

## GLOBAL STATISTICAL DESCRIPTION OF TEMPORAL FEATURES

**Jamie D. Shutler, Mark S. Nixon and Chris J. Harris**  
ISIS research Group, Dept. of Electronics and Computer Science,  
University of Southampton, Southampton. SO17 1BJ, UK  
{jds98r,msn,cjh}@ecs.soton.ac.uk

**KEY WORDS:** Moments, Moment Reconstruction, Temporal Description, Statistical Description, Image processing

### ABSTRACT

In order to describe a temporal image sequence, global information about each separate image and their interrelations is required. Traditional moments, as applied to single images, are designed to capture global information about an object, but are unable to describe temporal changes. In this paper a new global statistical description is proposed - velocity moments, to compress and characterise information from temporal (moving) features. These new velocity moments are based on traditional moment theory. The sequence of images to be characterised is treated as a single entity and the moments aim to describe how different images are interrelated. The new method shows promising recognition properties when used with simple synthetic sequences of moving shapes. Due to their structure they exhibit similar properties to traditional moments when reconstruction is applied using a technique called moment matching. The noise performance of the velocity moments and a set of traditional Hu invariant moments are evaluated. Experimental results show that the velocity moments exhibit a higher resilience to perimeter noise, which suggests that they are better suited to describing real-world, extracted temporal features. As such, a new statistical method has been proposed to statistically describe moving shapes, which combines information about the shape's structure with information about its movement.

### 1 INTRODUCTION

Initially this work began by using traditional statistical moment theory to describe the motion of a shape through a sequence of images. However, this does not include detailed information regarding the motion, as there is no information linking the images together. A method has been developed using the general theory of moments, which not only contains information about the pixel structure of the shape, but also how the shapes' movement flows between images, producing a metric which describes the complete temporal sequence.

The application of classical moments to two dimensional images was first shown in the early sixties by Hu (Hu, 1962). Hu tested their validity using a simple experiment to recognise written characters. Hu was only concerned with images without noise, but further work (Teh and Chin, 1988), showed that traditional moment performance degrades where the view is occluded or noisy. A survey of moment based techniques with respect to computer vision (Prokop and Reeves, 1992) details many of the current techniques regarding representation and recognition. There have been many other studies using two dimensional moments for image recognition purposes. For example Dudani (Dudani et al., 1977) used moments to recognise aeroplane silhouettes with results that were more successful than the human eye. Local moments were used to recognise hand poses (Takamatsu, 1997), whilst Cartesian moments have been used to produce simple hand control of a toy robot (Beardsley et al., 1998). But all of these are only interested in processing single images. This was built upon (Little and Boyd, 1998) where moment based features were used to characterise optical flow between images for automatic gait recognition. However this approach only linked adjacent images, not the complete sequence.

Here we propose the new method of velocity moments, which are based on moment theory to characterise a sequence of images. We are specifically interested in the movement throughout the sequence of images but we are also eager to include information about the structure of the object that is moving.

### 2 STATISTICAL MOMENT THEORY

Moments when applied to images describe the image content with respect to its axes. They are designed to capture global information about the image. Here we are using them to characterise a grey level image so as to extract properties which have analogies in statistics or mechanics. However only rectangular moment descriptions are considered, as opposed to polar based descriptions. The moment expressions use basis functions which have a range of useful properties that are passed onto the moments. This produces descriptions which are invariant under rotation, scale, translation and/or

orientation. Hu (Hu, 1962), stated that the continuous two dimensional  $(p + q)$ th order Cartesian moment is defined in terms of Riemann integrals as:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q p(x, y) dx dy \quad (1)$$

It is assumed that  $p(x, y)$  is a piecewise continuous, bounded function and that it can have non-zero values only in the finite region of the  $x - y$  plane. If this is so then moments of all orders exist and the following uniqueness theorem holds (Hu, 1962):

**Theorem 1** *Uniqueness theorem* : the moment set  $m_{pq}$  is uniquely defined by  $p(x, y)$  and conversely,  $p(x, y)$  is uniquely defined by  $m_{pq}$ .

This means that the original image can be described and reconstructed, if high order moments are used. The discrete version of the Cartesian moment for an image consisting of pixels  $P_{xy}$ , replacing the integrals with summations is:

$$m_{pq} = \sum_{x=1}^M \sum_{y=1}^N x^p y^q P_{xy} \quad (2)$$

Where  $M$  and  $N$  are the image dimensions and the monomial product  $x^p y^q$  is the basis function.

**2.0.1 Zeroth Order Moment - Area** The zero order moment  $m_{00}$  is defined as the total mass (or power) of the image. If this is applied to a binary  $M \times N$  image (ie a silhouette) of an object, then this is literally a pixel count of the number of pixels comprising the object.

$$m_{00} = \sum_{x=1}^M \sum_{y=1}^N P_{xy} \quad (3)$$

**2.0.2 First Order Moments - Center of Mass** The two first order moments are used to find the Centre Of Mass (COM) of an image. If this is applied to a binary image and the results are then normalised with respect to the total mass ( $m_{00}$ ), then the result is the center co-ordinates of the object. Accordingly the centre co-ordinates  $\bar{x}, \bar{y}$  are given by :

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (4)$$

The COM describes a unique position within the field of view which can then be used to compute the centralised moments of an image.

**2.0.3 Centralised Moments** The definition of a discrete centralised moment as described by Hu(Hu, 1962) is:

$$\mu_{pq} = \sum_{x=1}^M \sum_{y=1}^N (x - \bar{x})^p (y - \bar{y})^q P_{xy} \quad (5)$$

This is essentially a translated Cartesian moment, which means that the centralised moments are invariant under translation. To enable invariance to scale and rotation, normalised moments  $\eta_{pq}$  are used, given by:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad (6)$$

where :

$$\gamma = \frac{p + q}{2} + 1 \quad \forall p + q \geq 2 \quad (7)$$

## 2.1 Moment Reconstruction

Moment reconstruction for moments with orthogonal basis functions (such as Legendre and Zernike moments) has been developed extensively, (Teague, 1979, Teh and Chin, 1988, Prokop and Reeves, 1992, Pawlak, 1992). However in the case where the basis set is non-orthogonal (such as Cartesian and Centralised moments), only one method has appeared. This is the method of moment matching for non-orthogonal moment reconstruction (Teague, 1979). The method is based upon creating a continuous function which has identical moments to that of the original function. Here it has been applied first to Cartesian moments and then Centralised moments. (It must be noted that in applying the theory to sampled images, the continuous conditions are replaced by discrete versions, reducing the accuracy of the final function).

**2.1.1 Cartesian Moment Matching** Assuming that all moments  $M_{jk}$  of a function  $f(x, y)$  and of order  $N = (j + k)$  are known from zero through to order  $N_{max}$ . It is then possible to obtain the continuous function  $g(x, y)$  whose moments match those of the original function  $f(x, y)$ , up to order  $N_{max}$ . Assuming that the given continuous function can be defined as:

$$g(x, y) = g_{00} + g_{10}x + g_{01}y + g_{20}x^2 + g_{11}xy + ..g_{jk}x^jy^k \tag{8}$$

which reduces to:

$$g(x, y) = \sum g_{jk}x^jy^k \tag{9}$$

then the constant coefficients  $g_{jk}$ , are calculated, so that the moments of  $g(x, y)$  match those of  $f(x, y)$ . Assuming that the image is a continuous function bounded by:

$$x \in (-1, +1) , y \in (-1, +1) \tag{10}$$

These limits can be achieved by normalising the value used to calculate the Cartesian moments, which for the image are defined as:

$$\int_{-1}^{+1} \int_{-1}^{+1} g(x, y)x^jy^k dx dy \equiv M_{jk} \tag{11}$$

Substituting Equation 8 into Equation 11 and then solving the integration produces a set of Coupled Linear Equations (CLE), the number of which is determined by the order  $(j + k)$  of reconstruction. These can then be solved for the coefficients  $g_{jk}$ , (in terms of the moments  $M_{jk}$ ) by using matrix inversion. For order three, the CLEs in matrix form are:

$$\begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{3} & \frac{1}{9} & \frac{1}{9} \end{bmatrix} \begin{bmatrix} g_{00} \\ g_{20} \\ g_{02} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} M_{00} \\ M_{20} \\ M_{02} \end{bmatrix} \tag{12}$$

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{5} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{1}{15} \end{bmatrix} \begin{bmatrix} g_{10} \\ g_{30} \\ g_{12} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} M_{10} \\ M_{30} \\ M_{12} \end{bmatrix} \tag{13}$$

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{7} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{1}{15} \end{bmatrix} \begin{bmatrix} g_{01} \\ g_{03} \\ g_{21} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} M_{01} \\ M_{03} \\ M_{21} \end{bmatrix} \tag{14}$$

with finally:

$$g_{11} = \frac{9}{4}M_{11} \tag{15}$$

Applying matrix inversion to the first matrix, Equation 12 produces:

$$\begin{bmatrix} 14 & -15 & -15 \\ -15 & 45 & 0 \\ -15 & 0 & 45 \end{bmatrix} \begin{bmatrix} M_{00} \\ M_{20} \\ M_{02} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} g_{00} \\ g_{20} \\ g_{02} \end{bmatrix} \tag{16}$$

By repeating this, it is possible to calculate all the coefficients. If they are then substituted back into Equation 8 an expression for  $g(x, y)$  is produced. This expression is then used to reconstruct an approximation of the original image.

The reconstruction function  $g(x, y)$  is now in terms of weighted sums of the moments  $M_{jk}$ , which have been previously calculated from the original image ( $f(x, y)$ ). The resultant function  $g(x, y)$  for order three is:

$$\begin{aligned}
 16g(x, y) &= (14M_{00} - 15M_{20} - 15M_{02}) \\
 &+ (90M_{10} - 105M_{30} - 45M_{12})x \\
 &+ (90M_{01} - 105M_{03} - 45M_{21})y \\
 &+ (-15M_{00} + 45M_{20})x^2 + 36M_{11}xy \\
 &+ (-15M_{00} + 45M_{02})y^2 + (-105M_{10} + 175M_{30})x^3 \\
 &+ (-45M_{01} + 135M_{21})x^2y + (-45M_{10} + 135M_{12})xy^2 \\
 &+ (-105M_{01} + 175M_{03})y^3
 \end{aligned} \tag{17}$$

Implementing this method to order  $(j + k) = 8$  for binary images of simple shapes produces recognisable results, as shown in Figures 1 and 2. Figure 1(a) is the original image from which the moments were calculated and Figure 1(b) is the image reconstructed from the moments. The borders of the shape appear unclear, but they appear when the reconstructed image is thresholded, Figure 1(c). This analysis is then repeated for the square in Figure 2. The thresholded shape will converge to the original shape as the number of moments increases. However for more complex shapes, higher accuracy  $(j + k) \gg 8$  is needed. (This is analogous to the high frequency information needed to reconstruct stepped time domain waveforms, using methods like Fourier series.) As the order (and accuracy) increases, so does the number of CLEs that need to be solved. (Reconstruction for order of eight resulted in forty five CLE's.) The reconstruction function produced can be thresholded for easy comparison with the original image, as shown in Figures 1(c) and 2(c). The level of the applied threshold was adjusted by visual comparison with the original image.

**2.1.2 Centralised Moment Matching** By assuming the same constraints as for Cartesian moment matching, the theory can be extended to Centralised moments. The continuous function  $g(x, y)$  is now defined as:

$$g(x, y) = \sum g_{jk}(x - \bar{x})^j (y - \bar{y})^k \tag{18}$$

similarly, Equation 11 becomes:

$$\int_{-1}^{+1} \int_{-1}^{+1} g(x, y)(x - \bar{x})^j (y - \bar{y})^k dx dy \equiv M_{jk} \tag{19}$$

where  $\bar{x}$  and  $\bar{y}$  are the  $x$  and  $y$  COM's, respectively. Solving for  $g(x, y)$  is then achieved in the same manner as already described for the Cartesian case.

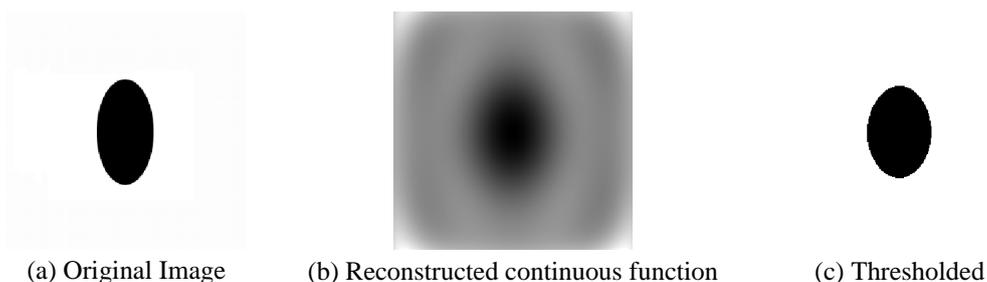


Figure 1: Order 8 Cartesian reconstruction of an ellipse.

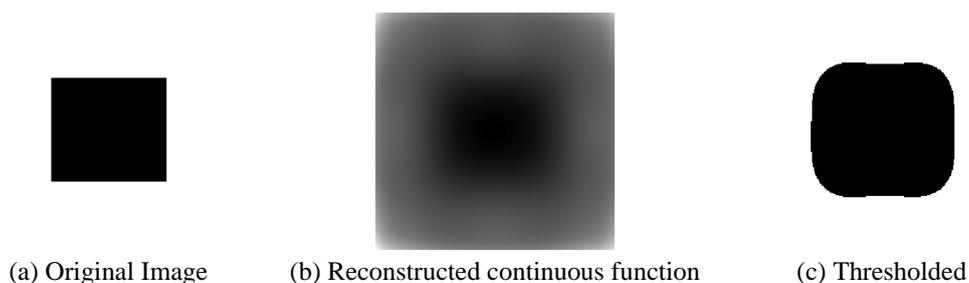


Figure 2: Order 8 Cartesian reconstruction of a square.

### 3 VELOCITY MOMENTS

#### 3.1 Velocity Moment Description

The new velocity moments (VMs) are computed from a sequence of images as:

$$vm_{pq\mu\gamma} = \sum_{i=2}^{Images} \sum_{x=1}^M \sum_{y=1}^N V \cdot C \cdot P_{i_{xy}} \tag{20}$$

$C$  arises from the centralised moments:

$$C = (x - \bar{x}_i)^p (y - \bar{y}_i)^q \tag{21}$$

$V$  introduces velocity as:

$$V = (\bar{x}_i - \bar{x}_{i-1})^\mu (\bar{y}_i - \bar{y}_{i-1})^\gamma \tag{22}$$

where  $\bar{x}_i$  is the  $i$ 'th COM for the  $x$  direction and  $\bar{x}_{i-1}$  is the COM for the  $(i - 1)$ 'th image. Respectively  $\bar{y}_i$  is the  $i$ 'th COM for the  $y$  direction and  $\bar{y}_{i-1}$  is the COM for the previous, or  $(i - 1)$ 'th image. The equation is not strictly a moment in its purest form, as it would initially appear unfeasible to reconstruct an original shape from this description. The name has been used in the context of the overall compressed description it produces. It can be seen that the equation can easily be decomposed into averaged central moments ( $vm_{1100}$ ), and then further into an averaged Cartesian moment ( $vm_{1100}$  with  $\bar{x}_i = \bar{y}_i = 0$ ). The zero order VMs for which  $\mu = 0$  and  $\gamma = 0$  are then:

$$vm_{pq00} = \sum_{i=2}^{Images} \sum_{x=1}^M \sum_{y=1}^N (x - \bar{x}_i)^p (y - \bar{y}_i)^q P_{i_{xy}} \tag{23}$$

which are the averaged centralised moments. The zero order components for which  $p = 0$  and  $q = 0$  are:

$$vm_{00\mu\gamma} = \sum_{i=2}^{Images} \sum_{x=1}^M \sum_{y=1}^N (\bar{x}_i - \bar{x}_{i-1})^\mu (\bar{y}_i - \bar{y}_{i-1})^\gamma P_{i_{xy}} \tag{24}$$

which is a summation of the difference between COM's of successive images (velocity). These results are averaged by normalising with respect to the number of images and the average area of the object. This results in pixel values for the velocity terms, where the velocity is expressed in pixels per image. The normalisation is expressed as:

$$\overline{vm_{pq\mu\gamma}} = \frac{vm_{pq\mu\gamma}}{A \cdot I} \tag{25}$$

where  $A$  is the average area (in numbers of pixels,  $p$ ) of the moving object,  $I$  is the number of images and  $\overline{vm_{pq\mu\gamma}}$  is the normalised VM. The structure of Equation 20 allows the image structure to be described together with velocity information from both the  $x$  and  $y$  directions. In order to assess the performance of the VMs on extracted images, tests were run on synthetic images. Applying the moments to three different sequences the recognition capabilities were examined. The sequences were a moving square, circle and triangle moving at 5, 5 and 7 pixels/image respectively in the  $x$  direction while the circle had a small deviation of 0.1 pixels/image in the  $y$  direction. These produced significantly different second order moments, Table 1, and the  $x$  velocity term  $vm_{0010}$  was estimated correctly in each case.

| Shape  | Index       | Value/p | Shape    | Index       | Value/p | Shape  | Index       | Value/p |
|--------|-------------|---------|----------|-------------|---------|--------|-------------|---------|
| square | $vm_{2201}$ | 0.00e00 | triangle | $vm_{2201}$ | 0.00e00 | circle | $vm_{2201}$ | 1.32e04 |
|        | $vm_{2010}$ | 0.16e04 |          | $vm_{2010}$ | 6.65e02 |        | $vm_{2010}$ | 0.26e04 |
|        | $vm_{2210}$ | 0.55e06 |          | $vm_{2210}$ | 2.66e04 |        | $vm_{2210}$ | 0.74e06 |
|        | $vm_{0010}$ | 5.00e00 |          | $vm_{0010}$ | 5.00e00 |        | $vm_{0010}$ | 7.00e00 |

Table 1: Recognition results

The tests were then repeated where the perimeter of the shape (Figure 3(a) ) was degraded by noise, simulating poor extraction. A basis for comparison is needed to analyse the effect of perimeter noise on the VMs. For this reason a set of averaged invariant Hu moments (Hu, 1962) were used. The variance of the added perimeter noise took values from 0.0 (no noise) to 4.0, in 0.1 steps. The performance of the two methods was then plotted and compared, example results are shown in Figure 3, (where here the invariant moment values have been scaled for visual reasons only). Clearly the new VM (here  $vm_{2010}$ ) is much less affected by noise, changing at most by 10%, whereas the traditional moments (here

$M3$ ) even when averaged can deviate by a great amount, Figure 3(b) shows over 100% variation. (A similar result can be found through theoretical analysis of the effect of the perimeter noise.) The effects of the perimeter noise can be seen in Figure 3(c), where the increasing VM value reflects the spread of the image in the  $x$  axis, with respect to the  $x$  direction velocity. As such a statistical technique has been developed which describes an object and its motion, and which appears less sensitive to noise than extant approaches. Further to this the VMs have been applied to automatic gait recognition (Shutler et al., 2000). Here they have been tested on a database of 4 subjects, with 4 sequences of each subject, a total of 16 sequences. Each subject is first extracted using a simple spatial template method, once extracted the subject's movement (or optical flow) is analysed producing temporal templates. These temporal templates are then compressed by the VMs. Then by using 3 different VMs and a  $k$  nearest neighbour classifier, recognition rates of over 90% have been achieved.

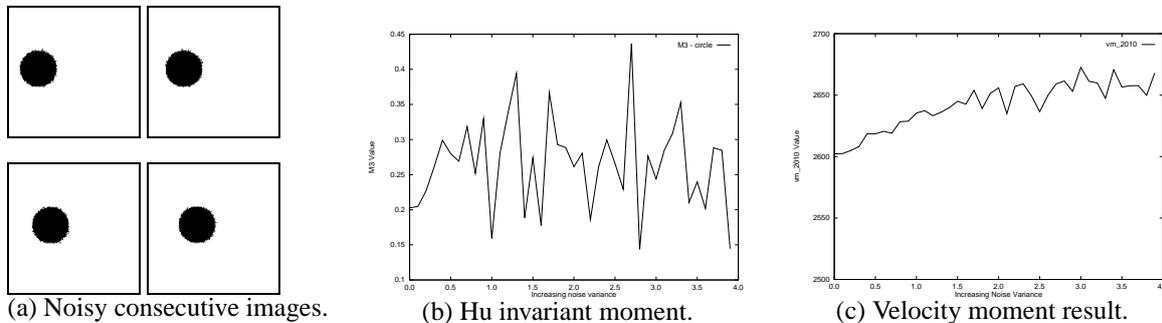


Figure 3: Test images and results.

### 3.2 Velocity Moment Matching

The velocity moments are non-orthogonal, which means that the application of the moment matching method (Section 2.1.1) is applicable. If we consider the velocity moment case for a single image, then this is actually a Centralised moment for a single image, so the method described in Section 2.1.2 holds. To consider the case of the image index  $I$ , for  $I > 1$ , the problem can be simplified by assuming that the resulting velocity moments produced describe a single image, even though they are in fact derived from a sequence of images. To consider attempting to reconstruct the complete image set, the COM's for each separate image would be needed, along with the velocity moment value before it is summed and averaged over the complete sequence. So to reproduce a single image,  $g(x, y)$  can be defined as:

$$g(x, y) = \sum g_{jklm} (x - \bar{x})^j (y - \bar{y})^k a^l b^m \quad (26)$$

where:

$$a = \text{average } x \text{ velocity} = vm_{0010}, \quad b = \text{average } y \text{ velocity} = vm_{0001} \quad (27)$$

Here  $\bar{x}$  and  $\bar{y}$  are assumed to be the averaged COM's in the  $x$  and  $y$  directions. Similarly Equation 11 becomes:

$$\int_{-1}^{+1} \int_{-1}^{+1} g(x, y) (x - \bar{x})^j (y - \bar{y})^k a^l b^m dx dy \equiv vm_{jklm} \quad (28)$$

Applying this in practice means that an increasing number of large CLEs need to be solved, even if only low order ( $(j + k) < 4$ ) reconstruction is required. However an order of at least  $(j + k) \gg 8$  is needed to produce a meaningful reconstruction (as found in section 2.1.1). As such, the computation involved appears excessive though it certainly would appear that the new velocity moments do allow for reconstruction, should it be desired. However the fact that reconstruction is possible certainly suggests that the velocity moments do indeed possess the requisite uniqueness properties, as evidenced by the results.

## 4 CONCLUSIONS AND FURTHER WORK

New velocity moments have been developed which give basic recognition properties for moving shapes, producing distinct results for different synthetic temporal test sets. The results have reflected the structure of the velocity moment equation. They have indicated that the expression holds information about the structure of the moving object as well as temporal information, hence producing unique results for different objects moving at the same constant velocity. The method has been shown to be more immune to perimeter noise (in comparison to the averaged normalised invariant moments) and non-orthogonal velocity moment reconstruction appears feasible. These results suggest that the method would prove useful when applied to poorly extracted sequences, or possibly those where incomplete perimeter contours are apparent.

The Hu invariant moments have been seen to be inherently sensitive to perimeter noise when considering simple shapes, such as triangles and squares. This is quite often due to the original value of the moment being zero for the higher orders, such as  $M_5$ ,  $M_6$  and  $M_7$ . Once noise is added the values for these moments begin to oscillate considerably. These results parallel those found in a previous study (Teh and Chin, 1988). If the numerical value of a moment (velocity or invariant) is zero with no perimeter noise, then this term is inherently more sensitive to the perimeter noise. The magnitude of this problem is greater in the invariant set of moments. The effects of the noise are amplified by the presence of the squared terms in the invariant moment equations. From this it can be seen that the higher order invariant moments are inherently noisy.

This promising new statistical technique is currently being applied to automatic gait recognition. Here we are characterising temporal and spatial sequences of subjects which are then compressed using the velocity moments. The work to date (Shutler et al., 2000) has produced encouraging results with recognition rates of over 90%.

## REFERENCES

- Beardsley, P. A., Freeman, W. T., Anderson, D. B., Dodge, C. N., Roth, M., Weissman, C. D. and Yerazunis, W. S., 1998. Computer vision for interactive computer graphics. *IEEE Computer Graphics I/O Devices* **May-June 98**, pp. 42–53.
- Dudani, S. A., Breeding, K. J. and McGhee, R. B., 1977. Aircraft identification by moment invariants. *IEEE Transactions on Computers* **C-26**(1), pp. 39–46.
- Hu, M., 1962. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* **IT-8**, pp. 179–187.
- Little, J. J. and Boyd, J. E., 1998. Recognising people by their gait: The shape of motion. *Videre :Journal of computer vision research* **1**(2), pp. 2–32.
- Pawlak, M., 1992. On the reconstruction aspects of moment descriptors. *IEEE Transactions on Information Theory* **38**(6), pp. 1698–1708.
- Prokop, R. J. and Reeves, A. P., 1992. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP Graphical models and Image Processing* **54**(5), pp. 438–460.
- Shutler, J. D., Nixon, M. S. and Harris, C. J., 2000. Statistical gait recognition via velocity moments. *IEE Colloquium - Visual Biometrics* pp. 11/1 – 11/5.
- Takamatsu, R., 1997. A pointing device gazing at hand based on local moments. *Proc. SPIE* **3028**, pp. 155–163.
- Teague, M. R., 1979. Image analysis via the general theory of moments. *Journal of the Optical Society of America* **70**(8), pp. 920–930.
- Teh, C. and Chin, R. T., 1988. On image analysis by the method of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**(4), pp. 496–513.