

Chapter 2



2 Literature Review

2.1 Introduction

An initial literature review was performed which surveyed the field of 3D computer vision. The review covered types of image data from 2D images, range data and depth maps to volumetric segmentation. Acquisition methods, reconstruction and image segmentation were also covered and conclusions drawn to support the remainder of the research. This initial survey was too general for inclusion within this manuscript and hence is available as a separate technical report [Bowden 96].

The conclusions of the report were that contour or surface based approximations (specifically statistical contour models) are important for the following reasons:

- Image searching is localised along contour boundaries and hence provides significant computational savings over more traditional low level image processing techniques. This benefit is more apparent where real-time processing of image sequences or large volumetric datasets are considered.
- The ability to introduce *a priori* knowledge about object shape and deformation into a contour provides a robust deformable template

which can be applied to an image where the absence or occlusion of object features and cluttered/complex backgrounds would result in the failure of other techniques.

- The ability to accurately segment objects from images or sequences provides smoothed object boundaries.
- The ability to aid in the classification of objects under affine transformation.

The remainder of this chapter will present a more specific review of related literature, namely in the area of statistical models of deformation and associated approaches.

2.2 Contour Models

The seminal work of Kass *et al* on *Snakes* or the *Active Contour Model* (ACM) presented a class of semi-automatic methods for segmentation using energy minimising curves [Kass, 1988; Kass, 1987]. In these methods, a user draws the approximate boundary of the region of interest in an image. Then, an elastic contour is fitted to the boundary points and the curve is iteratively refined until its internal energy defined by its curvature is minimised while responding to external forces derived by image edges. Many researchers have shown how these active contour models can be used to locate and track an object in an image [Etoh, 1992; Ueda, 1992; Cipolla, 1992].

Zhou and Pycock segment cells from 2D images using statistical models applied like snakes [Zhou, 1995; Zhou, 1995]. Models are built up for different forms of cells; the interpretation process optimises the match between models and the data using a Bayesian distance measure. Lobregt and Viergever extend upon this model, presenting solutions to the problems of unwanted deformation like shrinking and vertex clustering [Lobregt, 1995]. There is a wealth of published work on variations on the basic model proposed by Kass *et al*, all use the same basic model with small constraints added to allow *a priori* knowledge of shape to be imposed upon the model and hence provide better performance.

Terzopolous and Vasilescu [Terzopoulos 91] extended the snake model to include an inflation force that helps remove the need for initial contour placement and thus avoid convergence on local minima. The inflation force drives the snake model outwards towards the object boundary like an inflating balloon. Terzopolous and Vasilescu formulated the model as a finite element mesh and later extended the model to a thin plate spline, demonstrating successful results in the reconstruction of range data and volumetric CT data surface representations [McInery 93]. Bowden *et al* extended this work further and is discussed in more detail in Appendix 2 [Bowden 97].

Several researchers have proposed B-Spline variations of the active contour model [Rueckert, 1995; Schnabel, 1995; Blake 1998]. Schnabel and Arridge looked at the problems associated with high curvature in active contour models, proposing a curvature matching technique for isophoto curvature matching. They look at the applications of using this approach to segment high curvature contours of the brain from medical images. Blake and Isard have combined many of their publications on the subject in the text 'Active Contours' which covers the construction, tracking and applications of B-spline contour approximations [Blake 1998].

It has been shown that these 2D models can be used to reconstruct 3D surfaces from volumetric data by applying snakes to individual slices to extract contours that can then be reconstructed into a 3D model [Carlbom, 1994; Goshtasby, 1995]. A typical implementation of such a system uses the final model from one slice as an initial estimate for the next to reduce user intervention.

Ivins and Porrill presented Active Region Models [Ivins 98], an adaptation to Kass's Active Contour Models where colour regions within an image are used to locate and track the boundaries of regions within the image.

A Neural network approach was proposed by Chiou *et al* called the neural network based stochastic active contour model (NNS-SNAKE) which integrates a neural network classifier for systematic knowledge building, and an active

contour model for automated contour location, using energy functions to stochastically locate the most probable contour.

2.3 Statistical Models of Deformation

A Point Distribution Model (PDM)[Cootes 95] gets its nickname of ‘Smart Snake’ from its obvious similarity to elastic snakes (Active Contour Models, ACM [Kass, 1987]). The major difference is that while snakes retain shape information in the elasticity and rigidity of their constituent points, a PDM uses a statistical model to specify allowable deformations. This not only makes the PDM less computationally expensive than the ACM but deformation is easier to build into the model.

Since they were proposed by Cootes *et al*, a wealth of research has been undertaken into Point Distribution Models. A PDM (the underlying mathematical model) or Active Shape Model (the model’s applied name) is a statistical model which can be constructed from a training set of correctly labelled images. A PDM represents an object as a set of labelled points, giving their mean positions and a small set of modes of variation which describe how the object’s shape can change. These modes of variation are gained from Principal Component Analysis (PCA) on the training set and represent the largest eigenvectors of the covariance matrix. An Active Shape Model exploits the linear formulation of PDMs in an iterative search procedure capable of rapidly locating the modelled structures in noisy, cluttered images, even when partially occluded [Cootes, 1995].

Turk and Pentland [Turk 91] present a method for extracting only the number of eigenvectors equal to the number of training examples and not the dimensionality of the set, in a similar manner to that of Cootes *et al* [Cootes 95] and this is discussed in more detail in Chapter 3.

It has been shown by Bowden *et al* that the PDM provides sufficient dimensional reduction inherent to the model to enable the simple classification of static shape [Bowden, 1995; Bowden, 1996]. These authors outline a simple method for using this dimensional reduction to classify shape deformation from the variation

weights from the mean. They show how static gestures can be recognised in real-time for a PDM of the human hand.

Lantis, Taylor and Cootes have also extended their initial work from contour models to shape and grey-level models [Lantis, 1994]. They use a combined PDM that uses both shape and a grey scale maps to locate and identify human faces.

Turk and Pentland use principal component analysis to describe face images in terms of a set of basis functions or 'eigenfaces'. Though valid modes of variation are learnt from a training set, and are more likely to be more appropriate than a 'physical' model, the eigenface is not robust to shape changes, and does not deal well with variability in pose and expression. However, the model can be matched to an image easily using correlation-based methods [Turk 91].

Magee and Bole presented Vector Distribution Models, where points around a connected contour are converted into a vector, and these vectors are concatenated into a final training vector on which PCA is performed [Magee 98]. These authors went on to discuss the use of Canonical Analysis, a similar procedure to PCA where two co-variance matrices are formed, one describing Intra class variation and one Inter class variation. After extraction of a generalised eigen system a new eigen space is extracted. Although this space may not necessarily be optimised for dimensional reduction, it is useful for data classification as the first components of the model represent inter-class variation [Magee 99].

Swets and Weng [Swets 96] presented a technique called a combined eigen-canonical transform which combined canonical analysis with PCA to give data reduction and improved classification. Canonical analysis was performed on data after it had been projected down into the lower eigen space gained from PCA similar to that outlined in section 6.

Initial work of extending the PDM (Active Shape Model) to 3D has already been proposed by [Hill, 1995].

Ferryman *et al* use PCA on 3D rigid models to build a deformable model for various different car shapes which is used to locate and track moving traffic [Ferryman, 1995]. The process is very similar to that of the PDM. However, instead of modelling the object as points that make up the boundaries of the object, points are chosen at landmarks such as corners, and the model built up from the known interconnection of these points.

O'Toole *et al* presented work for 3D models of faces represented as a mean face with weightings that can be used to deform the model [O'Toole 96]. Faces were built up as 3D surfaces from a set of 65 male and 65 female heads. PCA analysis was performed to provide a compact model. They show that the primary mode of variation of the eigenface data set provides the mapping from a male head to a female head.

2.4 Non Linear PDMs

The linear formulation of the PDM relies on the assumption that similar shapes will produce similar vectors. This being the case, it is a fair assumption that the training set will generate a cluster in some shape space. However, it is unfair to assume that this cluster will be uniform in shape and size. As more complex models are considered the training set may even generate multiple, separate clusters in the shape space.

Under these circumstances the linear PDM will begin to fail as non-linear training sets produce complex high dimensional shapes which, when modelled through the linear mathematics of PCA, produce unreliable models. The nature of non-linear shape spaces will be discussed in depth in later chapters but a number of authors have addressed the problems associated with the construction of non-linear PDMs.

Where rotational non-linearity is known to be present within a model this can be removed/reduced by mapping the model into an alternative linear space. Heap and Hogg suggested using a log polar mapping to remove non-linearity from the training set [Heap 95]. This allows a non-linear training set to be projected into a

linear space where PCA can be used to represent deformation. The model is then projected back into the original space. Although a useful suggestion for applications where the only non-linearity is pivotal and represented in the paths of the landmark points in the original model, it does not provide a solution for the high non-linearity generated from other sources.

Higher order non-linearity is often the result of incorrect labelling of training examples. By carefully selecting landmark points by hand, a near optimum labelling can be achieved which will minimise the non-linearity of a training set. However, for all but the most simple of cases this is not a feasible solution. Often semi-automated procedures are used where a user can speed up the process of labelling example shapes for analysis. Fully automated procedures are rarely used due to the problems of correctly assigning landmarks and the highly non-linear models that this produces.

Work done by Baumberg and Hogg goes some of the way to solving non-linearity in deformable models by using a B-Spline representation. Landmark points for the Spline are represented as a PDM [Baumberg, 1995]. The curvature of the B-Spline takes on some of the non-linearity of the model and therefore reduces the problems presented with linear PDM representing non-linear models.

It has been proposed by Kotcheff and Taylor that non-linearity introduced during assembly of a training set could be eliminated by automatically assigning landmark points in order to minimise the non-linearity of the corresponding training cluster [Kotcheff 97]. This can be estimated by analysing the size of the linear PDM that represents the training set. The more non-linear a proposed formulation of a training set, the larger the PDM needed to encompass the deformation. The procedure was demonstrated using a small test shape and scoring a particular assignment of landmark points according to the size of the training set (gained from analysis of the principal modes and the extent to which the model deforms along these modes, i.e. the eigenvalues of the covariance matrix). This was formulated as a minimisation problem, using a genetic algorithm. The approach performed well but at a heavy computation cost [Kotcheff 97].

As the move to larger, more complex models or 3D models is considered, where dimensionality of the training set is high, this approach becomes unfeasible. A more generic solution is to use accurate non-linear representations. As linear PCA is used for linear PDMs, so, non-linear PCA can be used to model non-linear PDMs and many researchers have proposed approaches to this end.

Sozou et al first proposed using polynomial regression to fit high order polynomials to the non-linear axis of the training set [Sozou 94]. Although this compensates for some of the curvature represented within the training set, it does not adequately compensate for higher order non-linearity, which manifests itself in the smaller modes of variation as high frequency oscillations. In addition, the order of the polynomial to be used must be selected and the fitting process is time consuming.

Sozou et al further proposed modelling the non-linearity of the training set using a backpropagation neural network to perform non-linear principal component analysis [Sozou 95]. This performs well, however the architecture of the network is application specific; also, training times and the optimisation of network structure are time consuming. What is required is a means of modelling the non-linearity accurately, but with the simplicity and speed of the linear model.

Several researchers have proposed alternatives, which utilise non-linear approximations, estimating non-linearity through the combination of multiple smaller linear models [Bowden 97; Bregler 94; Cootes 97; Heap 97]. These approaches have been shown to be powerful at modelling complex non-linearity in extremely high dimensional feature spaces [Bowden 97].

The basic principle behind all these approaches is to break up any curvature into piecewise linear patches, which estimate the non-linearity rather than modelling it explicitly. This is akin to the polygonal representation of a surface. A smooth curved surface can be estimated by breaking it down into small linear patches. In the field of computer graphics this technique is performed to reduce render time. There exists, of course, a trade off between visual accuracy and computation

speed (where the minimum numbers of polygons are used to achieve the desired appearance). The same problem is present in non-linear PDM estimation, where the minimum number of linear patches that accurately represent the model must be determined.

Bregler and Omohundro suggested modelling non-linear data sets of human lips using a *Shape Space Constraint Surface* [Bregler 94]. The surface constraints are introduced to the model by separating the space surface into linear patches using cluster analysis. However the dimensionality of these 'lip' shape spaces is low as is the non-linearity due to the simplified application of the work.

Cootes and Taylor suggested modelling non-linear data sets using a Gaussian mixture model, which is fitted to the data using Expectation Maximisation [Cootes 97]. Multiple Gaussian clusters are fitted to the training set. This provides a more reliable model as constraints are placed upon the bounds of each piecewise patch of the shape space, which is modelled by the position, and size of each Gaussian.

Both of these estimation techniques become unfeasible as dimensionality and training set size increase. However by projecting the training set down into the linear subspace as derived from PCA the dimensionality and therefore computation complexity of the non-linear analysis can be reduced significantly to facilitate statistical and probabilistic analysis of the training set. This projection relies upon the dimensional reduction of PCA while retaining the preservation of the important information, the shape of the training set [Bowden 97; Bowden 98] and will be discussed fully in the following Chapters.

2.5 Tracking

By treating the problem of model fitting and tracking as an optimization technique the problems of discontinuity can be overcome. Hill *et al* proposed using genetic algorithms to model the discontinuous changes in shape space/model parameters [Hill 91][Hill 92]. Cootes *et al* present the use of genetic algorithms for initial image search and initialisation of PDMs within the image

frame [Cootes 95]. The use of genetic algorithms to overcome the complexities of tracking with the piecewise non-linear model has been investigated. However, the performance of such an approach relies largely on the formulation and structure of the genetic algorithm itself.

Blake *et al* emphasised the advantage of using low-parameter descriptions of deformable models in terms of B-Splines [Blake 93]. In this method, a deformable model is regarded as a linear combination of basis templates, and the state of the model is specified by a vector of coefficients for these templates. The mode leads naturally to a Kalman filter formulation in which the model is driven by an explicit local search for edges lying perpendicular to its boundary. These suggested movements are then used to update the model via the Kalman filter. Ivinns and Porrill suggested a similar approach but proposed an alternative to the Kalman filter using an explicit least-squares approximation [Ivins 98].

Numerous approaches and variations exist on the subject of object tracking but a recent development is that of CONDENSATION [Blake 98][Isard 98]. Blake and Isard presented the **Stochastic Conditional Density Propagation** (CONDENSATION) algorithm in which the location of a contour or object is probabilistically tracked over time using a model of the object's dynamics to predict movement. Objects are not represented by a single parameterisation but instead by a probability density function (PDF) which represents all possible parameterisations of the model. By generating multiple hypotheses from this distribution at each iteration, and checking each hypothesis against the image for supporting information, CONDENSATION allows objects to be tracked which exhibit discontinues movement in complex noisy scenes.