Classifying Complex Human Motion Using Point Distribution Models

Ezra Tassone, Geoff West and Svetha Venkatesh
School of Computing
Curtin University of Technology
GPO Box U1987, Perth 6845 Western Australia
{tassonee,geoff,svetha}@computing.edu.au

Abstract

The Point Distribution Model (PDM) has been successfully used in modelling shape variations in groups of static images. It has also been effectively adapted to temporal image sets and used to track moving bodies such as hands and walking persons. However standard models do not consider the temporal characteristics of the data and are purely models of shape. This research proposes an extension to the PDM which explicitly considers the temporal sequencing of the images in the motion. The modified model can then be built from temporal quantities such as linear velocity and acceleration which are derived from the images. The new model formulation also enables movements to be tracked and classified according to their distinguishing temporal characteristics. This has been tested against distinct sets of arm movements under varying sets of experimental conditions.

1. Introduction

A number of computer vision techniques have been devised and successfully used to model variations in shape in large sets of images. Such models are built from the image data and are capable of characterising the significant features of a correlated set of images. One such model is the Point Distribution Model (PDM) [4] which builds a deformable model of shape for a set of images based on coordinate data of features of the object in the image. This research proposes an extension to the PDM in which the model is not solely constructed from the coordinate image data. The basic model has been adapted for use in temporal domains previously, but this is concentrated on the basic coordinate shapes. Results produced from these models would be the same regardless of the order of shape presentation.

The modified model of this research significantly differs from the standard in that it encodes the temporal sequencing of the data and not only the shape variation present. Hence it takes into account the order in which the shapes occur. This research focuses on the reparameterisation of the model by determining quantities such as acceleration and velocity from the spatial data, based upon an image sequence of a moving object. The model is tested experimentally on a number of arm motions and these motions classified based on comparison with a set of PDMs.

2. Background

The Point Distribution Model is built from the coordinates of points on objects in image sets and then performing PCA on this data. This enables the objects described by the model to be characterised by a linear model consisting of a mean shape plus a number of modes of variation which show the deviations from the mean for all points on the object [4]. The use of PCA enables the dimensionality of the model to be reduced as only the most significant modes of variation are added into the model, those which will allow for a high proportion of the variance in the shapes to be represented. Additionally, new shapes may be derived from the model which are representative of the shapes contained in the training set. Techniques such as the Active Shape Model [3] are also used to fit parameters to unseen images which are similar to those found in the training set.

The PDM has often been used on static images such as those found in medical imagery and objects such as circuit boards. Research with moving objects has been performed, such as the tracking of a walking person [2]. In this instance, a B-spline is used to model the shape of the walker with the control points of the spline forming the shape vectors for use in the PDM. A Kalman filter is used in conjunction with the Active Shape Model to accurately track the person over the sequence. PDMs have also been used in tracking people from moving camera platforms [5]. The human form is again represented by a B-spline with an application of the Condensation algorithm used to perform the tracking. The PDM has also been used to track and classify sequences of hand gestures [1].

Reparameterisations of the PDM have also been achieved in the domain of motion. One such application is the Cartesian-Polar Hybrid PDM which adjusts its modelling for objects which may pivot around an axis [7]. Points which undergo angular motion are mapped into polar coordinates, while other points remain as Cartesian coordinates and this allows for a more accurate representation of the motion. Other research has characterised the flock movement of animals by adding parameters such as flock velocity and relative positions of other moving objects in the scene to the PDM [9]. These parameters combined with standard image coordinates found in typical PDMs yield a richer, more useful description of the movement occurring.

3. The Point Distribution Model

3.1. Standard linear PDM

The construction of the PDM is based upon the shapes of images contained within a training set of data [4]. Each shape is modelled as a set of n "landmark" points on the object represented by xy-coordinates. The points indicate significant features of the shape and must be marked consistently across the set of shapes to ensure proper modelling. Hence for the 2D model, each shape is represented as a vector of the form:

$$\mathbf{x} = (x_1, y_1, x_2, y_2, x_3, y_3, \ldots, x_n, y_n)^T \tag{1}$$

To derive proper statistics from the set of training shapes, the shapes are aligned using a weighted least squares method in which all shapes are translated, rotated and scaled to correspond with each other. This technique is based upon Generalised Procrustes Analysis [6]. The mean shape $\overline{\mathbf{x}}$ is calculated from the set of aligned shapes, where N_s is the number of shapes in the training set:

$$\overline{\mathbf{x}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{x}_i \tag{2}$$

The difference $\mathbf{dx}_i$ of each of the aligned shapes from the mean shape is taken and the covariance matrix $\mathbf{S}$ derived:

$$\mathbf{S} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{dx}_i \mathbf{dx}_i^T \tag{3}$$

The modes of variation of the shape set are found from the derivation of the unit eigenvectors, $\mathbf{p}_i$, of the matrix $\mathbf{S}$:

$$\mathbf{S}\mathbf{p}_i = \lambda_i \mathbf{p}_i \tag{4}$$

The most significant modes of variation are represented by the eigenvectors aligned with the largest eigenvalues. The total variation of the training set is calculated from the sum of all eigenvalues with each eigenvalue representing a fraction of that value. Therefore the minimal set of eigenvectors that will describe a certain percentage (typically 95% or 99%) of the variation is chosen.

Hence any shape, $\mathbf{x}$, in the training set can be estimated by the equation:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{5}$$

where $\mathbf{P} = (\mathbf{p}_1 \mathbf{p}_2 \ldots \mathbf{p}_m)$ is a matrix with columns containing the m most significant eigenvectors, and $\mathbf{b} = (\mathbf{b}_1 \mathbf{b}_2 \ldots \mathbf{b}_m)^T$ is the set of linearly independent weights associated with each eigenvector. The set of weights may also be used as parameters to produce other shapes which are possible within the range of variation described by the PDM. As the variance of each $\mathbf{b}_i$ is λ_i, the parameters would generally lie in the limits:

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i} \tag{6}$$

3.2. Modified PDM for Motion Components

While prior research has shown it is possible to use the standard PDM for constructing models based on a temporal sequence of images, this paper instead proposes a reparameterisation of the PDM. The modified version of the model does not directly use image coordinates of the body but instead processes this data and derives other measures for input. Hence the utility of the model can be increased by using a feature space more suited to the analysis of motion.

To construct the PDM, a number of frames of the object in motion are taken as in temporal applications of the standard model. After extracting the boundary of the object, a subset of n points is selected for use in developing the model. The movement of the body from frame to frame and the subsequent boundary extraction generates a new image for input and processing. As the focus of this research is to describe

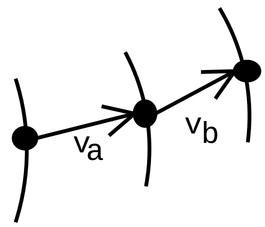


Figure 1. Frame triple and its vectors for modified PDM

and classify movements on the basis of motion components, the temporal sequencing of the shapes and the relative movement of the points on the shapes is used to reparameterise the PDM.

To achieve this a set of three temporally adjacent frames is considered at a time with the movement of a point from the first to second frame being one vector and the movement from the second frame to the third being a second vector. This is illustrated in Figure 1. These vectors are measured as the Euclidean norm between the xy coordinates of the points. From these vectors, the relevant motion components and thus the input parameters for the PDM can be calculated. There are many potential motion components that can be modelled, however this extension suggests the following four components:

1. **Angular velocity**, $\Delta\theta$ — the change in angle between the vectors, with a counter-clockwise movement considered a positive angular velocity and a clockwise movement a negative angular velocity.

2. **Acceleration**, a — the difference in the Euclidean norm between the vectors with the norm of the first vector being v_a and that of the second being v_b ie. $v_b - v_a$.

3. **Linear velocity**, v — this is the norm of the second vector v_b.

4. **Velocity ratio**, r — the ratio of the second velocity to the first, v_b/v_a. For a constantly accelerating body this measure will remain constant.

These parameters are calculated for every one of the n points of the object leading to a new vector representation for the PDM:

$$\mathbf{x} = (\Delta\theta_1, a_1, v_1, r_1, \Delta\theta_2, a_2, v_2, r_2, \ldots, \Delta\theta_n, a_n, v_n, r_n)^T$$

The user may also choose to focus on only one parameter for each point reducing the vector size and complexity of the model. Note also that performing this process for all points requires some algorithm for determining correspondences between points over the set of consecutive frames. Without this process the relationship between the points over the frames may be lost and the resultant model will not be accurate. This is analogous to the standard PDM in which landmark points should associate with the same features over all shapes in the training data to ensure correct modelling of the data.

This process is repeated for all triples of consecutive frames in the sequence. In this way information from all N frames in the sequence is included. However this reduces the number of temporal component shapes in the training set to be $N - 2$. After this reparameterisation of the model, the PDM can be built in the standard way. This characterisation encapsulates the temporal sequencing of the motion with the changes in parameters modelled on a frame to frame basis.

4. Classification and Tracking

To test the model, experiments were carried out on a set of sequences of a moving arm. The aim of this proportion of the research was to correctly track and classify movements through the building of the PDM and an application of a search strategy to match motions with a previously built PDM. The general method of experimentation used will now be described.

4.1. Video Capture and Image Processing

For all arm motion sequences, the same general sequence of preprocessing is applied in order to generate the coordinate image data for the building of the models. After video capture, the moving objects are segmented from the scene via thresholding. After the segmentation of the moving portion of the frame, images are chaincoded to derive the boundary of object as this is to be used in determining movements of points and tracking objects.

As described in Section 3.2, the modified PDM is built using a subset of n points on the boundary of the object rather than the traditional landmark points of the standard model. For the first frame of the sequence, this is done by selecting points spaced equidistant around the boundary. For subsequent frames, points are selected based on their correspondence with the points of the previous frame as is typically performed in motion analysis using various schemes. For this research a more simple method of choosing the closest point in terms of Euclidean distance in the next frame is used. This is possible as the motion between image frames is small and hence the likelihood of a poorly matched point is reduced. The matching point is searched for in a particular sized window on the boundary in a region matching where the corresponding point would be expected to be found. A check is also performed using Sobel edge detection to ensure that the corresponding point found is oriented in the same way as the point in the previous frame.

4.2. Training and Test Data

After deriving the set of N xy-coordinate shapes from the images, the modified PDM is built. The xy shapes are reparameterised into motion components for all points on all images. For this research, one parameter is modelled for each point rather than a combination of parameters and the feature of linear velocity is chosen as the motions were stable over time. This yields a vector of the following type for each frame:

$$\mathbf{x} = (v_1, v_2, v_3, \ldots, v_n)^T \tag{7}$$

The test data for use in tracking is prepared in the same manner as data for the model. Video footage of a movement is captured, the images processed into the correct format and a subset of n points on an image derived. Both the model and its test data must have the same number of points present in the temporal shape. As for the PDM, these points are reparameterised into a set of vectors using the same parameter of linear velocity. This yields an equivalent set of vectors such as were used in the construction of the PDM.

4.3. Classification

In order to evaluate the effectiveness of the PDM in classifying movement, the test sequence of vectors derived from the reparameterisation are tracked against the model. This involves deriving vectors that best fit the desired vector through adjustment of the b vector of the model. As described in Equation 6 these limits are typically taken to lie within three standard deviations.

As stated previously, the Active Shape Model [3] is a standard technique for fitting an instance of a PDM model to an example shape. However, this research differs in that temporal data rather than image data is present and modelled. Features such as gray level profiles and edge information are not available and so a standard ASM cannot be implemented. Hence a multidimensional version of Powell's method, as described in [8], is used. For this research it will find the combination of b values that will most closely approximate the required vector and hence minimise the error function between the actual and predicted vector. The vector of b values must however be minimised within the set limits stated previously. Hence any values that do not fall within the specified limits are adjusted to fit within the limits and so only allow predicted motion within the bounds of the PDM.

The methodology used to classify the movements was to match a test set of data against several prebuilt models, including a model built from a part of the sequence from which the test data came. The test data itself was "unseen" in that it did not form part of the data from which the PDM was built. The error between the matching and actual vectors of point velocities was measured over time as was the increasing total error over the series. The model which produced the lowest match error at the end of the tracking phase was classified as the matching model, and the test data classified as the same type of motion as the model.

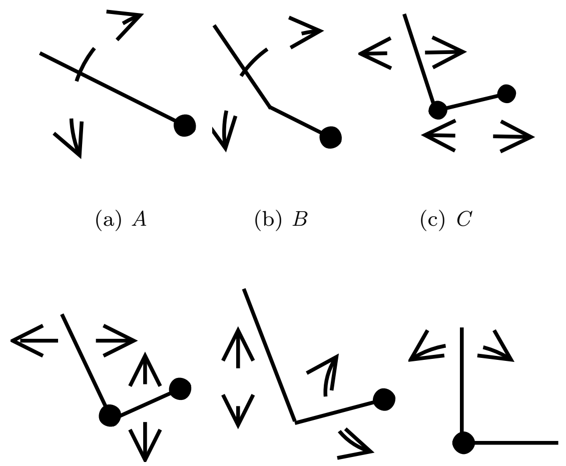
5. Experimental Results

5.1. Arm Motion

To test the research hypothesis, a set of six distinct arm movements were captured and linear velocity data derived from the images. Additionally two faster versions of the first two movements were captured given a total of eight movements to model and classify. The motions are illustrated in Figure 2.

There were 20 points identified on the boundary of the image for use in the model. Typically most sequences consisted of a few hundred frames of the demonstrated motion. Thus the PDM was built with the first 300 frames of motion of a sequence. The last 200 frames of motion of the sequence were used as the test data set.

Table 1 shows the final matching error value over all comparisons in matrix format for PDMs built to capture 95% of the variation in the training set. The boxed values are the lowest match error for a data set, as well as those along the diagonal of the matrix. An ideal set of results would place all the lowest errors down the diagonal of the matrix – that is the test data set matching most correctly with the PDM that is built from the same sequence. The results showed that in all but one of the motions the correct model was matched to the test data set. Motions A, B, C, D, E and H were unique and motions F and G were fast versions of A and B respectively. Only H was misclassified in that motion model E appears to be the correct match for the test data of H. The model for H has the second lowest final error and thus a correct solution was close to being found.



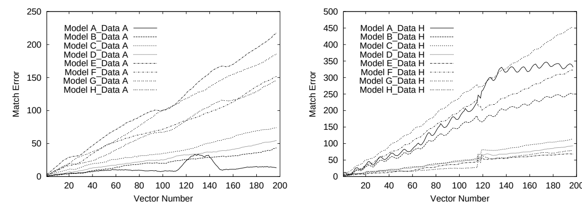
(a) A (b) B (c) C

(d) D (e) E (f) H

Figure 2. Six arm movements where each blob denotes a point of rotation. Arrows show allowable movement.

Data	Models							
	A	B	C	D	E	F	G	H
A	$\boxed{13.3}$	43.5	74.0	54.6	217.9	151.7	185.7	146.0
B	96.3	$\boxed{32.1}$	62.9	40.7	329.3	223.3	157.2	279.0
C	424.8	93.6	$\boxed{38.3}$	67.2	235.7	229.0	361.9	130.0
D	30.0	78.1	55.7	$\boxed{17.3}$	373.7	242.2	221.9	269.4
E	131.6	100.6	107.3	101.7	$\boxed{24.9}$	223.0	346.4	72.7
F	3658.0	2481.0	761.9	2341.1	1081.4	$\boxed{121.8}$	336.3	1077.0
G	3412.2	2320.8	537.8	2061.3	1273.0	477.0	$\boxed{130.0}$	1418.3
H	331.6	249.6	114.1	93.1	$\boxed{68.4}$	322.5	451.8	$\boxed{78.8}$

Table 1. Error matrix for arm with PDM variance of 95%



(a) Test data A (b) Test data H

Figure 3. Error plots for arm with PDM variance of 95%

Two of the match error plots are shown in Figure 3 displaying progressively error values over all frames in the sequences. For the correct match of test data A, the match error generally showed a consistently lower trend than the incorrect models which would obviously also lead to a lower total error at the end of testing. The only exception to this is for test data A in which a "bump" occurs in the error curve for a period of approximately 20 vectors, caused by an irregularity in the movement and poor segmentation. However, the tracking process is sufficiently robust to recover from this and validate the correct model at the end of the testing phase. In the set in which model E is incorrectly matched to motion H, Figure 3(b), the error plots C, D, E and H are close together and have similar gradients. This indicates these models provide more plausible matches for the motion with other models providing much higher error values.

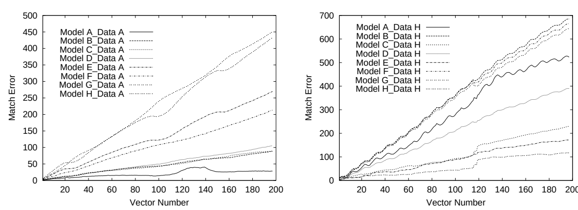
5.2. Decreasing variation of PDM

A further exploration of the efficacy of the model in classification was to investigate what effect reducing the variance modelled by the PDM had. This would reduce the number of eigenvectors contained in the models and limit the possible tracking that the models could apply to the test data, leaving the models less likely to cope with the more extreme variations in the motions. The variation represented by the models was reduced to 80%. For this experiment, all test motions were classified correctly against the models. The error that had previously occurred in the classification of motion H no longer existed. Therefore decreasing the variation reduced the possibility of incorrect results. The error matrix is shown in Table 2, with the lowest final match error highlighted. These coincide with the error values on the diagonal of the matrix, showing a full set of correct results. There is also a moderate upward trend in all error values as could be expected with the models less able to deal with larger fluctuations in the motion.

The same set of error plots as in the previous Section are shown for these experiments in Figure 4. While error values have increased, there is more distinction between the correctly classified error curves and the incorrect curves. The gap between the error values has increased over the models with 95% variation. The increased error over 20 vectors for test set A is still present, as is the step in some curves for H. Thus the irregularities in these motions still affect classification, although it is less noticeable than in the previous experiment. Hence while the decreased variation increased the match error, it has also reduced the likelihood of incorrect matches as models were less able to capture the features of motions which they were not directly modelled upon.

Data	Models							
	A	B	C	D	E	F	G	H
A	$\boxed{28.4}$	88.4	88.8	104.9	269.8	213.1	451.2	432.8
B	125.6	$\boxed{46.2}$	103.6	123.9	407.2	328.1	376.4	618.1
C	499.7	447.6	$\boxed{64.7}$	226.0	251.5	377.3	436.5	404.6
D	105.8	138.8	133.7	$\boxed{26.1}$	446.1	323.3	519.4	560.8
E	266.5	432.2	156.5	268.8	$\boxed{63.1}$	330.5	442.5	244.5
F	3681.3	2508.7	971.6	2366.1	1304.7	$\boxed{164.0}$	455.9	1918.4
G	3420.7	2327.9	670.8	2074.0	1507.3	606.4	$\boxed{155.0}$	2116.3
H	523.1	684.8	229.95	389.9	171.7	664.3	643.2	$\boxed{117.2}$

Table 2. Error matrix for arm with PDM variance of 80%



(a) Test data A (b) Test data H

Figure 4. Error plots for arm with PDM variance of 80%

6. Conclusion

This paper has described and illustrated an adaptation of the Point Distribution Model for a series of images taken of moving objects. The PDM is built to represent changes between temporally adjacent images which differs from standard models employed and thus incorporates temporal sequencing into the model parameters. The spatial image data is processed in order to derive spatiotemporal quantities such as linear velocity and acceleration for a set of points representing an image. Over a complete set of images, these vectors are used to build a PDM that represents the set of variation demonstrated by the motion features over time.

The model was then used to classify distinct types of movement of a person's arm. Continuing research will consider aims such as using the PDM to deal with more complex data such as human gait analysis, discovering the points of articulation on a moving body and also whether building separate models for each articulated part aids in classification.

References

[1] T. Ahmad, C. Taylor, A. Lanitis, and T. Cootes. Tracking and recognising hand gestures using statistical shape models. In *Proceedings of the British Machine Vision Conference, Birmingham, UK*, pages 403–412. BMVA Press, 1995.

[2] A. Baumberg and D. Hogg. An efficient method of contour tracking using Active Shape Models. In *1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects*, 1994.

[3] T. Cootes and C. Taylor. Active shape models - 'smart snakes'. In *Proceedings of the British Machine Vision Conference, Leeds, UK*, pages 266–275. Springer-Verlag, 1992.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proceedings of the British Machine Vision Conference, Leeds, UK*, pages 9–18. Springer-Verlag, 1992.

[5] L. Davis, V. Philomin, and R. Duraiwami. Tracking humans from a moving platform. In *15th. International Conference on Pattern Recognition, Barcelona, Spain*, pages 171–178, 2000.

[6] J. Gower. Generalized Procrustes Analysis. *Psychometrika*, 40(1):33–51, Mar. 1975.

[7] T. Heap and D. Hogg. Extending the Point Distribution Model using polar coordinates. *Image and Vision Computing*, 14:589–599, 1996.

[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, second edition, 1992.

[9] N. Sumpter, R. Boyle, and R. Tillett. Modelling collective animal behaviour using extended Point Distribution Models. In *Proceedings of the British Machine Vision Conference, Colchester, UK.* BMVA Press, 1997.