Parameterised Moving Shape Extraction

Michael G. Grant, Mark S. Nixon and Paul H. Lewis

Department of Electronics and Computer Science,
University of Southampton SO17 1BJ UK
[mgg97r/msn/phl]@ecs.soton.ac.uk

Abstract

Extracting moving objects from image sequences naturally depends on the nature of their motion. The anticipated trajectory of movement through the sequence can be described as a template. These motion templates have been combined with a recent approach that can extract arbitrary shapes moving within image sequences. The new combined approach is phrased around evidence gathering to accrue proven performance advantages in respect of noise and occlusion. Further, unlike tracking, the approach needs no initialisation. We show, by result, how the new technique can extract walking human subjects and analyse its performance attributes. In view of its capability to handle noise and occlusion the new technique should provide an appropriate primer for, say, statistical approaches to gait recognition.

1. Introduction

The robust and accurate extraction of arbitrarily moving arbitrary shapes is a difficult problem for machine vision. There is a substantial literature on the problem of tracking shapes in a sequence, for example the tracking of humans has recently been admirably surveyed [1]. However, it is far less usual to extract shapes from an image sequence, which is distinguished from tracking by a more holistic view of the sequence, and frequently *a priori* knowledge and an off-line approach. Here, we propose a method that maintains optimal extraction of targets moving in a specified arbitrary fashion even in high levels of noise or occlusion. In addition, there is no need for initialisation or training. These qualities are derived from the use of robust evidence gathering techniques and the exploitation of temporal correlation in the sequence.

In terms of Hough transform (HT) based developments for moving object extraction, the use of pre-defined motion is novel. Previously, evidence gathering algorithms have always had limited motion models that handle particular forms of parametric motion (e.g. linear velocity only, sinusoidal motion, etc). The reason for this limit is that as the motion model becomes more complex, more terms are required in the parametric description and thus more dimensions are added to the parameter space. This rapidly and catastrophically enlarges the search space to an unfeasible extent. There are parallels to this parameterisation of motion in the earlier parameterisation of shape, where increasingly complex shapes were represented by more complex parameterisations and a commensurately larger dimensionality. The solution to this dimensional explosion was found in the use of templates, which allowed an efficient and low dimensional parameterisation of any shape.

Similarly, we show that templates have a place in the analysis of motion using the HT. These motion templates extend the use of templates from the spatial domain into the temporal. This solves the problems associated with increasing dimensionality since the aim changes from finding the potentially unlimited set of parameters that characterise a particular motion to finding the limited set of parameters that locate the object undergoing the specified motion. This enables the extraction of arbitrarily moving shapes without a punitive computational burden.

2. Fourier descriptor template representation

In the Generalised HT (GHT) [2], the template shape is represented by an R-table – a discrete lookup table. The problems with this representation are well known [3], but essentially derive from the fact that it is a discrete representation sampled at a particular scale. When the template is scaled or rotated, there can be problems with aliasing and rounding errors. Figure 1(b) shows the effects of scaling and rotating the discrete set of points comprising the shape in Figure 1(a). Clearly, the new set of points has missing data and the effects of discretisation are evident. If the shape had been reduced, points in the original would merge, effectively oversampling the shape.

(a) original shape   (b) rotated and scaled versions of original shape
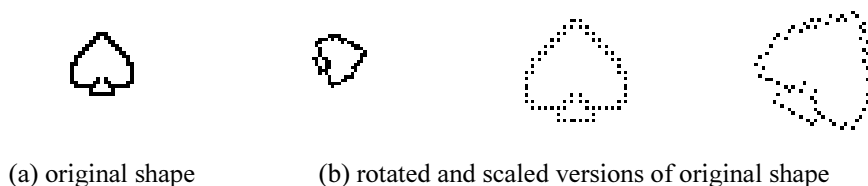
Figure 1 – Consequences of Discrete Shape Representation

Distortions are inevitable when working with discrete systems. Nevertheless, the worst effects can be avoided by using a continuous shape representation for as much of the process as possible. Elliptic Fourier descriptors [4] have been deployed in an adaptation of the GHT [5] giving a continuous representation that can be sampled at any resolution without aliasing problems. Elliptic Fourier descriptors were chosen for their completeness, simple geometric interpretation, access to frequency information and the fact that they can be easily produced from a chain code of the contour.

3. Temporal evidence gathering

It is well known that most image sequences contain significant correlation between frames - a fact commonly utilised by machine vision and video compression algorithms amongst many others. This correlation was first exploited in the evidence-gathering field by the Velocity Hough Transform (VHT) [6], which extracts the optimum parameters describing a conic section moving with linear velocity. In essence, the VHT accumulates evidence from the whole sequence, concurrently extracting optimal structural and motion parameters. Using the temporal correlation in a sequence makes the VHT more robust than a standard frame-by-frame tracking implementation, especially when the target is occluded or noisy. Any missing or damaged structural information in a frame can be compensated for by the redundant information in others. As an evidence-gathering technique, the VHT also requires no initialisation and avoids

the correspondence problem due to the global nature of the search. The motion model is parametric and can be extended from linear velocity by including extra terms to describe more complex motion. In this respect, an extension to the VHT to find walking subjects using an articulation model [7] required 13 parameters, when moving with constant velocity. However, any parametric extension to the motion model will increase the dimensionality of the accumulator and, thus, the computational resources required.

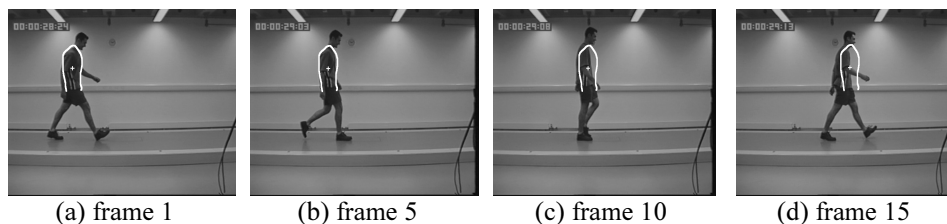| (a) frame 1 | (b) frame 5 | (c) frame 10 | (d) frame 15 |

Figure 2 – Problems with Extracting an Arbitrary Shape with Constant Velocity

In an earlier work [8], we extended the shape model in the VHT to handle arbitrary shapes. This was achieved by incorporating the Fourier descriptor based variant of the GHT (described above) into the VHT. Amalgamating these techniques unifies their unique and complementary advantages. The Fourier descriptors provide a continuous template representation, minimising discretisation error in the algorithm, and the VHT exploits the temporal correlation across a sequence, whilst also mitigating the effects of noise and occlusion. The new algorithm does not require initialisation or training and avoids the need to solve the correspondence problem, inheriting these characteristics from the VHT. However, this technique was still limited by the parameterised motion model. Figure 2 shows the extraction of a walking subject using a purely linear velocity motion model. As can be seen, the template overtakes the subject due to the variation in the gait motion which violates the (constant) assumptions of the motion model (as the walker rises up on his leg around frame 10, he also slows down).

4. Motion Templates

The earlier approaches to moving shape extraction are limited by the dimensionality of the parameterised motion model. As stated, our objective is to remove this limitation and enable the extraction of arbitrary shapes undergoing arbitrary motion by using motion templates to describe the movement of the target. With these templates it is no longer necessary to accumulate for the parameters describing the motion since they are already known. Naturally, the motion templates require *a priori* knowledge concerning the target object's path before analysis. Since tracking, by definition, precludes the possession of this information it is important to observe that the niche for motion templates is extraction or recognition. Motion templates make it possible to robustly and efficiently extract parameters describing a shape that is following a specified trajectory. However, a limited predictive capability is possible if the motion template covers a larger time than analysed (or for repetitive motion) since the subject will be assumed to continue the motion.

As a result of the requirement for detailed prior knowledge, the new algorithm will be of use in cases where the general path of motion is known (e.g. cars turning at traffic lights will follow roughly the same path). However, there may be difficulties in real

world imagery since not all objects will follow exactly the same path. The forgiving nature of an evidence-gathering approach should abate this concern provided the deviations are not excessive. Motion templates must encode the relative position of the target object at all times. This will automatically describe properties of the motion such as speed, and acceleration. Motion templates should also record changes in scale and rotation over time since, in motion, many objects rotate (e.g. a car, viewed from above, rotates as it corners) and scale (e.g. due to perspective effects). This additional detail only increases the complexity of the motion template rather than that of the algorithm.

The representation of the motion templates should be continuous so as to avoid the problems of discrete representations (section 2). As such, it is again prudent to use Fourier descriptors with the earlier advantages. Furthermore, we have a consistent framework for the representation of arbitrary shape and motion.

Using motion templates in the HT requires no additional parameters to be searched for in the accumulation phase. However, such a strict implementation would reduce the functionality excessively. Essential parameters are rotation of the motion template in its spatial dimensions and scaling in both spatial and temporal dimensions. The scaling in the spatial axes is uniform since we will initially handle affine transformations only. Scaling in the temporal axis adjusts the time taken to traverse the motion template and thus the speed with which an object must move to be identified. Finally, we must add an offset (phase) to separate the time encoded in a motion template from that used in a sequence. Otherwise, the frame time zero would be tied directly to time zero in the motion template preventing correct extraction for a sequence that began with the target object part-way through its trajectory. Hence, the time-scale parameter provides temporal scale invariance and the offset parameter gives temporal translation invariance. Rotation invariance is not required since time is one-dimensional.

Accordingly, using motion templates adds just four extra parameters, giving an eight–dimensional accumulator. Whilst this will increase computational requirements, it is much less than the dimensional explosion offered by polynomial extension. Clearly, the ability to extract robustly and optimally arbitrary shapes with an arbitrary motion is well worth the additional computational requirements.

4.1. Theory

The theory supporting motion templates is developed from the arbitrary-shape VHT [8], which itself was an extension to the Fourier descriptor variant [5] of the GHT. As an aid to visualisation, for a given feature point (e.g. an edge pixel in a frame), a locus of points is plotted through the eight-dimensional accumulator space. This locus is constructed from the shape and motion descriptors, which are used to trace scaled and rotated instances of the template shape, translated to compensate for the expected motion of the object relative to the time reference of each frame, in two-dimensional planes taken along the x- and y-axes of the accumulator. Once the voting process is complete, peaks in the accumulator indicate the location and other parameters (at time $t = 0$) of an instance of the target shape moving along the specified path. In order to develop the voting mechanism, we require an arbitrary-curve parameterisation for shapes. A curve defined by two sets of orthogonal co-ordinates, $c_x(s)$ and $c_y(s)$, parameterised by $s \in [0, 2\pi)$ has elliptic Fourier descriptors as follows:

$$a_{xk} = \frac{1}{\pi} \int_{-\pi}^{\pi} c_x(s) \cos(ks) \mathrm{d}s \quad \text{and} \quad b_{xk} = \frac{1}{\pi} \int_{-\pi}^{\pi} c_x(s) \sin(ks) \mathrm{d}s \qquad (1)$$

with similar equations for the *y* descriptors, where *k* is the harmonic number. *n* FDs can be converted to vectors (along *x*- and *y*-axes) from the origin to a point on the curve by:

$$\upsilon_x(s,\overline{FD}_x) = \sum_{k=1}^{n} \left(a_{xk}\cos(ks) + b_{xk}\sin(ks) \right) \qquad (2)$$

where $\overline{FD}_x = \{a_{x1}, b_{x1}, a_{x2}, b_{x2}, \ldots, a_{xn}, b_{xn}\}$,with matching equations in *y*. The DC terms have been omitted (by not summing for *k=0*), which translates the curve so that its centre is at the origin. The shape's initial scale and rotation is given by $\mathbf{a}_s = [l_g \ \rho_g]$ as

$$R_x(s,\mathbf{a}_s) = l_g \upsilon_x(s,\overline{FD}_x)\cos(\rho_g) - l_g \upsilon_y(s,\overline{FD}_y)\sin(\rho_g) \qquad (3)$$

with a similar equation for R_y. We must parameterise the motion template, encoding the changing *x*, *y*, rotation and scale parameters. The major difference in the definition of the motion template from that of the shape parameterisation is that the DC terms are retained and used in the reconstruction. Removing them would effectively translate reconstructed co-ordinates relative to some arbitrary origin (e.g. the centre), which would destroy the utility of the sequence by moving the start point.

Let $m_x(f)$, $m_y(f)$, $m_p(f)$ and $m_l(f)$ be functions that take a frame time *f* and recover a co-ordinate (for *x* position, *y* position, rotation or scale respectively) from a motion template. This motion template is now used to transform the co-ordinates calculated from the shape descriptors. These co-ordinates have already been globally scaled and rotated (i.e. we are using R_x and R_y from equation 3) to adjust for possible initial scales and orientations. These are then scaled and rotated as:

$$\xi_x(s,f,o,\mathbf{a}_s) = m_l(f-o)R_x(s,\mathbf{a}_s)\cos(m_\rho(f-o)) - m_l(f-o)R_y(s,\mathbf{a}_s)\sin(m_\rho(f-o)) \quad (4)$$

with a similar equation for *y* co-ordinates where *f* is the frame of the current image and *o* is a parameter that offsets any mismatch between the frames in the sequence and in the motion template. $m_l(f-o)$ and $m_p(f-o)$ respectively recover scale and rotation values for time *f - o* from the motion template. These rotated and scaled co-ordinates now represent the shape at the expected orientation and scale. Next, they are translated to compensate for the object's expected motion. However, the path of expected motion is also scaled and rotated (requiring parameters, $\mathbf{a}_m = [l_m \quad \rho_m]$):

$$\mu_x(s,f,o,\mathbf{a}_m,\mathbf{a}_s) = \xi_x(s,f,o,\mathbf{a}_s) + l_m m_x(f-o)\cos(\rho_m) - l_m m_y(f-o)\sin(\rho_m) \quad (5)$$

again with a similar equation in *y* and where $m_x(f\text{-}o)$ and $m_y(f\text{-}o)$ are the *x* and *y* offsets recovered from the motion template at time *f - o*. Now we can form the kernel that defines the shape of votes in the accumulator. This is a combination of curves at a number of translations, orientations and scales, and can be obtained from:

$$\overline{\omega}(s,f,o,l_f,\mathbf{a}_m,\mathbf{a}_s) = \mu_x\left(s,\left(f\cdot l_f\right),o,\mathbf{a}_m,\mathbf{a}_s\right)U_x + \mu_y\left(s,\left(f\cdot l_f\right),o,\mathbf{a}_m,\mathbf{a}_s\right)U_y \quad (6)$$

where U_x and U_y are two orthogonal vectors defining the *x* and *y* axes, respectively. One final parameter, l_f, is added to perform temporal scaling on the motion template, thereby allowing adjustment of the speed at which the path is traversed. This curve is inserted into the accumulator by offsetting it from the co-ordinates of each feature point in the image sequence *IS*, defined by:

$$IS = \left\{ \overline{\lambda}(t,f) \mid t \in D_t, f \in D_f \right\} \qquad (7)$$

Here, $\overline{\lambda}(t,f)$ is a parametric function that defines the points in the image sequence for a frame time *f*, where D_t is the domain of an image in the sequence and D_f is the domain of the frames of the sequence. To ensure that the locus drawn in the accumulator passes through the reference point, the kernel is offset from the image co-ordinates of each

feature point. Hence, the following expression defines how votes are placed in the accumulator for an image sequence *IS*:

$$A_t = \{\overline{\lambda}(t,f) - \overline{\omega}(s,f,o,l_f,\mathbf{a}_m,\mathbf{a}_s) \big| s \in D_s\} \forall t \in D_t, f \in D_f \qquad (8)$$

The equations above describe the concept of the HT but do not formalise the actual technique used, namely the accumulation phase. Parameter space can be mapped into an accumulator by using a matching function *M*, which determines whether a point in parameter space should be incremented. The simplest accumulation strategy is to increment by unity for each match. Changing *M* can accommodate more complex strategies. Applying the matching function to A_t over a range of parameter values gives rise to a mapping into an accumulator space. The parameter space S_{DF} is given by:

$$S_{DF}(\overline{b},l_f,o,\mathbf{a}_m,\mathbf{a}_s) = \sum_{s \in D_s} \sum_{t \in D_t} \sum_{f \in D_f} M(\overline{b},\overline{\lambda}(t,f) - \overline{\omega}(s,f,o,l_f,\mathbf{a}_m,\mathbf{a}_s)) \qquad (9)$$

where $\overline{b}$ is a vector of the image co-ordinates at time 0, l_f is the time scale-factor, o is the time offset (phase), l_m and ρ_m are scale and orientation that transform the motion template's path and l_g and ρ_g are scale and orientation parameters for the initial position of the target shape. $\overline{\lambda}(t,f)$ retrieves feature points from the sequence.

This expression gives an accumulation strategy for finding arbitrary shapes moving arbitrarily. It allows extraction of the optimal parameters describing an arbitrary (but specified) shape of unknown orientation, position and scale that is following an arbitrary (but specified) path of unknown orientation and scale, which takes an unknown time to traverse.

The basic algorithm requires an eight-dimensional accumulator, which imposes significant computational cost even for low-resolution sequences and small parameter ranges. However, most of the speed-up and memory-reduction modifications to HT-related techniques are applicable – e.g. parameter space decomposition, multi-stage processing, or genetic algorithms. Another important feature of the HT is that it is well suited to parallel implementation.

5. Results

It is usually prudent to compare new technique with contemporary or equivalent approaches. This appears unfeasible in this case as the most appropriate comparator techniques are the GHT, but with interpolation, or tracking. The suitable form of this interpolation for the GHT or motion model for tracking is actually the motion template, the very subject of these results. The alternative would be a fully representative parametric motion model that, as explained previously, is computationally intractable due to its infinite dimensionality. Consequently, the only practical comparison that can be made is that of a frame-by-frame extraction process with a non-analytic template representation against an integrated multi-frame extraction process with an analytic template representation. This comparison has already been made [8] in the earlier approach to moving arbitrary-shape extraction (without motion templates but with a linear motion model) and fails to test the subject of this paper - the motion template in an evidence gathering context. The study found that the frame-by-frame GHT had inferior performance in noise and occlusion since errors in a single frame would cause global estimation errors that were averaged out by the VHT's more integrated approach.

Comparison with other techniques that use similar knowledge of motion (such as a neural network based human motion tracker [9] or a spatio-temporal repetitive motion detector using temporal textures [10]) is not comparing like with like. In the case of techniques that are as dissimilar as a tracker and an extractor, the comparison is best made on a qualitative requirements or features basis (e.g. optimality vs. on-line performance) rather than a quantitative performance analysis. In light of these difficulties, we have examined the performance of the new technique in terms of noise affecting each component of the system rather than attempt to make direct comparisons with other distantly related approaches.
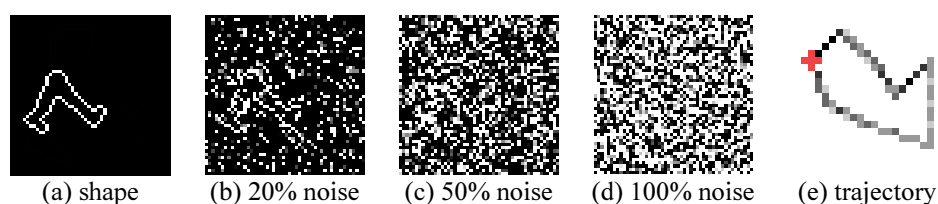
| (a) shape | (b) 20% noise | (c) 50% noise | (d) 100% noise | (e) trajectory |

Figure 3 –shape with 20, 50 and 100% added Gaussian noise, and motion trajectory

5.1. Performance analysis on synthetic images

The new algorithm was run on a nine-frame sequence based on a small (50×50) image, Figure 3(a), moving along the path shown in Figure 3(e). The motion template used was given perfect co-ordinates since we are examining the response to image noise in the input sequence, not noise in the motion template. A small image was chosen to make practical computation of large-scale tests. Noise was added at random to each frame of the sequence at eleven noise levels from 0% random coverage to 100% random coverage of the frame, with pixel values wrapping rather than being clipped when the addition of noise took the values out of range. The noise distribution was zero-mean Gaussian with a standard deviation of one. Examples of the effects of the increasing noise levels can be seen in Figure 4. The grey-level images produced are thresholded by the algorithm (shown before thresholding in Figure 3 (c)-(d)). Note that, at the maximum noise level, the shape is completely obliterated and that, at around 50% noise, the shape is barely visible to the eye.
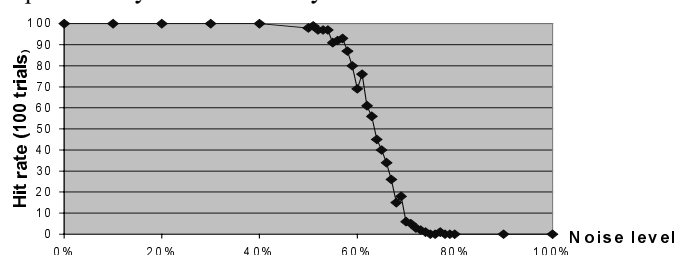
Figure 4 - Noise performance for synthetic imagery

The graph above shows the new technique can handle significant levels of noise. The performance curve is similar to those for previous VHT-derived techniques and, from those studies, by accumulating the additional temporally correlated evidence the VHT-based techniques are able to handle noise levels that are approximately 20 percent greater than a comparable GHT-based frame-by-frame technique.

The effect of noise corruption of the motion templates was also analysed, though space limits prevent inclusion here. Essentially, if the number of points contributing false evidence exceeded the number giving true evidence, then a false peak was extracted, as expected. The noise was arranged to affect motion templates co-ordinates at random, and with different variance. Peak spreading was observed in situations of high noise where false peaks were extracted.

5.2.　Performance analysis on real images

Our current intention is to apply the new technique to the analysis of human motion and, eventually, to produce biometrics from gait. For these purposes, and to examine real-world imagery, we have evaluated the algorithm's performance when locating a walking person viewed from the side. Since we currently require a rigid shape model as with the arbitrary shape VHT, we have chosen to use the head and torso, which is approximately a constant shape throughout the sequence. However, distortions in the shape of the body contour occur due to the motions of walking and, especially, due to clothes. Also, self-occlusion of the body due to the motion of the arms and legs is a problem that affects the performance of many person-tracking algorithms.

By the nature of evidence gathering, the new algorithm inherently copes with occlusions and inaccuracies that do not reduce the number of correct votes (from the remainder of the true contour) below the level of false peaks in the accumulator. Again, owing to the robustness of its formulation there is no need for exceptional precautions in the new technique. Consequently, it is possible to detect people using the technique in its current form by searching for the torso. However, no meaningful gait data can be gathered from just the location of the torso so this method of locating a human is only useful as a primer for another technique. Nonetheless, using the new technique to locate a human demonstrates that it is equally applicable to real world images.

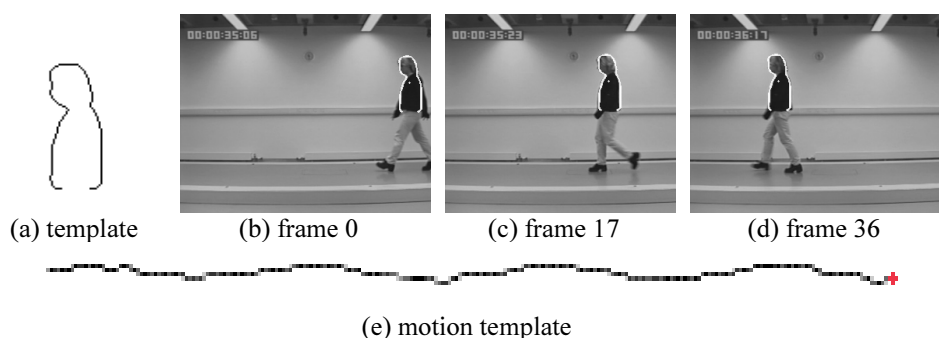| (a) template | (b) frame 0 | (c) frame 17 | (d) frame 36 |

(e) motion template

Figure 5 - Templates and extraction for frames 0, 17 and 36 of sequence CA1

Figure 5a shows a reconstructed template of a walker's torso and head, which was originally created by manually tracing from one of the frames in the sequence. Also shown is a plot of the x and y components of the motion template used (Figure 5e) and several frames of the walker sequence CA1 with the extracted template superimposed (Figure 5b-d). As noise is added the performance holds approximately to that of the synthetic case. The difference can be attributed to the imperfect conditions prevalent in the real world, i.e. cluttered and noisy background and imperfect shape and motion templates. The results of the Gaussian noise tests are shown in Figure 6. This curve is

similar in shape to that shown earlier for synthetic shapes, starting to lose accuracy at 40% noise, and missing the target completely above 80% noise.

A simple test of the effects of occlusion was carried out on the walker sequence CA1 described above. No noise was added to the sequence since this would be an unnecessary complicating factor. Instead, a number of vertical lines of pixels were blanked out and the algorithm was run on the resulting image. The results revealed that the new technique keeps correctly extracts the walker until the blanking is 175 pixels wide – which completely obscures the walker for approximately 70% of the duration of the entire sequence. As in evidence gathering, the new technique handles even extreme occlusion owing to the global integration of evidence across the entire sequence.
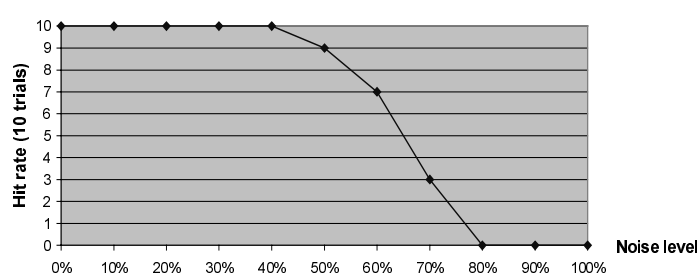

Figure 6 - Noise performance for sequence CA1

Figure 7 - MA1, SG1 and VH1 (top row) and MA3, SG3 and VH3 (bottom row) frame zero with template superimposed

5.3.   **Finding people**

Further tests were performed to check the repeatability of human walker extraction. Three extra image sequences of walkers (MA1, SG1 and VH1) were extracted correctly as shown in Figure 7. Also shown are the results of attempting to extract the same walker from a different sequence (MA3, SG3 and VH3 respectively) whilst using the same shape and motion templates. As can be seen, MA3 and SG3 are correctly extracted but VH3 is slightly in error. This is due to some problems towards the end of

the sequence with bad edges (and hence a bad motion template) resulting from height of the subject and the interaction with the lighting in the room.

Since the results are generally very accurate, this offers hope for a future attempt at locating humans with a "standard walk" motion template and a suitable range of scaling parameters. This has not yet been attempted due to the computational requirements but should be achievable with a genetic algorithm implementation.

6. Conclusions and Further Work

We have presented a new technique that robustly extracts optimal structural and motion parameters for arbitrarily moving arbitrary shapes in an image sequence. The technique requires no initialisation or training and has demonstrated excellent tolerance to noise and occlusion. Discretisation errors are minimised in the accumulator by using Fourier descriptors to represent the templates of shape and motion in continuous form, eliminating common problems with rotation and scaling. Also, temporal correlation between frames is exploited to maximise chances of correct extraction. Finally, motion templates are a novel development of the HT and allow for a wide range of applications that require a more general motion model. This new capability comes without the explosion of parameter space dimensionality inherent in current parametric approaches.

Future development is likely to occur in two areas. To further generalise the approach, it would be useful to partly relax the rigid shape requirement. A possible approach might allow the motion template to specify different models for different parts of the sequence, perhaps with some form of morphing between models loosely based on the HT for natural shapes. In application, we intend to integrate current gait research with motion templates. In combination with an articulated leg model, it might be possible to find specific individuals using a motion template to describe a person's unique walking motion. This will require investigation of whether a motion template of a walking pattern is unique to an individual.

References

[1] J. K. Aggarwal and Q. Cai, *Human Motion Analysis: A Review*, CVIU, **73**(3): 428-440, 1999.

[2] D. H. Ballard, *Generalising the HT to Find Arbitrary Shapes, CVGIP*, **13**: 111-122, 1981.

[3] A. S. Aguado, M. E. Montiel and M. S. Nixon, *Bias Error Analysis of the Generalised Hough Transform*, J of Math. Imaging and Vision, **12**, pp. 25-42, 2000.

[4] F. P. Kuhl and C.R. Giardina, *Elliptic Fourier Descriptors of a Closed Contour*, CGIP, **18**: 236-258, 1982.

[5] A. S. Aguado, M. S. Nixon and M. E. Montiel, *Parameterising Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Detection*, CVIU, **69**(2): 202-221, 1998

[6] J. M. Nash, J. N. Carter and M. S. Nixon, *Dynamic Feature Extraction via the Velocity Hough Transform*, Pattern Recog. Lett., **18**(10): 1035-1047, 1997.

[7] J. M. Nash, J. N. Carter and M. S. Nixon, *Extraction of Moving Articulated-Objects by Evidence Gathering*, Proc. BMVC98, **2**, pp. 609-618, 1998.

[8] M. G. Grant, M. S. Nixon and P. H. Lewis, *Finding Moving Shapes by Continuous-Model Evidence Gathering*, Proc. BMVC99, **2**: 554-563, 1999.

[9] Li-Qun Xu and D. C. Hogg, *Neural Networks in human motion tracking - An experimental study*, Image and Vision Computing, **15**, pp. 607-615, 1997.

[10] R. Polana and R. C. Nelson, *Detection and Recognition of Periodic, Nonrigid Motion*, International J. Computer Vision, **23(3)**, pp. 261-282, 1997.