Parameterisation Invariant Statistical Shape Models

Johan Karlsson, Anders Ericsson, and Kalle Åström Centre for Mathematical Sciences, Lund University

Abstract

In this paper novel theory to automate shape modelling is described. The main idea is to develop a theory that is intrinsically defined for curves, as opposed to a finite sample of points along the curves. The major problem here is to define shape variation in a way that is invariant to curve parameterisations. Instead of representing continuous curves using landmarks, the problem is treated analytically and numerical approximations are introduced at the latest stage. The problem is solved by calculating the covariance matrix of the shapes using a scalar product that is invariant to global reparameterisations.

An algorithm for implementing the ideas is proposed and compared to a state of the art algorithm for automatic shape modelling. The problems with instability in earlier formulations are solved and the resulting models are of higher quality.

1. Introduction

The basic idea behind statistical models of shape is to from a given training set of known shapes be able to describe new formerly unseen shapes, which still are representative. The shape is traditionally described using landmarks on a geometric object. A major drawback of this approach is that a dense correspondence between the shapes must be established. In practice this has been done by hand. A process that commonly is both time consuming and error prone.

There have been many suggestions on how to automate the process of building shape models, or more precisely, finding a dense correspondence among a set of shapes [9]. Many have stated the correspondence problem as an optimisation problem [1, 2, 4, 8, 10, 12, 6, 7, 5].

Minimum Description Length or MDL [11] is a paradigm that has been used in many different applications. In recent papers [3, 4, 7] this paradigm is used to locate a dense correspondence between the boundaries of shapes.

A problem when locating these correspondences, by optimizing over parameterisations, arises when continuous curves are represented by a finite sets of landmarks. Correspondences of landmarks do not necessarily imply correspondence of the entire curves. If many landmarks are placed in one point or a small region, the cost function measuring the correspondence may get a low value, but this value is based on a small part of the curve.

It has been suggested that this can be prevented by using a "master example", i.e. an example that is not reparameterised during optimisation. The idea is that each curve will be reparameterised to fit the "master example". This strategy breaks down if there are too many curves in the training set. A cost for bad reparameterisations can also be applied as suggested in [13]. The cost penalises that the parameterisations move in the same direction on all curves. The mean movement of corresponding landmarks should be zero during optimisation.

In this paper it is shown that the infimum of the description length (DL) in the standard formulation is zero. A new scalar product, which always considers the entire curves and is invariant to global reparameterisations, is defined. Using the new scalar product when calculating the covariance matrix makes the optimisation well defined and the infimum of the DL not equal to zero. With this formulation it is no longer an advantage to place all landmarks at one point or in a small region.

2 Preliminaries

When analysing a set of n_s similar shapes, it is convenient and usually effective to describe them using Statistical Shape Models. After the shapes \mathbf{x}_i $(i = 1 \dots n_s)$ have been aligned and normalised to the same size, a PCA-analysis of the covariance matrix for the shapes is performed. The *i*-th shape in the training set can now be described by a linear shape model of the form,

$$\mathbf{x}_i = \bar{\mathbf{x}} + \mathbf{\Phi} \mathbf{b},$$

where $\bar{\mathbf{x}}$ is the mean shape, the columns of Φ describe a set of orthogonal modes of shape variation and \mathbf{b}_i is the vector of shape parameters for the *i*-th shape.

Assume that a population of geometric objects, represented as continuous curves $\mathbf{c}_i(t), i = 1, \ldots, n_s, t \in [0, 1]$, is given and that the shape is to be modelled.

Each curve is represented using some arbitrary parameterisation. For simplicity it is assumed that they are parameterised by arc length,

$$\mathbf{c}_i: [0,1] \ni s \to \mathbf{c}_i \in \mathbf{R}^2$$

To model the shape it is necessary to solve the correspondence problem, i.e. to find parameterisations $\{\gamma_i\}_{i=1}^{n_s}$, where $\gamma_i : [0,1] \rightarrow [0,1]$ are strictly increasing functions so that $\mathbf{c}_i(\gamma_i(s))$ corresponds to $\mathbf{c}_j(\gamma_j(s))$ for all pairs (i,j)and all parameter values $s \in [0,1]$.

Correspondence between the curves $\mathbf{c}_i(\gamma_i(s))$ and $\mathbf{c}_j(\gamma_j(s))$ is denoted $\mathbf{c}_i(\gamma_i(s)) :=: \mathbf{c}_j(\gamma_j(s))$.

MDL has proven to be a successful algorithm for locating the parameterisation functions γ_i . The cost in MDL is derived from information theory and is, in simple words, the effort that is needed to transmit the model and the training shapes the model approximates bit by bit. The MDL - principle searches iteratively for the set of functions $\{\gamma_i\}_{i=1}^{n_s}$ that gives the cheapest model to transmit. The cost function can be interpreted as a tradeoff between a model that is general (can represent any instance of the object) and compact (can represent the variation with as few parameters as possible).

The goal of this paper is to derive a shape theory that is intrinsically defined for curves and is independent of parameterisations. It is reasonable to use a linear model, i.e.

$$\mathbf{c}_i(\gamma_i(s)) = \bar{\mathbf{c}}(s) + \sum_{k=1}^{n_m} \mathbf{b}_{i,k} \mathbf{\Phi}_k(s) = \bar{\mathbf{c}} + \mathbf{\Phi} \mathbf{b}_i$$

One way of determining Φ from experimental data $\{\mathbf{c}_i\}_{i=1}^{n_s}$, where the curves $\{\mathbf{c}_i\}_{i=1}^{n_s}$ have been aligned according to the Procrustes condition (similarity transformations), is to make a singular value decomposition of the matrix $\mathbf{C}_0 = \frac{1}{n_s-1}\mathbf{X}\mathbf{X}^T$, where $\mathbf{X} = [\mathbf{c}_1 - \bar{\mathbf{c}}, \dots, \mathbf{c}_{n_s} - \bar{\mathbf{c}}]$. This is straightforward in the case of finite point configurations, but for curves it is more difficult since the 'column vectors' are infinite dimensional. However, the covariance matrix $\mathbf{C} = \frac{1}{n_s-1}\mathbf{X}^T\mathbf{X}$ is finite dimensional and has the same singular values. The element c_{ij} is a scalar product of the curve $(\mathbf{c}_i - \bar{\mathbf{c}}) \circ \gamma_i$ with the curve $(\mathbf{c}_j - \bar{\mathbf{c}}) \circ \gamma_j$. The standard scalar product is

$$c_{ij} = \int_0^1 (\mathbf{c}_i(\gamma_i(s)) - \bar{\mathbf{c}}(s)) \cdot (\mathbf{c}_j(\gamma_j(s)) - \bar{\mathbf{c}}(s)) ds \quad . \tag{1}$$

A possible criterion for solving the correspondence problem ($\mathbf{c}_i(\gamma_i(s)) :=: \mathbf{c}_j(\gamma_j(s))$) for all i,j) is the description length of the shape model. The reparameterisation functions $\{\gamma_i\}_{i=1}^{n_s}$ are located by minimizing the description length.

Definition 1. The description length of a shape model \mathcal{M} is

$$DL(\mathcal{M}) = \sum_{\lambda_i \ge \lambda_c} (1 + \log \frac{\lambda_i}{\lambda_c}) + \sum_{\lambda_i < \lambda_c} \frac{\lambda_i}{\lambda_c} + K ,$$

where the scalars λ_i are the eigenvalues of the covariance matrix **C**, the scalar λ_c is a cutoff constant and K is a scalar, which is independent of the parameterisations. The constant K can be ignored during optimisation and the following cost function is used,

$$F(\mathbf{C}) = \sum_{\lambda_i \ge \lambda_c} (1 + \log \frac{\lambda_i}{\lambda_c}) + \sum_{\lambda_i < \lambda_c} \frac{\lambda_i}{\lambda_c} \quad . \tag{2}$$

3 Difficulties with Parameterisation Dependent Methods

One problem with optimising the description length as presented above is that it is not independent of curve parameterisations. By changing all parameterisations γ_i with a common parameterisation γ , so that $\tilde{\gamma_i} = \gamma_i \circ \gamma$, one effectively puts different weights at different parts of the curve. The same problem occurs when a discrete set of landmarks is used as in [3, 4]. The most weight is put on that part of the curve which has the most landmarks. The problem with this is that the landmark placement problem becomes ill-defined as illustrated by the following theorem.

Theorem 1. Using criterion (2),

$$\min_{\gamma_1,\ldots,\gamma_n} F(\mathbf{C}) = 0$$

The infimum is attained in the limit when all weight is put at a single point.

Proof. Assume that all landmarks are placed at one point on each curve. Then the covariance matrix \mathbf{C} will be the zero matrix. Hence all eigenvalues of \mathbf{C} are zero and $F(\mathbf{C}) = 0$.

This shows that the minimisation criterion is not well defined. A global optimiser could find the global minimum. Even if the global minimum is not found it is still possible to reduce $F(\mathbf{C})$ by concentrating the landmarks on parts of the curves with low variance. This means that it is possible to reduce the DL in two ways. Both by finding correspondences and by performing a global reparameterisation. By global reparameterisation is meant a reparameterisation function that is applied to all curves in the training set. Such a global reparameterisation function can put all the weight at one point or a small section of the curve. Therefore any algorithm for automatic landmark placement based on (2) tends to gather points together.

In order to avoid this several authors have presented preliminary solutions such as keeping the parameterisation of the first curve unchanged or penalising bad reparameterisations. But even such methods have difficulties. For example, if one keeps the parameterisation of the first curve unchanged it is still possible to reparameterise the other n-1 curves so as to put more weight to a particular point and ignore the mismatch with the first curve.

A new criterion based on invariance to global reparameterisations would be desirable.

4 A Parameterisation Invariant Method

To improve the algorithm discussed above, a new scalar product, which is invariant under global reparameterisations is proposed. This removes the undesired way to reduce the DL.

Definition 2. Let $\{\mathbf{c}_i\}_{i=1}^{n_s}$ be the curves with the mean curve subtracted and let them be parameterised with $\{\gamma_i\}_{i=1}^{n_s}$. A new scalar product is defined by

$$\mathbf{c}_i \cdot \mathbf{c}_j = \frac{1}{n_s} \sum_{k=1}^{n_s} \int_0^1 \mathbf{c}_i \left(\gamma_i \circ \gamma_k^{-1}(s) \right) \mathbf{c}_j \left(\gamma_j \circ \gamma_k^{-1}(s) \right) ds$$

Intuitively, what happens is that the weight is distributed by what could be seen as a sort of mean arc length parameterisation.

This also means that the entire curves are considered, since it can be seen that there is one term that gives arclength parameterisation for c_i and one term that gives arclength parameterisation for c_j .

Theorem 2. The scalar product in definition 2 is invariant under global reparameterisations.

Proof. Let $\{\gamma_i\}_{i=1}^{n_s}$ be a set of parameterisation functions such that $\mathbf{c}_i(\gamma_i(s)) :=: \mathbf{c}_j(\gamma_j(s))$. Let γ be an arbitrary reparameterisation function and let $\tilde{\gamma}_i(s) = \gamma_i \circ \gamma$ and $\tilde{\gamma}_j(s) = \gamma_j \circ \gamma$. Then $\mathbf{c}_i(\tilde{\gamma}_i(s)) :=: \mathbf{c}_j(\tilde{\gamma}_j(s))$ still holds.

$$\mathbf{c}_{i}(\tilde{\gamma}_{i}(s)) \cdot \mathbf{c}_{j}(\tilde{\gamma}_{j}(s)) =$$

$$\frac{1}{n_{s}} \sum_{k=1}^{n_{s}} \int_{0}^{1} \mathbf{c}_{i} \left(\tilde{\gamma}_{i} \circ \tilde{\gamma}_{k}^{-1}(s) \right) \mathbf{c}_{j} \left(\tilde{\gamma}_{j} \circ \tilde{\gamma}_{k}^{-1}(s) \right) ds =$$

$$\frac{1}{n_{s}} \sum_{k=1}^{n_{s}} \int_{0}^{1} \mathbf{c}_{i} \left(\gamma_{i} \circ \gamma \circ \gamma^{-1} \circ \gamma_{k}^{-1}(s) \right) \cdot$$

$$\cdot \mathbf{c}_{j} \left(\gamma_{j} \circ \gamma \circ \gamma^{-1} \circ \gamma_{k}^{-1}(s) \right) ds =$$

$$\frac{1}{n_{s}} \sum_{k=1}^{n_{s}} \int_{0}^{1} \mathbf{c}_{i} \left(\gamma_{i} \circ \gamma_{k}^{-1}(s) \right) \mathbf{c}_{j} \left(\gamma_{j} \circ \gamma_{k}^{-1}(s) \right) ds =$$

$$\mathbf{c}_{i}(\gamma_{i}(s)) \cdot \mathbf{c}_{j}(\gamma_{j}(s)) \quad .$$

The scalar product would actually be invariant to global parameterisation using only one term in the sum in Definition 2. However, even though totally global reparameterisations are prevented this way it is still possible to find parameterisations that descrease the DL by gathering landmarks together in different ways on differenct curves. Theoretically it could still be possible to get some effect of this type even when using all the terms but the parameterisations would be hard to find and even if found the effect would be small.

Another motivation for using all parameterisations in each scalar product is that otherwise the scalar products would not be compareable since the weight would be distributed differently on the curves.

To evaluate the scalar product numerically it is of course necessary to sample the curves. These sample points are just a step in the evaluation and should not be confused with landmarks in the traditional sense.

5 Experimental Validation

In this section the algorithm is validated on three data sets.

Hands: 23 contours of a hand segmented out semiautomatically from a video stream. To simplify the segmentation the hand was filmed against a dark background. **Femurs:** 32 contours of femurs taken from X-rays in the supine projection.

The letter g: 17 curves of the letter g. The curves of the letter g are sampled using a device for handwriting recognition.

Models built using the proposed scalar product are compared with models built using the standard scalar product. For the standard scalar product the algorithm uses node cost penalties [13] to prevent clustering of landmarks and the other problems discussed earlier.

Thodberg's efficient implementation of MDL [13] has been used for the comparison. MATLAB source code and test data are available at hht@imm.dtu.dk.

Seven control nodes have been used for the reparameterisations in all our simulations. For the standard scalar product 2048 landmarks were placed on each curve and for the proposed scalar product the number of sample points were 65 times the number of curves in the training set (between 1105 and 2080). The initialisations and parameters for the optimisation algorithm were identical in all cases.

By specificity of the models is meant that shapes generated by reasonable parameter values (± 3 std.) should be representative for the training set.

In Figure 1, the mean shape and the mean shape plus and minus three standard deviations of the first shape mode is plotted for each model. In this figure, the specificity of the models can be compared qualitatively.

The results for the most difficult dataset, the one of gshapes, are presented. In the figure, it can be seen that the shapes generated by the proposed algorithm are more representative. Look at the left figure and note, for example, the sharp bend in middle of the dashed curve and the not sharp enough corner of the solid curve. For the other two datasets the difference in the results of the two algorithms are harder to distinguish. The generality of the models is

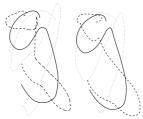


Figure 1. The mean (solid line) and the first mode of variation (dashed lines) of the optimised models. The model on the left is optimised using the standard scalar product and the model on the right is optimised using the proposed scalar product.

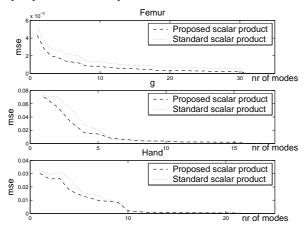


Figure 2. The mean square approximation error of the six models is plotted against the number of modes used. The top figure shows the models of femurs, the middle figure shows the models of g:s and the bottom figure shows the models of hands.

measured as the mean square error in leave-one-out reconstructions. Leave-one-out means that a model is built with all but one example. The model is then fitted to this unseen example. The error between the original curve and the model approximation curve is calculated by integrating the squared distance between the curves by arc-length. This is shown in Figure 2. It can be seen that the new scalar product gives models that generalise better than models built using the standard scalar product, even if node cost penalties are used.

6 Summary and Conclusions

In this paper a new scalar product that is invariant to global reparameterisations has been proposed. The optimisation of the shape model becomes invariant to global reparameterisations and therefore focuses on finding correspondences. This gives a more robust and stable algorithm. The algorithm is compared to a state of the art algorithm, which uses ad hoc solutions to prevent clustering of landmarks. The comparison shows that the achieved models are more specific and general.

In the future, it would be interesting to explore these principles further. A similar approach with a new scalar product using inverse parameterisation functions should be able to extend to 3D models of surfaces. Also, extentions to segmentation algorithms should be possible.

References

- S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(24):509–522, 2002.
- [2] F. Bookstein. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 3:225–243, 1999.
- [3] R. Davies, T. Cootes, and C. Taylor. A minimum description length approach to statistical shape modeling. In *Information Processing in Medical Imaging*, 2001.
- [4] R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor. Shape discrimination in the hippocampus using an mdl model. In *Information Processing in Medical Imaging*, 2003.
- [5] A. Ericsson. Automatic shape modelling and applications in medical imaging. Technical report, Mathematics LTH, Centre for Mathematical Sciences, Box 118, SE-22100, Lund, Sweden, nov 2003.
- [6] A. Ericsson and K. Åström. An affine invariant deformable shape representation for general curves. In *Proc. Int. Conf.* on Computer Vision, ICCV'03, Nice, France, volume 2, pages 1142–1149, 2003.
- [7] A. Ericsson and K. Åström. Minimizing the description length using steepest descent. In *Proc. British Machine Vision Conference, Norwich, United Kingdom*, volume 2, pages 93–102, 2003.
- [8] A. Hill and C. Taylor. Automatic landmark generation for point distribution models. In *Proc. British Machine Vision Conference*, pages 429–438, 1994.
- [9] C. Kambhamettu and D. Goldgof. Points correspondences recovery in non-rigid motion. In *Proc. Conf. Computer Vi*sion and Pattern Recognition, CVPR'92, pages 222–237, 1992.
- [10] A. Kotcheff and C. Taylor. Automatic construction of eigenshape models by direct optimization. *Medical Image Analy*sis, 2:303–314, 1998.
- [11] J. Rissanen. Modeling by shortest data description. Automatica, 14:465–471, 1978.
- [12] T. Sebastian, P. Klein, and B. Kimia. Constructing 2d curve atlases. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 70–77, 2000.
- [13] H. Thodberg. Minimum description length shape and appearance models. In *Image Processing Medical Imaging*, *IPMI 2003*, 2003.