An Integrated Approach for Object Shape Registration and Modeling

Tomasz Adamek
Centre for Digital Video Processing
Dublin City University
Dublin 9, IRELAND
adamekt@eeng.dcu.ie

Noel E. O'Connor
Centre for Digital Video Processing
Dublin City University
Dublin 9, IRELAND
oconnorn@eeng.dcu.ie

Gareth Jones
Centre for Digital Video Processing
Dublin City University
Dublin 9, IRELAND
gjones@computing.dcu.ie

ABSTRACT

In this paper, an integrated approach to fast and efficient construction of statistical shape models is proposed that is a potentially useful tool in *Information Retrieval*(IR). The tool allows intuitive extraction of accurate contour examples from a set of images using a semi-automatic segmentation approach. The user is allowed to draw on the scene by simply dragging a mouse over the image and creating a set of labelled scribbles for the objects to be segmented. An automatic segmentation algorithm uses the scribbles to partition the scene and extract objects' contour. A set of labelled points (landmarks) is identified automatically on the set of examples thereby allowing statistical modeling of the objects' shape. The main contribution of this paper is the new approach to automatic landmark identification eliminating the burden of manual landmarking. The approach utilizes a robust method for pairwise correspondence proposed originally in [1, 2]. The landmarks are used to train statistical shape models known as Point Distribution Models (PDM) [11]. Qualitative results are presented for 3 classes of shape which exhibit various types of nonrigid deformation.

1. INTRODUCTION

Recently there has been significant interest in the content-based information retrieval community in object-based search and retrieval, where objects correspond to arbitrarily-shaped image segments that represent a semantic entity [31, 17]. However, object-based retrieval requires important underpinning technology for both object extraction (i.e. segmentation) and indexing in terms of features such as shape, color and texture. To date, segmentation and indexing have usually been performed independently. In this paper, we present an integrated approach for shape registration and modeling that is useful for generating shape models for different classes of objects that could subsequently be used for both segmentation and shape-based indexing in a coherent manner. In other words, we present a generic approach to generating models for particular classes of objects (e.g. human head and shoulders) that are useful for segmentation and recognition purposes and also as an effective basis for indexing. In this paper, we focus on the issues of model generation and not on their subsequent use in a retrieval context.

More than a decade ago Cootes and Taylor introduced one of the more influential ideas within the the image analysis community, the so-called *Active Shape Model*(ASMs) or 'Smart Snakes' [11]. Active Shape Models are deformable models with global shape constraints learned through observations. Objects are represented by a set of labelled points. Each point is placed on a particular part of the object. By examining the statistics of the position of the labelled points a 'Point Distribution Model' (PDM) is derived. The model gives the average position of the points, and a description of the main modes of variation found in the training set. The statistical analysis of shapes is widely applicable to many areas of image analysis including: analysis of medical images [7, 28], industrial inspection tools, modeling of faces, hands and walking people [12, 25]. The use of PDMs to automatically identify examples of the model object in unseen images and the relations of PDMs with other forms of flexible templates has been presented in [12]. For detailed studies on PDMs the reader can refer to the vast amount of literature available [15, 19, 7, 10, 28, 14].

Statistical shape modeling methods are based on examining the statistics of the coordinates of the labelled points over the training set. The only requirement is a labelled set of examples upon which to train the model. Correspondence is often established by manual annotation, which is subjective, labor intensive and for improved accuracy, intra and inter-annotator variability studies are required. The goal of this work is to develop algorithms necessary for fast and efficient construction of PDMs eliminating the burden of manual landmarking. In this way, the approach could easily be used for many different classes of objects in an IR context. The proposed approach is based on two key technologies: a semi-automatic segmentation approach for fast extraction of examples of closed contours from a set of images and a method for automatic identification of landmarks on the set of examples. The typical scenario would be a situation when the user collects several images containing instances of the object under consideration and constructs its statistical shape model, using as small a number of interactions as possible. The process would involve extraction of several object contours by the supervised segmentation. The tool would then automatically place a set of landmarks on each contour and build the required statistical shape model.

The remainder of this paper is organized as follows: The next section reviews briefly the previous work in automatic landmarking. Section 3 gives a short introduction to the field of statistical shape analysis by describing alignment of sets of points using Procrustes analysis and modeling of shape variation using PDMs. The technique used for supervised segmentation is discussed in section 4. The proposed landmarking framework is described in detail in section 5 and section 6 presents selected qualitative results. Possible alternative approaches and directions for future research in this area are discussed in section 7. Finally, conclusions are formulated in section 8.

2. PREVIOUS WORK

In the past, semi-automatic landmarking systems have been developed where the image containing a new example is searched using the model built from the current set of examples and the result can be edited before adding the new example to the training set. Although such approaches considerably reduce the required effort, many researchers recognize the need for fully automatic landmark placement.

The methods for finding correspondences across sets of shapes can be classified into two classes [10]: *pair-wise* methods [20, 5, 23] which perform a sequence of individual pairwise correspondences optimizing a pairwise function and *group-wise* methods [24, 14, 37, 16] which explicitly aim to optimize a function defined on the whole set of shapes.

An elegant solution was proposed in [33], where salient image features such as corners and edges are used to calculate a Gaussian-weighted proximity matrix measuring the distance between each feature. A Singular Value Decomposition (SVD) is performed on the matrix to establish a correspondence measure for the features and effectively finds the minimum least-squared distance between each feature. However, this approach is unable to handle rotations of more than 15 degrees and is generally unstable [10]. This method was further extended in [34] and [32]. Although, the above extensions provides good results on certain shapes, stability problems were reported [20] in cases when the two shapes are similar as well as an inability to deal with loops in the boundary. An interesting system for automated 2D shape model design by registering similar shapes, clustering examples and discarding outliers was described in [27].

Other methods [5, 23] assume that contours (usually closed) have already been segmented and use curvature information to select landmark points. However, they perform well only if the corresponding points lie on boundaries that have the same curvature. Hill et. al. proposed an alternative method for determining non-rigid correspondence between pairs of closed boundaries [20] based on generating sparse polygonal approximations for each shape. The landmarks were further improved by an iterative refinement step.

Since the above pair-wise methods may not find the best global solution, recently group-wise methods were proposed. In [24], landmarks are place on sets of closed curves using direct optimization. The approach is based on a measure of compactness and specificity of a model which is a function of the landmark position on the training set. Although in some cases the method produces results better than manual annotation, this is a large, nonlinear optimization problem and the algorithm does not always converge. Recently, Davies et.al. [13] proposed an objective function for group-wise correspondences of shapes based on the Minimum Description Length (MDL) principle. The landmarks are chosen so the data is represented as efficiently as possible. This approach was further extended in [37] by adding curvature measures. A steepest descent version of MDL shape matching was proposed in [16]. Currently, methods based on the MDL principle are seen by many as the state of the art solution: They are fully automatic and in many cases provide more meaningful models than manual annotations. However, the quality of the model directly relies on the choice of the objective function [13, 16] and optimization method which has to take into account many local minima. Often, convergence is slow and scales poorly with the number of examples. The duration of one model-evaluation-tuning loop can take hours or even days. Also, implementation and tuning of the MDL framework requires a lot of experimentation and parameter tweaking [21].

For a more detailed review of methods finding correspondences across contour sets one may refer to [14, 10, 9].

3. POINT DISTRIBUTION MODEL

This section aims at giving a brief introduction to the field of statistical shape analysis by describing two basic techniques: alignment of sets of points using Procrustes analysis and modeling of shape variation using PDMs.

3.1 Pairwise Alignment

According to the definition from [15], shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object. Therefore, equivalent points from different shapes can be compared only after establishing a reference pose (position, scale and rotation) to which all shapes are aligned.

A commonly used procedure for aligning corresponding point sets is *Procrustes Analysis*. The Procrustes analysis requires one-to-one correspondence of points often called *landmarks*[1], i.e. points which identify a salient feature on an object and which are present on every example of the class. Each planar shape is characterized by a set of n landmarks (common for the whole set of examples) and its vector representation is denoted as $\mathbf{x} = [x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n]^T$. Procrustes analysis is based on minimization of the Procrustes distance between two shapes, $\mathbf{x}_1$ and $\mathbf{x}_2$, which is simply the sum of the squared point distances:

$$P_d^2 = \sum_{j=1}^{n} [(x_{j1} - x_{j2})^2 + (y_{j1} - y_{j2})^2] \tag{1}$$

The alignment of two shapes $\mathbf{x}_1$ and $\mathbf{x}_2$ involves three steps:

```
1.Align position of the two centroids,
2.Re-scale each shape to have equal size,
3.Align orientation by rotation.
```

The centroid is defined as the center of a mass of the

[1]Synonyms for landmarks include homologous points, nodes, vertices, anchor points, markers, model points and key points.

physical system consisting of unit masses at each landmark:

$$(\bar{x}, \bar{y}) = \left(\frac{1}{n} \sum_{j=1}^{n} x_j, \frac{1}{n} \sum_{j=1}^{n} y_j \right) \quad (2)$$

Size normalization is performed using a scale metric called *centroid size*:

$$S(\mathbf{x}) = \sum_{j=1}^{n} \sqrt{(x_j - \bar{x})^2 + (y_j - \bar{y})^2)} \quad (3)$$

Orientation alignment is achieved by maximizing the correlation between the two sets of landmarks. The shape $\mathbf{x}_1$ is optimally superimposed upon $\mathbf{x}_2$ by performing a Singular Value Decomposition (SVD) of matrix $\mathbf{x}_1^T \mathbf{x}_2$ and using $\mathbf{VU}^T$ as the rotation matrix.

3.2 Generalized Procrustes Analysis

The alignment of a set of shapes is typically performed in an iterative[2] manner [7]:

```
1.Chose an initial estimate of the mean shape
  (e.g.the first shape in the set),
2.Align all the shapes to the mean shape,
3.Re-calculate the estimate of the mean from
  the aligned shapes,
4.If the mean has changed return to step 2.
```

The most frequently used estimate of the mean shape is the *Procrustes mean shape*:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \quad (4)$$

where N denotes the number of shapes.

It is important to fix the size and orientation at each iteration by normalization in order to avoid any shrinking or drifting of the mean shape.

3.3 Modeling Shape Variation

Once a set of aligned shapes is available the mean shape and variability can be found. In this work, we used statistical shape models known as Point Distribution Models (PDM) [11].

Consider the case of having N planar shapes consisting of n points, covering a certain class of shapes and all aligned into a common frame of reference. Training PDMs relies upon exploiting the inter-landmark correlation in order to reduce dimensionality. It involves calculating the mean of the aligned examples $\bar{\mathbf{x}}$ (according to equation 4), and the deviation from the mean of each aligned example $\delta \mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}}$, and calculating the eigensystem of the $2n \times 2n$ covariance matrix of the deviations $\sum_{\mathbf{x}} = 1/N \sum_{i=1}^{N} (\delta \mathbf{x}_i)(\delta \mathbf{x}_i)^T$. The modes of variation are described by $\mathbf{p}_k$ $(k = 1, \ldots, 2n)$, the unit eigenvectors of $\sum_{\mathbf{x}}$ such that:

$$\sum_{\mathbf{x}} \mathbf{p}_k = \lambda_k \mathbf{p}_k \quad (5)$$

where λ_k is the k^{th} eigenvalue of $\sum_{\mathbf{x}}$ and $\lambda_k \geq \lambda_{k+1}$. This results in an ordered basis where each component is ranked according to variance. Modifying one component at a time gives the *principal modes* of variation. Any shape in the training set can be approximated using the mean shape and a weighted sum of these deviations obtained from the first t modes:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb} \quad (6)$$

where $\mathbf{P} = (\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_t})$ is the $2n \times t$ matrix of the first t eigenvectors and $\mathbf{b} = (b_1, b_2, \ldots, b_t)^T$ is a t element vector of weights (shape parameters), one for each eigenvector. New examples of shapes can be generated by varying the parameters $\mathbf{b}$ within suitable limits, so the new shapes will be similar to those in the training set. t is typically chosen as the smallest number of modes such that the sum of their variances explain a sufficiently large proportion of the total variance of all the variables.

The modes are often comparable to those a human would select to design a parameterized model. However, as explained in [10], they may not always separate shape variation in an obvious manner since they are derived directly from the statistics of a training set.

4. SEMI-AUTOMATIC SEGMENTATION

In the proposed approach, contour examples are extracted using a semi-automatic segmentation approach. Since a typical model requires tens of examples, special consideration has been devoted to ensure easy and intuitive user interactions. The user is allowed to draw on the scene by simply dragging a mouse over the image creating a set of labelled *scribbles* for the object and its background[3] – similar to the interaction scheme proposed in [8, 29]. Two or more different scribbles can have the same label, indicating different parts of the same object. Each time a new drag is added an automatic process uses the scribbles to produce partitioning of the image.

An elegant and efficient solution to image partitioning based on scribbles was proposed in [30] where the input image is pre-segmented into regions and represented as Binary Partition Tree (BPT). The tree structure is used to encode similarities between regions pre-computed during the automatic segmentation. The process starts by assigning the labels from scribbles to the leafs of the tree. Then, the labels are propagated to remaining nodes(regions). A label associated to a node is propagated to its sibling and parent only if none of the sibling's descendant has been assigned to a different label. This results in the creation of the zones of influence(connected subtrees) for the two types of labels. Also, a certain number of nodes remain without labels, judged as being "too different" with respect to the regions defined by the scribbles.

In our implementation of the above method, the BPT was created using the automatic segmentation approach described in [3]. In this approach, the segmentation is performed using Recursive Shortest Spanning Tree(RSST) algorithm with additional incorporation of the so-called syntactic features[6] which, in most cases, lead to more meaningful partitions. The complete BPT, starting from the level of pixels, is used to ensure that every part of the image can be split if required by the user. Additionally, in order to simplify the user interactions, the labelling algorithm was extended to guarantee that all parts of the image are classified as part of a known object (foreground or background).

[2]An analytic solution can be found in [15].

[3]Left and right mouse buttons are associated with foreground and background respectively.

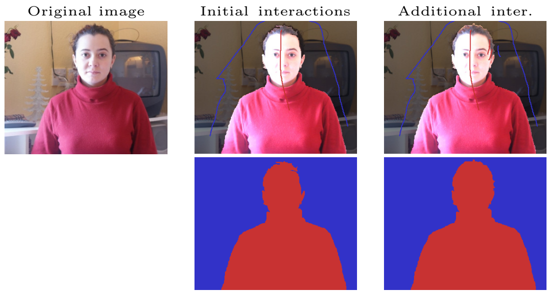


Figure 1: Typical user interactions required for contour extraction.

The above approach has proven to be extremely practical and efficient. The pre-segmentation of a CIF (352x288) image takes under 3 seconds on a standard PC with Pentium III 600 MHz processor. The segmentation mask is updated in real-time whenever a new scribble is added to the original image. The typical time spent by the user obtaining a required segmentation result is between 5-10 seconds. However this may depend on complexity and size of the object under consideration and also in some cases on the artefacts introduced by compression of the input image. An example of typical user interactions and corresponding results are shown in Figure 1.

In a retrieval context, this semi-automatic segmentation approach could be used for both off-line shape template construction and on-line object extraction for query formulation.

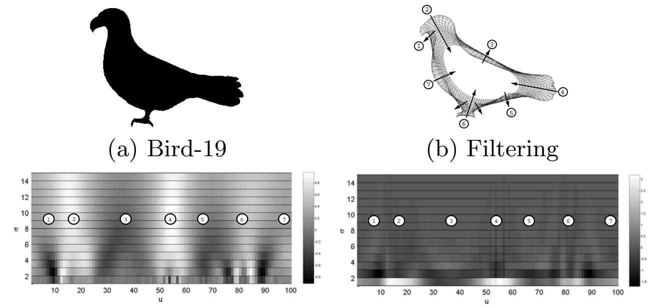
5. LANDMARK IDENTIFICATION

The strategy used in the framework for automatic identification of landmarks on a set of contours is similar to the scheme proposed by Fleute and Lavalée for landmarking sets of closed 3D surfaces [18]. Each training example is initially matched to a single reference template and a mean is built from these matched examples. Then, iteratively, each example is matched to the current mean and the procedure is repeated until convergence.

5.1 Pair-wise Matching

The above framework relies upon the ability to match pairs of shapes (in order to match each shape with the mean) and to measure the quality of the match (in order to decide the initial reference shape). In other words, the mean cannot be built and subsequently updated if there is no dense correspondence between contour points of the whole set. In order for this process to be successful, an accurate, robust method of pairwise correspondence is required. It is also required that the pairwise matching is invariant in respect to pose (translation, scale and rotation). Also, the method has to perform well in cases where the two boundaries represent different examples from the same class of objects and a nonrigid transformation is required to map one boundary onto the other.

In this paper, the usefulness of a technique proposed originally by the authors in [1, 2] for comparing general shapes will be studied. In this approach, a rich shape descriptor, termed Multi-scale Convexity Concavity (MCC) representation, stores information about the amount of convexity/concavity at different scale levels for each contour point (Figure 2(b)). A Dynamic Time Warping (DTW) technique [26] is used to find an optimal refined alignment along the contours upon which the dissimilarity measure is defined. The approach is robust to several transformations including translation, scaling, rotation, modest occlusion and symmetric transformation. The method is particulary attractive for the landmarking framework as it has been demonstrated that it performs well in cases of elastic deformations and where the similarity between curves is weak [1]. An example of pair-wise matching is shown in Figure 3. A more detailed description of the MCC representation and the associated matching algorithm can be found in [1].



(a) Bird-19 (b) Filtering

(c) MCC (d) MCC-DCT

Figure 2: Extraction of MCC representation.

For the purpose of this work, a MCC-DCT version [1, 2] of the contour representation was used. In this approach, a one dimensional Discrete Cosine Transform (DCT) is applied to each multi-scale contour point feature vector de-correlating information from different scale-levels and placing most of the energy in low frequency coefficients –see example from Figure 2d. MCC-DCT allows utilization of an iterative optimization framework for determining the relative proportions in which DCT coefficients should be combined in the final similarity measure. The matching procedure was optimized using the shape collection from the MPEG-7 Core Experiment "CE-Shape-1" (part B) [22].

It should be noted that the matching algorithm requires the two contours to be represented by an equal number of equidistant points. This will lead to an additional resampling step in the landmarking algorithm. The extension of the matching algorithm to a non-uniformly sampled contour is straightforward and will be implemented in the future. Also, the landmarking scheme requires dense correspondence (ideally between every pixel). To allow such fine matching without increasing the computational cost, faster versions of the matching algorithm will be developed in the future and reported elsewhere.

5.2 Correspondence for a Set of Curves

Before the landmarking process begins, all examples are down-sampled and represented by a fixed number of equidistant contour points (300 in all presented experiments). Then, the position and size of the contours are normalized according to equations 2 and 3. Next, an initial reference contour is selected from the set in such a way that the total value of dissimilarities between the reference and all other examples is minimized. This ensures that the initial reference is chosen close to the center of the modelled shape space. Subsequently, each example is matched to the reference us-

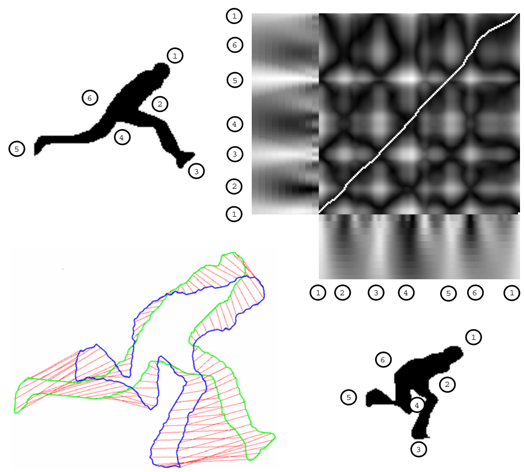


Figure 3: Matching example. Finding the optimal match between two contour representations corresponds to finding the lowest cost diagonal path through the table containing pairwise distances between contour point features.

ing the pairwise matching method described in section 5.1. This provides a common basis for identification of points from all examples. In other words, a contour point in any of the examples is identified by the point from the reference contour to which it is best matched. This allows optimal rotation of examples to superimpose with the reference (as explained in section 3.1) and finally construction of the mean contour according to equation 4. If the distance between the current reference and the newly estimated mean, computed using equation 1 and normalized by the centroid size of the reference, exceeds a predefined threshold the mean is taken as the new reference and the procedure is repeated. As a result, a dense correspondence between each example and the mean is obtained. Using these dense correspondences the landmarks placed on the mean shape can be projected back to each example. The main steps of the proposed framework are outlined below:

```
1.Sample uniformly all examples.
2.Normalize all examples:
  a)Translate centroids to position (0,0);
  b)Re-scale to unit size.
3.Select the initial reference contour.
4.Find correspondences between examples
  and the reference.
5.Rotate each example to superimpose with
  the reference.
6.Create a mean contour according to Eq. 4.
7.If not converged take the mean as
  the new reference and go to step 4.
8.Generate landmarks in the reference contour.
9.Project landmarks back to each example.
```

To avoid any drifting of the mean shape its orientation is adjusted at each iteration by optimal superposition with the initial reference. Additionally, the mean contour is resampled before a new iteration begins to ensure equidistance of its contour points. This additional step could be eliminated by adapting the pairwise matching algorithm to match non-uniformly sampled contours.

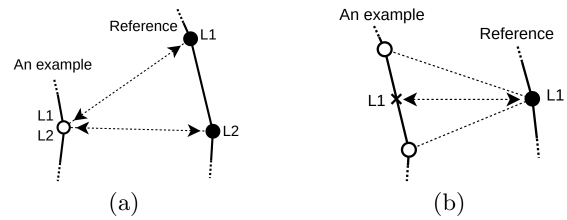


Figure 4: Updating of the mean and landmark propagation in cases of multiple assignments of points.

The objective of step 4 is to identify, on each example, positions corresponding best to the contour points from the reference mean. However, the matching algorithm must compensate for nonrigid transformations as well as small differences in the position of contour points caused by sampling. Moreover, the matching is symmetrical – both contours (an example and the reference mean) are treated equally. Therefore, a single contour point from an example can be matched with two reference points from the mean, and vice versa, a reference point from the mean can be matched with two points from an example. The first case, depicted in Figure 4a, does not require any special treatment. The position of the single point from the considered example can be used to update the positions of both matched points from the reference, e.g. to construct the new mean. Also, during stage 8, if both reference points from the mean are selected to become landmarks, the position of the single point can be used for both of them. In contrast, the second case requires special consideration since the single landmark cannot be used for two different points. Therefore, only the center of the segment connecting both points is used to update the position of the matched reference point and can potentially become a landmark – see Figure 4b.

The number of contour points should be considerably large according to the required precision in localization of the final landmarks. For the maximum precision the number of contour points should be comparable with the average number of pixels in each example.

The set of landmarks can be generated automatically on the mean using any sensible method, for example the approach based on detection of *Critical Points* presented in [38]. In the current implementation all points from the mean are selected as landmarks. However, choosing the extrema of convexity/concavity measure from the MCC representation as major landmarks (due to high matching reliability at those points) and placing the minor landmarks equally between them could further improve specificity of final PDMs.

The selection of the initial reference contour requires computation of pairwise dissimilarities between all examples. Since a coarse measure of the dissimilarity is sufficient for the selection, the computational load of this step can be reduced by using a reduced number of contour points.

5.3 Modeling Symmetrical Shapes

Often one would like to model object shapes which exhibit a mirror symmetry, e.g. face or head & shoulder. In such cases, extending the training set by mirrored versions of the boundaries usually leads to more intuitive modes of variations – especially if there is a discrepancy between the numbers of original examples from each view (e.g. more faces looking to the left than to the right) and the number of original examples is small. Clearly such an approach does not require any modification of the proposed landmarking method, however it implies additional cost of matching of these additional examples with the mean. An alternative strategy requires adaptation of steps 4-6 of the proposed framework. In such a case, each iteration begins by matching only the original set to the mean. Based on this match, a new mean and its mirrored version are created. Then, both versions of the mean (original and mirrored) are matched and their correspondence is used to establish correspondences between the mirrored set and the mean obtained from the original set. Finally, the new symmetrical version of the mean, including original and mirrored sets, is created. This approach avoids the need for individual matching of mirrored examples, explicitly ensures that each mirrored boundary is matched with the reference in exactly the same way as its original version[4], and further improves convergence.

6. RESULTS

This section presents qualitative results of training PDMs for 3 classes of shape exhibiting various types of nonrigid deformation. In all three cases, the supervised segmentation of the set took less than 5 minutes. The landmarking process converged after 2-5 iterations in all cases, requiring typically less than 30 seconds on a standard PC.

6.1 Cross-sections of Pork Carcasses

The data in this experiment consists of 14 gray scale images (768x576 pixels) [35]. An example image is shown in Figure 5a. The results are comparable with the manual annotations presented in [35]. It can be observed from Figure 5b that the model is compact (the three first eigenvectors explain more than 77% of all variations) and qualitative results, shown in Figure 5c, indicate good specificity [5].

6.2 Head & Shoulders

The data in this experiment consists of 19 color images of CIF size. Example images (with extracted object boundaries) are shown in Figure 6a. The training set was extended using mirrored versions of the examples as described in section 5.3. The final model is compact (three first eigenvectors explain more than 76% of all variations) and qualitative results shown in Figure 6c indicate good specificity.

6.3 Synthetic Example

In this experiment we use synthetic data designed to "stress test" the model, in order to illustrate some potential limitations of the proposed approach. The training set shown in Figure 7 is a synthetic "Bump" example introduced and discussed originally in [14]. The examples were generated from a synthetic object that exhibits a single mode of shape variation where the "bump" moves along the top of the box. Therefore, it should be possible to describe the variation with a single parameter.

Qualitative results for our approach are shown in Figure 7. The landmarking algorithm failed to find "ideal" correspondences for the set as some of the "bump" end-points lie on

[4]with the precision limited by the contour sampling.

[5]A specific model should only generate instances of the object class that are similar to those in the training set.

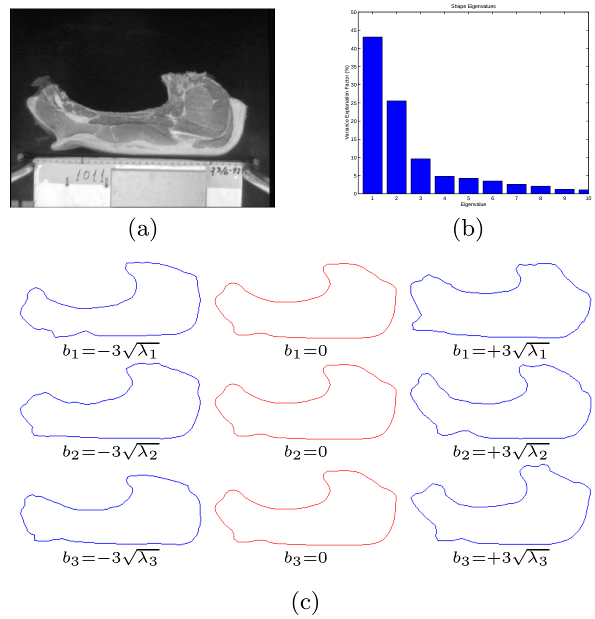


Figure 5: Pork carcasses: (a) Example image of pork carcass cross-section, (b) Eigenvalues, (c) Deformation using 1st, 2nd and 3rd principal mode.

boundaries with concavity while others correspond to convex parts. The matching algorithm warped the parts with different curvature attempting to find optimal global correspondence. The poor localization of the end-points of the "bump" for some examples compromised the specificity of the model as shown in Figure 7b. It should be noted however, that the first eigenvector still describes almost 98% of the total model variations. This example suggest that making the approach fully functional (applicable to all classes of shape deformations) would require allowing manual corrections of some landmarks.

7. DISCUSSION

The main limitation of the approach is related to the fact that the matching relies on the assumption that the corresponding points lie on boundaries that have similar values of curvature (convexity/concavity). As shown in section 6.3, this may compromise the compactness and specificity of the final model. One solution would be to allow the user to correct the position of misplaced landmarks. Alternatively, a MDL-based method would have to be used as it does not rely on pairing any boundary features. Ultimately, the proposed method could provide initial correspondences which could be then refined using a MDL-based technique. This could reduce the sensitivity of the MDL method to the search parameters and reduce the time required for convergence. Also, neither MDL methods or the proposed method are suitable for objects represented by multiple open/closed boundaries, e.g. face expressions. Addressing this problem will be part of our future research. One approach would be to integrate a method based on *Shape Context* [4] or its recently proposed extension with a figural continuity constraint [36].

The extension of the proposed approach to open curves would require extension of the user interface to allow extraction of open curves (e.g. by marking end-points of the

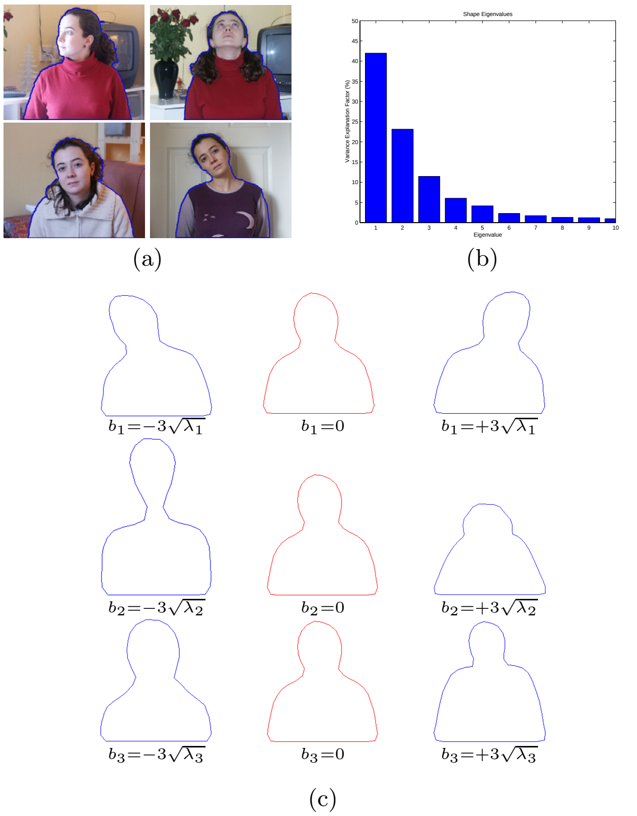


Figure 6: (a) Examples of contours extracted using user-driven segmentation tool for the head & shoulders model, (b) Eigenvalues, (c) Deformation using 1st, 2nd and 3rd principal mode.

closed curves) and a relatively straightforward modification of the pairwise matching algorithm. In fact, the matching problem could be simplified by assuming that the end-points of the curves correspond. Summarizing, the tool would have to integrate multiple modules allowing different types of interaction and generated models for full flexibility in different applications.

Our future work will include utilization of models of human Head & Shoulder for detection of close-up shots for automatic generation of highlights of TV sports programs.

8. CONCLUSIONS

The proposed approach allows rapid creation of statistical shape models for classes of objects represented by single closed boundaries. The tool allows intuitive semi-automatic segmentation of objects' examples from a set of images, automatically identifies correspondences for the extracted set of contours and creates Point Distribution Models. The proposed landmarking method is fast and does not require any tweaking of the parameters. In all presented examples, it took less than 5 minutes for an unexperienced user to segment and create a final model.

9. ACKNOWLEDGMENTS

The support of the Enterprise Ireland Informatics Initiative is gratefully acknowledged. Part of this work was supported by Science Foundation Ireland under grant 03/IN.3/I361.

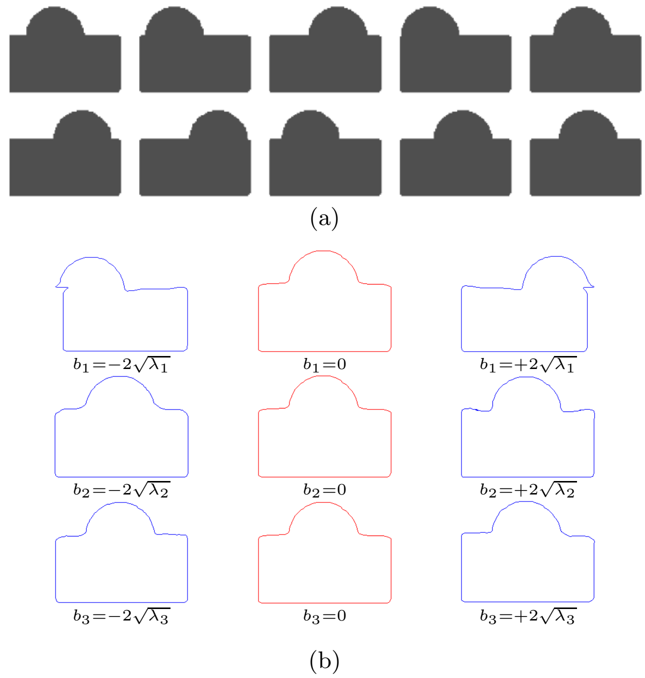


Figure 7: "Bump": (a) The "Bump" training set, (c) Deformation using 1st, 2nd and 3rd principal mode.

Dr. Hans Henrik Thodberg, Danish Meat Research Institute, provided the meat and Dr. Nette Schultz recorded the images.

10. ADDITIONAL AUTHORS

Noel Murphy, Centre for Digital Video Processing, Dublin City University, IRELAND, email: `murphyn@eeng.dcu.ie`

11. REFERENCES

[1] T. Adamek and N. E. O'Connor. A multi-scale representation method for non-rigid shapes with a single closed contour. *Special Issue on Audio and Video Analysis for Multimedia Interactive Services in IEEE Transactions on Circuits and Systems for Video Technology*, 5 2004.

[2] T. Adamek, N. E. O'Connor, and N. Murphy. Multi-scale representation and optimal matching of non-rigid shapes. In *submitted to CBMI 2005*, 2005.

[3] T. Adamek, N. E. O'Connor, and N. Murphy. Region-based segmentation of images using syntactic visual features. In *WIAMIS*, 2005.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intell.*, 24(4):509–522, April 2002.

[5] A. Benayoun, N. Ayache, and I. Cohen. Human face recognition: From views to models - from models to views. In *Int'l Conf. on Pattern Recognition*, pages 225–243, September 1994.

[6] C. F. Bennstrom and J. R. Casas. Binary-partition-tree creation using a quasi-inclusion criterion. In *IEEE Computer Society Press, in the*

proceedings of the Eighth International Conference on Information Visualization (IV). London, UK, 2004.

[7] F. L. Bookstein. Landmark methods for forms without landmarks: localizing group differences in outline shape. *Medical Image Analysis*, 1(3):225–244, 1997.

[8] E. Chalom and V.M.Bove. Segmentation of an image seguence using multi-dimensional image attributes. In *IEEE Int'l Conf. on Image Processing, ICIP'96, Lausanne*, volume 2, pages 525–528, September 1996.

[9] T. Cootes. Timeline of developments in algorithms for finding correspondences across sets of shapes and images. Technical report, University of Manchester, `http://www.isbe.man.ac.uk/~bim`, June 2004.

[10] T. Cootes and C.J.Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, `http://www.isbe.man.ac.uk/~bim`, March 2004.

[11] T. F. Cootes and C. J. Taylor. Active shape models - smart snakes. In *Proc. British Machine Vision Conf., Springer Verlag*, page 266, 1992.

[12] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 95.

[13] R. Davies, T. Cootes, C. Twining, and C. Taylor. An information theoretic approach to statistical shape modelling. In *12th British Machine Vison Conf., BMVA Press*, pages 3–11, September 2001.

[14] R. H. Davies. *Learning Shape: Optimal Models for Analysing Shape Variability*. PhD thesis, University Of Manchester, 2002.

[15] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.

[16] A. Ericsson and K. Åström. Minimizing the description length using steepest descent. In *14th British Machine Vision Conf., BMVA Press*, volume 2, pages 93–102, 2003.

[17] B. Erol and F. Kossentini. Shape-based retrieval of video objects. *IEEE Trans. Multimedia*, 7(1):179–182, February 2005.

[18] M. Fleute and S. Lavallée. Building a complete surface model from sparse data using statistical shape models: Application to computer assisted knee surgery system. In *MICCAI-Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*, pages 879–887, 1998.

[19] C. Goodall. Procrustes methods in the statistical analysis of shape. *Jour. Royal Statistical Society, Series B*, 53:285–339, 1991.

[20] A. Hill and C. J. Taylor. A method of non-rigid correspondence for automatic landmark identification. In *7th British Machine Vison Conf., BMVA Press*, pages 323–332, September 1996.

[21] J. Hladuvka. Establishing point correspondence on training set boundaries. Technical Report 2003-040, VRVis Technical Report # 2003-040, 2003.

[22] S. Jeannin and M. Bober. Description of core experiments for mpeg-7 motion/shape. *MPEG-7, ISO/IEC/JTC1/SC29/WG11/MPEG99/N2690, Seoul*, March 1999.

[23] C. Kambhamettu and D. Goldgof. Point correspondence recovery in non-rigid motion. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 222–227, September 1992.

[24] A. C. W. Kotcheff and C. J. Taylor. Automatic construction of eigenshape models by direct optimisation. *Medical Image Analysis*, 2(4):303–314, September 1998.

[25] J. Marchant and C. Onyango. Fitting grey level point distribution models to animals in scenes. *Image and Vision Computing*, 13(2):3–12, February 1995.

[26] C. M. Myers, L. R. Rabiner, and A. E. Rosenberg. Performance tradeoff in dynamic time warping algorithms for isolated word recognition. *IEEE Trans. Acoust. Speech Signal Processing*, ASSP-28(12), 1980.

[27] N.Duta, A. Jain, and M. Dubuisson-Jolly. Automatic construction of 2d shape models. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(5):433–446, 2001.

[28] A. Neumann and C. Lorenz. Statistical shape model based segmentation of medical images. *Computerized Medical Imaging and Graphics*, 22(2):133–143, 1998.

[29] N. E. O'Connor. *Video Object Segmentation for Future Multimedia Applications*. Phd, Dublin City University, School of Electronic Engineering, 1998.

[30] P. Salembier and F. Marqués. Region-based representations of image and video. *IEEE Trans. Circuits Syst. Video Technol.*, 9(8):1149–1167, December 1999.

[31] S. Sav, H. Lee, A. F. Smeaton, and N. E. O'Connor. Using segmented objects in ostensive video shot retrieval. In *3rd Int'l Workshop on Adaptive Multimedia Retrieval*, July 2005.

[32] S. Sclaroff and A. P. Pentland. Modal matching for correspondence and recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(6):545–561, 1995.

[33] G. Scott and H. Longuet-Higgins. An algorithm for associating the features of two images. *Phil. Trans. Roy. Soc. London*, B 244:21–26, 1991.

[34] L. S. Shapiro and J. M. Brady. A modal approach to feature-based correspondence. In *P. Mowforth, editor, 2nd British Machine Vison Conference, Springer-Verlag*, pages 78–85, September 1991.

[35] M. B. Stegmann. Active appearance models: Theory, extensions and cases. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, `http://www.imm.dtu.dk/~aam/`, 2000.

[36] A. Thayananthan, B. Stenger, P. Torr, and R. Cipolla. Shape context and chamfer matching in cluttered scenes. In *Conf. Computer Vision and Pattern Recognition, Madison, USA*, June 2003.

[37] H. Thodberg and H. Olafsdottir. Adding curvature to minimum description length shape models. In *14th British Machine Vision Conf., BMVA Press*, volume 2, pages 251–260, 2003.

[38] P. Zhu and P. Chirlian. On critical point detection of digital shapes. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(8):737–748, August 1995.