IDIAP RESEARCH REPORT

IDIAP

Martigny - Valais - Suisse

Optimal Parameterization of Point Distribution Models

Georg Thimm Juergen Luettin

IDIAP–RR 98-01

revised in August 98

submitted for publication

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 – 27 – 721 77 11
fax +41 – 27 – 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

Optimal Parameterization of Point Distribution Models

Georg Thimm Juergen Luettin

revised in August 98

submitted for publication

Abstract. We address the problem of determining the *optimal model complexity* for shape modeling. This complexity is a compromise between model specificity and generality. We show that the error of a model can be split into two components, the model error and the fitting error, of which the first one can be used to optimize the model complexity based on the specific application. This strategy improves over traditional approaches, where the model complexity is only determined by vague heuristics or trial-and-error. A method for the determination of optimal active shape models is proposed and its efficiency is validated in several experiments. Furthermore, this method gives an indication on the range of valid shape parameters and on whether or not an increased number of training data will reduce the number of shape parameters further.

Acknowledgements: This work has been performed with financial support from the Swiss National Science Foundation under Contract No. 21 49 725 96.

1 Introduction

In computer vision, it is a standard problem to model deformable objects (lips, hands) or classes of objects with similar shapes (faces, cars). The appearance of these objects is often subject to additional variability due to affine transformation. These affine transformations are usually caused by a rotation or translation of an object relative to a camera. Several techniques to describe the general shape of such objects are known: contours are modeled as snakes [Kass-88], B-splines [Blake-94], *box models* [Black-96], *minimum description length* based shape models [Li-93], polygons [Jaynes-94], *deformable templates* [Yuille-92] [Jain-96] [Zhong-98], *flexible templates* [Hill-92], or *point distribution models* [Cootes-92.2]. Among these, we choose *point distribution models* (**PDM**s) [Cootes-92.2] as they are very generic, highly specific, and can be automatically build from (hand) labeled data. Ideally, shape models should only generalize to shapes that can be observed for some object in the class of objects subject to modelization. Models which fulfill this property are called *specific*, or *more specific* than others if they describe fewer shapes not belonging to the target object class. On the other hand, PDMs are statistical models and therefore it is unlikely that they specify exactly all possible deformations or instances of the object, and hence will be affected by an error.

Generally, the error committed by a shape recognition tool is composed from two parts: the error caused by an insufficient flexibility of the shape model, in the following called *model error*, and the *fitting error* caused by the fitting algorithm. Both errors combined are the observed error when a model is fitted to an image and therefore are *ad hoc* not distinguishable. Consequently, the main reason for some unacceptably high error can not be determined easily. As the error obtained by fitting a model to a certain image is at best the model error, we propose a method that optimizes the model complexity only.

Point distribution models are a very generic shape modeling tool among the approaches mentioned above as they can be automatically build from labeled data. These models have further the advantage that they are potentially more specific than others (i.e. use less free parameters) as the number of free parameters can easily be changed, even after the model was build (it is not contested here that other models have other advantages).

However, the core algorithm for the calculation of the point distribution models does not determine the number of free parameters to use. Only a method that is of limited use is described in [Cootes-92.1]. It uses the eigenvalues to calculate the percentage of the total model deformation described if a certain number of non-zero shape parameters is used. The problem is that such a percentage is hard to imagine in terms of a difference between an ideal shape and the shape produced by a restricted model.

This is an important problem, as the performance of the model and the fitting algorithm, that adapts a model to a given image depends on this number. The optimal choice of this number is therefore the main subject addressed in this paper. The latter proposed method balances between more specific models with a potentially high model error and less specific models with a low model error. This is achieved by varying the number of free parameters combined with a statistical analysis of the model error.

2 Point Distribution Models

Roughly said, a point distribution model consist of a selection of orthogonal vectors describing variations of a mean shape. The *mean shape vector* $\bar{\mathbf{x}}$ and the set of vectors describing the variations, which are represented by a matrix $\mathbf{P}$ of eigenvectors, are calculated from labeled data. In the selection of the eigenvectors, preference is given to eigenvectors with the highest eigenvalues (the *strongest eigenvectors*) of variations relative to a mean shape vector $\bar{\mathbf{x}}$ calculated over all training examples. The *training data* or *labels* are a set of shape vectors of equal length, each a concatenation of coordinates of selected points in an image $(x_1, y_1, x_2, y_2, \dots)$. These selected points are usually placed by hand (using a simple graphical tool) on contours or other outstanding points in the images describing the shape of an object. Examples will be given later in section 3.

By definition, *valid shape vectors* $\mathbf{x}$ with respect to a certain point distribution model and a certain *number of shape parameters* k are those which can be obtained by multiplying the eigenvector matrix $\mathbf{P}$ with a vector of *shape parameters* $\mathbf{b}$ with $b_i \in \Re$ (or some interval $[-a, a] \subset \Re$) for $1 \leq i \leq k$ and $b_i \equiv 0$ for $i > k$. It is presumed that the eigenvectors in the eigenmatrix are sorted with decreasing strength. A *valid shape vector* is then represented as:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{1}$$

This approach is based on the assumption that the strongest eigenvectors, and only these, characterize the important deformations of the mean model $\bar{\mathbf{x}}$, whereas weak eigenvectors correspond to noise in the training data. In the resulting model,

each eigenvector corresponds ideally to a deformation that is independent from all others. However, this is not the case if some deformations are nonlinear or if the training data is insufficient.

Note, that the *optimal shape vector*, the valid shape vector with the smallest mean square distance to the given shape vector $\mathbf{x}$, with respect to some number of shape parameters can be calculated efficiently with the cost of a matrix multiplication[1]:

$$\mathbf{b} = \mathbf{P}^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{2}$$

Using this shape vector for a given k, the *mean model error* $E_{\mathrm{Mean},k}$ is defined as the mean square distance between a label $\mathbf{x}$ and the valid shape vector closest to it:

$$E_{\mathrm{Mean},k}(\mathbf{x}) = ||\mathbf{x} - (\bar{\mathbf{x}} + \mathbf{P}\mathbf{b})|| \quad \text{with } b_i \equiv 0 \text{ if } i > k \tag{3}$$

The *maximal model error* $E_{\mathrm{Max},k}$, which is defined as the largest error of all nodes (points) is defined by:

$$E_{\mathrm{Max},k}(\mathbf{x}) = \max_j((\tilde{x}_j)^2) \quad \text{with } \tilde{\mathbf{x}} = \mathbf{x} - (\bar{\mathbf{x}} + \mathbf{P}\mathbf{b}) \quad \text{and } b_i \equiv 0 \text{ if } i > k \tag{4}$$

Either model error is not necessarily zero for the shape vectors in the training database, especially if k is small as compared to the size of the training set.

The creation of a point distribution model and the adjustment of a model to a given image are sensitive to the number and choice of labeled data used for calculating the model. Furthermore, the model and the fitting algorithm may be wrongly parameterized. In practice, this means that a considerable amount of experience is necessary to successfully tune the whole system as the cause of bad performance is difficult to identify. This problem can be reduced, if the model is evaluated and parameterized before it is actually fitted to images.

The method described in section 4 relies on an estimation of the model error and evaluates a given point distribution model independently from all other procedures that are involved in adapting a shape to an image. Using this estimation, it is straight forward to select a number of shape parameters that makes a model just flexible enough to satisfy some condition on the model error. In other words, before a fitting algorithm is involved, the user knows that a certain model is appropriate for the task. This approach aims at problems with rather different character than those described and solved by T.F. Cootes and C.J. Taylor in [Cootes-96]. They replace in a certain sense incomplete training data by some flexibility of the model that can not be derived from the training data.

[1] This formula can be derived from equation (2) and the fact that $\mathbf{P}^T = \mathbf{P}^{-1}$ as $\mathbf{P}$ consists of orthogonal vectors.

3 The Databases

3.1 The X-ray Database

The images in this database were extracted from a database transfered on video disc in collaboration at the *Queen's University (Kingston, Canada)* and the *ATR Human Information Processing Labs (Kyoto, Japan)* which consists of X-ray movies of the vocal tract shot in the 1960's. Video discs with this database are provided at no cost ([Munhall-95]). This database is corrupted by a considerable amount of noise and variations of the average grey level. The latter makes it particularly difficult to label smooth or blurred contours consistently - even for humans. Furthermore, the contours of organs, such as the tongue, the lips, the velum, and the teeth are in some cases invisible due to low contrasts (lips, tongue), shadowing (tongue behind teeth fillings), or touching (tongue against velum or palate, lip against lip). Figure 1 shows three images from this database.

In consequence to the low quality of the images, it is very important that the point distribution model is highly specific. This can help to extrapolate from visible contours to occasionally invisible ones and to reduce the sensitivity of the fitting algorithm to contours not belonging to the object.

Note that fitting a point distribution model to an image is not discussed here. This can for example be done by the means of active shape models using grey-level models with gradient descend on the shape parameters [Luettin-97] or moving points along lines orthogonal to the curve [Cootes-92.2] [Cootes-95] [Cootes-96], active shape models using geometric histograms [Di-Mauro-96], Gabor filters [McKenna-97], or the *condensation* algorithm using edges [Isard-96]. In [Davis-97] and [DeCarlo-96] it is shown how to combine optical flow calculations with shape models (see [Barron-94] for an overview an evaluation of optical flow algorithms). However, optical flow algorithms are very sensitive to noise and variations of the average grey level, which makes their usage on the database of X-ray images unfruitful.

Figure 1: Three examples from the database of X-ray images.

3.2 The Tulips1 Database

This publicly available database consists of video recordings of lips during the isolated pronunciation of the first for English digits [Movellan-96]. The recordings were mad from 9 men and 3 women. From this database, 190 images were selected and the outer rim of the lips labeled by hand as shown in figure 2. Each label consists of 2 points placed on the junctions of the upper and lower lip and 10 points placed on the either lips.

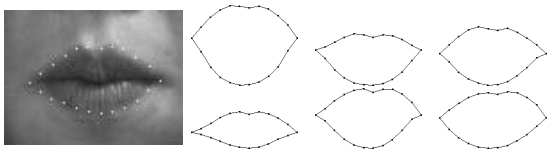


Figure 2: An image from the Tulips database and some example labels.

3.3 The M2VTS Database

This database is very similar to the Tulips1 database, recordings of more speakers are included (25 male, 12 female). Furthermore, all digits were pronounced in sentences [Pigeon-97.2]. From this database 526 examples were selected and labeled as shown in figure 3. In addition to the Tulips1 database, also the inner rim of the lips are labeled. The labels consist of 22 points for the outer rim of the lips and 16 for the inner rim.

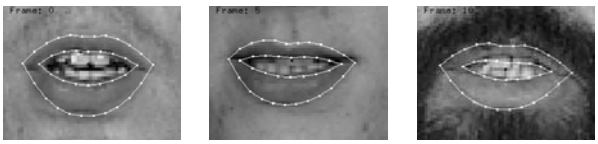


Figure 3: Labeled examples from the M2VTS database.

4 The Proposed Method

In order to find a point distribution model which is, both, very specific and affected only by a low model error, it is necessary to estimate the maximal or mean model error (compare equations (3) and (4)). This can be done using a cross-validation like procedure:

1. Determine for the specific application the *acceptable mean or maximal model error* $\mathbf{E}_{\mathrm{Mean}}$, respectively $\mathbf{E}_{\mathrm{Max}}$.

2. Split the set of labels $\mathcal{L}$ randomly into a number of training $\mathcal{T}_i$ and test sets $\mathcal{S}_i$, where all training sets, and therefore the corresponding test sets, have to have the same size and $\mathcal{T}_i \cup \mathcal{S}_i = \mathcal{L}$.

3. For each training set $\mathcal{T}_i$, build a point distribution model (that is $P^{(i)}$ and $\bar{\mathbf{x}}^{(i)}$).

4. For each model $(P^{(i)}, \bar{\mathbf{x}}^{(i)})$ and label in test set $\mathcal{S}_i$, and for all possible numbers of shape parameters, calculate the optimal shape vector according to equation (2) for $1 \leq k \leq N$ with N equal to the elements in a shape vector (twice the number of points per label for the database of x-ray images[2]). Then, calculate the mean square error between these vectors and the corresponding original labels (that is $E_{\mathrm{Mean},k}$, respectively $E_{\mathrm{Max},k}$, for all possible k and all labels).

5. Average the mean model error $E_{\mathrm{Mean},k}$ per test set for all possible numbers of shape parameters. This gives the *estimated mean model error* $\bar{E}_{\mathrm{Mean},k}$. Similarly, the *estimated maximal model* $\bar{E}_{\mathrm{Max},k}$ is calculated from the maximal model error $E_{\mathrm{Max},k}$.

[2] If the shapes are scaled and a certain point translated to the origin of the coordinate system, the effective N is reduced by 3.

6. Compute the estimated mean (maximal) model error $\bar{E}_{\mathrm{Mean},k}$ ($\bar{E}_{\mathrm{Max},k}$) as a function of the number of non-zero shape parameters k.

7. Search the number $\hat{k}$ of active shape parameters so, that the *acceptable mean or maximal model error* $\mathbf{E}_{\mathrm{Mean}}$, respectively $\mathbf{E}_{\mathrm{Max}}$, is equal or just inferior to the estimated model mean error $\bar{E}_{\mathrm{Mean},\hat{k}}$, respectively the estimated maximal model error $\bar{E}_{\mathrm{Max},\hat{k}}$:

   choose $\hat{k}$ so that

$$\begin{cases} \bar{E}_{\mathrm{Mean},\hat{k}} \leq \mathbf{E}_{\mathrm{Mean}} < \bar{E}_{\mathrm{Mean},\hat{k}+1} \\ \quad \text{respectively} \\ \bar{E}_{\mathrm{Max},\hat{k}} \leq \mathbf{E}_{\mathrm{Max}} < \bar{E}_{\mathrm{Max},\hat{k}+1} \end{cases} \tag{5}$$

   A point distribution model using $\hat{k}$ eigenvectors (or equivalent: $\hat{k}$ non-zero shape parameters) is optimal in the sense of having a model error inferior or equal to $\bar{E}_{\mathrm{Mean},k}$, respectively $\bar{E}_{\mathrm{Max},k}$, and being as specific as possible.

Example: 350 images of the X-ray image database were labeled. Each label consists of 37 coordinates, which corresponds to a 74 element feature vector. Each label marks the outline of the chin (5 points), the upper jaw (2 points), the upper and lower lips (3 points each), the upper and lower teeth with one point on the front teeth (4 points each), the palate (3 points), the velum (4 points), and the tongue (9 points). An example for such a label is shown in figure 4.

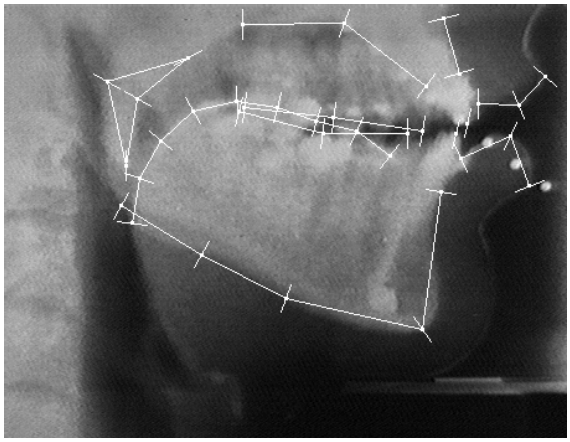


Figure 4: A labeled image of the database of X-ray images.

Before a point distribution model is build, the labels are scaled and the coordinates of the upper front teeth are chosen to be the origin of the coordinate system. Therefore, a point distribution model with 71 shape parameters can theoretically describe all possible labels, unfortunately also all impossible ones. For a model, which is calculated from 300 labels, a selected label and the valid shape vectors closest to the true label are calculated for $k \in \{1, 3, 10, 30\}$ (see equation (2)). That those restricted models are not necessarily able to describe exactly a given label, can be seen in figure 5. The points of the label are connected by a line with the corresponding point in the valid shape (the length of the lines correspond to the error committed for this point). It can be seen that increasing the number of non-zero shape parameters b_i decreases the mean and maximal model error.

30 pairs of training and test sets are likely to be sufficient for reducing the statistical error of the estimated model error to a reasonable value. To be on the safe side, the statistical error and the variance of the averaged errors should be taken into consideration and the number of training and test sets increased if the statistical error is too high.

A user may find number $\hat{k}$ of shape parameters to be unrealistic high. This can be the cause of the number of training samples being to low, as it will be shown in section 5. As an enlargement of the training data base will decrease the number of shape parameters only in the presence of insufficient or inaccurate training data. It does not help if the shape deformations are non-linear. In order to cope with this case, one would use a non-linear type of model, for example a *mixture model* [Cootes-97], a non-linear point distribution model using neural networks [Sozou-95] or polynomials [Sozou-94]. Approaches that use point distribution models locally are described in [Bregler-94] and [Heap-97].

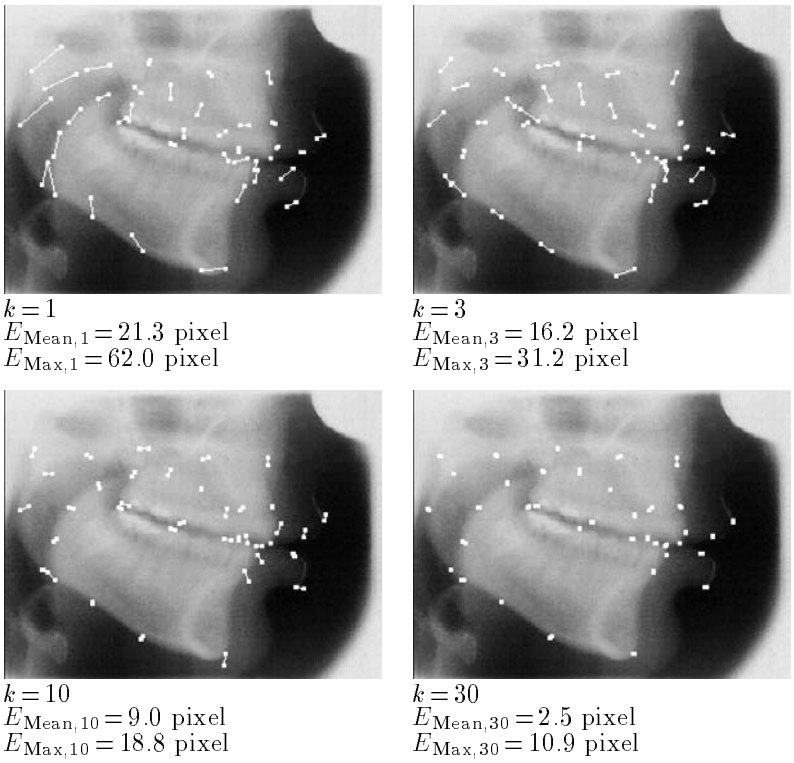
5 Experiments

5.1 ... with the X-ray Database

For the testing procedure, the set of 350 labels was split randomly in six groups of 30 pairs of training and test sets: in training sets of sizes 50, 100, 150, 200, 250, or 300 labels and corresponding test sets of sizes[3] 300, 250, 200, 150, 100, or 50. Then, for each group of training and test sets, step 1 to 5 of the method described in section 4 are performed.

The six curves shown in figure 6 reflect the changes of the estimated maximal model error in image pixels if the number of non-zero shape parameters is varied (the superscript of the $\bar{E}$ indicates the size of the training set). The statistical error

[3] The different sizes of the test sets will, without harm, result in different confidence intervals for the $\bar{E}_{\mathrm{Mean}}$, and $\bar{E}_{\mathrm{Max}}$.



$k=1$
$E_{\mathrm{Mean},1}=21.3$ pixel
$E_{\mathrm{Max},1}=62.0$ pixel

$k=3$
$E_{\mathrm{Mean},3}=16.2$ pixel
$E_{\mathrm{Max},3}=31.2$ pixel

$k=10$
$E_{\mathrm{Mean},10}=9.0$ pixel
$E_{\mathrm{Max},10}=18.8$ pixel

$k=30$
$E_{\mathrm{Mean},30}=2.5$ pixel
$E_{\mathrm{Max},30}=10.9$ pixel

Figure 5: The mean and maximal model error for a varying number k of non-zero shape parameters for a test label. Lines connect points of the label with corresponding points in the closest valid shape vector. The model was constructed from 300 labels.

for the maximal error is typically smaller than 0.015 pixel (assuming a student-t distribution of the error and a confidence of 95%), whereas the variance of the error is typically smaller than 0.3 pixel (both with one exception for $k \equiv 1$).

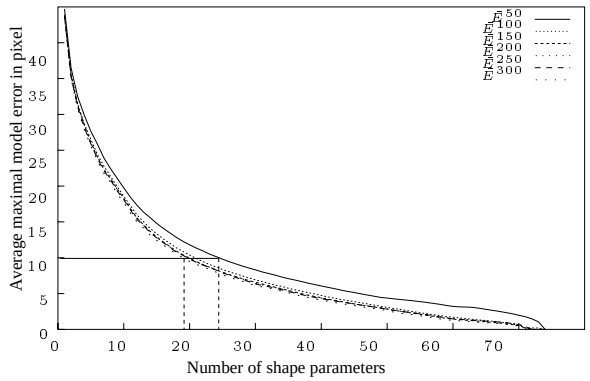


Figure 6: The variation of the *estimated maximal model error* for point distribution models depending on the number of training samples and non-zero shape parameters for the X-ray data.

Similarly, figure 7 shows the estimated mean errors calculated in the same way. It can be easily seen that the curves are similarly shaped with the $\bar{E}_{\mathrm{Mean},k}$ having naturally lower values. The remainder of the paper therefore discusses only the maximal model error; results are in general valid for the mean model error too.

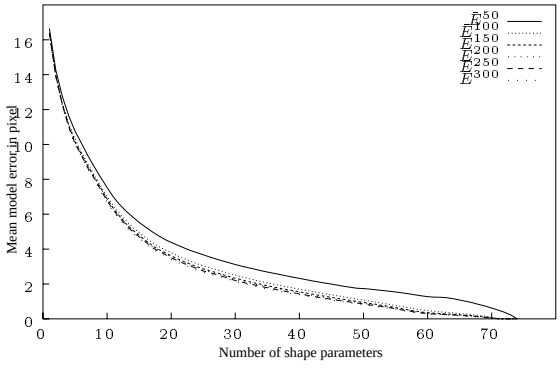


Figure 7: The variation of the *estimated mean model error* for point distribution models depending on the number of training samples and non-zero shape parameters for the X-ray data.

Example: for the database of X-ray images, we assume that labels of a given image made by hu-

mans will differ up to 10 pixel per point (in other words, the human will commit an error of up to 10 pixels), due to the reasons explained in section 3.1. It is therefore appropriate to permit a maximal model error of 10 pixel. Then, using figure 6, it can be determined that a point distribution model build from 50 samples requires on average 24 shape parameters to meet this condition, whereas a model build from 300 samples requires only 19 parameters. Clearly, using more training examples improves the quality of the point distribution model.

For better observation of the changes caused by varying the size of the training set, the curve corresponding to the point distribution models build from 300 samples is used as reference and subtracted from the others. The resulting curves are displayed in figure 8. It can be seen that, with the increase of the number of training samples, the point distribution models improve progressively: the lower the curve for a predefined model, the less shape parameters are required to perform at a certain maximal error (disregarding the uninteresting shape parameters higher than 50). An obvious, but *a priori* unanswerable question is, how far can the quality of the model be improved by augmenting the number of training samples.

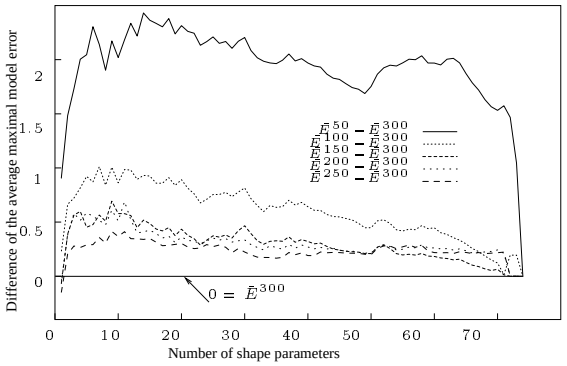


Figure 8: Relative differences of the estimated maximal error for point distribution models for the X-ray data.

Some information can be drawn from such plots: theoretically, not taking statistical errors into account, curves approach with an increasing size of the training set asymptotically a lower bound. This lower bound is the curve corresponding to the unknown optimal model. This fact may be used with some risk to estimate whether or not the number of samples for the calculation of the point distribution model is sufficient: if two consecutive curves are almost identical, probably no further important improvement is possible. From figure 8, one might estimate that increasing the training set size from 300 to 350 labels, the average maximal model error can be reduced by 0.1 to 0.3 pixel.

5.2 Experiments with the Tulips1 Database

Similar to the experiments with the X-ray image database, the total of 189 labels was split in 100 randomly selected labels with 25, 50, 75, 100, and 125 items for the training and the remaining items as test set. Figure 9 shows that the general shape of the maximal error as a function of the used shape parameters and the size of the training set is the same as for the X-ray database. Although the deformation of the lips and the vocal tract are nonlinear, it is obvious that the deformation of the lips are less complex. This coincides with the fact, that the number of required shape vectors corresponding to the same error are much lower for the lip data.

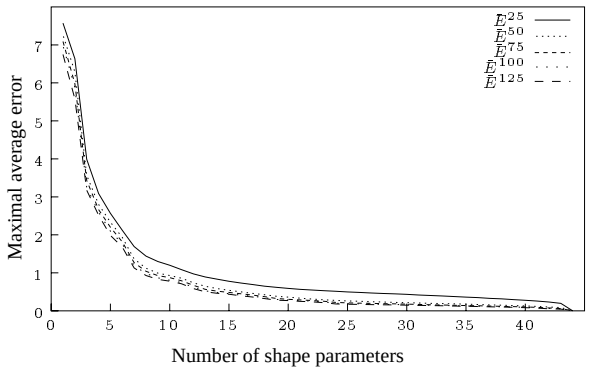


Figure 9: The maximal error as a function of the number of used shape parameters and the size of the training set for the Tulips1 database.

5.3 Experiments with the M2VTS Database

Following the same scheme, the total of 522 labels was split into training and test sets with the training sets including 50, 100, 200, 300, 400, and 500 items. Comparable to the other two data sets, the maximal error decreases with the number of increasing shape parameters and an increasing size of the training set (figure 10). In a, here omitted for space reasons, difference plot, it can be seen that the error decreases with an increasing size of the training set. Comparing the number of shape parameters required to obtain a certain error (*e.g.* 1 pixel) with those for the Tulips1 and X-ray database results in an intermediate value, close to the value of the Tulips1 database. This is sensible, as the inner rim of the lips performs similar - although not

totally dependent - movements, and therefore requires more parameters for their description.

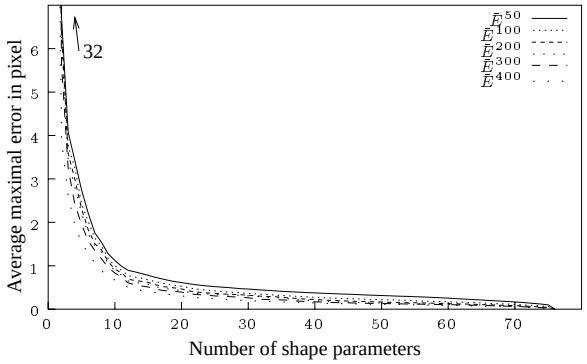


Figure 10: The maximal error as a function of the number of used shape parameters and the size of the training set for the M2VTS database.

6 Restricting Shape Vectors Further

Although the restriction of shapes to some hyperplane defined by the shape vectors reduces the amount of possible shapes, it does, for example, not limit the absolute size. Therefore, while fitting a point distribution model to a given image, the user may wish to limit the domain of valid shape vectors further by restricting each shape parameter to some subset of $\Re$, usually some finite interval. An indication of an useful size of these intervals are the eigenvalues: they are the variance of the corresponding shape parameter b_i if the model is adapted to the various samples in the database. In other words, the eigenvalue λ_i is also the variance of valid values of the corresponding shape parameter. Assuming that the distribution D_i is Gaussian-like with a zero mean, an interval $[-c, c]$ with some constant $c \gg \sqrt{\lambda_i}$ will virtually include all valid shape parameters. For example, T.F. Cootes and C.J. Taylor consider to use $c_i = 3\sqrt{\lambda_i}$ (see [Cootes-96]). Another approach to limit the b_i, is to add a penalty term for large b_i to the cost function used to fit the shape model to an image [Cootes-94]. This approach reacts more gracefully in cases where a shape corresponds best to a shape vector with some b_i slightly outside the interval $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$.

Although both approaches are simple and straightforward, and therefore easily applicable, they are not always satisfactory as the D_i are not necessarily Gaussian-like (as it is assumed by T.F. Cootes *et al.*). The reasons for this are numerous:

- An unequal distribution of object classes. Example: the shape of cars. A realistic database with images of cars will only contain a few images with Ferraris, but may showing Fords. Therefore the shape vector "Ford" will be more probable.

- An unequal distribution of object deformations. Example: randomly chosen images from movies are searched for faces and labeled for facial expressions. In this database, expressions like "laugh" or "pain" will be rare, whereas "attentive listening" will be frequent.

- A bad selection of training samples.

- Non-linear deformations of a shape. Assume a point of an object is moving along a **T**-shaped path with a relative long vertical component and an equal probability for the point being on a certain location on this path. The first eigenvector corresponding to this movement will be parallel to the dominant movement on the vertical bar of the **T**. The second eigenvector is consequently to the horizontal bar. Figure 11 idealizes this situation.

  It can be seen that the distribution of the first shape parameter is biased to the location of the horizontal bar and the second shape parameter to the vertical bar. Both distributions are non-Gaussian. A model using a restriction of possible point positions with symmetric intervals to limit the b_1 will necessaryly permit shapes with points above the horizontal bar of the **T**. Furthermore, if the standard limits are used, the model will falsely forbid the three ends.

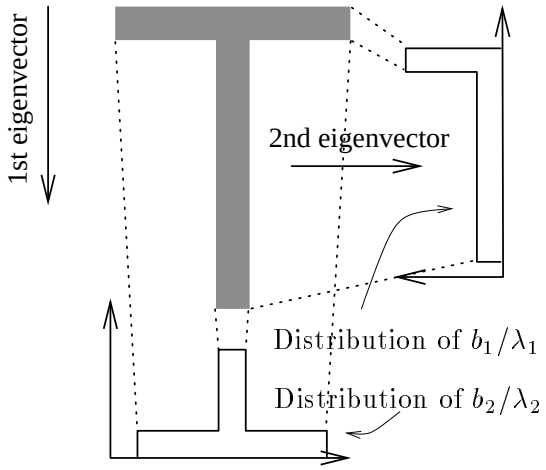


Figure 11: A non-Gaussian distribution of shape parameters can be caused by non-linear deformations.

Fortunately, the shape parameters $\mathbf{b}$, which were already calculated for the estimation of the model errors, can be used to estimate the distributions D_i of the shape parameters, or more important for this purpose, the maxima and minima of these distributions. That the distributions D_i for the example of the database of X-ray images differ considerably from a Gaussian distribution is shown in figure 12. The histograms in this figure represent the estimated and normalized distribution $\bar{D}_i$ for the two strongest eigenvectors averaged over the 30 point distribution models build from 300 samples. The normalized distributions $\bar{D}_i$ are estimated by dividing the total range of all b_i/λ_i into intervals of length $\frac{1}{10}$, counting the observed b_i/λ_i per interval, and dividing these counts by the total number of observed $\mathbf{b}$.

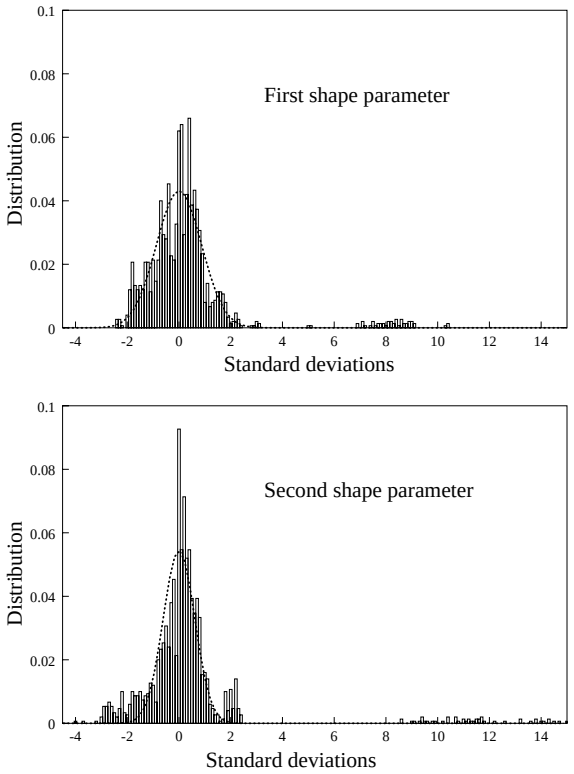


Figure 12: The *estimated distribution* $\bar{D}_i$ for the two strongest shape parameters. The dotted lines are the Gaussian curves that fit the $\bar{D}_i$ best.

A comparison of $\bar{D}_1$ and $\bar{D}_2$ with Gaussian curves (the dotted lines in figure 12) that are fitted[4] to $\bar{D}_1$ and $\bar{D}_2$, respectively, show important differences. Given the observed distributions, it is straight forward to chose for each i the maxima and minima of $\bar{D}_i$ as limits for the b_i/λ_i (or an interval that is slightly bigger).

According to this strategy, b_1/λ_1 is limited to the interval $[-3, 11]$ and b_2/λ_2 to $[-4.5, 15.5]$ for the database of X-ray images. These intervals are a better choice than any symmetric interval, as are less likely to exclude some valid shape, respectively to permit invalid ones. To give an example, figure 6 shows the increase in pixels of the average maximal model error for the point distribution model build from 300 samples if the b_i are instead to the observed $\bar{D}_i$ limited to the interval $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$. The augmentation of the average error is limited (maximal 0.6 pixel), but this difference is caused by only a few image points for very small portion of the data, for which it is therefore much more important.

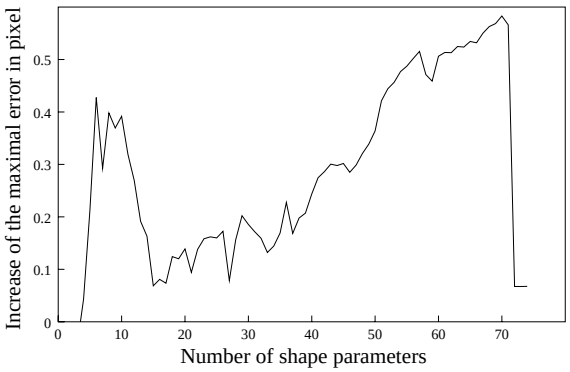


Figure 13: The increase of the average maximal model error if the b_i are $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$ for the X-ray data.

This behavior can also be observed for the other two data sets. Figure 14 and 15 show the minimal and maximal observed shape parameters if the point distribution model is optimally fitted to the shapes in the test set (the models were constructed from the training sets with 125, respectively 500 items).

As compared to the experiments performed with the X-ray data, the distribution of the most important shape vectors for the Tulips1 and the M2VTS database are closer to a Gaussian distribution. For example, figure 16 show such distributions of the first three, normalized distributions and the best fitting Gaussian distribution (the distributions were calculated using the point distribution models calculated from 500 labels). The first and second shape parameter have a non-neglectable number of observed shape parameters with values superior to 3 standard deviation, and below -3 for the second shape parameter.

[4] That is: $\sum_{b_i} (\bar{D}_\ell(b_i) - \alpha \exp(-\beta(b_i)^2))^2$ minimal for some α and β.

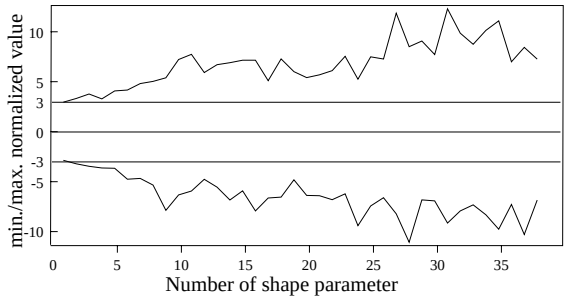


Figure 14: Minimal and maximal values of observed normalized shape parameters $b_i/\sqrt{\lambda_i}$ for the Tulips1 database. The horizontal lines correspond to 3 standard deviations.

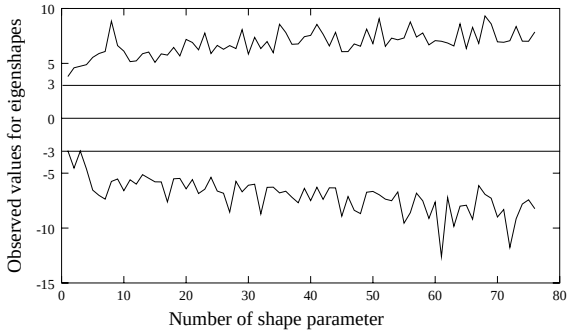


Figure 15: Minimal and maximal values of observed normalized shape parameters $b_i/\sqrt{\lambda_i}$ as compared to the 3 standard deviation limit for the M2VTS database.

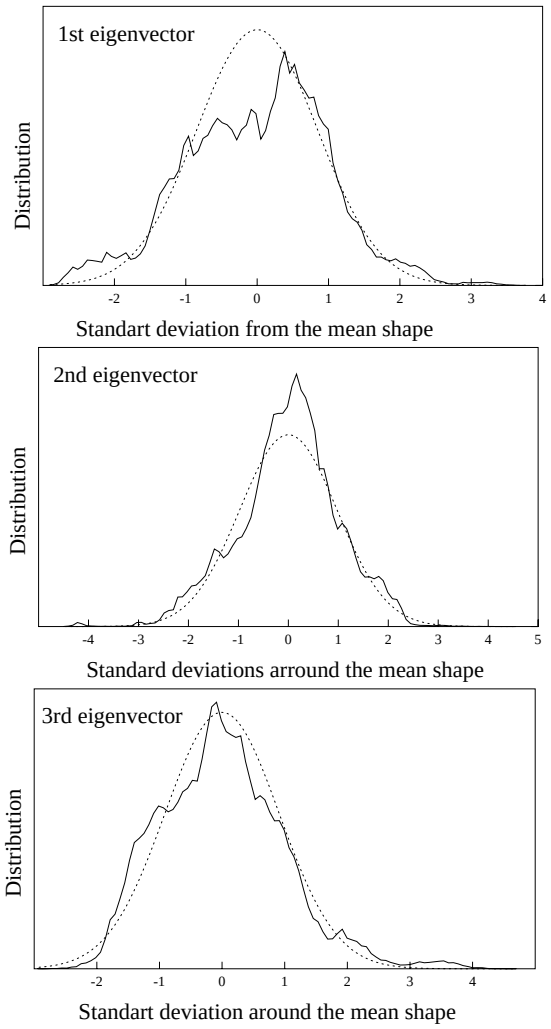


Figure 16: The distribution of observed values for the first three shape parameters for the M2VTS data set.

7 Using the Shape Parameter Distributions

Besides for defining valid shape parameters, the distributions $\bar{D}_i$ can be used outside the core recognition algorithm. This is possible, as an *a priori* probability of shapes can be estimated from the $\bar{D}_i$ under the assumption that the shape parameters b_i are independent and that the training samples are drawn randomly from the whole database (*i.e.* the distributions of the shapes of the labeled images is the same as for the whole database):

$$P(\mathbf{b}) = \prod_{i=1}^{k} P(b_i) = \prod_{i=1}^{k} D_i(b_i/\lambda_i) \qquad (6)$$

This shape probability may used in various ways, for example in tracking algorithms using probabilities [Isard-96], or in classification tasks.

In contrast to this, a user might not be interested in *a priori* probabilities of shapes, but in shape models with a very low model error. This requires a training database that represents well all possible shapes. The distributions D_i can be used to find underrepresented shapes by searching for b_i which correspond to missing values in distribution D_i.

Then, in some sample shape vector(s), a vector element is changed to a missing or underrepresented value. The resulting vector is then transformed into a shape using equation (1). A simple graphical tool can the visualize this shape, and if this shape is likely a shape of a target object, the whole database can be searched for images showing similar shapes. Labeling the found images and recalculating the shape model will result in an improved shape model.

Example: for the X-ray images database, the first shape parameter assumes almost never values

in the range of $b_1/\sqrt{\lambda_1} \in [3,7]$, and the second never values in the range $b_2/\sqrt{\lambda_2} \in [2.5,9]$ (compare figure 12). Shapes corresponding to these shape parameters must exist, as to the deformations of the vocal tract are continuous.

8 Conclusion

The proposed method for the optimal parameterization of a point distribution model permits to select an optimal number (within a small statistical error) of shape parameters. The analysis of several experiments confirms the consistence of theory and practice and show that this approach is more precise than previously used approaches.

Besides this, the method simplifies the use of point distribution models, as it establishes a straight forward criteria for the selection of the number of shape parameters. As a by-product of this method, the actual distribution of the shape parameters can be estimated. An artificial example and three real-world data sets showed that these distributions can be very asymmetric and can differ greatly as compared to a Gaussian distribution. Using the maxima and minima of the actual distribution, each shape parameter can be limited to values that are more appropriate for the application and lead to more specific models than any zero-mean symmetric interval.

If the method for the estimation of the model error is repeated with different training set sizes, it can be predicted whether or not a larger training set will increase the quality of a model. Depending on this prediction, a decision on labeling more data or instead using a another (*e.g.* non-linear) point distribution model is simplified.

The proposed method is generalizable to non-linear point distribution models and/or higher dimensional shapes: an intermediate representation of a shape with a variable number of free parameters, as well as an inverse transformation from a shape to these parameters are required. The reasonable computational complexity of the method and the small additional effort for its implementation justifies its use.

References

[Barron-94] S.S. Beauchemin J.L. Barron and D.J. Fleet. "On Optical Flow." In *Int. Conf. on Artificial Intelligence and Information-Control Systems of Robots*, pages 3–14. Bratislava, Slovakia, September 12-16, 1994.

[Black-96] S.X. Ju, M.J. Black, and Y. Yacoob. "Cardboard people: A parameterized model of articulated motion." In *2nd Int. Conf. on Automatic Face- and Gesture-Recognition*, pages 38–44. Killington, Vermont, October 1996.

[Blake-94] A. Blake, M.A. Isard, and D. Reynard. "Learning to track curves in motion." In *Proc. IEEE Int. Conf. Decision Theory and Control*, pages 3788–3793. 1994.

[Bregler-94] C. Bregler and S. Omohundro. "Surface learning with applications to lipreading." In J.D. Cowan, G. Tesauro, and J. Alspector (editors), *Advances in neural information processing systems 6*. 1994.

[Cootes-92.1] T.F.Cootes, C.J.Taylor, D.H.Cooper, and J.Graham. "Training Models of Shape from Sets of Examples." In *Proc. British Machine Vision Conference*, pages 9–18. Springer-Verlag, 1992.

[Cootes-92.2] T.F. Cootes and C.J. Taylor. "Active Shape Models - 'Smart Snakes'." In *Proc. British Machine Vision Conference*, pages 266–275. Springer-Verlag, 1992.

[Cootes-94] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam. "Use of Active Shape Models for Locating Structures in Medical Images." *Image and Vision Computing*, volume 12, number 6, pages 355–365, July/August 1994.

[Cootes-95] T.F. Cootes, C.J. Taylor, D.H. Copper, and J. Graham. "Active Shape Models — Their Training and Application." *Computer Vision and Image Understanding*, volume 61, number 1, pages 38–59, 1995.

[Cootes-96] T.F. Cootes and C.J. Taylor. "Data Driven Refinement of Active Shape Model Search." In *British Machine Vision Conference*. Edinburgh, Sept. 9-12, 1996.

[Cootes-97] T.F. Cootes and C.J. Taylor. "A Mixture Model for Representing Shape Variation." In *British Machine Vision Conference*. 1997.

[Davis-97] Malcolm Davis and Muhran Tuceryan. "Coding of Facial Image Sequences by Model-Based Optical Flow." In Nikos Sarris and Michael G. Strintzis (editors), *IWSNHC3DI'97 International Workshop on Syntetic-Natural Hybrid Coding and Three Dimensional Imaging*, (European project ACTS 057-VIDAS), pages 192–195. September 5-9, 1997.

[DeCarlo-96] Douglas DeCarlo and Dimitris Metaxas. "The Integration of Optical Flow and Deformable Models with Applications to Human Face Shape and Motion Estimation." In *CVPR'96*, pages 231–238. 1996.

[Di-Mauro-96] E.C. Di Mauro, T.F. Cootes, C.J. Taylor, and A. Lanitis. "Active Shape Models search using Pairwise Geometric Histograms." In *British Machine Vision Conference*. Edinburgh, Sept. 9-12 1996.

[Heap-97] Tony Heap and David Hogg. "Improving Specificity in PDMs using a Hierarchical Approach." In Adrian F. Clark (editor), *British Machine Vision Conference*, (BMVW). 1997.

[Hill-92] A. Hill, C.J. Taylor, and T.F. Cootes. "Object Recognition by Flexible Template Matching using Genetic Algorithms." In *2nd European Conference on Computer Vision*, pages 852–856. Santa Margherita Ligure, Italy, May 1992.

[Isard-96] Michael Isard and Andrew Blake. "Contour tracking by stochastic propagation of conditional density." In *ECCV'96*. 1996.

[Jain-96] Anil K. Jain, Yu Zhong, and Sridhar Lakshmanan. "Object Matching Using Deformable Templates." *IEEE Trans. PAMI*, volume 18, number 3, pages 267–278, March 1996.

[Jaynes-94] C. Jaynes, F. Stolle, and R. Collins. "Task Driven Perceptual Organization for Extraction of Rooftop Polygons." In *IEEE WACV*. Sarasota, FL, 1994.

[Kass-88] M. Kass, A. Witkin, and Terzopoulos. "Snakes: Active Contour Models." *International Journal of Computer Vision*, , pages 321–331, 1988.

[Li-93] Mengxiang Li. "Minimum Description Length Based 2-D Shape Description." In *International Conference on Computer Vision ICCV93*. 1993.

[Luettin-97] Juergen Luettin and Neil A. Thacker. "Speechreading using Probabilistic Models." *Computer Vision and Image Understanding*, volume 65, number 2, pages 163–178, 1997.

[McKenna-97] Stephen J. McKenna, Shaogang Gong, Rolf P. Würtz, Jonathan Tanner, and Daniel Bain. "Tracking Facial Feature Points with Gabor Wavelets and Shape Models." In *Proc. of the 1st Int. Conf. on Audio- and Video-based Biometric Person Authentication*, Lecture Notes in Computer Science. 1997.

[Movellan-96] J.R. Movellan and G. Chadderdon. "Channel Separability in the Audio-Visual Integration of Speech: A Bayesian Approach." In David G. Stork and Marcus E. Hennecke (editors), *Speechreading by Humans and Machines*, volume 150 of *NATO ASI Series, Series F: Computer and Systems Sciences*, pages 473–488. Springer Verlag, Berlin, 1996.

[Munhall-95] K.G. Munhall, E. Vatikiotis-Bateson, and Y. Tokhura. "X-ray film database for speech research." *Journal of the Acoustical Society of America*, volume 98, number 2, pages 1222–1224, 1995.

[Pigeon-97.2] S. Pigeon and L. Vandendorpe. "The M2VTS Multimodal Face Database." In J. Bigün, G. Chollet, and G.Borgefors (editors), *Lecture Notes in Computer science: Audio- and Video based Biometric Person Authentication*, volume 1206, pages 403–409. Springer, 1997.

[Sozou-94] P.D. Sozou, T.F. Cootes, C.J. Taylor, and E.C. Di Mauro. "A non-linear Generalisation of PDMs using Polynomial Regression." In *Proceedings of the 5th BMVC*, pages 397–406. 1994.

[Sozou-95] P.D. Sozou, T.F. Cootes, C.J. Taylor, and E.C. Mauro. "Non-linear Point Distribution Modelling using a Multi-layer Perceptron." In David Pycock (editor), *British Machine Vision Conference*, volume 1, pages 107–116. September 1995.

[Yuille-92] A.L. Yuille, P. Hallinan, and D.S. Cohen. "Feature Extraction from Faces Using Deformable Templates." *International Journal of Computer Vision*, volume 8, pages 99–112, 1992.

[Zhong-98] Yu Zhong, Anil K. Jain, and M.P. Jolly. "Object Tracking Using Deformable Templates." In *International Conference on Computer Vision (ICCV98)*. 1998.