# Automated Pivot Location for the Cartesian-Polar Hybrid Point Distribution Model

Tony Heap and David Hogg
School of Computer Studies, University of Leeds,
Leeds LS2 9JT, UK
email: ajh@scs.leeds.ac.uk

**Abstract**

The Point Distribution Model (PDM) has already proved useful for many tasks involving the location or tracking of deformable objects. A principal limitation is that non-linear variations must be approximated by combining linear variations, which sometimes results in a non-optimal model producing implausible object shapes. The Cartesian-Polar Hybrid PDM helps to overcome this limitation; selective use of polar geometry allows bending or pivotal deformation to be modelled more accurately; model components which exhibit no such trend remain in the Cartesian domain.

Use of the Hybrid PDM currently requires the identification of pivot points by hand. In this paper we present a method for automatically identifying rigid model parts and pivot points from the training data. Experimental results are given for real data from human hands and for synthetic data from a simple jointed object.

**Keywords:** Deformable models, Point Distribution Model, polar coordinates.

## 1 Introduction

Models are used widely in computer vision; image features can be located, tracked or classified using *a priori* knowledge of object shape. Many objects are non-rigid, and thus require some sort of *deformable* model in order to capture shape variability.

One such model is the Point Distribution Model (PDM) [1], which has already been used as the basis for several vision applications [2, 3, 4, 5, 6]. An object is modelled in terms of landmark points positioned strategically on object features, and at regular intervals in between. By extracting such points from a set of training examples of the object, a statistical approach can be used to discover the mean object shape, and the major modes of shape variation.

The standard PDM is based purely on linear statistics; for any particular mode of variation, the positions of landmark points can vary only along straight lines.

BMVC 1995 doi:10.5244/C.9.10

Non-linear variation is achieved by combining two or more modes. This situation is not ideal, firstly because the most compact representation of shape variability is not achieved, and secondly because implausible shapes can be generated, due to the incorrect assumption that the variation modes are independent. The effect is particularly bad when the object being modelled can bend or pivot.

Attempts have been made to combat this problem. The Polynomial Regression PDM [7] allows landmark points to move along combinations of polynomial paths; however the computational complexity required is high, and structures experiencing large amounts of pivotal or bending motion still cause problems.

The Cartesian-Polar Hybrid PDM [8] helps to overcome these limitations. By selectively reparameterizing landmark points into polar coordinates, bending and pivotal deformation can be 'linearized' and thus modelled more accurately. For maximum flexibility, polar-mapped points are allowed to pivot about any other model point, with an axis fixed by a third point; thus pivot chains are possible.

One of the major strengths of the standard PDM is that training is fully automatic (apart from landmark point generation). The Hybrid PDM requires landmark points to be classified as either Cartesian or polar, and polar points must have a pivot and axis reference chosen. This can be done by hand but ideally it should be automated.

We previously used a classification algorithm which found mappings by maximizing the model's 'compactness', (measured quantitatively) (see [8] for details). This worked well, but assumed that all the necessary pivot points had been marked on the training examples, which is not always the case.

In this paper we propose a new classification method which discovers any pivotal deformation present via statistical analysis of the training data, and automatically inserts pivot points on training examples as and when required. We show the algorithm operating on real data, and use synthetic data to study the system's robustness to varying input data.

## 2 The Point Distribution Model

A PDM is built purely from the statistical analysis of a number of examples of the object to be modelled [1]. Given a collection of training images of an object, the Cartesian coordinates of $N$ strategically-chosen landmark points are recorded for each image. Training example $e$ is represented by a vector $\mathbf{x_e} = (x_{e1}, y_{e1}, \ldots, x_{eN}, y_{eN})$ (for a 2D model).

The examples are aligned (translated, rotated and scaled) using a weighted least squares algorithm, and the mean shape $\overline{\mathbf{x}}$ is calculated by finding the mean position of each landmark point. The modes of variation are found using Principal Component Analysis (PCA) on the deviations of examples from the mean, and are represented by $N$ 'variation vectors' $\mathbf{v_j}$. An object shape $\mathbf{x}$ is generated by adding linear combinations of the $t$ most significant variation vectors to the mean shape:

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j=1}^{t} b_j \mathbf{v_j} \tag{1}$$

where $b_j$ is the weighting for the $j^{th}$ variation vector.

By ensuring $t \ll N$, we extract only the important deformations, discarding training data noise, and thus we can compactly capture object shape and variation. In most cases, deformation is restricted to plausible shapes, making the PDM an ideal tool for robust location or tracking of non-rigid objects. However, the linear nature of the PDM can limit its usefulness; models are not as compact as they could be, and implausible shapes can occur.

# 3 The Cartesian-Polar Hybrid PDM

The Cartesian-Polar Hybrid PDM provides an improvement on the standard PDM for objects where significant pivoting or bending occurs. Before PCA is performed, landmark points which appear to pivot about some other landmark point are mapped into polar coordinates, with the suspected pivot point as origin and a third point chosen sensibly to provide an axis reference. Landmark points which have no such angular behaviour remain in Cartesian form. In this way, what was a non-linear variation has become linear simply by altering its coordinate frame.

Full details of the process are given in [8], however two important points should be noted:

- The hybrid model does not in any way constrain the allowable shapes; polar-mapped points are still able to vary in two dimensions, hence the hybrid model is more than just a simple articulated model.
- The increase in computation time is small; for example, to generate an instance of the anglepoise lamp as shown below required 221105 machine cycles using a Hybrid PDM, compared to 215621 cycles for the standard PDM (an increase of 2.5%).

The Hybrid PDM performs well in practice. Figures 1 and 2 show how the two models compare when used to represent the variation of an anglepoise lamp.

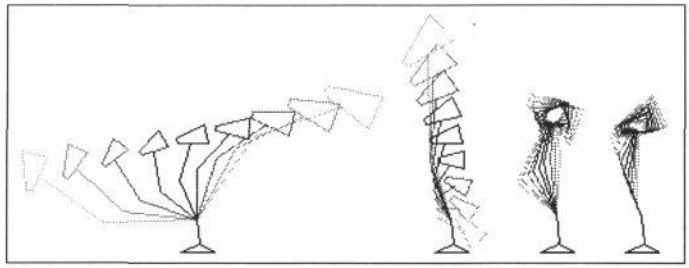

Figure 1: The first four modes of variation under the standard PDM

Statistically, the Hybrid PDM was more compact. Mapping points into the polar domain for the anglepoise removed 58.5% of the variation as represented by the standard PDM, suggesting that in this case the Hybrid PDM was at least twice as compact.

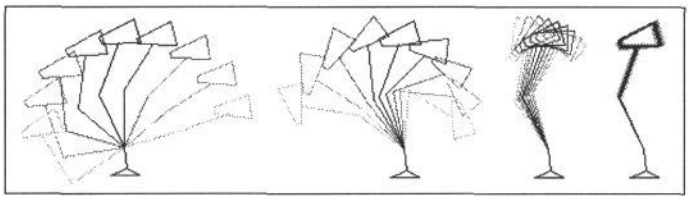

Figure 2: The first four modes of variation under the Hybrid PDM

# 4 Automatic Pivot Generation

The Hybrid PDM requires model landmark points to be classified as being represented in either the Cartesian or polar domain, and in the polar case, a choice of origin and axis is also needed. In order to make the Hybrid PDM useful in the general case, this process should be automated.

The technique suggested here uses statistical analysis of the training data to spot trends in the movement of points. Broadly speaking, if a point appears to move in an arc then it should be classified as polar, and its centre of rotation marked. To allow for chains of pivots this idea is extended; polar movement *relative to other points or sets of points* is detected. The method employed is to construct groups of points which appear to represent different roughly-rigid parts of the object, and look for pivotal relationships between pairs of groups.

## 4.1 Finding Rigid Groups of Points

We are interested in finding groups of landmark points which appear to be rigid in the sense that they appear to 'move together' over the training set. In order to achieve this, we first calculate the $N$ by $N$ normalized distance variance matrix $\mathbf{D}$ for the training set, where:

$$[\mathbf{D}]_{ij} = \mathrm{Var}|_{e=1}^{E} \left( \frac{\sqrt{(x_{ej} - x_{ei})^2 + (y_{ej} - y_{ei})^2}}{S_e} \right) \tag{2}$$

where Var is the variance operator, and $S_e$ is a measure of the overall size of example $e$, given by:

$$S_e = \frac{1}{E} \sum_{i=1}^{N} \sum_{j=1}^{N} \sqrt{(x_{ej} - x_{ei})^2 + (y_{ej} - y_{ei})^2} \tag{3}$$

A small value for $[\mathbf{D}]_{ij}$ indicates that points $i$ and $j$ move together over the training set. Hence we can use $\mathbf{D}$ as a basis for 'growing' rigid groups as follows:

1. Each point starts off as a singleton group.

2. For each group, an attempt is made to add a new point. If $G$ is the set of points in the group under consideration, then we choose the point $k \notin G$ which minimizes $S_k$ in:

$$S_k = \max(d_{ij} | i, j \in G \cup \{k\}) \quad (4)$$

If $S_k$ is below a fixed threshold then point $k$ is added to the group, otherwise no point is added.

3. Remove any duplicate groups formed as a result of step 2.

4. Iterate steps 2 and 3 until groups have stopped growing.

5. Finally, groups with two points or less are discarded. It follows that any points which do not form part of a rigid area are not assigned a grouping, and are excluded from the pivot search.

The threshold used for grouping is deliberately set quite tight, making groups smaller than they might otherwise be; two groups which should ideally be one can be merged later when they are discovered to have a similar pivot point, but one group which should ideally be two can give poor results at the pivot-finding stage.

The use of thresholding can, in some cases, cause problems, since different model components often require different thresholds for good results. For example, in objects such as the human hand, which consist of a large main body (eg. the palm) and smaller, pivoting sections (eg. the fingers, and in particular the thumb, which has two separate rigid portions), a larger threshold is generally required for the main body than for the smaller sections. The method we use to combat this involves a three-stage process:

1. The training set is aligned using a weighted least squares technique [1] (the same alignment process is used prior to PCA in the training process). Points whose standard deviation from their mean position is less than a fixed threshold (1.5% of the model size) are deemed fairly static and are thus included in a *base group* of points.

2. Group-growing is then performed using the subset of points not in the base group.

3. When the iteration has converged, base group points are reintroduced, and the iterative process is again allowed to converge. This final step allows overlap between the base group and other groups.

## 4.2 Finding Potential Pivots Between Groups

Having identified rigid groups of landmark points, the next step is to find plausible pivot points and augment training examples with additional landmarks at the pivots. Pivot points will generally have different coordinates in each training example. Even if we globally align the examples, pivots in chains will not occupy a fixed position. It is therefore necessary to construct a local coordinate frame for each rigid group in each training example, such that points in the group are roughly static *in local coordinates* over all training examples. We can look for static pivot points in the *local* coordinate frames, and then map them back into global coordinates. The details of the procedure are as follows:

1. For each group $g$ in each training example $e$, we must find a suitable coordinate frame transformation $C_{g,e} = (s_{g,e}, t_{g,e}, p_{g,e}, q_{g,e})$, where a point $(x, y)$ in local coordinates is transformed onto $(x', y')$ in global coordinates by:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} p_{g,e} & -q_{g,e} & s_{g,e} \\ q_{g,e} & p_{g,e} & t_{g,e} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{5}$$

The transformation matrix in (5) is labelled $\mathbf{R}_{g,e}$.

For each group $g$, we perform a least squares alignment [1] of its member points over the training set. $\mathbf{R}_{g,e}$ is defined as the inverse of the transformation required to align example $e$.

2. For each pair of groups $a$ and $b$, the best pivot point is found. We define this point to be at $(u_a, v_a)$ in $a$'s local coordinates and at $(u_b, v_b)$ in $b$'s local coordinates. Ideally, for each training example, these two points would transform onto roughly the same global position, ie. if $\mathbf{R}_{g,e}(u_g, v_g, 1)^T = (u_g^e, v_g^e, 1)^T$ then $\forall e.(u_a^e, v_a^e) \approx (u_b^e, v_b^e)$.

Hence we find the values for $u_a$, $v_a$, $u_b$ and $v_b$ which minimize $\mathcal{E}$, the sum of the squares of the distances between the pairs of global coordinates:

$$\mathcal{E} = \sum_{e=1}^{E} (u_b^e - u_a^e)^2 + (v_b^e - v_a^e)^2 \tag{6}$$

Using $(u_g^e, v_g^e, 1)^T = \mathbf{R}_{g,e}(u_g, v_g, 1)^T$, we obtain:

$$\mathcal{E} = \sum_{e=1}^{E} \begin{array}{l} ((p_{b,e}u_a - q_{b,e}v_b + s_{b,e}) - (p_{a,e}u_a - q_{a,e}v_a + s_{a,e}))^2 + \\ ((q_{b,e}u_a + p_{b,e}v_b + t_{b,e}) - (q_{a,e}u_a + p_{a,e}v_a + t_{a,e}))^2 \end{array} \tag{7}$$

To minimize $\mathcal{E}$, we equate the partial derivatives $\partial\mathcal{E}/\partial u_a$, $\partial\mathcal{E}/\partial v_a$, $\partial\mathcal{E}/\partial u_b$ and $\partial\mathcal{E}/\partial v_b$ to zero, and hence obtain four simultaneous linear equations, the solution of which simplifies to:

$$(u_a, v_a, u_b, v_b)^T = \sum_{e=1}^{E} (J_e^T J_e + K_e^T K_e)^{-1} \sum_{e=1}^{E} (s_{b,e} - s_{a,e}) J_e + (t_{b,e} - t_{a,e}) K_e \tag{8}$$

where $J_e = (p_{a,e}, -q_{a,e}, p_{b,e}, -q_{b,e})$ and $K_e = (q_{a,e}, p_{a,e}, q_{b,e}, p_{b,e})$.

From here, the pivot point in global coordinates $(c_{a,b,e}, d_{a,b,e})$ can be found for each training example $e$ by transforming the two points into global coordinates and averaging them:

$$\begin{pmatrix} c_{a,b,e} \\ d_{a,b,e} \end{pmatrix} = \frac{1}{2} \left( \begin{pmatrix} u_a^e \\ v_a^e \end{pmatrix} + \begin{pmatrix} u_b^e \\ v_b^e \end{pmatrix} \right) \tag{9}$$

The variability $V_{a,b}$ of the pivot can be measured as the average separation of these two points over the training set, when mapped into global coordinates:

$$V_{a,b} = \frac{1}{E} \sum_{i=1}^{E} \sqrt{(u_b^e - u_a^e)^2 + (v_b^e - v_a^e)^2} \tag{10}$$

3. A pivot is declared between groups $a$ and $b$ if *all* of the following are true:

   - The variability $V_{a,b}$ of the pivot is less than 5 pixels.
   - The standard deviation of the distribution of angles between the coordinate frames for $a$ and $b$ over the training set is at least 5°.
   - The normalized deviation (standard deviation divided by mean) in the distribution of the size ratios of the two coordinate frames is less than 0.1 (ie. the relative size of the two groups remains fairly constant).

   The thresholds have been chosen to be tolerant, as false pivots can still be rejected in the final stage.

### 4.3 Constructing the Pivotal Structure

Once we have found a set of potential pivots, we need to construct a mapping for use by the Hybrid PDM, noting that the formulation of polar mappings relies on having no cyclic dependencies amongst points. Hence, groups are organized into one or more tree-like structures, where child groups pivot off parent groups, and the root group(s) provide(s) a relatively stable reference.

The algorithm we have implemented caters for most cases; a single tree is constructed. The root is taken to be the largest rigid group found, and a breadth first search is used to attach child groups which have a common pivot with the parent group. Using the breadth first search ensures that a structure is found with the shortest pivotal chains.

The point mapping is then constructed in an obvious fashion; points which are in a group with no parent (ie. at the tree root or not included in the tree), and points in no group, are assigned a Cartesian mapping.

Points in parented groups are assigned a polar mapping. New landmark points are generated in all training examples to represent the pivot points, as described in 4.2 above. If a new pivot is consistently sufficiently close (5 pixels) to any existing landmark point, it is removed and the existing point used instead. A suitable point from the parent group is chosen as an axis reference point.

## 5 Results and Discussion

Results are presented for experiments using training data from human hands. Thirty training examples were captured, all of hands positioned with the palm down and fingers outstretched, and the positions of 61 landmark points were extracted from each one. Figure 3 summarizes the results.

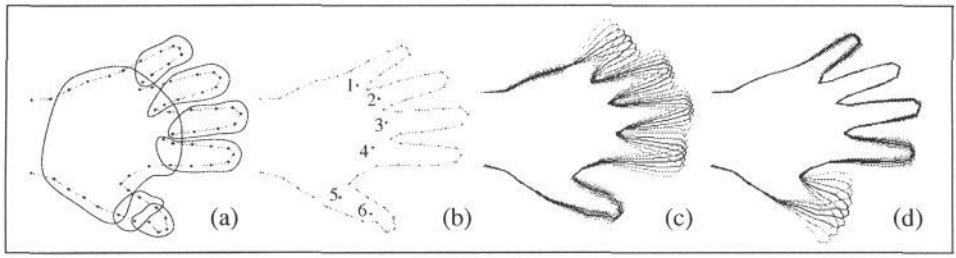


Figure 3: Automatic pivot generation for a hand model

Plot (a) shows how the points have been grouped and plot (b) shows the pivot points thus generated. Each finger has a single group, and the thumb has been split into two groups, giving two rigid sections. Points near the wrist have failed to be grouped because the angle of the wrist varies significantly over the training set, and the groups produced were deemed too small to be significant enough for pivoting.

All pivots (1 to 6) have been placed roughly where we would expect them; the pivotal structure of the thumb is such that pivot 6 is parented by pivot 5.

Plots (c) and (d) show examples of the variation modes produced after training using the annotated data. Pivotal motion is clearly visible, especially in the thumb.

These results are encouraging; the natural structures present in the human hand have been discovered accurately, and the pivot points have been positioned roughly as expected.

Quantitative data has been obtained in a second experiment. A synthetic model was constructed, comprising a fixed body and a pivoting arm (see Figure 4). Training sets were constructed with each example having the arm pivoted at a different angle, and with additive Gaussian noise applied to each landmark point. The idea was to study the accuracy of finding the pivot point whilst varying the following input parameters:

- The range of angles of the pivoted arm (a Gaussian distribution with standard deviation of $\sigma_a$ degrees).
- The amount of noise used to displace landmark points (additive Gaussian, with standard deviation $\sigma_n$).
- The number of training examples used, $E$.
- The total number of landmark points in the model, $P = 6n - 4$.

Running the pivot finder on the synthetic data produces the coordinates of a single pivot. The distance of this pivot from ground truth can be calculated as a measure of accuracy. Several hundred trials were performed for each choice of parameters to give the *mean output noise* with tight confidence limits.

The angle range $\sigma_a$ is an inherent property of the object being modelled and the noise $\sigma_n$ is dependent on the accuracy of the image capture technique. Figure 5(a) shows how they both affect accuracy of the pivot position (control values used were

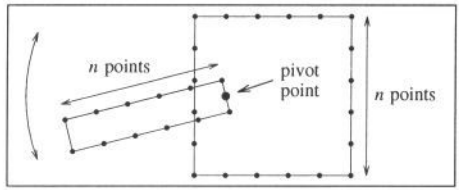


Figure 4: A synthetic model of an object with a single pivot

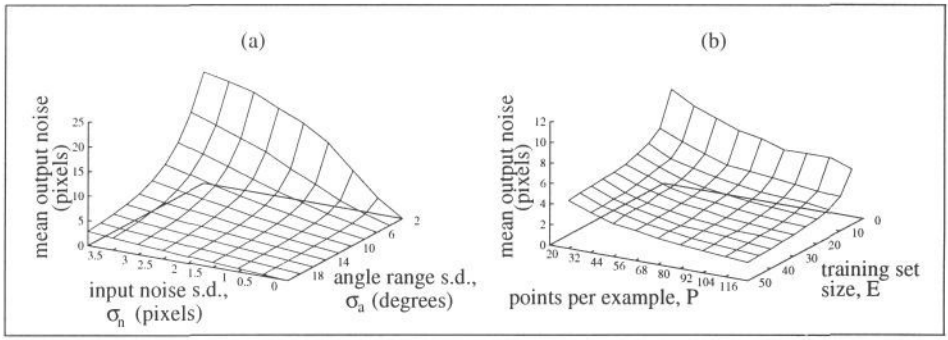


Figure 5: Pivot location accuracy under varying conditions of (a) input noise and angle range; (b) training set size and number of points per training example.

$P = 32$ and $E = 20$). As expected, the accuracy decreases as the angle decreases, and as the level of noise increases.

The remaining parameters $E$ and $P$ can be seen as measures that could be employed to improve accuracy. As can be seen in Figure 5(b), accuracy improves with an increase in both the training set size and the number of points (control values used were $\sigma_n = 3$ pixels and $\sigma_a = 8°$). More specifically, increasing the training set size will only help to a certain extent (in this example, above 25 examples there is little improvement), but increasing the number of points per example gives a more steady increase in performance.

# 6 Conclusions

We have presented a method which improves the usefulness of the Cartesian-Polar Hybrid PDM. We can automatically discover pivotal motion present in training examples, augment training examples with the necessary pivot points, and correctly classify model landmark points as either Cartesian or polar mapped.

The motivation for this work is in anticipation of modelling human hands in 3D using PDMs; the standard PDM would be unsuitable for modelling the non-linearity present in, for example, the bending of the fingers. The automated Hybrid PDM should provide a better alternative.

There are improvements we would like to make to the system; in particular, the construction of the pivot 'tree' is based on a very simple breadth-first search algorithm, in which only one group can act as a root. This caters for most cases, but objects with several distinct pivoting structures will cause problems. Also,

the manner in which the various thresholds are fixed rigidly is unsatisfactory; a method requiring no thresholding would be preferable in order to cope with a wider range of tasks.

## Acknowledgements

This work is supported by an EPSRC grant. Thanks are due to members of the Leeds University Vision Group for general help and ideas.

## References

[1] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proc. BMVC*, pages 9–18, Leeds, UK, 1992. Springer-Verlag.

[2] A. Hill, T.F. Cootes, and C.J. Taylor. A generic system for image interpretation using flexible templates. In *Proc. BMVC*, pages 276–285, Guildford, UK, 1993. BMVA Press.

[3] A. Hill, A. Thornham, and C.J. Taylor. Model-based interpretation of 3D medical images. In *Proc. BMVC*, pages 339–348, Leeds, UK, 1992. Springer-Verlag.

[4] A. Lanitis, C.J. Taylor, and T.F. Cootes. A generic system for classifying variable objects using flexible template matching. In *Proc. BMVC*, pages 329–338, Guildford, UK, 1993. BMVA Press.

[5] A. Baumberg and D. Hogg. Learning flexible models from image sequences. In *Proc. 3rd ECCV*, pages 299–308, Stockholm, Sweden, 1993. Springer-Verlag.

[6] A.J. Heap. Real-time hand tracking and gesture recognition using Smart Snakes. In *Proc. Interface to Human and Virtual Worlds*, Montpellier, France, June 1995.

[7] P.D. Sozou, T.F. Cootes, C.J. Taylor, and E.C. Di-Mauro. A non-linear generalisation of PDMs using polynomial regression. In *Proc. BMVC*, volume II, pages 397–406, York, UK, 1994. BMVA Press.

[8] A.J. Heap and D.C. Hogg. Extending the Point Distribution Model using polar coordinates. In *Proc. CAIP*, Prague, Czech Republic, September 1995.