Novel Temporal Views of Moving Objects for Gait Biometrics

Stuart P. Prismall, Mark S. Nixon, and John N. Carter

Department of Electronics and Computer Science
University of Southampton, Southampton, SO17 1BJ

Abstract. There is increasing interest in novel view reconstruction but less for new time-based views of moving objects as needed for gait biometric deployment. Our interests concern reconstructing moving shapes from their moment history with a view to constructing new temporal views. This paper shows how the moment description through an object sequence can be used to predict missing or intermediate frames within the sequence. Additionally, this highlights generic aspects of moment reconstruction which rarely receive more than scant attention. We use Zernike moments for the convenience of reconstruction, although the framework is applicable to all types of moments. As an example, we show that by interpolating the moment history of a moving human silhouette, a missing frame can be constructed with accuracy, providing a practical basis for the construction of new temporal views of moving objects.

1 Introduction

Statistical moments have a long history in computer vision and 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. in [2]. More recently moments have been applied to image sequences, to describe moving shapes, for recognition purposes [5], 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, and entirely new temporal views. 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, while an alternative for objects is to deploy tracking approaches for prediction [8]. Both methods require relatively fast sampling to ensure either sufficient accuracy in the estimates of optical flow or sufficient tracked history to ensure that the prediction is valid. Using the new moment based approach can benefit from the compactness of the moment description

J. Kittler and M.S. Nixon (Eds.): AVBPA 2003, LNCS 2688, pp. 725-733, 2003.

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 effect, the changing shape of the object is being followed by a well-recognised shape description method. We have previously shown the viability of this new approach [4], and here we demonstrate new factors to improve the reconstruction and extend the analysis of the prediction of intermediate frames, thus improving potential for biometric deployment.

Section 2 describes orthogonal Zernike moments and reconstruction. 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.), though it has been demonstrated in [7] that the orthogonal Zernike moments are highly uncorrelated with little redundancy.

2.1 Zernike Theory

The orthogonal Zernike moments, first proposed in [6], use the Zernike polynomial as basis function and are defined over the unit disc (in polar coordinates) by:

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

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 and is defined elsewhere, e.g. in [6]. Strictly, the Zernike polynomial should be normalized first (using the root of the normalization factor shown in (1)), though historically the normalization in (1) has been used. For a discrete square image (size $N \times N$), Z_{mn} can be calculated with:

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

using a suitable translation in order to map the image into the unit disc.

2.2 Reconstruction from Zernike Moments

Moments 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 is calculated. In the case of orthogonal moments like Zernike, the reconstruction is simple, by virtue of the orthogonality of the basis functions [3]. If all the Zernike moments up to and including order *p* are known, the reconstructed function, *g*(*x,y*), is given by:

$$g(x,y) = \sum_{m=0}^{p} \sum_{n} Z_{mn} V_{mn}(r,\theta) \tag{3}$$

with the same constraints on the repetition index, *n*, as before. Notice that the Zernike polynomials used in (3) are not normalised. In the limit when *p* approaches infinity the reconstructed function *g*(*x,y*) approaches the original function *f*(*x,y*). Most previous work on reconstruction from Zernike moments has concentrated on recognition, and here we concentrate on accuracy.

The reconstruction of a function in this way can be seen as a summation of weighted (by the moment value) Zernike basis functions. Since these basis functions are continuous, it is clear that the reconstructed function will also be continuous. Therefore, any representation of the reconstructed function needs to be approximated to discrete values. In the case of a binary object this is achieved by using an appropriate threshold. The selection of this threshold has received very little attention and is usually set at the mid-point between the minimum and maximum values of the reconstructed function as in [3]. This is perhaps intuitive and recent work [4] has confirmed that this threshold performs well as a generic value. However, it should be noted that for a binary image there is information available which can be used to steer the selection of a threshold, since the zero order moment is a representation of the mass of the object. Thus the reconstructed object can be thresholded such that number of pixels in the reconstructed object matches that of the original object. We refer to this as *adaptive thresholding*.

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 [1], and video sequences are sampled above the Nyquist rate. Since the moments of an image are shape descriptors, it follows that any particular moment will vary periodically across a sequence. It is therefore possible that the moments from a corrupted or missing frame in a sequence can be predicted. We can also consider a missing frame to be one that cannot be captured in data acquisition, i.e. a new temporal view.

Fig. 1 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). As individual (independent) objects no normalisation is necessary for moment calculation and subsequent reconstruction. However, if we wish to use the objects in relation to each other then we require them to have a common centre of mass (COM). This normalisation of the 64×64 silhouettes was achieved by calculating the COM in the x- and y-directions using geometric moments (0th and 1st order). The silhouette was then translated (to retain a binary image) such that the COM occurs in one of the four pixels that make up the centre of the image.

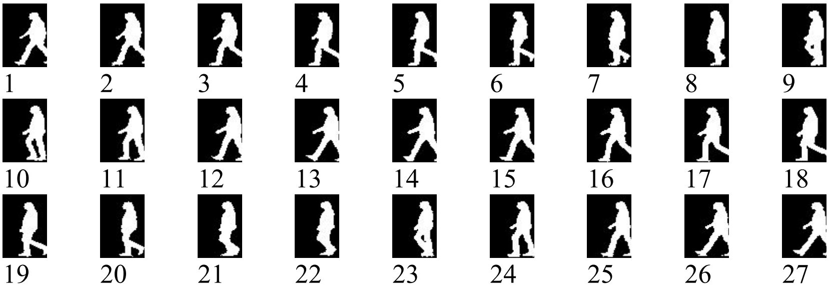
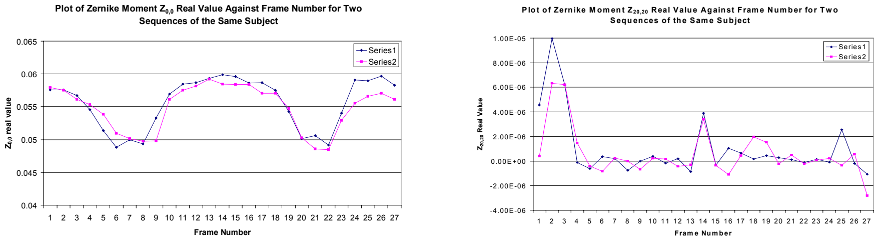


Fig. 1. An example gait cycle sequence in silhouette

Fig. 2 shows plots of particular moment values across two sequences of the same subject, with linear interpolation between each value of the same sequence. The moment values show periodicity, with varying degrees of smoothness. The shape of each plot for a moment reflects how the particular level of detail changes through the sequence. As we would expect, the low frequency moment shows a smooth change reflecting how the general shape and size shows little variation over the sequence. The higher frequency moment shown in Fig. 2b is much less smooth demonstrating how the detail can change rapidly between the frames. Both plots show little inter-subject variability. By using these plots to describe the moving silhouette through time, it is possible to predict new temporal views of the silhouette. In particular, this has applications in the normalization of gait sequences.

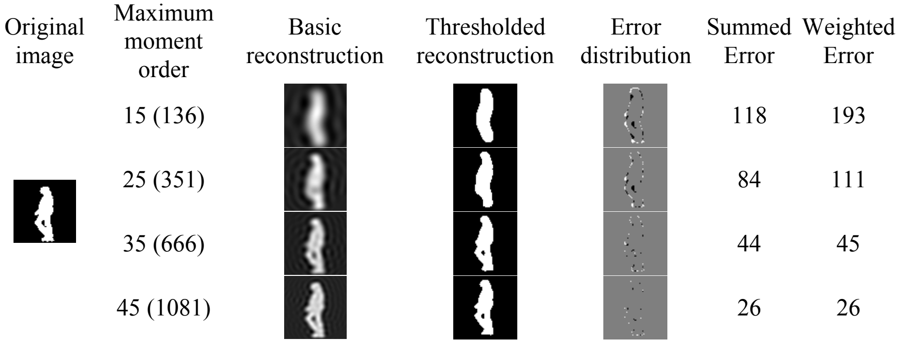


(a) $Z_{0,0}$ (b) $Z_{20,20}$

Fig. 2. Moment value plots for two sequences of the same subject

4 Results

For each silhouette, a set of moments was calculated, from which reconstructions to various orders were conducted. The basic reconstructions were thresholded to create a binary image, which was then compared to the original silhouette. The reconstruction error was determined by summing the pixels that differed between the two images. In addition, a weighted error was calculated in which incorrect pixels were weighted by the square of the distance to the nearest 'correct' pixel, giving additional information as to the accuracy of a reconstruction.

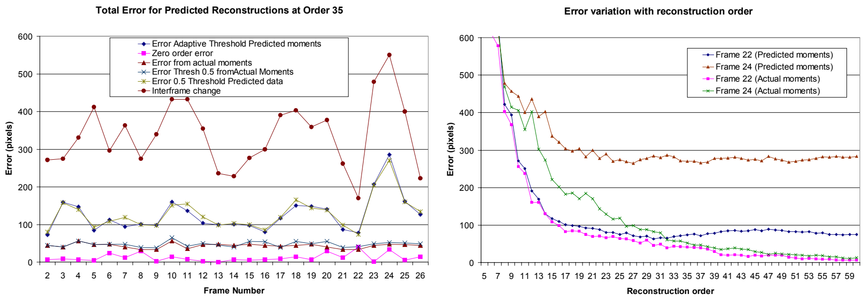


Original image	Maximum moment order	Basic reconstruction	Thresholded reconstruction	Error distribution	Summed Error	Weighted Error
	15 (136)				118	193
	25 (351)				84	111
	35 (666)				44	45
	45 (1081)				26	26

Fig. 3. Reconstructions of a 64×64 human silhouette (silhouette 23 in Fig. 1) at various orders (total components). The error distribution images show grey pixels for correct reconstruction, black for pixels incorrectly added, and white for pixels incorrectly removed

The reconstruction of silhouette 23 from Fig. 1 using different numbers of moments is shown in Fig. 3. The thresholded images in Fig. 3 are derived from the basic reconstructions by using the adaptive thresholding described in Section 2. Here, as expected, the quality of the reconstruction improves with increase in the number of moments used. In particular, the level of detail improves with the higher order moments and that the weighted error becomes comparable to the summed error implying that the incorrect pixels are less grouped. After applying the threshold, the reconstructions through order 35 show an approximately 6.3% error over the original (44 pixels incorrectly assigned in the 695 pixels of the object), while using orders through 45, the error is less than 3.7% (26 pixels incorrect).

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. frames 1 and 3 to predict frame 2, 2 and 4 to predict frame 3, etc.) by linear interpolation of the equivalent moment values. Fig. 4a shows how the error in the reconstruction varies with the frame. Of particular note is the adaptive threshold generally has similar or better performance than the fixed threshold. In some respects this may be surprising since the adaptive threshold relies on accurate prediction of the zero order moment. As we can see in Fig. 4a, even when there is quite a large error in the zero order moment (e.g. in frame 22), the adaptive threshold still performs well. Obviously, the predicted reconstructions are less accurate than the equivalent reconstructions from the actual moment data. This is only to be expected since we have only used a simple method to interpolate the data. In Fig. 4 the best performing prediction at order 35 (frame 2) shows an approximately 9.7% error over the original (72 pixels incorrectly assigned out of 741 pixels in the object), while the worst performing (frame 24) the error is 37.5%. The reconstructions from actual moment data show a fairly consistent error across the sequence, but the error in the predicted reconstructions is much more erratic and in fact seems to have more in common with the underlying interframe change (in terms of pixels). It is easy to surmise from this that where we have a large interframe change, there is the greatest change of the moment values, and hence the prediction of the values will be subject to the most error, leading to a poorly predicted frame. This would then lead to the hypothesis that a large average error in the predicted moments would lead to a large error in the reconstructed frame. However, the true situation appears to be more complex. For while it is indeed true that the worst performing frame in Fig. 4a (frame 24) has the largest average error in the predicted moment values, the second worst performing (frame 23) is only the 13th worst performing of the 25 frames being considered. Referring to equation (3), we see that the reconstruction is a summation of the Zernike polynomials, weighted by the moment values. It is therefore clear that large moment values have a greater influence on the reconstruction than small values. Therefore, a large relative error in the prediction of a small-valued moment will probably have less effect on the final reconstruction than a small relative error on a large-valued moment. Thus, it is unlikely that there would be a direct correlation between the average error in the predicted moment values and the error in the resulting reconstructed image.



(a) (b)

Fig. 4. a) Reconstruction error for each frame (predicted and actual) for fixed and adaptive thresholds, and b) Reconstruction errors with predicted and real value moments showing the error variation with increasing reconstruction order

Fig. 4b shows how the reconstructions perform over a range of orders. In the case of predicted reconstructions there is a level of accuracy that is reached that is not improved upon with the inclusion of higher orders. This lack of improvement of the reconstruction error is probably a direct result of the inaccuracies of using linear interpolation for the prediction. For frame 24 where the linear interpolation performs badly (from Fig. 4), the predicted reconstruction diverges from the actual reconstruction at a low order leading to a poor result at higher orders. In the case of frame 22 this divergence between the predicted and actual reconstructions happens at around order 30 giving a more accurate reconstruction. In effect, the low level detail is incorrect and adding the high level detail has little effect on the overall accuracy.

To further demonstrate the validity of the interpolation technique, linear interpolation was used to predict frames using larger interframe distances (i.e. more than two frames). The same gait sequence in Fig. 1 was used, and this time we chose only to predict frame 14 with increasingly large interframe distances (2 to 26) with the predicted frame falling at the midpoint. In Fig. 5, a plot of the error in the reconstructed image (using moments through order 35) against the interframe distance is shown. Here, as we would expect, we find that the image reconstructed from the predicted moments becomes less accurate as the gap increases. However, note that once the interframe distance becomes greater than half the length of the gait cycle sequence (approximately 14), the predicted object starts to become more accurate. This is because at these distances the reference frames become increasingly similar to the frame we wish to predict, hence the interpolation becomes more accurate.

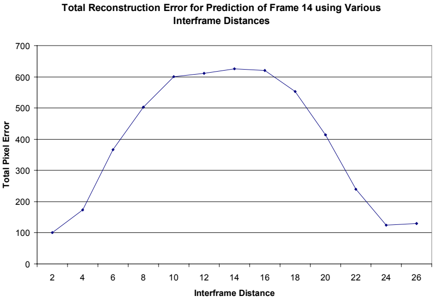


Fig. 5. Reconstruction error for frame 14 using increasing interframe distance for interpolation

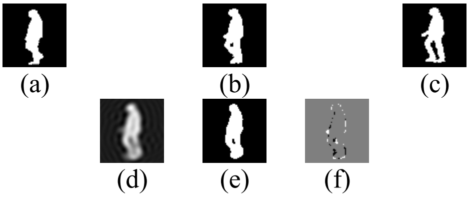


Fig. 6. Silhouette reconstruction: a) first reference silhouette, b) silhouette to be predicted, c) second reference silhouette, d) crude reconstruction at order 35, e) reconstruction with adaptive thresholding, and f) error image by comparison with (b) (92 incorrect pixels)

In terms of the actual images Fig. 6 shows a reconstruction example from linear interpolated data. Fig. 6b (frame 9 from Fig. 1) is the frame to be predicted from the moment data of the frames in Figs. 6a (frame 8) and 6c (frame 10). The frames in Figs. 6d, 6e, and 6f display the results of reconstruction up and including moment order 35. Whilst this frame is one of the better performing reconstructions (see Fig. 4), of particular interest in this example is how the occluded leg in Fig. 6a (which is unoccluded in Fig. 6c) can be seen by the moment interpolation technique (Fig. 6e). 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 Fig. 6f that many of the errors are in the area undergoing most change (i.e. the legs) as expected. The basic reconstruction in Fig. 6d reflects how these errors arise, where we can see that the leg area is less bright. In effect, the predicted moment values have conflicted with each other (due to interpolation errors), leading to the blurred (i.e. more uncertain in binary terms) image. However, it should be noted that we have used linear interpolation to demonstrate that using interpolation is viable. This suggests that using more sophisticated interpolation (such as cubic splines) could improve matters.

5 Conclusions and Further Work

We have shown how moments can be used to predict missing data within object sequences and to produce new temporal views which has application in gait biometrics. Zernike moments have been used but the fundamental idea is applicable to any type of moment, although preferably an orthogonal one since this gives practicable reconstruction. It is has been shown that binary images can be accurately reconstructed from Zernike moments and has highlighted some generic factors rarely discussed in the literature. 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 a more accurate form of interpolation such as cubic splines will significantly reduce the error and lead to an efficient method to obtain accurate new temporal views. 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. In the light of this, using Fourier components in a similar fashion is also worthy of investigation.

Acknowledgments

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

References

[1] 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.

[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] 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.

[4] S.P. Prismall, M.S. Nixon and J.N. Carter. On moving object reconstruction by moments. *BMVC 2002, Proceedings of the 13th British Machine Vision Conference*, 73-82, 2002.

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

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

[7] 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.

[8] T.R. Tuinstra and R.C. Hardie. High-resolution image reconstruction from digital video by exploitation of nonglobal motion. *Optical Engineering*, **38**(5): 806-814, 1999.