On Moving Object Reconstruction By Moments

Stuart P. Prismall, Mark S. Nixon and John N. Carter
Image, Speech and Intelligent Systems
Department of Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ
{spp00r,msn,jnc}@ecs.soton.ac.uk

Abstract

Recent research using statistical moments to describe moving shapes through an image sequence has led to an interest in reconstructing moving shapes from their moment description. This paper discusses how the moment description through a series of frames might be used to predict missing or intermediate frames within a sequence. Additionally, this highlights generic aspects of moment reconstruction which rarely receive more than scant attention. The ideas presented use Zernike moments, although the general framework is applicable to all types of moments. We show how a moving human silhouette can be reconstructed with accuracy by interpolation from a moment history.

1 Introduction

Statistical moments have a long history in computer vision since the original work of Hu [1] on moment invariants in the early 1960's. They are particularly popular due to their compact description, their capability to select differing levels of detail and their known performance attributes. The ability to reconstruct a shape from its moment description is often cited as justification for their deployment, but has received much less attention than their descriptive capabilities for object recognition, e.g. Dudani *et al.*, [2]. More recently moments have been applied to image sequences, to describe moving shapes, for recognition purposes (Shutler *et al.*, [3]), thus prompting an interest in their use for reconstructing (moving) shapes. In this new scenario, an object's moment description allows for the prediction of intermediate or missing appearances. One potential application is for the synchronisation of the source imagery in multi-view 3D human reconstruction.

One approach to moving-object, or even frame, prediction (known as 'in-betweening' in broadcast technology) is to use optical flow. An alternative for objects is to deploy tracking approaches for prediction. Both approaches require relatively fast sampling to ensure either sufficient accuracy in the estimates of optical flow or sufficient tracked

BMVC 2002 doi:10.5244/C.16.5

history to ensure that the prediction is valid. Using the new moment based approach can benefit from the compactness of the moment description and for a potentially lower sampling rate, given sufficient samples for accurate interpolation of the moments' history. This is especially true when reconstructing humans and their movement as the motion of the limbs can be faster than video-rate sampling.

In Section 2, orthogonal Zernike moments are described and reconstruction is considered. Section 3 describes how moments in a sequence can be interpolated, and, in particular, how this can be applied to reconstructing moving people. In Section 4, we present some early results that demonstrate the validity of our new approach, while in Section 5, we assess the preliminary results and outline the future directions of the research.

2 Zernike Moments

There are many different types of moments that have been applied to computer vision problems (geometric, Legendre etc.), but it has been demonstrated in [4] that the orthogonal Zernike moments offer a set of moments which are highly uncorrelated with little information redundancy.

2.1 Zernike Theory

The orthogonal Zernike moments, first proposed by Teague [5], utilise the Zernike polynomial as basis function and are defined over the unit disc (in polar coordinates) by:

$$Z_{mn} = \frac{m+1}{\pi} \int_0^{2\pi} \int_0^1 V_{mn}^*(r,\theta) f(r,\theta) \, r \, dr d\theta \tag{1}$$

where m is the order of the moment (with $m \geq 0$) and n represents the repetition (where $|n| \leq m$, and $m + n$ is even). $V_{mn}(r,\theta)$ is the complex-valued Zernike polynomial with * indicating the complex conjugate. For a discrete square image (size NxN), Z_{mn} can be calculated with:

$$Z_{mn} = \frac{m+1}{\pi} \frac{1}{(N-1)^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} V_{mn}^*(r,\theta) f(x,y) \tag{2}$$

given that $r = (x^2 + y^2)^{1/2} / \sqrt{2}N$ and $\theta = \tan^{-1}(y/x)$, in order to map the image into the unit disc. Note that Equation 2 is only orthogonal over the unit circle. The Zernike polynomial, $V_{mn}(r,\theta)$, is defined as:

$$V_{mn}(r,\theta) = R_{mn}(r) e^{-jn\theta} = R_{mn}(r)(\cos n\theta - j \sin n\theta) \tag{3}$$

where the radial polynomial, $R_{mn}(r)$, is:

$$R_{mn}(r) = \sum_{s=0}^{m-|n|/2} \frac{(-1)^s (m-s)!\, r^{m-2s}}{s!\left(\frac{m+|n|}{2}-s\right)!\left(\frac{m-|n|}{2}-s\right)!} \tag{4}$$

This polynomial is such that over the unit disc, $|R_{mn}(r)| \le 1$ and that $R_{mn}(1) = 1$, for any values of m and n. The definition of the radial polynomial also leads to $R_{mn}(r) = R_{m,-n}(r)$. It can then easily be shown that:

$$V^*_{mn}(r,\theta) = V_{m,-n}(r,\theta) \tag{5}$$

From which it follows that:

$$Z^*_{mn} = Z_{m,-n} \tag{6}$$

The number of Zernike moments for any order, m, is given by $m + 1$, while the number of moments up to and including order m is $(m/2 + 1)(m + 1)$, (although because of the relationship between Z_{mn} and $Z_{m,-n}$ given above, only the moments with $n \ge 0$ need to be known).

2.2 Reconstruction from Zernike Moments

It is a well-recognised property of moments that they can be used to reconstruct the original function, i.e. none of the original image information is lost in the projection of the image on to the moment basis functions, assuming an 'infinite' number of moments are calculated. For non-orthogonal moments, the reconstruction is not straightforward and requires a moment-matching technique [5]. In the case of orthogonal moments like Zernike, the reconstruction is simple, by virtue of the orthogonality of the basis functions [6]. For an image for which all the Zernike moments up to and including order p, the reconstructed image, $g(x,y)$, is given by:

$$g(x,y) = \sum_{m=0}^{p} \sum_{n} Z_{mn} V_{mn}(r,\theta) \tag{7}$$

with the same constraints on the repetition index, n, as before. As with non-orthogonal moments, it remains true that as p approaches infinity the reconstructed function $g(x,y)$ approaches the original function $f(x,y)$. More simply, more moments suggests a better reconstruction.

It is worth noting that this reconstruction formula gives a discrete approximation to a continuous function, i.e. while the values of x and y are discrete, the values of $g(x,y)$ are from a continuous range. Another way of considering this point is that if the original image is binary, then the reconstructed function, $g(x,y)$, will not simply take the values 0 and 1. The reconstruction effectively gives the simplest (smoothest) function whose moments match the given set. Therefore, when reconstructing binary images such as those in Figure 1a, the reconstructed image needs to be thresholded to reproduce a

(a) (b) (c) (d)

Figure 1: a) original image, b) reconstruction order 10 (66 moments), c) reconstruction order 15 (136), and d) reconstruction order 20 (231).

binary image. An appropriate threshold, used in [6], would appear to be the mid-point between the maxima and minima of the reconstructed function, but there appears to be little discussion in the literature on thresholding reconstructions. This will be discussed further in Section 4. Using an example of a 64x64 binary letter 'A', Figure 1 illustrates how increasing the number of moments in the reconstruction improves the resulting image (which are shown here after thresholding as described above).

It is quite clear from Figure 1 that the letter 'A' is recognisable from the reconstruction up to and including order 15 (a total of 136 moments). However, it is also clear that the reconstructed image does not match the original. Thus, we see that recognition requires fewer moments than reconstruction. Most previous work [5,6] on reconstruction from moments has concentrated on recognition rather that accuracy, but here we wish to concentrate on the accuracy alone.

It is known and confirmed in Figure 1 that the higher order moments capture increasingly higher frequencies within a function and in the case of an image the higher frequencies represent the detail of the image. This is also consistent with work on other types of reconstruction, such as eigenanalysis where it has been found that increasing numbers of eigenvectors are required to capture image detail [7] and again exceed the number required for recognition. Thus, when considering image reconstruction from moments, the number of moments required for accurate reconstruction will be related to the frequencies present within the original image. For a given image size it would appear that there should be a finite limit to the frequencies that are present in the image and for a binary image that limiting frequency will be relatively low. As the higher order moments approach this frequency the reconstruction will become more accurate.

3 Interpolating Gait

Articulated motions (such as human gait) are periodic, and it is this periodicity that can be exploited to predict frames within a sequence. It is known that the highest angular frequency within human walking gait is approximately 5Hz [8], while video sequences are recorded at a rate well above the Nyquist one.

Since the moments of an image are shape descriptors, it follows that any particular moment will vary periodically across a sequence, since the shapes will repeat themselves. It is therefore possible that the moments from a corrupted or missing frame in a sequence can be predicted from the values in the sequence.

Figure 2 shows silhouettes for a full human gait cycle (a heel strike through to the next heel strike of the same foot). These images are derived from a sequence of a human subject walking in a laboratory environment. However, it should be noted that these images have not undergone any normalisation. In particular, it can be seen that the

silhouettes are not centralised in the image space (e.g. compare silhouette 11 with silhouette 19). Normalisation of the silhouettes was achieved by calculating the centre of mass (COM) in the *x*- and *y*-directions using geometric moments (0^{th} and 1^{st} order). The silhouette was then translated, but by a whole number of pixels to retain a binary image. In the case of these 64x64 images, this means that the COM occurs in one of the four pixels that make up the centre of the image.

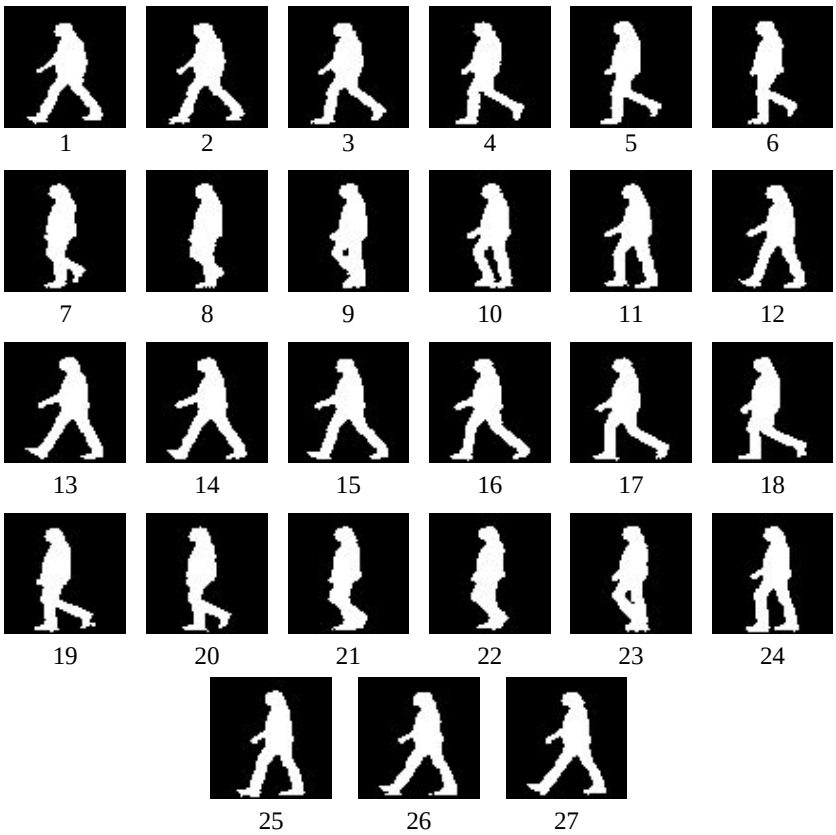
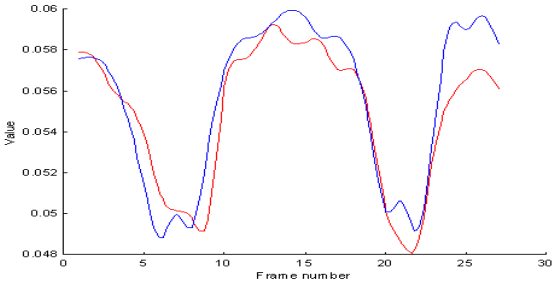
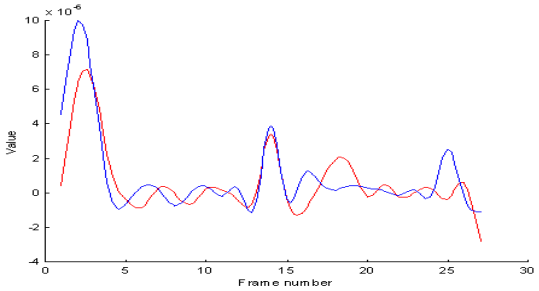


Figure 2: An example gait cycle sequence in silhouette.

Figure 3 shows plots of particular moment values across two sequences of the same subject, the curves produced by cubic spline interpolation. It is clear from these plots that the moment values show periodicity, with varying degrees of smoothness. In addition, we see that the inter-subject variability is fairly small. Figure 4 shows the same plots but for two different subjects. Again the periodicity is present (as expected), but we also see that there is some intra-subject variability. However, the general shape of the plots is similar, which is to be expected since the general shape patterns between subjects are similar. The curve of each moment for a sequence reflects how the particular level of detail changes through the sequence.

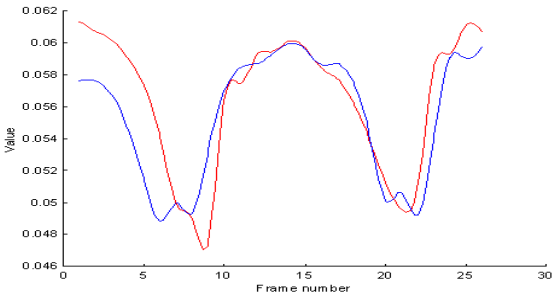


(a)

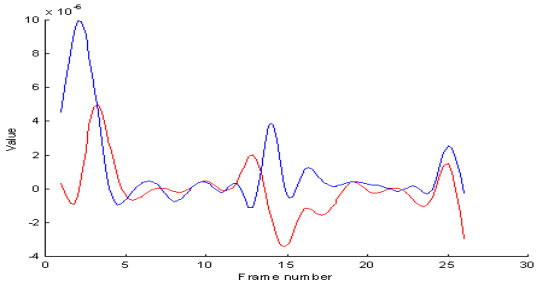


(b)

Figure 3: Moment value plots for two sequences of the same subject, showing a) Z_{00}, and b) $Z_{20,20}$.



(a)



(b)

Figure 4: Moment value plots for two different subjects, showing a) Z_{00}, and b) $Z_{20,20}$.

We expect that the moment value prediction from these plots will benefit from interpolation using cubic splines. However, in Section 4 we present some initial findings using linear interpolation to determine basic properties.

4 Results

The reconstruction of silhouette 23 from Figure 2 using different numbers of moments is shown in Figure 5. Here we can see that, as in Figure 1, there is a general improvement in the quality of the reconstruction as the number of moments used is increased. The reconstructions through order 35 show an approximately 1% error over the original (44 pixels incorrect in 4096 pixels), while using orders through 55, the error is less than 0.6% (24 pixels incorrect).

The effect of using different thresholds for the reconstructed images was also investigated. The basic reconstructions were mapped to values between 0 and 1. Various thresholds (0.3, 0.4, 0.5, 0.6 and 0.7) were then applied to the mapped reconstructions. The differences between the reconstructed image and the original were

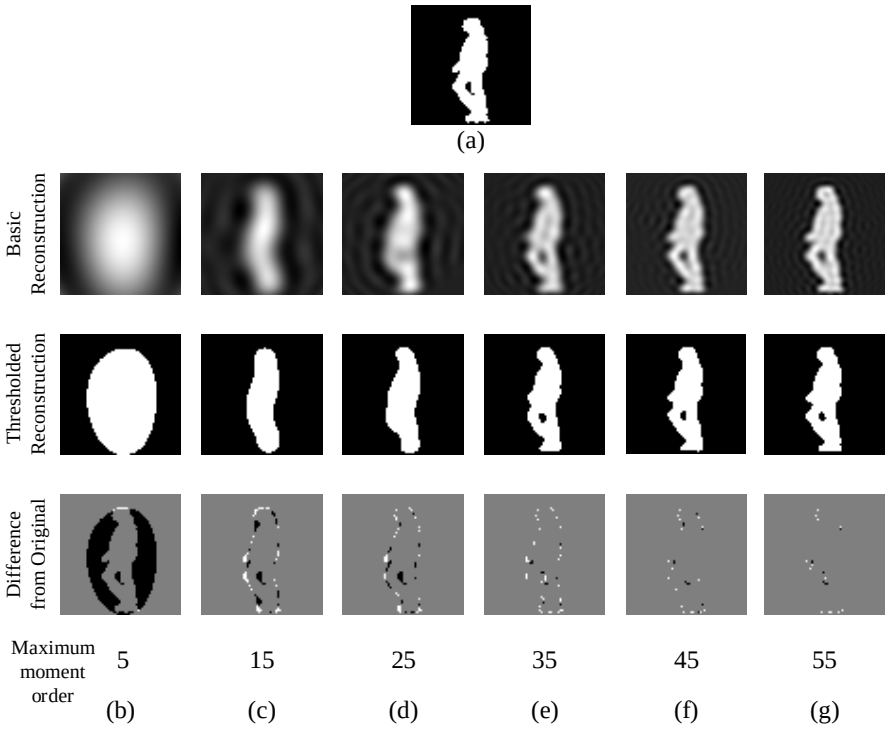


Figure 5: Reconstructions of a 64x64 human silhouette, a) original image (silhouette 23 in Figure 2), b) using maximum order 5 (21 moments), c) order 15 (136), d) order 25 (351), e) order 35 (666), f) order 45 (1081), and g) order 55 (1596). The thresholded images were thresholded at the mid-point between the maximum and minimum values. The difference images show grey pixels for correct reconstruction, black for pixels incorrectly added, and white for pixels incorrectly removed.

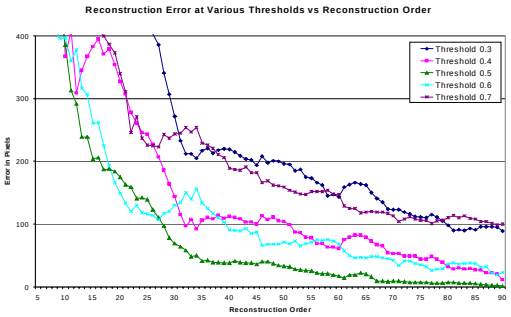


Figure 6: Reconstruction error plots using various thresholds.

measured. The plot in Figure 6 displays the reconstruction error (in pixels) for the different thresholds against the reconstruction order, for silhouette 2 in Figure 2. These plots show that the threshold at 0.5 offers the best general performance, although at lower orders it appears that other thresholds can offer similar or even improved performance. However, the error metric is simply a total of incorrectly assigned pixels, and does not take account of the distribution of the error pixels. The error at order 35 is 41 pixels (i.e. approximately 1% of the 64x64 image). Note also that reconstruction through order 90 (a total of 4186 moments) yields a perfect reconstruction.

The silhouettes in Figure 2 were normalised and then used to test the moment value prediction hypothesis with regard to reconstruction. Every other frame was used to predict the intermediate frame in the sequence (i.e. frame 1 and frame 3 to predict frame 2, 2 and 4 to predict frame 3, etc.) by linear interpolation of the equivalent moment values. Figure 7 shows the reconstruction error for each predicted frame when reconstructing using moments orders through 35, together with the reconstruction error at the same order using the moments from the actual frame. Naturally, reconstruction by the real moments of a frame is better, but in some cases the interpolated reconstruction is quite close. On average the interpolated reconstruction shows a 3.1% error over the original image. At best (in frame 22) the error is 1.8%, while the worst performer (frame 24) has an error of 6.6%. Frame 24 probably has the largest inter-frame movement so this may explain its poorer performance.

Figure 8 illustrates how the reconstruction error of the predicted frame reaches a minimum error which does not improve with increasing moment order, and compares the

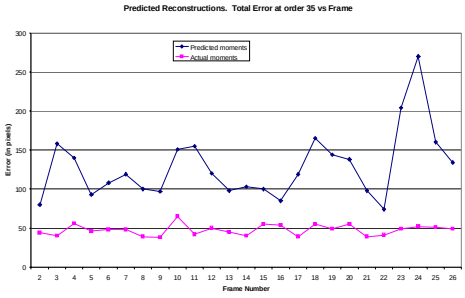


Figure 7: Reconstruction errors with predicted and real value moments showing the error for each frame using reconstruction through order 35.

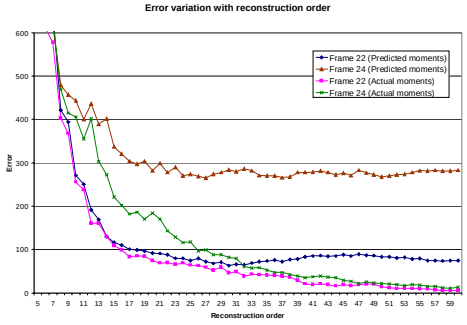


Figure 8: Reconstruction errors with predicted and real value moments showing the error variation with increasing reconstruction order.

reconstruction error of the silhouette from its actual moments. This lack of improvement of the reconstruction error is probably a direct result of the inaccuracies of using linear interpolation for the prediction. The degree to which the inaccuracy hampers reconstruction can be seen by how much the error of the predicted frame diverges from the error in reconstruction from the actual moments. For Frame 22 in Figure 8 we can see that the error is approximately the same in both cases up to approximately order 30, at which point reconstruction from the actual moments becomes significantly better.

Figure 9 shows a reconstruction example from linear interpolated data. Figure 9b (frame 9 from Figure 2) is the frame to be predicted from the moment data of the frames in Figures 9a (frame 8) and 9c (frame 10). The frames in Figures 9d, 9e, and 9f display the results of reconstruction up and including moment order 35. Whilst this frame is one of the better performing interpolated reconstructions (see Figure 7), of particular interest in this example is how the occluded leg in Figure 9a (which is unoccluded in Figure 9c) can be seen by the moment interpolation technique (Figure 9e). The retention of the general shape of the silhouette demonstrates that the interpolation approach is valid. However, it is clear from the error image in Figure 9f that many of the errors are in the area undergoing most change (i.e. the legs). The crude reconstruction in Figure 9d reflects how these errors arise, where we can see that the leg area is less bright. This can be attributed to inaccuracies in the interpolated moment values causing some values to

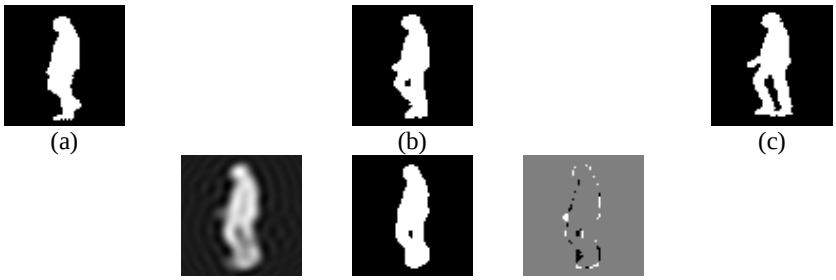


Figure 9: Reconstruction of a silhouette by linear interpolation with a) the first reference silhouette, b) the silhouette to be predicted, c) the second reference silhouette, d) showing the crude reconstruction at order 35, e) reconstruction thresholded at 0.5, and f) its error image by comparison with (b) (97 incorrect pixels).

conflict with each other, leading to a more blurred (uncertain) image. However, it should be noted that the linear interpolation is a very crude form of interpolation. A more appropriate form of interpolation should lead to quantifiably better results.

5 Conclusions and Further Work

This paper shows how moments can be used to predict missing data within image sequences. Zernike moments have been used but the fundamental idea is applicable to any type of moment, although preferably an orthogonal one since this gives a practical route to reconstruction. It is has been shown that binary images can be accurately reconstructed from Zernike moments if high orders are used but has highlighted some generic factors rarely discussed in the literature. However, further work is needed on reconstruction, and in particular, the relationship of the moment order to image frequency content, and how this can be used to improve the accuracy of reconstruction and reduce the number of moments required.

Additionally, it has been shown that the moment values can be interpolated in a sequence. Moment values that have been predicted by linear interpolation have been successfully used to predict a missing frame in a sequence, and it is expected that cubic spline interpolation will further enhance performance. In addition, the recent velocity moment descriptions may provide a route towards the same aim.

Acknowledgments

We gratefully acknowledge partial support by the European Research Office of the US Army under Contract No. N68171-01-C-9002.

References

[1] M.K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, **IT-8**:179-187, 1962.

[2] S.A. Dudani, K.J. Breeding and R.B. McGhee. Aircraft identification by moment invariants. *IEEE Trans. on Computers*. **C-26**(1): 39-46, 1977.

[3] J.D. Shutler, M.S. Nixon and C.J. Harris. Zernike velocity moments for description and recognition of moving shapes, *BMVC 2001*. 705-714, 2001.

[4] C-H. Teh and R.T. Chin. On image analysis by the method of moments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **10**(4): 496-513, 1988.

[5] M.R. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*. **70**(8): 920-930, 1979.

[6] A. Khotanzad and Y.H. Hong. Invariant image recognition by Zernike moments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **12**(5): 489-497, 1990.

[7] D.C. Schuurman and D.W. Capson. Video-rate eigenspace methods for position tracking and remote monitoring. *Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*. 45-49, 2002.

[8] C. Angeloni, P.O. Riley and D.E. Krebs. Frequency content of whole body gait kinematic data. *IEEE Trans. Rehabilitation Engineering*, **2**(1): 40-46, 1994.