouettes

2.3 **Video Processing Algorithm**

In this work, we analyze videos captured by a fish-eye camera, fixed on the ceiling of a living environment. The recorded videos have been foreground segmented, while empty frames are being discarded. Then, the mask shown in Fig. 7(a) is applied to suppress noisy segmented pixels outside the field of view. The initial estimation of the real human position in the room coordinates is accomplished by the recently proposed algorithm in [7] based on the segmented frame pixels. For this purpose, we

employ the calibration of the acquiring fish-eye camera that provides the spherical coordinates (θ, φ) for each pixel of the current frame, as well as the frame pixel that corresponds to any real world point (x,y,z), according to [7]:

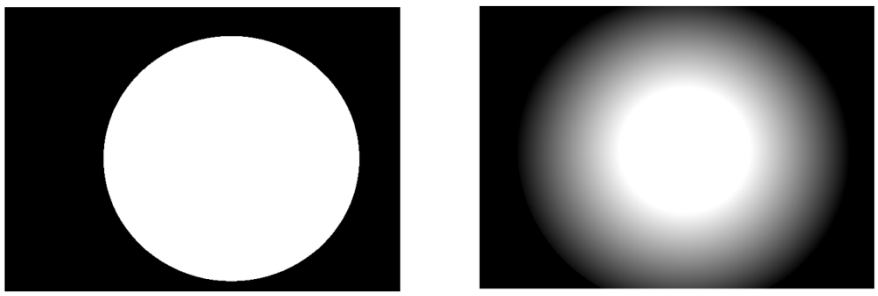
$$(j,i) = M(x,y,z) \tag{6}$$

$$(\theta,\varphi) = M_1(i,j) \tag{7}$$

Let $PHI(i,j)$ hold the value of φ for pixel (i,j), as obtained by (7) and shown in Fig. 7(b). Thus, for any pose of the 3D parametric model, we can obtain the binary image-I_M of the human model, rendered by the fish-eye model using (6). Fig.8illustrates image I_M combined for various standing and sitting models, as rendered by the calibrated fish-eye camera. Let I_S be segmented binary frame, after using the binary mask in Fig.7(a). The initial estimation of the person's position is obtained by locating the non-zero pixel (i_0,j_0) of I_S that holds the minimum value of angle φ. The objective function, which quantifies the match between the model and the segmented human silhouette as a function of its real world position (x,y) and its orientation θ_0, is defined as the intersection of the segmented silhouette $I_S$and the rendered human modelI_M:

$$f(x,y,\theta_0) = \sum_{\substack{image \\ domain}} I_M \cap I_S \tag{8}$$

where I_M (defined above) is shown in red, I_S shown in green and their intersection $I_M \cap I_S$ is shown in yellow. Fig. 9 presents graphically the calculation of the objective function for one instance of each class (sitting and standing). Subsequently, the simplex [10] multidimensional unconstrained maximisation algorithm is utilised to optimise the objective function. The initial position (x,y) of the human is approximately computed from the first frame by finding the segmented human silhouette pixel (i,j) with maximum PHI, (as described in [7]) and used to initialize the simplex method. Thus, the maximisation algorithm computes the human model parameters that best match the segmented figure and returns the coordinates x and y, as well as the orientation θ_0 of model.



(a) (b)

Fig. 7. (a) Binary mask used to exclude out of filed-of-view pixels in video frames (b) Visualization of the PHI angle for each frame pixel as resulted from the camera calibration [7].

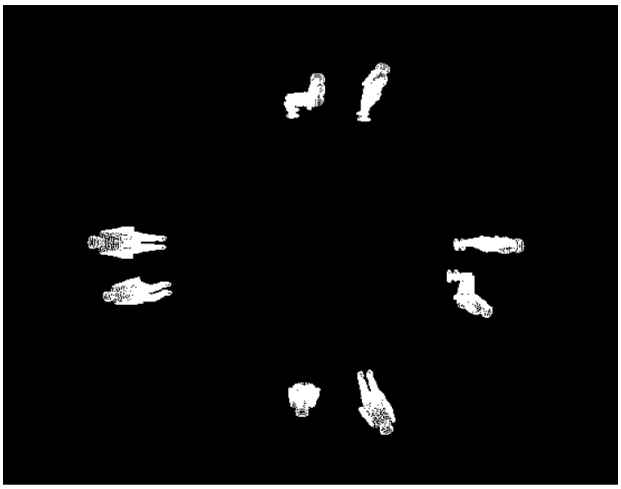
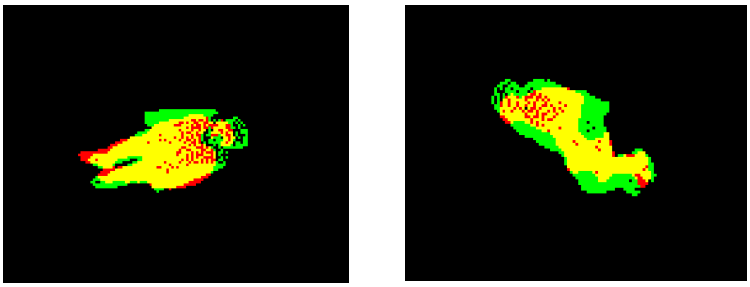


Fig. 8. Rendered standing and sitting human model in various angles and positions in the room.



(a) (b)

Fig. 9. Visualization of the calculation of the objective function (see text for details) for the standing (a) and sitting human (b).

3 Experimental Results

Classification results of two different videos are presented in Table 1, considering the sitting as "positive" and the standing as "negative" status. The ground truth was established by manually annotating the video with the human pose, as "sitting" or "standing".

Table 1. Pose classification results

	TP	TN	FN	FP	Accuracy	Sensitivity	Specificity
Video 1	110/278	112/278	23/278	33/278	0.7986	0.8271	0.7724
Video 2	98/164	54/164	5/164	7/164	0.9268	0.9515	0.8852

The proposed model-based algorithm is able to estimate the trajectory of the human silhouette, as well as its orientation. Fig. 10 shows the estimated positions and orientation of the human silhouette for video 1, as recovered by the optimization described above. The person moves from left to right at the bottom of the frame (A to B), then vice versa at the top of the frame and finally sits down at the point designated by E, where it rotates. Between points B and C the segmentation fails temporarily, however, the model-based tracking algorithm successfully recovers the silhouette's new position. The experimental results indicate that the simplex optimization method is robust and efficient. It needs approximately 30 iterations for each frame in order to converge. Each objective function evaluation requires approximately 30msec on an Intel i5 laptop with 4 GB Ram using the Matlab environment. Running time approaches the 900 milliseconds per frame.

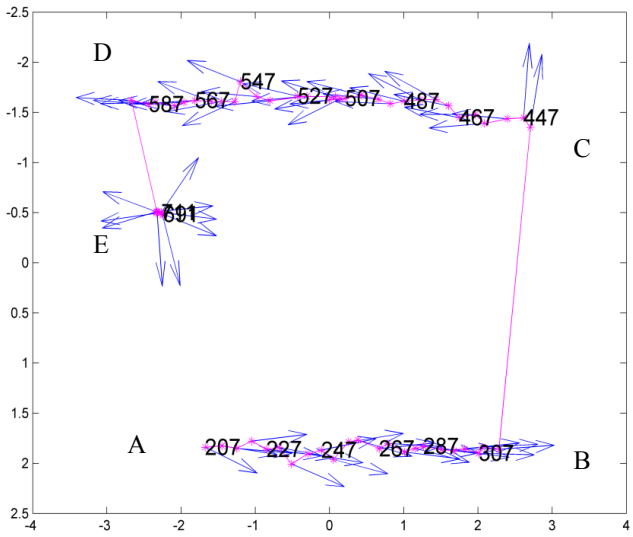


Fig. 10. Graphical representation of the results for model based trajectory and orientation estimation for video 1.

4 Discussion and Conclusions

In this paper, an algorithm for estimating the trajectory of a human silhouette in indoor videos acquired by an omni-directional camera has been presented. The algorithm is based on a parametric 3D human model and it recovers the model parameters (translation and orientation) by optimizing a suitable defined objective function. Initial results show that the proposed algorithm can estimate the trajectory and orientation and discriminate between two different postures: sitting and standing. The proposed methodology may improve the accuracy of more complex activity recognition algorithms usually found in ambient assisted living environments. This methodology

can be adopted for detecting higher level activity events and understand behavioural patterns.

Acknowledgment

The authors would like to thank the European Union (European Social Fund ESF) and Greek national funds for financially supporting this work through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: \Thalis \ Interdisciplinary Research in Affective Computing for Biological Activity Recognition in Assistive Environments.

References

1. Bottino A., Laurentini A.: A silhouette-based technique for the reconstruction of human movement. In: Computer Vision and Image Understanding (CVIU), 83(1), pp. 79–95 (2001).
2. German K.M. Cheung, Baker S., Kanade T.: Shape-from silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'03), vol. 1, pp. 77–84, Madison, WI (2003).
3. Mikic I., Trivedi M., Hunter E., Cosman P.: Human body model acquisition and tracking using voxel data, International Journal of Computer Vision 53(3), 199–223 (2003).
4. Pl*ä*nkers R., Fua P.: Tracking and modeling people in video sequences, Computer Vision and Image Understanding (CVIU) 81(3), 285–302 (2001).
5. Haritaoglu I., Harwood D., Davis L. S., W4s: A real time system detecting and tracking people in 2 1/2D In: Proceedingsof the European Conference on Computer Vision (ECCV'98),Lecture Notes in Computer Science, vol. 1406, 877–892, Freiburg, Germany (1998).
6. Jojic N., Gu J., Shen H., S. Huang T.S: 3-Dreconstruction of multipart, self-occluding objects. In: Proceedingsof the Asian Conference on Computer Vision (ACCV'98), 455–462, HongKong, China (1998).
7. Delibasis K.K., Goudas T., Plagianakos V.P., Maglogiannis I.: Fisheye Camera Modeling for Human Segmentation Refinement in Indoor Videos. In: PETRA '13, May 29 - 31 2013, Island of Rhodes, Greece Copyright 2013 ACM 978-1-4503-1973-7/13/05.
8. K.K. Delibasis K., Kechriniotis A., Maglogiannis I.: A novel tool for segmenting 3D medical images based on generalized cylinders and active surfaces, Computer Methods and Programs in Biomedicine, 111, 148–165 (2013).
9. http://www.3dmodelfree.com/models/20966-0.htm
10. Nelder J.A. and Mead R.: A simplex method for function minimization, Computer Journal, 7, pp. 308-313 (1965).