

Calculation of Complex Zernike Moments with Geodesic Correction for Pose Recognition in Omni-directional Images

K. Delibasis, Spiros Georgakopoulos, Vassilis Plagianakos, Ilias Maglogiannis

► **To cite this version:**

K. Delibasis, Spiros Georgakopoulos, Vassilis Plagianakos, Ilias Maglogiannis. Calculation of Complex Zernike Moments with Geodesic Correction for Pose Recognition in Omni-directional Images. 10th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2014, Rhodes, Greece. pp.375-384, 10.1007/978-3-662-44654-6_37 . hal-01391339

HAL Id: hal-01391339

<https://hal.inria.fr/hal-01391339>

Submitted on 3 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Calculation of Complex Zernike Moments with Geodesic Correction for Pose Recognition in Omni-directional Images

K.K. Delibasis¹, S. Georgakopoulos¹, V. Plagianakos¹, I. Maglogiannis²

¹ University of Thessaly, Dept. of Computer Science and Biomedical Informatics, Lamia, Greece

² University of Piraeus, Dept. of Digital Systems, Piraeus, Greece
kdelibasis@yahoo.com, {spyrosgeorg ; vpp}@dib.uth.gr; imaglo@unipi.gr

Abstract

A number of Computer Vision and Artificial Intelligence applications are based on descriptors that are extracted from imaged objects. One widely used class of such descriptors are the invariant moments, with Zernike moments being reported as some of the most efficient descriptors. The calculation of image moments requires the definition of distance and angle of any pixel from the centroid pixel. While this is straightforward in images acquired by projective cameras, it is complicated and time consuming for omni-directional images obtained by fish-eye cameras. In this work, we provide an efficient way of calculating moment invariants in time domain from omni-directional images, using the calibration of the acquiring camera. The proposed implementation of the descriptors is assessed in the case of indoor video in terms of classification accuracy of the segmented human silhouettes. Numerical results are presented for different poses of human silhouettes and comparisons between the traditional and the proposed implementation of the Zernike moments are presented. The computational complexity for the proposed implementation is also provided.

Keywords: Pattern recognition, computer vision, image descriptors, moment invariants, Omni-directional image/video, fish-eye camera, silhouette pose recognition.

1 Introduction

Pattern recognition in images is a very common task in artificial intelligence. Among other methods [1], invariant moments have been used extensively for providing regional descriptors to be used in pattern recognition problems [2] – [6]. The calculation of invariant moments of a region of an image requires a valid distance metric defined in the image domain (e.g. [3]). In omnidirectional images, the distance between two pixels cannot be defined in terms of their coordinates in the image frame, since the image is acquired not through a simple projection, but using a spherical element with 180 degrees field of view.

This need has been recently addressed in a few publications. In [7] the implementation of the SIFT algorithm ([8]) is presented for omni-directional images, using the

convolution operators on a sphere in the Fourier domain, whereas it has also been studied for wide angle images [9]. In other approaches, the SIFT algorithm has been applied to the unwrapped omni-directional images [10]. Pixel distance was redefined in conic sensor images [11.]

The contribution of this work focuses on an efficient way of computing Zernike moment invariant (or any other invariant moments) for calibrated omni-directional images in the time domain. We propose the measurement of geodesic distance and angles between image pixels, based on the calibrated model of the acquiring fish-eye camera. The geodesically corrected implementation of Zernike moments presents reduced variability when applied to known geometric shapes and it achieves increased accuracy when applied to pose classification of segmented human silhouettes. The proposed implementation is computationally quite efficient, since it allows the processing of a high number of frames per second.

2 Proposed Methodology

2.1 Zernike moment invariant

Zernike moments of order m , n are defined by a set of radial polynomials $R_{nm}(r)$, which are orthogonal inside the unit circle:

$$R_{nm}(r) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n+|m|}{2} + s\right)!} r^{n-2s}, \quad (1)$$

where n positive integer and m integer such that $n-|m|$ is even and $|m| < n$. Orthogonality is preserved since $\int_0^1 R_{nm}(r) R_{km}(r) dr = \frac{1}{2(n+1)} \delta_{nm}$.

Zernike polynomials are defined as

$$V_{nm}(r, \theta) = R_{nm}(r) e^{jm\theta} \quad (2)$$

The Zernike moments of a bivariate function $f(x, y)$ are defined as:

$$Z_{nm} = \frac{n+1}{\pi} \int_0^1 \int_0^{2\pi} f(r, \theta) V_{nm}(r, \theta) r dr d\theta \quad (3)$$

The above Eq. is easily adapted to be used in discrete images.

2.2 Omni-directional image formation and camera calibration

The formation of omni-directional image using a spherical camera, as presented in detail in [12], is shown in Fig. 1, using only two dimensions to facilitate understanding. In Fig. 2, neighbouring pixels are shown in different positions on the image sen-

sor. It is clear that the distance of two pixels is different when measured on the sensor and on the position of the image points on the spherical optical element. Therefore, in order to produce accurate results, image processing algorithms that use pixel distances have to be re-implemented for omni-directional images. In this work, we utilize the calibration of the specific fish-eye camera, proposed in [12], according to which, for each pixel (i, j) of the image, the corresponding vector (θ, φ) on the spherical element is precalculated

$$(\theta, \varphi) = M(j, i) \quad (4)$$

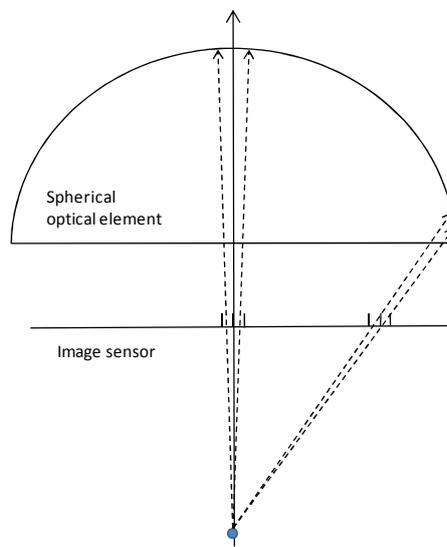


Fig. 1. Model of image formation for spherical omni-directional camera.

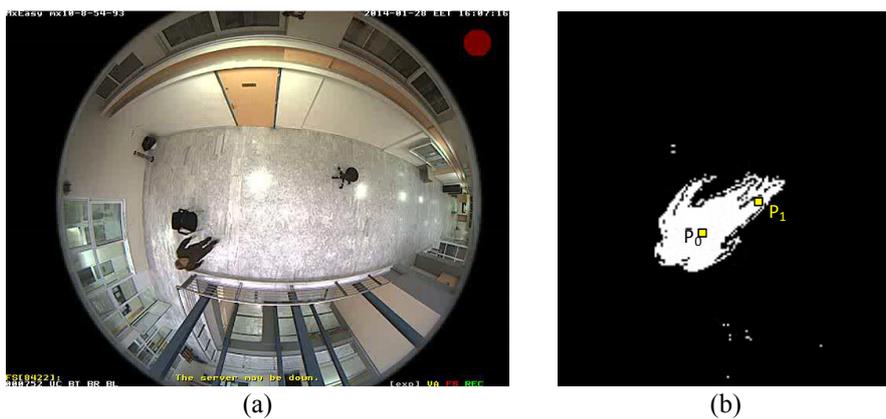


Fig. 2. A typical omni-directional image (a) and the segmented human silhouette (b). The strong deformation imposed by the fish-eye spherical lens is evident.

2.3 Calculation of Zernike moment invariants for calibrated omni-directional images – Geodesic correction

The calculation of Zernike moments requires the distance and orientation with respect to the centre of mass for each pixel of the segmented object / pattern to be classified. Let us assume that we need to calculate Zernike moments in the case of the segmented human silhouette of Fig. 2(b). If P_0 is the centroid pixel, the distance and angle to any other pixel P_1 needs to be calculated. Let us use the calibration of the camera to obtain the spherical coordinates (θ_0, φ_0) and (θ_1, φ_1) on the unit sphere, of points P_0 and P_1 , respectively. Now, the distance and relative angle of P_1 with respect to P_0 can be measured on the unit sphere.

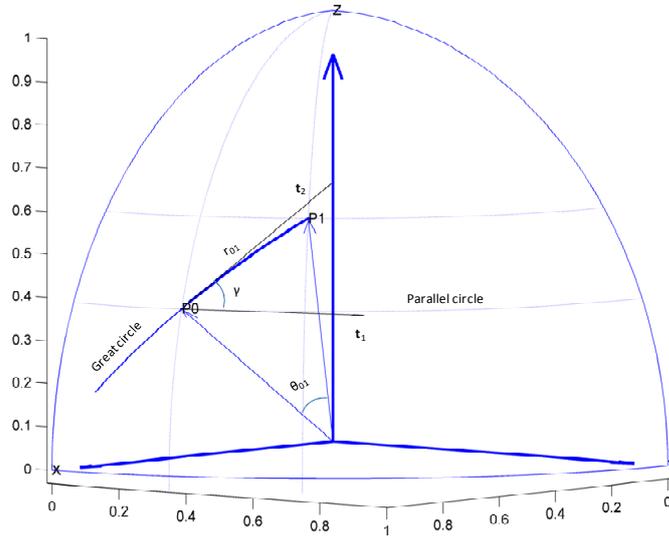


Fig. 3. The mapping of pixels P_0 and P_1 on the unit sphere and the calculation of their geodesic distance and angle. The great circle and the parallel circles through P_0 and P_1 are also shown.

It is well known that the geodesic curve of a sphere is a great circle. Thus, the distance between any two points on a sphere is the length of the arc that is defined by the two points and belongs to a circle that passes through the two points and has the same centre with the sphere. We will call this circle “great circle”. Let \mathbf{v}_0 and \mathbf{v}_1 be the position vectors pointing to P_0 and P_1 . The distance r_{01} of points P_0 and P_1 on the unit sphere is easily calculated, (assuming that \cos^{-1} returns the result in radians):

$$\mathbf{v}_0 \cdot \mathbf{v}_1 = (\cos \theta_0 \cos \varphi_0, \sin \theta_0 \cos \varphi_0, \sin \theta_0) \cdot (\cos \theta_1 \cos \varphi_1, \sin \theta_1 \cos \varphi_1, \sin \theta_1) =$$

$$(\cos \theta_0 \cos \varphi_0 \cos \theta_1 \cos \varphi_1 + \sin \theta_0 \cos \varphi_0 \sin \theta_1 \cos \varphi_1 + \sin \theta_0 \sin \theta_1) \quad (5)$$

$$r_{01} = \cos^{-1}(\mathbf{v}_0 \cdot \mathbf{v}_1). \quad (6)$$

Zernike moments, like any other definition of moment invariants, require the distance between any two pixels as well as their angle γ . If we assume the parallel circle passing through P_0 and the great circle passing through P_0 and P_1 , then angle γ is calculated as the angle between the tangent vectors $\mathbf{t}_0, \mathbf{t}_1$ of the two circles at P_0 . The great circle through P_0, P_1 is calculated as following:

$$\mathbf{v} = \lambda_0 \mathbf{v}_0 + \lambda_1 \mathbf{v}_1, \text{ with } |\mathbf{v}| = 1. \quad (7)$$

After some algebraic operations we obtain:

$$\lambda_0^2 + 2 \cos(\mathbf{v}_0 \cdot \mathbf{v}_1) \lambda_0 \lambda_1 + \lambda_1^2 - 1 = 0. \quad (8)$$

After solving the above equation and requiring λ_0, λ_1 to be real numbers, we obtain:

$$\begin{aligned} \lambda_0^2 (\cos^2(\mathbf{v}_0 \cdot \mathbf{v}_1) - 1) &\geq 0, \\ \lambda_1 &= -\cos^2(\mathbf{v}_0 \cdot \mathbf{v}_1) \lambda_0 \pm \sqrt{(1 - \cos^2(\mathbf{v}_0 \cdot \mathbf{v}_1)) \lambda_0^2 + 1} \end{aligned} \quad (9)$$

Since execution speed is essential and taking into consideration that the circle is a very smooth curve (with constant curvature), the tangent vector is calculated by finite differences, as following. It is obvious that for $\lambda_0 = 1, \lambda_1 = 0, \mathbf{v} = \mathbf{v}_0$. We set $\lambda_0 = 0.9$ and use Eq. (9) to calculate λ_1 . Then, $\mathbf{t}_1 = 0.9 \mathbf{v}_0 + [\lambda_1]_{\lambda_0=0.9} \mathbf{v}_1 - \mathbf{v}_0$. The tangent vector of the parallel circle at P_0 is trivially calculated: $\mathbf{t}_2 = (-v_{1y}, v_{1x}, 0)$. The angle γ is now easily obtained as

$$\gamma = \cos^{-1} \left(\frac{\mathbf{t}_1 \cdot \mathbf{t}_2}{|\mathbf{t}_1| |\mathbf{t}_2|} \right). \quad (10)$$

The sign of angle γ will be set equal to the sign of $(\theta_1 - \theta_0)$. Thus, the distance and angle (r, θ) in Eq. (3) are replaced by r_{01} defined in (6) and γ defined in Eq. (10). This correction will be called ‘‘geodesic’’ in the rest of the paper.

3 Experimental Results

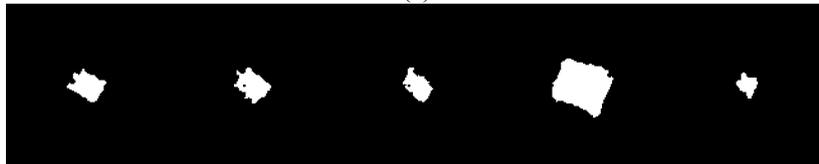
3.1 Experiments with a geometric shape

A rectangular shape of known constant RGB values was captured in a video sequence of approximately 500 frames with different distances and positions from the fish-eye camera. Fig. 4(a) shows portions from 5 such frames of the original RGB video and the corresponding segmented ones are shown in Fig. 4(b). Segmentation was performed by thresholding, based on the known RGB values of the test pattern. The

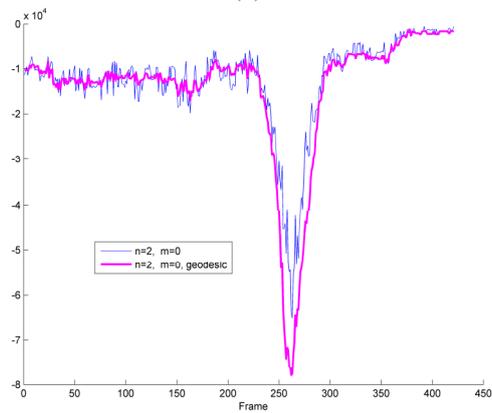
Zernike moments of the binary image were calculated, using the traditional definition and the geodesic corrections. Fig. 4(c) plots the Zernike for $n=2, m=0$. It is evident that the geodesic correction Zernike is less noisy than the traditional implementation. The negative peak is caused by a drastic change in scale. Fig. 4(d) and 4(e) show the fraction of Z_{60}/Z_{30} and the fraction Z_{40}/Z_{20} (see Eq. (3)). It can be observed that these fractions are immune to the sudden scale change. Furthermore, the fractions of the geodesically corrected Zernike are less noisy than the traditional ones.



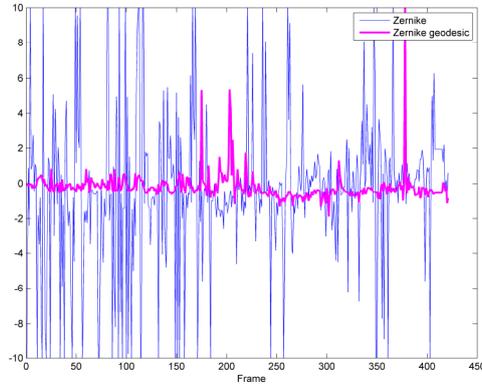
(a)



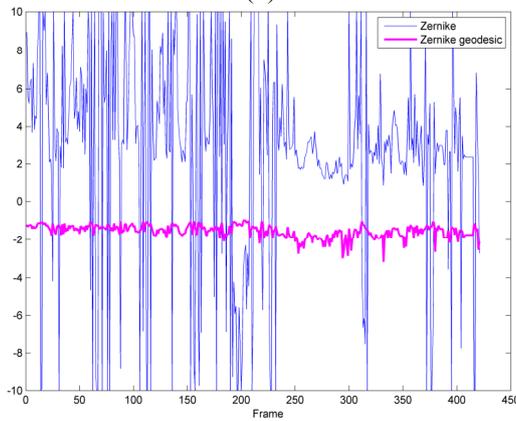
(b)



(c)



(d)



(e)

Fig. 4. Zernike moments before and after geodesic correction for the rectangular geometric shape (see text for details).

3.2 Classification of the human silhouette pose

In this experiment, we acquired 2 different videos of a human with two different classes of poses: walking/standing and sitting. In each video the subjects sits in 2 or 3 different places. The human silhouette is segmented using an unsupervised segmentation algorithm presented in [12] and the Zernike moments are calculated twice: without and with the geodesic correction. The simple k-Nearest Neighbour (k-NN) classifier with $k=3$ was employed to classify each frame into the two classes, using the Euclidean distance metric between the Zernike features.

The classical Zernike moments have been used to classify the human action of each frame, utilizing moments of the first n orders, where $n = 3, 4, 5$. In Table 1 the results for each video are presented. The classification results for both videos are quite similar for moments of the same order. More specifically, in video 1 the classification accuracy increases for higher order Zernike moments. On the other hand, in

the more complex video 2, the best results are obtained using $N = 4$. It is evident that higher order Zernike moments, results in an increase in computational complexity, as well as the CPU time needed to process each frame. The average number of frames per second (FPS) processed by applying the Zernike moments and the k-NN classifier is also reported.

Table 1. The resulting classification accuracy of the segmented silhouette pose for different orders of the traditional Zernike implementation. The frame processing rate is also provided.

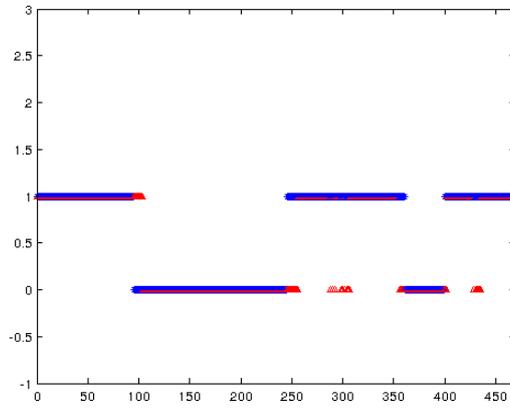
Zernike order	Classification Accuracy (%)		FPS
	Video 1	Video 2	
$n = 3, m = -1, +1$	84.22	85.77	16 fps
$n = 4, m = -2, 0, +2$	92.11	92.08	15 fps
$n = 5, m = -3, -1, +1, +3$	94.88	91.82	12 fps

To improve the classification results, the subset of the Zernike moments with radial symmetry were used. Table 2 exhibits the classification results for n up to 30, while $m=0$ for the classical and the geodesically corrected Zernike moments, respectively. The rate of frames processed is also provided. It can be observed that the classification results are improved by the introduction of geodesic correction, with minimal increase to the measured execution time.

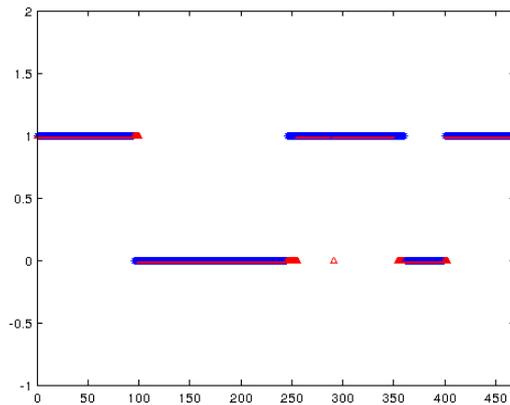
Table 2. The resulting classification accuracy of the segmented silhouette pose for different orders of the traditional radial Zernike implementation. The rate of frame processing is also provided.

Zernike order	Geodesic correction	Classification Accuracy (%)		Execution speed
		Video 1	Video 2	
$n = 2, 4, 6, 8, 10, m = 0$	NO	93.39	94.53	17 fps
	YES	94.09	96.13	17 fps
$n = 2, 4, 6, 8, \dots, 20, m = 0$	NO	91.68	94.14	12 fps
	YES	94.24	95.94	11 fps
$n = 2, 4, 6, 8, \dots, 30, m = 0$	NO	92.11	94.01	8 fps
	YES	92.24	95.75	9 fps

Fig. 5 graphically exhibits the classification results, using $n \leq 20, m = 0$, without and with geodesic correction. Class 0 corresponds to sitting silhouette and class 1 corresponds to standing/walking silhouette. The algorithmic results are shown in blue and the ground truth in red. It can be seen that the geodesic correction produces less incorrectly classified poses in the video frames, thus improves classification accuracy.



(a)



(b)

Fig. 5. Graphical representation of the results for pose estimation. Class 0: “sitting”, Class 1: “standing/walking”, using Central Zernike with $n \leq 10$, without geodesic correction (a) and with geodesic correction (b).

4 Discussion and Conclusions

As shown in the previous section the geodesic correction enhances the descriptive properties of Zernike moments, when applied to omni-directional images. The Zernike descriptors appear to be more stable (less variability for the same pattern). The accuracy of a classifier in terms of classifying two human postures was also increased when using Zernike moments with the proposed geodesic correction. Finally, the execution time was only marginally increased. The proposed methodology can

significantly improve the accuracy of more complex activity recognition algorithms usually found in ambient assisted living environments.

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. Mikolajczyk K., Schmid C.: Scale and affine invariant interest point detectors, *International Journal of Computer Vision*, 1(60), 63–86 (2004)
2. M.K. Hu, Visual pattern recognition by moment invariants, *IRE Trans. Inform. Theory* IT-8 (1962) 179–187.
3. M. Teague, Image analysis via the general theory of moments, *J. Opt. Soc. Am.* 70 (8) (1980) 920–930.
4. Y. S. Abu-Mostafa and D. Psaltis, "Recognitive Aspects of Moment Invariants", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 698-706, 1984
5. Flusser, B. Zitova, and T. Suk, "Invariant-based registration of rotated and blurred images," in *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International*, 1999.
6. M. Schlemmer, M. Heringer, F. Morr, I. Hotz, M. H. Bertram, C. Garth, W. Kollmann, B. Hamann, and H. Hagen, "Moment Invariants for the Analysis of 2D Flow Fields," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, pp. 1743-1750, 2007.
7. Javier Cruz-Mota, Iva Bogdanova, Benoît Paquier, Michel Bierlaire, Jean-Philippe Thiran, *Scale Invariant Feature Transform on the Sphere: Theory and Applications*, *Int J Comput Vis* (2012) 98:217–241.
8. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
9. Hansen, P., Corke, P., Boles, W., & Daniilidis, K., Scale invariant features on the sphere. In *International conference on computer vision (2007)*, pp. 1–8.
10. Tamimi, H., Andreasson, H., Treptow, A., Duckett, T., & Zell, A. (2006). Localization of mobile robots with omnidirectional vision using particle filter and iterative sift. *Robotics and Autonomous Systems*, 54(9), 758–765.
11. Bruno MARHIC, El Mustapha MOUADDIB, Claude PEGARD, *A Localisation Method with an Omnidirectional Vision Sensor Using Projective Invariant*, *Proceedings of the 1998 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems Victoria, B.C., Canada October 1998*
12. K.K. Delibasis, T. Goudas, V.P. Plagianakos, I. Maglogiannis, *Fisheye Camera Modeling for Human Segmentation Refinement in Indoor Videos, PETRA '13, May 29 - 31 2013, Island of Rhodes, Greece Copyright 2013 ACM 978-1-4503-1973-7/13/05.*