Automatic feature point correspondences and shape analysis with missing data and outliers using MDL

Kalle Åström, Johan Karlsson, Olof Enquist, Anders Ericsson, Fredrik Kahl
Centre for Mathematical Sciences, Lund University, Sweden

Abstract. Automatic construction of Shape Models from examples has recently been the focus of intense research. These methods have proved to be useful for shape segmentation, tracking, recognition and shape understanding. In this paper we discuss automatic landmark selection and correspondence determination from a discrete set of landmarks, typically obtained by feature extraction. The set of landmarks may include both outliers and missing data. Our framework has a solid theoretical basis using principles of Minimal Description Length (MDL). In order to exploit these ideas, new non-heuristic methods for (i) principal component analysis and (ii) Procrustes mean are derived - as a consequence of the modelling principle. The resulting MDL criterion is optimised over both discrete and continuous decision variables. The algorithms have been implemented and tested on the problem of automatic shape extraction from feature points in image sequences.

1 Introduction

Inspired by the successful methods for finding feature correspondences along curves and on surfaces for deriving shape variation models, [4], this paper develops these methods further for the problem of shape modelling of unordered point sets with outliers and possibly missing data. The main idea is that an information criterion, such as Minimum Description Length (MDL), is well suited for determining which points should be considered as outliers and which points should be considered to be in correspondence.

One motivation for this study is that, although it is straightforward to detect interesting features in images, e.g., using corner detectors such as [8], it is not at all straightforward to solve the correspondence problem. Many methods for finding correspondences are based on either continuity assumptions, e.g., [18] or that a model is a priori known, e.g., [16].

For unordered points sets, as opposed to what is usually assumed for curves, it is often the case that points are frequently missing and that there are outliers. Traditional methods for shape analysis have problem with missing data. As a by-product of the development here, we derive novel methods for Procrustes analysis and principal component analysis for missing data, based on principles for model selection. Although generalisations to missing data have been done before both for Procrustes analysis, [9] and for principal component analysis, [11], in the present formulation they are shown to be a logical consequence of the modelling principle.

The underlying modelling principle we will use is MDL, that is, choose the representation or model with shortest description, cf. [15]. It has previously been successfully applied in computer vision to model selection problems, e.g., [12] and as mentioned curve and surface modelling [4].

MDL has also been used on sets of images to unify groupwise registration and model building, [20], although these mostly work on appearance based models. This is related to our approach, but here we include decisions about what to include in the modelling and what to consider as outliers explicitly in calculating the description length. The algorithm therefore decides what in the images to build the model from and what to consider as irrelevant background data.

There are several algorithms for matching one point cloud to another by determining both a deformation and correspondences, see for example [2], and many handle outliers in some way. However, these algorithms do not work with a whole sequence of point sets at once. Although it is conceivable that such methods could be used to construct shape variation models, the model built from the corresponding points might not be optimal at all. In this paper the model itself is an integral part of the algorithm for determining the correspondences.

Another line of research that is related to our work is the area of non-rigid factorisation methods, e.g., [3]. In addition to reconstructing a model for the point set, they also try to estimate the camera motion. However, they assume that feature correspondences are given or obtained by some heuristics. Similarly, outliers and missing data are handled with ad-hoc techniques. Our framework is well-suited to be applied to these problems as well.

2 MDL for Feature Point Selection and Correspondence

The main problem that we formulate and solve in this paper is the following. Assume that a number of examples $S = \{S_1, \ldots, S_n\}$ are given, where each example S_i is a set of unordered feature points

$$S_i = \{z_{i,1}, \ldots, z_{i,k_i}\}, \tag{1}$$

and each point $z_{i,j}$ lies in $\mathbf{R}^p$ for some dimension p, typically $p = 2$ or $p = 3$. An example of such sets is the output of a feature detector, e.g., [10].

(i) How should one determine which points in each set is considered to be outliers?
(ii) How should one determine the correspondences of feature points across the examples?
(iii) What is a suitable shape variation model for the inliers?

Such a procedure would be useful for unsupervised learning of shape variation in image sequences. The learnt models could then be useful for a number of applications, including *tracking*, *recognition*, *object pose* etc.

The idea here is to transfer similar ideas from shape variation estimation for curves and surfaces, [19, 4], where the correspondence problem is a crucial problem, to the concept investigated here.

As opposed to the theory for curves and surfaces, however, we do not here have the problem of how to weight different parts of the curves relative to the other. On the other hand we will here allow outliers and missing data. In this paper we will use the minimum description length paradigm, [1], to select a suitable model.

We pose the problem as that of selecting: (i) outliers, (ii) correspondences and (iii) model complexity, so that the description length is minimised. As we shall see this becomes a mixed combinatorial and continuous optimisation problem, where for each of a discrete set of possible outliers/correspondences, there is a continuous optimisation problem which has to be solved. These continuous optimisation problems involve both the problem of missing data Procrustes and missing data principal component analysis. In contrast to other ad-hoc methods dealing with outliers and missing data, the way we define missing data Procrustes and missing data principal component analysis is a natural consequence of the way we model the whole problem.

3 Unordered Point Set Shape Analysis

The formulation of the problem is as follows. Assume a set S of n unordered point sets, $S = \{S_1, \ldots, S_n\}$ is given. For simplicity we order the points in each set S_i of k_i points arbitrarily, cf. (1).

The object is now to find a reordering of such points. Assuming that the model contains N points, such a reordering can be represented either as a matrix O of size $n \times N$ whose entries are either 0 - representing that a model points is not visible in an image or the identity number between 1 and k_i telling which image point correspond to the model points, i.e. $O_{i,j} = 0$ if model point j is not visible in image i or $O_{i,j} = k$ if model point j in image i is $z_{i,k}$. Also introduce the set I of indices (i, j) such that model point j is visible in image j, i.e. $I = \{(i, j) \,|\, O_{i,j} \neq 0\}$. Outliers are then not represented in O.

Given an ordering O the data can be reordered, possibly with missing data into a structure T of N points in n images, i.e.

$$T_{i,j} = \begin{cases} z_{i,O_{i,j}} & \text{if } (i,j) \in I \\ \text{undefined} & \text{if } (i,j) \notin I. \end{cases}$$

For such a ordered point set T with missing data one can do a Procrustes mean shape analysis with respect to a transformation group G. In loose terms the aim is to find a mean shape m and a number of transformations $\{g_1, \ldots, g_n\}$ with $g_i \in G$ such that $g_i(m) \approx T_i$.

The usual method is then to perform a Principal Component Analysis on the residuals between $g_i^{-1}(T_i)$ and m, from which a number of shape variational modes, denoted v_l, can be determined. New shapes can then be synthesised as $g(m + \sum_{l=1}^{d} \lambda_l v_l)$, where λ_l are scalar coordinates.

Here we need to assess a number of different choices: the number of model points N, the ordering O, the mean shape m, the transformation group G, the transformations g_i, the number of variational modes d, the shape variation

modes v_l and the coordinates λ_l. The approach we make here is that a common framework such as the minimum description length framework could be used to determine all of these choices [13]. This would put the whole chain of difficult modelling choices on an equal footing and would make it possible to use simple heuristics for making fast and reasonable choices, while at the same time have a common criterion for evaluating different alternatives.

The whole process can thus be seen as an optimisation problem

$$\min_{\mathcal{M}} \mathrm{dl}(\mathrm{S}, \mathcal{M}),$$

over the unknowns $\mathcal{M} = (O, m, \{g_i\}, d, \{v_l\}, \{\lambda_{li}\})$ given data S.

4 Calculating the Description Length

A number of sets are given. Each set of points comes typically from images, where a number of interesting points have been detected. In order to determine a model that explains points that can be seen in many of the images, the goal is to minimise the description length that is needed to transmit all the interesting points of all views, in hope that a model will be able to make a cheaper description than simply sending the data bit by bit. Here we will derive the description length for the data and the model. For the outliers one must simply send the information bit by bit. For the points that are modelled, the idea is that it is cheaper to send the model with parameters and residuals etc. to explain the data. For the modelled points one must send: the model, the model parameters, information if a certain point is missing, the transformation and the residuals.

Preliminaries on information theory To transmit a continuum value α it is necessary to quantify the value. The continuum value α quantified to a resolution of Δ is here denoted $\hat{\alpha}$, $\alpha_{min} \leq \hat{\alpha} \leq \alpha_{max}, \hat{\alpha} = m\Delta, m \in \mathbf{Z}$.

The ideal coding codeword length for a value $\hat{\alpha}$, encoded using a statistical model $\mathcal{P}(\hat{\alpha})$ is given by the Shannon codeword length [17]. Using Shannon's codeword length the description length of a given value, $\hat{\alpha}$, encoded using a probabilistic model, is $-log(\mathcal{P}(\hat{\alpha}))$, where $\mathcal{P}$ is the probability-density function.

Coding data with uniform distribution Assume α is uniformly distributed and quantified to $\alpha_{min} \leq \hat{\alpha} \leq \alpha_{max}$, $\hat{\alpha} = m\Delta$, $m \in \mathbf{Z}$. Then α can take $\frac{\alpha_{max}-\alpha_{min}}{\Delta}$ different values. Since uniform distribution is assumed, the probability for a certain value of $\hat{\alpha}$ is $\mathcal{P}(\hat{\alpha}) = \frac{\Delta}{\alpha_{max}-\alpha_{min}}$. This gives Shannon's codeword length for $\hat{\alpha}$, $-log(\mathcal{P}(\hat{\alpha})) = -log(\frac{\Delta}{\alpha_{max}-\alpha_{min}})$. If the parameters α_{min}, α_{max} and Δ are unknown to the receiver, these need to be coded as well.

Coding data with assumed Gaussian distribution Since the mean of our data μ is zero, the 1-parameter Gaussian function can be used. The frequency function of the 1-parameter Gaussian function is $f(x;\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{x^2}{2\sigma^2})$. The derivation for sending a number of normally distributed 1 dimensional data sets

was done in Davies [5]. The derivation gives the following expression: $\mathcal{L}_{guassian} = \tilde{F}(n_s, R, \Delta) + \sum_{i=1}^{n_g}(n_s - 2)\log(\sigma_i) + \frac{n_s}{2} + \sum_{j=n_g+1}^{n_g+n_{min}}(n_s - 2)\log(\sigma_{min}) + \frac{n_s}{2}(\frac{\sigma_j}{\sigma_{min}})^2$, where σ_{min} is a cutoff constant, n_g is the number of directions where $\sigma > \sigma_{min}$ holds and n_{min} is the number of directions where $\sigma \leq \sigma_{min}$ holds. $\tilde{F}(n_s, R, \Delta)$ is a function that only depends on the number of shapes n_s, the range of the data R, and the resolution Δ. It is assumed constant for a given training set, i.e. it does not depend on decisions about outliers or correspondences.

$$\mathcal{L}_{guassian} = F(n_s, R, \Delta) + \sum_{i=1}^{n_g}(n_s - 2)\log(\sigma_i) + \frac{n_s}{2} + \sum_{j=n_g+1}^{n_g+n_{min}}(n_s - 2)\log(\sigma_{min}) + \frac{n_s}{2}(\frac{\sigma_j}{\sigma_{min}})^2, \quad (2)$$

The total description length of the interesting points in all images The description length for a point $\hat{x}$ equally distributed over the image is

$$dl_{rect} = -log(\mathcal{P}(\hat{x})) = -2log(\frac{dx}{X}) \ .$$

Here X is the range, typically 100 pixels in our examples, and dx is the resolution, which has been set to 0.5 pixels. The factor 2 comes from that an image point is two-dimensional.

The outliers are assumed to be uniformly distributed over all the image, so with n_o number of outliers $dl_{outliers} = n_o dl_{rect}$. For each model point we need to know if the point is missing in an image or not. This means one bit for each n_p model points in all n_v images. Conversion to nats gives $dl_{index} = log(2)n_v n_p$, where n_p is the number of landmarks in the model and n_v is the number of images. For each image the transformation g of the model has to be encoded. The transformations are assumed equally distributed within the size of the image. For translations this gives the following expression $dl_{trans} = (n_v - 1)n_{dof} dl_{rect}$, where n_v is the number of images and n_{dof} is the degrees of freedom in the transformation group, e.g. with 10 images and using 2D translations, 18 translation parameters has to be encoded. The coordinates of the mean shape and the coordinates of the shape modes are also assumed equally distributed within the size of the image, thus the cost is

$$dl_{meanshape} = n_p dl_{rect}$$

$$dl_{shapemodes} = n_p n_m dl_{rect} \ ,$$

where n_m is the number of shape modes used by the model. The residuals and the λ-parameters are assumed Gaussian. The cost for these are

$$dl_\lambda + dl_{res} = \mathcal{L}_{guassian} \ .$$

So the full cost for sending the data is

$$DL_{tot} = dl_\lambda + dl_{res} + dl_{meanshape} + dl_{shapemodes} + dl_{trans} + dl_{index} + dl_{outliers}$$

Given a shape model that describes part of the data for a situation we can now calculate the description length for this data and model. For each suggested model one needs to calculate the description length of sending all the outliers and all the data modelled with that particular model. The number of shape modes can vary between zero to $n_s - 1$ and all these models must be evaluated. Note here that since the shape modes calculated when using missing data PCA depends on the number of modes used, the model needs to be calculated over and over as the number of shape modes increase. In the optimisation procedure the tested model with least description length is then compared to previous solutions.

5 Optimising DL

The whole optimisation process over all unknowns can be divided into two parts: (1.) Optimisation over the discrete ordering matrix O and (2.) optimisation over the remaining parameters $\tilde{\mathcal{M}} = (m, \{g_i\}, d, \{v_l\}, \{\lambda_{li}\})$.

Assume that a reordering O is given, then it is straightforward to reorder the inlier points into the data structure T as described above. Each ordering also determines the number of inliers n_{inlier} and the number of outliers $n_{outlier}$. The description length for the outliers is then independent of $\tilde{\mathcal{M}}$. Assume now also that a transformation group G and the number of shape variational modes d are given. The description length now depends more or less on the remaining residuals of the inliers. Minimising the description length is then a question of minimising

$$\min_{m,\{g_i\},\{v_l\},\{\lambda_{li}\}} \sum_{(i,j)\in I} \left| T_{i,j} - g_i \left(m_j + \sum_{l=1}^{d} v_{j,l}\lambda_{l,i} \right) \right|.$$

We solve this minimisation problem explicitly. To get an initial estimate we solve first for missing data Procrustes by

$$\min_{m,\{g_i\}} \sum_{(i,j)\in I} |T_{i,j} - g_i(m_j)|^2$$

and then use the residuals missing data residuals $r_{i,j} = g_i^{(-1)}(T_{i,j}) - m_j$ to obtain initial estimates on v and λ. These initial estimates are used as a starting point in a Gauss-Newton optimisation scheme to find the nearest local minima.

It is straightforward to search through the number of nodes d and the different transformation groups G and as described above, it is then possible to find optimal Procrustes, missing data PCA and model order that gives minimal description length, thereby determining the solution with the best description length for this particular choice of point reordering O. Thus, the minimal description length can be seen as a function of the ordering O.

Optimising description length with respect to O is a combinatorial optimisation problem. We suggest the following algorithm that (1) finds a reasonable

initial guess by heuristics and (2) searches for a local minima in a combinatorial optimisation sense by adding/removing inliers and adding/removing model points.

We approach this optimisation by a local search methods with the following four types of perturbations: (i) Change of a point from outlier to inlier, (ii) Change of a point from inlier to outlier, (iii) Deletion of a model point, (iv) Addition of a model point.

The final algorithm for determining minimal description length is then

1. Make an initial guess on point ordering O based on heuristics or randomness.
2. Calculate optimal description length for that ordering.
3. See if any of the perturbations above lowers the description length.
4. – If it does, make those changes and continue with step 3.
   – If not, then we are at a local minima, stop.

5.1 Initial guess

One way of picking a reasonable initial guess is the following. The initial guess is made by using the points in one image as the model. For each new image matching is done as follows. Given n points in the model and m points in the new image. Form an $n+1 \times m+1$ cost matrix C whose first $n \times m$ elements C_{ij} is the Euclidean distance between model point i and feature point j after the model points are translated or transformed according to the best fit in the previous frame. This step of the algorithms, thus assumes that there is a smooth motion of the feature point. The last row $C_{n+1,j}$ is set to a constant representing the cost of not associating a feature point j with a model point. Similarly the last column $C_{i,m+1}$ is set to a constant representing the cost of not associating a model point j to any of the feature points. The matching is then done by solving the transport problem with supply of $s = [1, \ldots, 1, m]$ and demand $t = [1, \ldots, 1, n]$. Here we used standard algorithms for solving the transport problem, cf. [14].

6 Experimental Validation

6.1 Feature point selection, model extraction and shape recognition

In this experiment we have taken a digital film recording of a persons face as it moves in the scene. A sequence of 944 frames was captured and a standard interest point detector [10], was run on all of the frames. In each frame a face detector was run and those interest points that were within the rectangular frame of a face detector [6], was kept.

The first 100 frames were used for model estimation. This gave roughly 880 feature points (between 5 and 13 points in each frame). Three such frames are shown in Figure 1 together with the extracted feature point shown as small rectangular points.

The initial guess ordering resulted in 584 of the 880 feature points being associated with any of the 9 model points. The description length for this ordering was 10 826 bits.

After local optimisation the description length lowered to 9 575 bits for a model with 12 model points. Here 740 of the 880 points were associated to a

model point. In Figure 2 is shown three frames out of the 100 overlayed with feature points and best fit of the 12 model points obtained after minimising description length. Notice that certain points in Figure 1 are classified as outliers and are not shown in Figure 2.

Fig. 1. Three out of 100 frames used for testing. Detected feature points are shown as white squares

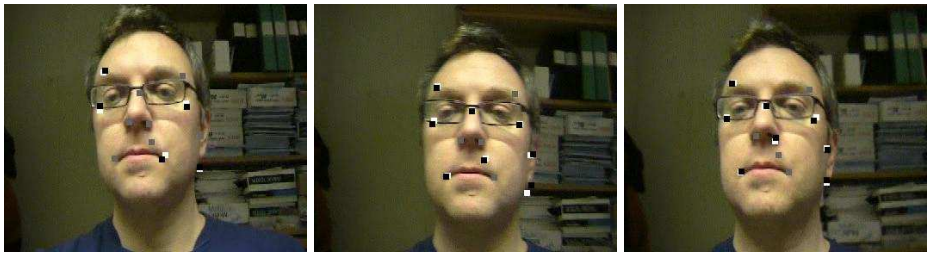

Fig. 2. Three out of 100 frames used for testing. For measured points (in white) the fitted model points are shown in black. For missing data the fitted model points are shown in gray.

Recognition using the shape model

The model can also be used to find the object in a new image without any prior knowledge about its position. This is accomplished by the RANSAC algorithm [7], where the consensus was based on the description length of the matching. For simple transformation groups, such as translation, it is enough to randomly match one point in the model to a feature point in the image.

Current limitations: Although the theory presented in this paper is quite general in the sense that any transformation group G can be used, our current implementation handles only the cases of 'no transformation' and 'pure translation'. Another limitation is the way the combinatorial optimisation scheme is implemented. It happens that the algorithms gets stuck in local minima, so that some points that are inliers are associated to the wrong object point or are considered as outliers. More perturbation types could be allowed, for example

moving an image point from one model point to another. Yet another limitation of the scheme is that it is relatively slow. Each evaluation of a selection of inliers and outliers involves several steps, including a singular value decomposition with missing data for the PCA.

7 Conclusions

In this paper we have studied the problem of automatic feature point correspondence determination and shape analysis with missing data and outliers using MDL.

The modelling problem is posed as a combined combinatorial/continuous optimisation problem. The continuous part involves missing data Procrustes and missing data PCA. The combinatorial part is solved by an initial guess based on heuristics followed by local search. Although not the main focus of this paper, missing data Procrustes and missing data principal component analysis are defined and algorithms for their determination are developed. The definitions are natural consequences of the modelling principles followed.

The result is an algorithm that given a number of unordered point sets determines (i) the number of model points, (ii) the mean shape and shape variational modes of the model, (iii) the outliers in the data sets, (iv) the transformations g, and (v) the inliers with correspondences. We envision this algorithm being used on a set of images after extraction of feature points. The fact that the model can be learnt automatically makes it possible to acquire models on the fly, without manual interactions. Such models can then be used for tracking, pose determination, recognition.

In this paper we have focused on the point positions. This makes the method relatively stable to lighting variations. However, it would be interesting to extend the ideas to that of feature points including local descriptors, that capture the local variations in intensity in patches around the feature points. This would probably make the system better at recognising and tracking under most circumstances.

Acknowledgements

This work has been supported by the Swedish Knowledge Foundation through the Industrial PhD programme in Medical Bioinformatics at Karolinska Institutet, Strategy and Development Office, the European Commission's Sixth Framework Programme under grant no. 011838 as part of the Integrated Project SMErobot, by the Swedish Foundation for Strategic Research (SSF) through the programme Vision in Cognitive Systems (VISCOS) and by the Swedish Road Administration (Vägverket) and by the Swedish Governmental Agency for Innovation Systems (Vinnova).

References

1. A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE trans. on information theory*, 44(6):2743–2760, 1998.

2. 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.
3. C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, pages 690–6 vol.2, 2000.
4. R.H. Davies, C.J. Twining, T.F. Cootes, J.C. Waterton, and C.J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Trans. medical imaging*, 21(5):525–537, 2002.
5. Rhodri H. Davies, Tim F. Cootes, and Chris J. Taylor. A minimum description length approach to statistical shape modeling. In *Information Processing in Medical Imaging*, 2001.
6. A. P. Eriksson and K. Åström. Robustness and specificity in object detection. In *Proc. International Conference on Pattern Recognition, Cambridge, UK*, 2004.
7. M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–95, 1981.
8. W. Förstner. A feature based correspondence algorithm for image matching. In *ISP Comm. III, Rovaniemi 1986, International Archives of Photogrammetry, 26-3/3*, 1986.
9. J.C. Gower. Generalized procrustes analysis. *Psychometrica*, 40:33–50, 1975.
10. C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. of the 4th Alvey Vision Conference*, pages 147–151, 1988.
11. D. Jacobs. Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 206–212, 1997.
12. K. Kanatani. Geometric information criterion for model selection. *Int. Journal of Computer Vision*, 26(3):171–189, 1998.
13. J. Karlsson and A. Ericsson. Aligning shapes by minimising the description length. In *Scandinavian Conf. on Image Analysis, Juuensu, Finland*, 2005.
14. D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 1984.
15. J. Rissanen. Modeling by shortest data description. *Automatica,*, 14:465–471, 1978.
16. K. Rohr. Recognizing corners by fitting parametric models. *Int. Journal of Computer Vision*, 9(3):213–230, 1992.
17. C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37, 1949.
18. J. Shi and C. Tomasi. Good features to track. In *Proc. Conf. Computer Vision and Pattern Recognition, CVPR'94*, 1994.
19. H. H. Thodberg. Minimum description length shape and appearance models. In *Image Processing Medical Imaging, IPMI*, 2003.
20. C.J. Twining, T.F. Cootes, S. Marsland, V.S. Petrovic, Schestowitz R.S., and C.J. Taylor. Information-theoretic unification of groupwise non-rigid registration and model building. *Proceedings of Medical Image Understanding and Analysis*, 2:226–230, 2006.