# On Properties of Active Shape Models

Mikkel B. Stegmann[1], Rune Fisker & Bjarne. K. Ersbøll
Department of Mathematical Modelling
Technical University of Denmark
DTU Building 321, DK-2800 Lyngby, Denmark

Technical Report
IMM-REP-2000-12

March 2nd 2000

[1] Corresponding author: aam@imm.dtu.dk

# Chapter 1

# On Properties of Active Shape Models

## Abstract

*Contrary to many other deformable models Active Shape Models (ASM) represents a general way of performing non-rigid object segmentation. Shape variation is extracted from a training set by applying principal component analysis to point distribution models, rather than hand crafting a priori knowledge into the model.*

*In this paper we investigate different properties of ASM. Topics treated are the generation of plausible shapes, tangent space transformation and model to image fitting assisted by statistical models of gray level variation in the training set. Finally a method for automatic initialization and a comparison of four model to image fitting methods are presented. The initialization part indicates that completely automatic segmentation could be done by ASMs. The comparison part shows an improved fit for model to image fit methods based on gray level variation in the training set.*

**Keywords:** Deformable Models, Active Shape Models, Snakes, Principal Component Analysis, Statistical Models of Gray Level Variation, Model Initialization.



## 1.1 Introduction

In the field of non-rigid object segmentation the group called Deformable Models has achieved much attention. These models has proven efficient in many applications: object segmentation, appearance interpretation, motion tracking etc.

A deformable model can be characterized as a model, which under an implicit or explicit optimization criterion, deforms the shape to match a known object in a given image. For a general review of the most commonly used models refer to [10, 9]. One of the earliest and most popular deformable model, *Snakes*, is proposed by Kass, Witkin and Terzopoulos [2].

Jain et al. [8, 10] classifies deformable models as either being *free form* or *parametric* where the first denotes model probability dependent on local constraints on the shape[1] and latter global shape constraints. By building a statistical model of the global shape variation from a training set, Active Shape Models (ASM) [6] qualifies in being a parametric deformable model. The model is called a Point Distribution Model (PDM). In this way the object segmentation can verify the detected shape against the PDM and thus only allowing shape variation consistent with the training set. This kind of shape variability inside the same object class can for example be found in medical imaging.

As an extension to the basic ASM this paper investigates different properties of the model. Most of this are inspired from work done by Cootes et al. [3, 6, 15].

The actual implementation is written in C++ as a plug-in to the windows-based deformable template framework, DMT.

## 1.2 Active Shape Models

The Active Shape Model represents a parametric deformable model where a statistical model of the global shape variation from a training set is built. This model, called the point distribution model (PDM), is then used to fit a model (or template) to unseen occurrences of the object earlier annotated in the training set. We will briefly describe the construction of the PDM using

[1] Such as Snakes [2].

a principal component analysis (PCA). For a more detailed description of ASMs refer to [6, 15].

The shape itself is represented as an $n$-point polygon in images coordinates:

$$\mathbf{X} = (x_1, y_1, ..., x_{n-1}, y_{n-1}, x_n, y_n)^{\mathrm{T}} \tag{1.1}$$

To measure the true shape variation the shape $\mathbf{X}$ is transformed into a normalized frame of reference with respect to the pose parameters: $t_x, t_y$ (translation), $s$ (scaling) and $\theta$ (rotation).

$$\mathbf{x} = T_{t_x,t_y,s,\theta}(\mathbf{X}) \tag{1.2}$$

The mean shape in this aligned domain is given as:

$$\overline{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i \tag{1.3}$$

And the deviation of each shape from the mean shape:

$$d\mathbf{x}_i = \mathbf{x}_i - \overline{\mathbf{x}} \tag{1.4}$$

The estimate of the covariance matrix can now be written as:

$$\mathbf{\Sigma} = \frac{1}{m}\sum_{i=1}^{m} d\mathbf{x}_i d\mathbf{x}_i^{\mathrm{T}} \tag{1.5}$$

The principal axis of the $2n^{th}$ dimensional point cloud are now given as the eigenvectors of the covariance matrix $\mathbf{p}_i$. If the $i^{th}$ eigenvalue is denoted $\lambda_i$, the following identity holds true:

$$\mathbf{\Sigma}\mathbf{p}_i = \lambda_i \mathbf{p}_i \tag{1.6}$$

The matrix $\mathbf{P}$ are then built from each eigenvector ordered in descending order of the corresponding eigenvalues.



$$\mathbf{P} = \begin{bmatrix} \\ \mathbf{p}_1 & \dots & \mathbf{p}_{2n} \\ \\ \end{bmatrix} \tag{1.7}$$

A shape instance can then be generated by deforming the mean shape by a linear combination of eigenvectors:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{1.8}$$

The $2n^{th}$ dimensional shape space is their by spanned using it's principal axis, i.e. the dimensions are ordered according to their level of shape variance explanation.

This results in a very convenient way to compare a candidate shape $\mathbf{x}'$, to the training set by performing the orthogonal transformation into $\mathbf{b}$-parameter space and evaluating the shape probability.

A model instance is now defined by it's model vector $\mathbf{v}$, which consists of the pose and shape parameters.

$$\mathbf{v} = \{t_x, t_y, s, \theta, \mathbf{b}\} \tag{1.9}$$

### 1.2.1 Choosing Modes of Variation

The primary goal of applying a principal component analysis (PCA) to the training set is to reduce the number of parameters in our model.[2] In this way the model parameters can be limited to only generate shapes similar to the ones contained in the training set.

By organizing the eigenvalues of the covariance matrix from the training shapes in descending order, $t$ modes of variation can be chosen to explain $V \times 100\%$ of the shape variation using:[3]

[2]This approach has several other advantages: An analysis of the major shape characteristics are often of great interest and thus the PCA enables easy object discrimination based on the training set.

[3]Since the variance of $i^{th}$ principal component equals $\lambda_i$ [1].

$$\sum_{i=1}^{t} \lambda_i \geq V \sum_{i=1}^{2n} \lambda_i \qquad (1.10)$$

The remaining $2n - t$ modes are then considered shape noise. A suitable value for $V$ could be 0.98; hence 98% of the shape variation can be modelled.[4]

### 1.2.2 Alignment of Training Shapes

To obtain a frame of reference for the alignment of shapes our previous work [16] translated all shapes from their center of gravity to origo and scaled them to unit scale, $|\mathbf{x}| = 1$. In this way the corners of a set of aligned rectangles with varying aspect ratio forms a unit circle (see fig. 1.2, the unaligned shapes are shown on fig. 1.1). Due to this non-linearity the PCA must use two parameters to span the shape space: $\lambda_1 = 99.6\%$, $\lambda_2 = 0.4\%$ even though variation only exists on one parameter (the aspect ratio). A closer look on figure 1.2 also shows that the overlaid mean shape doesn't corresponds to an actual shape in the training set. To avoid these non-linearities in the aligned training set the shape can be transformed into *tangent space* by scaling by $1/\mathbf{x}.\overline{\mathbf{x}}$ [3, 4].

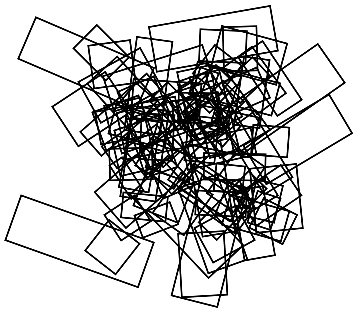

Figure 1.1: Training set of 100 unaligned artificially generated rectangles containing 16 points each.

[4] Our previous work was hard coded to use the first 5 modes of variation.



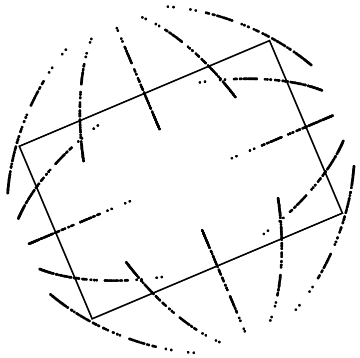

Figure 1.2: Point cloud from aligned rectangles sized to unit scale, $|\mathbf{x}| = 1$. The mean shape is fully shown.

The transformation into tangent space aligns all rectangles with corners on straight lines (see fig. 1.3) and thus enabling modelling of the training set using only linear displacements.

Notice how the mean shape is contained in the training set since the PCA now only uses one parameter, $\lambda_1 = 100\%$ to model the change in aspect ratio.

In this way the distribution of $b$-parameters can be kept more compact and non-linearities can be reduced. This leads to better and simpler models.

### 1.2.3 Generation of Plausible Shapes

In the process of matching a model to an unseen image only plausible shapes compared to the training set are of interest.

One way to determine this is to impose hard limits on the shape parameters, $\mathbf{b}$, under the model-assumption that the $b$-parameters are independent gaussian distributed with zero mean. Since the variance of $i^{th}$ principal component is $\lambda_i$ – and 98% of distribution of $b_i$ is covered in the range $\pm 3\sigma$ – the limits can be chosen as:

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i} \qquad (1.11)$$

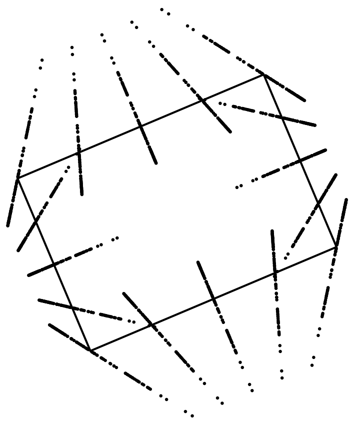

Figure 1.3: Point cloud from aligned rectangles sized to unit scale, $|\mathbf{x}| = 1$, and transformed into tangent space. The mean shape is fully shown.

This was the approach used in the our previous work. Due to the simple hypercube restriction it allows every $b$-parameter simultaneously to take the value of $\pm 3\sqrt{\lambda_i}$ which is highly unlikely.

To avoid this the $b$-parameters can be restricted to a hyperellipsoid using the Mahalanobis distance.

$$D_m^2 = \sum_{i=1}^{t} \frac{b_k^2}{\lambda_k} \leq D_{max}^2 \tag{1.12}$$

such that a $D_m$ is smaller than a suitable $D_{max}$ corresponds to a plausible shape. As a suitable value for $D_{max}$, 3.0 could be used.

If the shape fails this test, $\mathbf{b}$ is rescaled to lie on the closest point of the hyperellipsoid. This is illustrated in the two-dimensional case in figure 1.4.

$$\mathbf{b} = \mathbf{b} \cdot \frac{D_{max}}{D_m} \tag{1.13}$$

If the shape class in question is separated in distinct subclasses of which we need no discrimination[5] more complex methods must be used to model the

[5]For example due to a part of the object that only can reside in discreet positions.



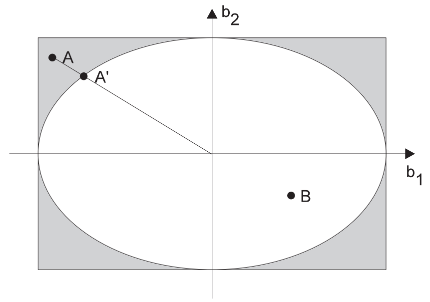


Figure 1.4: The effect of using the Mahalanobis distance in two dimensions. Shape B is valid, shape A is considered illegal and rescaled to A'

distribution of the $b$-parameters. One approach is to use a approximation to the distribution through a mixture of gaussians [4]. This approach can be used to represent any non-linear shape variations in the training set and their by control the generation of plausible shapes in a much more general way.

## 1.3 Model to Image Fit

Once the point distribution model has been built from the training set; a functional is needed for evaluating the fit of current shape model to the current (unseen) image. That is the probability of a shape model identified by it's shape and pose parameters $\mathbf{v}$ given an image $\mathbf{I}$; $P(\mathbf{v}|\mathbf{I})$.

Given $P(\mathbf{v}|\mathbf{I})$ a general purpose optimization method could be used to optimize $\mathbf{v}$ under the given constraint of plausible shapes. In practice however; this isn't a simple optimization problem. Under the assumption of a reasonable initialization of the model, the Active Shape Models attacks the problem more locally by using the image appearance around each model point to calculate an independent movement and fit quality. In this way the shape $\mathbf{X}$ can be transformed into the locally optimal shape $\mathbf{X}'$ and used in an iterative optimization scheme [6, 15]. This scheme substitutes an explict optimization criteria, $P(\mathbf{v}|\mathbf{I})$, contrary to many other deformable models.

The area around each model point; e.g. *the point neighborhood* is in this paper defined as a line normal to the model boundary, also called a 1D pixel profile (shown at figure 1.5, 1.6). This requires a definition of the connectivity of each model point. In figure 1.5 this is rather simple since this template consists of one closed path. A little more book keeping is needed when models consists of open and/or multiple paths.

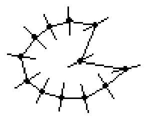

Figure 1.5: A shape model with overlaid 1D pixel profiles normal to the boundary.

### 1.3.1 Simple Edge Detection

A very simple assumption concerning the optimal placement of our model points is to look for strong image edges in the 1D pixel profiles of the points. In this way the points in the training set are all assumed to lie on the strongest edges.

In the following we call this assumption concerning the optimal model points placement an *image matching criteria*.

Like [3] the derivative is approximated as a convolution of pixel profile with a [1 0 -1] kernel. If $\mathbf{P}$ denotes the pixel profile the $i^{th}$ element of the derived profile is:

$$g_i = P_{i+1} - P_{i-1} \tag{1.14}$$

Each model point is now moved to the position of $[\max | \min](\mathbf{g})$ to form the optimal shape $\mathbf{X}'$ depending on the *a priori* knowledge of edge orientation. If the orientation is unknown the maximum absolute value of $\mathbf{g}$ is used.

Since the 1D pixel profile is far from aligned to the pixel-grid an interpolation method must be considered. In figure 1.6 we observe the unpleasant



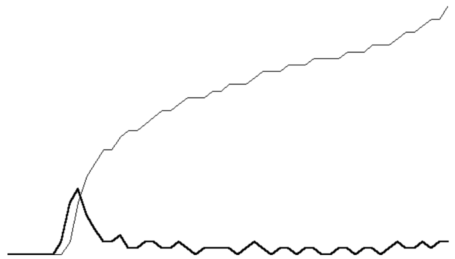

Figure 1.6: Example of a pixel profile (thin line) using nearest neighbour sampling and the approximated edge strength $w_{edge}$ (thick line) using [1 0 -1] convolution.

stair casing of a nearest neighbor interpolation. To avoid this bilinear interpolation of the pixel values are used at the cost of a small performance penalty.

To estimate *the goodness of fit*, Cootes et al. propose the following error measure for this simple image matching criteria [15]:

$$E(\mathbf{v}, \mathbf{I}) = |\mathbf{X}' - \mathbf{X}|^2 \tag{1.15}$$

where $\mathbf{X}$ denotes the shape generated from the pose and shape parameters in $\mathbf{v}$ and $\mathbf{X}'$ denotes $\mathbf{X}$ fitted to nearby strong edges in $\mathbf{I}$.

We stress that this error measure implicitly demands that all model points in the training set really are placed on the strongest nearby edges and the edge detector doesn't fail in the current image $\mathbf{I}$ due to clutter, noise etc. Otherwise the error measure will yield a false minimum.

### 1.3.2 Statistical Models from Gray Level Variation

The *strongest edge* assumption above imposes severe constraints on the types of images in the training set. In general model points could lie on a secondary edge or, more difficult, on the border of a change in texture etc.

By using the a priori knowledge in terms of pixel information in the training set we can build statistical models of the gray level variation around model points and their by archive a more robust and generalized model fit in an unseen image. This has previously been done successfully by many others [3, 5, 7, 15].

In the current work gray level models are built from a 1D pixel profile normal to the actual model point, although any area around the image point could be considered.

Consider the model point $p$ $(1 \leq p \leq n)$ in the $i^{th}$ training image where $k$ pixels are samples on the normal in an equal amount on each side. To make the model less sensitive to global changes in intensity, the derivative of the pixel profile is used:

$$g'_{pi} = \frac{dP_{pi}}{dx} \tag{1.16}$$

followed by a scale normalization:

$$\mathbf{g}_{pi} = \frac{1}{\sum_{j=1}^{k} |g_{pij}|}\mathbf{g}'_{pi} \tag{1.17}$$

If all profiles for point $i$ are assumed to be multivariate gaussian distributed and $m$ denotes the total number of training images[6] the estimates of the mean profile and the covariance matrix for the $p^{th}$ point is then given as:

$$\overline{\mathbf{g}}_p = \frac{1}{m}\sum_{i=1}^{m} g_{pi} \tag{1.18}$$

$$\mathbf{S}_p = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{g}_{pi} - \overline{\mathbf{g}}_p)(\mathbf{g}_{pi} - \overline{\mathbf{g}}_p)^T \tag{1.19}$$

Their by we have a statistical model for the gray level variation of each model point and can determine the quality of a new profile, $s_p$, as the probability that this profile, $s_p$, comes from the $p^{th}$ distribution. Maximizing this probability is equal to minimizing the Mahalanobis distance of the sample from the mean:

$$f(\mathbf{s}_p) = (\mathbf{s}_p - \overline{\mathbf{g}}_p)^T \mathbf{S}_p^{-1} (\mathbf{s}_p - \overline{\mathbf{g}}_p)^T \tag{1.20}$$

[6] Assuming that each image contains one shape.



In practice $K$ pixels are sampled on the normal in the unseen image ($K > k$). Then we extract a subrange of $k$ pixels, form the derivative, perform the normalization and then evaluate the $f(\mathbf{s}_p)$ term $K - k$ times to find the best fit.

Regarding the choice of $K$ we stress the importance of size-invariance both in the model fit and the generation of the gray level model. To achieve this the profile must be defined in terms of a number of samples, $K$ and a sample density (samples/pixel) estimated on basis of the current template size. This enhancement is neither implemented in our work, nor mentioned in any of Cootes et al. descriptions of ASMs [3, 6, 15].

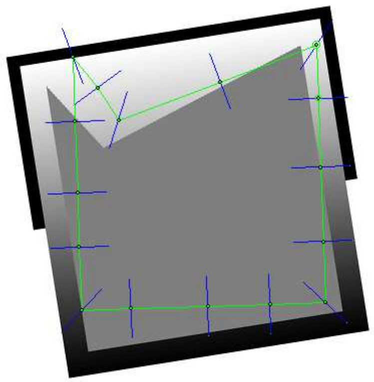

Figure 1.7: Initial shape.

Figure 1.7 – 1.9 demonstrates on a synthetic data set the difference between using the strongest edges and a statistical model of gray level variation in the ASM optimization process where the optimal shape doesn't lie on the strongest edges.

As observed on figure 1.8 the optimal shape doesn't lie on the strongest edges due to the conflict between the locally optimal point placements and the maximum allowable shape variation[7].

[7] The training set in this case consist only of very subtle shape variation.

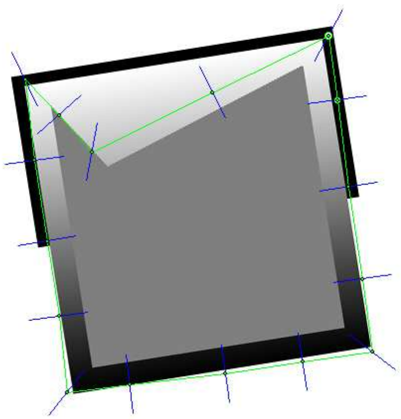

Figure 1.8: Optimization using maximum absolute edges as image matching criteria.

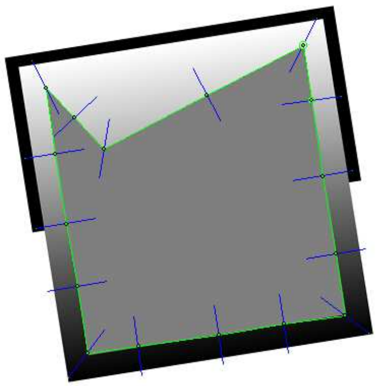

Figure 1.9: Optimization using a statistical model of gray level variation obtained from the training set as image matching criteria.

### 1.3.3 Specific Statistical Models from Gray Level Variation

As variations on the above section (1.3.2) consider three alternatives to building the statistical model of gray level variation concerning the type of



pixel profiles used:

**Type 1** Normal pixel profiles (no derivation)
**Type 2** Derived pixel profiles using a [1 0 -1] convolution.
**Type 3** Derived pixel profiles using convolution with the derivative of a gaussian.

Type 1 uses the absolute pixels values and type 2 uses a simple approximative derivation like the one used in the previous section.

Since type 2 is very sensitive to noise a gaussian convolution is used to enhance the large scale trends of the profile.

If $g(x, \sigma^2)$ denotes a gaussian distribution with variance $\sigma^2$ and mean $\mu = 0$:

$$g(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} \tag{1.21}$$

the pre-smoothing can then be expressed as:

$$L(x, \sigma^2) = I(x) * g(x, \sigma^2) \tag{1.22}$$

To find the derivative of $L(x, \sigma^2)$ the following identity is used:

$$\partial_x L(x, \sigma^2) = (\partial_x g(x, \sigma^2)) * I(x) \tag{1.23}$$

The gaussian is then derived symbolic and thus avoiding the discrete approximative derivative.

As error measures for the model to image fit of type 1-3 the sum of the minimum mahalanobis distances for each point could be used.

$$E(\mathbf{v}, \mathbf{I}) = \sum_{p=1}^{n} \min(\ f(\mathbf{s}_p)\ \forall\ \mathbf{s}_p\ ) \tag{1.24}$$

In other words; we generate all profiles for the $p^{th}$ point, calulates the mahalanobis distances and add the minimum distance to the sum.

## 1.4 Initialization

To perform the initialization of Active Shape Models, a search strategy [10, 11, 12, 13, 14] is applied. This section follows the initialisation part by Fisker et al. in [14].

The concept of the search strategy is to perform a sparse search in the parameter space of $\boldsymbol{v}$. In practice the search is done by shifting different configurations of the template around the image $\boldsymbol{I}$, and calculate the posterior energy $U(\boldsymbol{v}|\boldsymbol{I})$ at each position. These configurations are defined as the search configurations. The initial configurations are then extracted from the calculated energies. The full initialization algorithm can be summarized as:

1. Create relevant search configurations $(s_1, \theta_1, \boldsymbol{b}_1), \ldots, (s_k, \theta_k, \boldsymbol{b}_k)$ and let $i = 1$.
2. Calculate $U(\boldsymbol{v}_i|\boldsymbol{I})$ for the center of the template $(r, c)$ corresponding to each pixel, where $\boldsymbol{v_i} = (r, c, s_i, \theta_i, \boldsymbol{b}_i)$.
3. $i = i + 1$. Go to 2 if $i \leq k$.
4. Extract the initial configurations from the calculated values of $U(\boldsymbol{v}|\boldsymbol{I})$.

The actual choice and number of search configurations $(s_1, \theta_1, \boldsymbol{b}_1), \ldots, (s_k, \theta_k, \boldsymbol{b}_k)$ is determined by the amount of variation in scale, orientation and shape variation of the actual ASM combined with the overall demand for precise initialization.

The final step is to extract the initial configurations from the calculated $U(\boldsymbol{v}|\boldsymbol{I})$. Based on the assumption that only one object of interest is situated in the search area, the initial configuration is determined as the global minima in posterior energy $U(\boldsymbol{v}|\boldsymbol{I})$.

A disadvantage of the search strategy is the high computational cost, but the computation time is significantly reduced by a filter interpretation of the posterior function [12], such that the posterior probability is calculated for the center of the filter equal to each pixel in the image by a correlation of the filter and the image $I$.

In this way all repeated calculations: pixel profile normal points, bilinear interpolation of the normals etc. can be cached efficiently and presumably significant decrease the computational cost. Since filter convolutions are highly optimized and often done in hardware this will presumably overcome the penalty induced by the fact that the area of interest around a model



(i.e. the point normals) are relatively sparse compared to the total filter size.

As a suitable alternative to filter convolutions with large kernels, Fisker et al. uses a Fourier space approach [13] in their seach strategy.

We notice that this linear filter approach calls out for a simple posterior estimator like the oriented edge strength estimator.

### 1.4.1 Initialization using Model to Image Fit Functions

As a posterior energy function for initialization of ASMs; the model to image fit functions could be used. With a future filter representation in mind, the model points are not moved along the profiles. The posterior estimate is thus only based on the fixed model points.

#### Simple Edge Detection

Here the posterior energy is estimated as a sum of an approximated edge strength in each model point in the actual image:

$$U(\boldsymbol{v}|\boldsymbol{I}) = \sum_{p=1}^{n} |p_{left} - p_{right}| \tag{1.25}$$

#### Statistical Models from Gray Level Variation

Since the statistical models from gray level variation captures image appearence from $k$ pixels centred on model point profile (from the training set), the posterior are estimated as the mahalanobis distance of a $k$ pixels profile $\mathbf{s}_p$ centered on the model point in the actual image:

$$U(\boldsymbol{v}|\boldsymbol{I}) = \sum_{p=1}^{n} f(\mathbf{s}_p) \tag{1.26}$$

## 1.5 Experimental results

As a demonstration of the implementation at work, this paper includes examples of autonomous initialization of a metacarpal model on x-ray images and segmentation of pork carcasses using the four different model to image fit types. This is not an exhaustive verification but rather a brief presentation of the possibilities induced by the enhancements to the previous work.

### 1.5.1 Initialization in X-ray Images

From a set of 20 x-ray images with 20 different human hands an Active Shape Model of the metacarpal bones (2-5) has been built. In the following we will consider metacarpal 5 only.[8]

The model was trained on 20 images of metacarpal-5 bones annotated with 200 points each. The principal component analysis reduced the number of parameters from 400 to 14 modes of shape variation under a 98% variation constraint. The shape variation from the first three modes was $\lambda_1 = 32.2\%$, $\lambda_2 = 25.7\%$, $\lambda_3 = 12.4\%$.

A model of gray level variation was then built from the training set using 10 pixels on each side of each model point.

The model was tested by presenting it with an unseen image and a marked subregion seen in figure 1.11. This region was then searched using the mean shape calculated from the training set and the initialization technique described.

The mean shape from the training set was chosen as the only search configuration $(s, \boldsymbol{\theta}, \boldsymbol{b})$ using The Simple Edge Detection posterior energy from section 1.4.1. As mentioned earlier the initialization method is based on the optimization of a posterior energy map (see fig. 1.10) connected to the current shape (template) and it's pose.

During the optimization 30 pixels on each side of each model point were searched using type 2 as the image matching criteria. A fixed number of 50 iterations was used.

[8] Roughly speaking the bone that connects the pinky finger to the wrist.



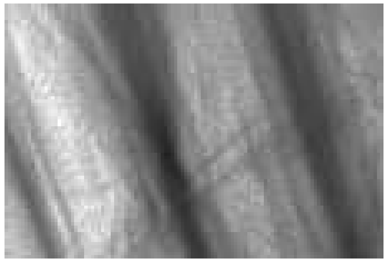

Figure 1.10: Posterior energy image, $U(\boldsymbol{v}|\boldsymbol{I})$. Generated by moving the template in the area shown in fig. 1.11.

Figure 1.11: Search region of the total bone image and initial template placed using the posterior map seen on fig. 1.10. The posterior image region is marked with a white rectangle.

Notice that the mean shape on fig. 1.11 is very similar to the actual occurrence of the metacarpal 5 bone; hence a rather good initialization is gained. Only the upper left part and the lower part of the bone is noticeably off. This is corrected nicely after optimization (see in fig. 1.12). The mean distance from model points to associated border of a ground truth annotation

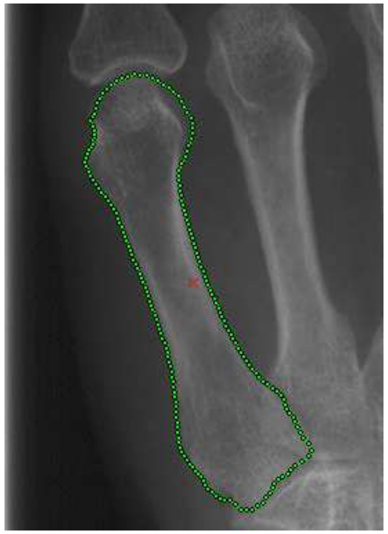

Figure 1.12: Optimized template using the initialization from fig. 1.11.

was 0.92 pixels.

To supplement the above experiment; the training set was broken down into 12 training images and 9 images with ground truth annotations. Furthermore the annotations were subsampled with factor of 4 yielding 50 model points on each shape. The remaining setup was similar to the above. The shape variation from the first three modes was $\lambda_1 = 42.5\%$, $\lambda_2 = 22.9\%$, $\lambda_3 = 10.1\%$. Initialisation, optimization and average point to associated border measurements were then performed. Results can be seen below.

In image 6 and 8 the initialization failed causing no convergence in the optimization. The improved values were archived by contracting the search area to engulf only metacapal 5. The high value in image 4 was caused by poor optimization of the lower part of the bone (at the wrist) where the x-ray is rather diffuse.

The above experiments has shown that even with only one search configuration and a simple posterior energy estimator the described search strategy



| **Image** | **Average point to associated border distance (pixels)** |
|---|---|
| 1 | 1.63955 |
| 2 | 2.07456 |
| 3 | 2.89751 |
| 4 | 7.24049 |
| 5 | 1.72735 |
| 6 | 86.66139 / 7.24290 |
| 7 | 3.18973 |
| 8 | 55.95096 / 3.06622 |
| 9 | 2.04075 |

Table 1.1: Average point to associated border distance measurement for 9 optimizations of an automatic initialized Metacarpal 5 ASM.

succesfully enables automatic initialisation of Active Shape Models.

Further experiments has shown that the subsampling of model points has no significant impact on the quality of the initialization and optimization. Thus further work should concentrate on verifing that a larger training set – that will lead to more shape flexibilty – also will improve the quality of the initialization and optimization.

### 1.5.2 Segmentation of Pork Carcasses

To stress the difference between the model to image fit functions based on edge assumptions and model to image fit functions based on gray level variation in the training set we have applied the Active Shape Model on a set of ten images of pork carcasses.

The pork carcasses was marked with 83 points each, in ten images, using the method proposed by Duta et al. [17]. The principal component analysis reduced the number of parameters from 166 to 9 modes of shape variation under a 98% variation constraint. The shape variation from the first three modes was $\lambda_1 = 38.4\%$, $\lambda_2 = 23.9\%$, $\lambda_3 = 9.6\%$.

The aligned pork carcasses can be seen on fig. 1.13.

Our model was then presented to four unseen images of pork carcasses containing a ground truth annotation. Each of the four model to image fit

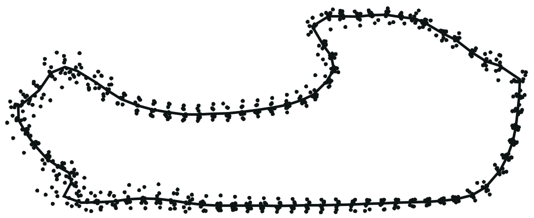

Figure 1.13: Point cloud from aligned pork carcasses. The mean shape is fully shown.

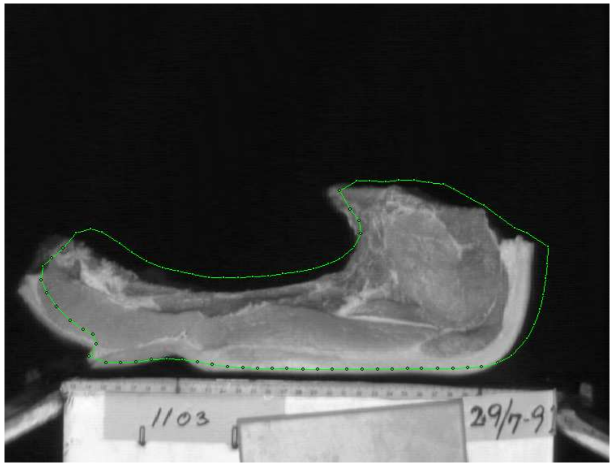

Figure 1.14: Initial pose and shape.

functions: max. abs. edge and type 1-3 was applied in turn to measure the final quality of fit after optimization using 50 iterations. Since the variance on the pose parameters in the training set was rather low, the mean pose and shape was used as initial configuration (see fig. 1.14 and 1.15).

As seen on figure 1.16 the fit error measured as the mean pixel distance from point to associated border is rather high. Even though we have a high distance it is obvious that the model to image fit functions based on gray level variation in the training set performs significantly better than the max. abs. edge fit function. The outlier, type 3 in image F1103, was due to poor initialization; the optimization did not converge. A translation



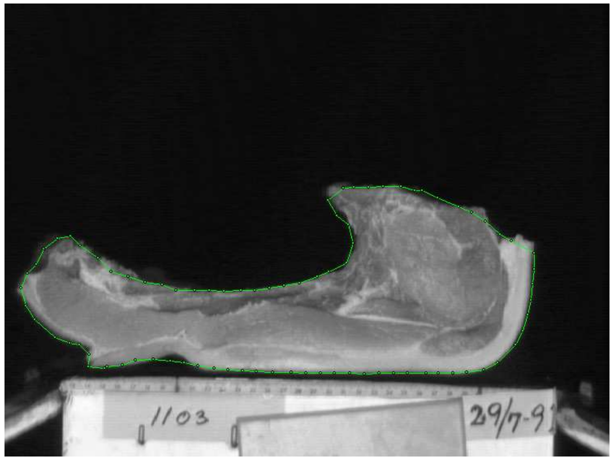

Figure 1.15: Optimized template using a type 2 model to image fit function.

of the initial pose 20 pixels to the left, converged resulting in an error of 2.82 pixels.

To investigate the model to image fit functions further we suggest:

- Relax the hyperellipsoid $b$-parameter constraints. It was obvious that the relatively small training set didn't contain enough shape variation to allow the shapes in the 4 new images, or:
- Use a larger training set with more shape variation.
- Use a training set containing differences in the global illumination. This will test the claimed robustness of model to image fit functions: type 2 & 3.

## 1.6 Conclusion

We have in this paper presented an investigation of various properties of Active Shape Models and extended our original ASM implementation [16].

All of these has been implemented in a Windows-based C++ deformable template framework and used to produce illustrations and experimental results.

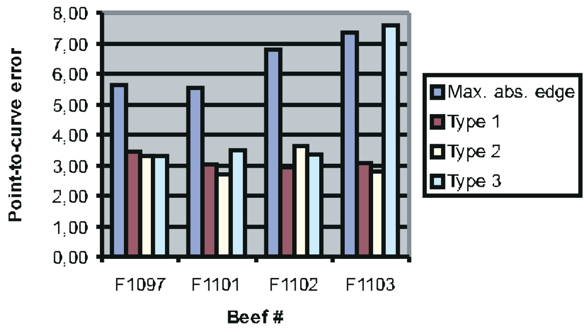


Figure 1.16: Mean pixel distance from point to associated border in four unseen images for each of the four model to image fit functions.

Topics treated were the choice of shape modes based on variation constraints, generation of plausible shapes using a hyperellipsoidal restriction on the **b**-parameters, compression of the **b**-parameter space using tangent space transformation and model fitting assisted by different statistical models of gray level variation in the training set.

This investigation has lead to a better understanding of the Active Shape Models and to theoretical and practical enhancements of the model in terms of performance, robustness and flexibility. Further more major parts of the original implementation has been revised.

Concerning model to image fitting methods we conclude that gray level model fitting leads to an improved fit compared to model fitting based on simple edge assumptions.

Futher more a scheme for automatic initialization of Active Shape Models based on model to image fit functionals has been presented. Preliminary results using only one search configuration indicates that this method could be sufficient to successfully initialize Active Shape Models on a completely general and automatic basis.



## Acknowledgements

Metacarpal images were provided by M.D. Lars Hyldstrup of Hvidovre Hospital and annotated by M.Sc., Ph.D. Hans Henrik Thodberg of Pronosco, Torsana Osteoporosis Diagnostics A/S.

Cross-section images of pork carcasses were provided by the Danish Slaughter-Houses and annotated by M.Sc. Eng. Rune Fisker, DTU and clustered by M.Sc. Nicolae Duta, Michigan State University.

# Bibliography


[1] K. Conradsen, *En introduktion til statistik - bind 2B.* IMSOR Textbook, Technical University of Denmark, 1984.

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

[3] T. F. Cootes, A. Hill, C. J. Taylor, J. Haslam, *The Use of Active Shape Models For Locating Structures in Medical Images*, Image and Vision Computing, Vol. 12(6), July 1994 pp. 355–366.

[4] T. F. Cootes, C. J. Taylor, *A mixture model for representing shape variation.* Image and Vision Computing, Vol. 17, pp. 567-573.

[5] J. Haslam, C. J. Taylor, and T. F. Cootes. *A probabalistic fitness measure for deformable template models.* In E. Hancock, editor, 5th British Machine Vision Conference, pages 33-42, York, England, Sept. 1994. BMVA Press.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, *Active Shape Models - Their Training and Application.*, Computer Vision and Image Understanding, Vol. 61, No. 1, January, pp. 38-59, 1995.

[7] T.F.Cootes, G.J.Page, C.B.Jackson, C.J.Taylor, *Statistical Grey-Level Models for Object Location and Identification*, In Proc. British Machine Vision Conference. (Ed. D.Pycock), BMVA Press, 1995, pp. 533-542.

[8] A. K. Jain, Y. Zhong, S. Lakshmanan, *Object matching using deformable models.* IEEE Transactions on Pattern Analysis and Machine Vision, 18(3) pp. 267-278, 1996.

[9] T. McInerney, D. Terzopoulos, *Deformable Models in medical image analysis: a survey*, Medical Image Analysis, vol. 2(1), pp. 91-108, 1996.

[10] Jain, A. K. and Zhong, Y. and Dubuisson-Jolly, M.-P., *Deformable template models: A review*, Signal Processing, vol. 71(2), pp. 109-129, 1998.

[11] G. J. Edwards, T.F. Cootes, C. J. Taylor, *Advances in Active Appearance Models*,Proc. Int. Conf. on Computer Vision, pp. 137-142, 1999





[12] R. Fisker, J. M. Carstensen, K. Madsen, *Initialization and Optimization of Deformable Models*, Proc. 11th. Scandinavian Conf. on Image Analysis, pp. 295-302, 1999.

[13] Fisker, R. and Schultz, N. and Duta, N. and Carstensen, J.M., *A General Scheme for Training and Optimization of the Grenander Deformable Template Model*, Vol 1, pp. 698-705, IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2000.

[14] R. Fisker, J. M. Carstensen, M. F. Hansen, S. Mørup and F. Bødker, *Automatic Segmentation and Analysis of Nanoparticle Images using an Elliptic Deformable Template Model*, under preparation.

[15] T. F. Cootes, C. J. Taylor, *Statistical Models of Appearance for Computer Vision*, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, University of Manchester – http://www.wiau.man.ac.uk, May 20th 1999.

[16] M. B. Stegmann, H. P. Palbøl, *ASM - En Active Shape Model implementation*, Department of Mathematical Modelling, Technical University of Denmark, August 1999.

[17] N. Duta and A. K. Anil and M.-P. Dubuisson-Jolly, *Learning 2D shape models*,Proc. Conf. on Computer Vision and Pattern Recognition, vol. 2, pp. 8-14, 1999.