On parameter estimation in deformable models

Fisker, Rune; Carstensen, Jens Michael

Published in:
Proceedings of the 14th International Conference on Pattern Recognition

Link to article, DOI:
10.1109/ICPR.1998.711258

Publication date:
1998

Document Version
Publisher's PDF, also known as Version of record

On parameter estimation in deformable models

Rune Fisker and Jens Michael Carstensen
Department of Mathematical Modelling
Technical University of Denmark
DTU Building 321, DK-2800 Lyngby, Denmark
rf@imm.dtu.dk and jmc@imm.dtu.dk

Abstract

Deformable templates have been intensively studied in image analysis through the last decade, but despite its significance estimation of model parameters has received little attention. We present a method for supervised and unsupervised model parameter estimation using a general Bayesian formulation of deformable templates. In the supervised estimation the parameters are estimated using a likelihood and a least squares criterion given a training set. For most deformable template models the supervised estimation provides the opportunity for simulation of the prior model. The unsupervised method is based on a modified version of the EM algorithm. Experimental results for a deformable template used for textile inspection are presented.

1. Introduction

The general idea of deformable models is that a structure embedded in the image can be considered as a deformation of a given template. The deformable model is a Bayesian combination of two parts. One part which represents the prior knowledge about the structure, i.e. the deformable template, and a second part which represent the interactions with the observations (the image), the observation model. Deformable templates can roughly be separated into 2 groups: Free form and Parametric. Free form deformable templates have no explicit global structure because the prior only contains local continuity and smoothness constrains [3, 8, 9, 10]. In parametric deformable models prior knowledge of the global structure is included using a parameterized template of a specific structure [2, 5, 6, 7, 11].

Another characteristic of deformable models is that a number of model parameters, which gives the relative influence of different terms in the model, have to be selected. Though all deformable models contains model parameters the selection of these model parameters has received very little attention. [8, 11] don't comment on the selection and [2, 3, 5, 7] selects the model parameters based on empirical observations. [10] gives guidelines for choosing the optimal parameters in the prior model based on bounds for the parameters. Only [9] uses an unsupervised method based on a minimax criterion to determine the regularization parameter which gives the relative influence of the prior and observation part, respectively. We present a new method for supervised and unsupervised selection of all model parameters in a deformable model. This method also gives the opportunity for simulating the prior model for most deformable models. Section 2 contains a general Bayesian formulation of deformable models. Then the method for supervised and unsupervised model parameter estimation is presented, followed by experimental results and a conclusion.

2. Deformable models

A structure is modelled by a template that is uniquely described by a set of template parameters $\boldsymbol{v} = (v_1, v_2, ..., v_p)$. Let $\boldsymbol{\Omega}$ denote the parameter space of $\boldsymbol{v}$. Using Bayes theorem deformable models can be described by the posterior probability $P(\boldsymbol{v}|\boldsymbol{y})$ of a realization of the template parameters $\boldsymbol{v}$ given an image $\boldsymbol{y}$. The posterior probability $P(\boldsymbol{v}|\boldsymbol{y})$ is defined as:

$$P(\boldsymbol{v}|\boldsymbol{y}) = \frac{P(\boldsymbol{v})P(\boldsymbol{y}|\boldsymbol{v})}{\sum_{\boldsymbol{v}\in\Omega} P(\boldsymbol{v})P(\boldsymbol{y}|\boldsymbol{v})} \tag{1}$$

where $P(\boldsymbol{v})$ is the prior model and $P(\boldsymbol{y}|\boldsymbol{v})$ is the observation model. [3, 8, 9, 10, 11] talks about energy and not about probabilities, but basically the formulations are analogous.

2.1. The prior model

The prior model represents the prior knowledge of the template parameter distribution. Typical the prior model consists of a number terms which represent different types of

prior knowledge. Let the prior probability be Gibbs distributed and given by:

$$P(\boldsymbol{v}) = \frac{1}{Z_p(\boldsymbol{\alpha})} exp\{-U(\boldsymbol{v},\boldsymbol{\alpha})\} \quad (2)$$

where $Z_p(\boldsymbol{\alpha}) = \sum_{\boldsymbol{v}\in\Omega} \exp\{-U(\boldsymbol{v},\boldsymbol{\alpha})\}$ is the normalizing constant, $U(\boldsymbol{v}) : \boldsymbol{\Omega} \mapsto \boldsymbol{R}$ is the energy function representing the prior knowledge and $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n) \in \boldsymbol{\Gamma_\alpha}$ are the prior parameters which gives the relative influence of the different terms.

Theorem 1: Let $U(\boldsymbol{v},\boldsymbol{\alpha}) = \boldsymbol{v}'\boldsymbol{A}(\boldsymbol{\alpha})\boldsymbol{v} + \boldsymbol{b}(\boldsymbol{\alpha})\boldsymbol{v} + c(\boldsymbol{\alpha})$ depending only on $\boldsymbol{v}$ and $\boldsymbol{\alpha}$. If $\boldsymbol{A}(\boldsymbol{\alpha})$ is symmetric and positive semi-definite and $\boldsymbol{b}(\boldsymbol{\alpha})$ belongs to the subspace defined by a linear mapping with $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ then $P(\boldsymbol{v}) \in N(\boldsymbol{\mu}(\boldsymbol{\alpha}), \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$ and:

$$\begin{array}{l} P(\boldsymbol{v}) = \frac{1}{\sqrt{2\pi^p}} \frac{1}{\sqrt{det\boldsymbol{\Sigma}(\boldsymbol{\alpha})}} \times \\ exp\{-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{\mu}(\boldsymbol{\alpha}))'\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}(\boldsymbol{v}-\boldsymbol{\mu}(\boldsymbol{\alpha}))\} \end{array} \quad (3)$$

where $\mu(\boldsymbol{\alpha}) = -\boldsymbol{\Sigma}(\boldsymbol{\alpha})\boldsymbol{b}(\boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1} = 2\boldsymbol{A}(\boldsymbol{\alpha})$.

Proof: $U(\boldsymbol{v},\boldsymbol{\alpha}) = \frac{1}{2}(\boldsymbol{v}-\boldsymbol{\mu}(\boldsymbol{\alpha}))'\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}(\boldsymbol{v}-\boldsymbol{\mu}(\boldsymbol{\alpha})) = \frac{1}{2}\boldsymbol{v}'\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}\boldsymbol{v} - \boldsymbol{\mu}'\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}\boldsymbol{v} + \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}\boldsymbol{\mu}$ i.e. $\mu(\boldsymbol{\alpha}) = -\boldsymbol{\Sigma}(\boldsymbol{\alpha})\boldsymbol{b}(\boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1} = 2\boldsymbol{A}(\boldsymbol{\alpha})$. □

If the prior (3) is rank deficient the conditional distributions $p(\boldsymbol{v} \setminus \boldsymbol{v}_s|\boldsymbol{v}_s)$ for $\boldsymbol{v}_s \subset \boldsymbol{v}$ will be rank sufficient. Theorem 1 implies that all $v_i \in \boldsymbol{v}$ has a prior distribution, i.e. $P(\boldsymbol{v})$ should be a function of all $v_i \in \boldsymbol{v}$. If this is not fulfilled then is the subset $\boldsymbol{v}_p \subset \boldsymbol{v}$, which have a prior distribution, used instead of $\boldsymbol{v}$. The assumption about $U(\boldsymbol{v},\boldsymbol{\alpha})$ made in theorem 1 covers almost all deformable templates presented in the literature and all cited here.

In the case where it can be justified that $U(\boldsymbol{v},\boldsymbol{\alpha})$ fulfill the assumption made in theorem 1 the mean $\boldsymbol{\mu}(\boldsymbol{\alpha})$ can also be found as the solution to $\frac{\partial U(\boldsymbol{v},\boldsymbol{\alpha})}{\partial \boldsymbol{v}} = 0$ and the inverse covariance $\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}$ as the Hessian $\frac{\partial^2 U(\boldsymbol{v},\boldsymbol{\alpha})}{\partial^2 \boldsymbol{v}} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}$. This method can be a fast alternative for determining $\boldsymbol{\mu}(\boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}$ if $U(\boldsymbol{v},\boldsymbol{\alpha})$ is not directly in the form $\boldsymbol{v}'\boldsymbol{A}(\boldsymbol{\alpha})\boldsymbol{v} + \boldsymbol{b}(\boldsymbol{\alpha})\boldsymbol{v} + c(\boldsymbol{\alpha})$

2.2. The observation model

The observation model gives the probability for a given realization of $\boldsymbol{v}$ corresponds to the observations $\boldsymbol{y} \in \boldsymbol{\Psi}$ - the image. In many cases the interaction corresponds to image intensity [2, 5, 6, 11] and/or edge information [3, 7, 8, 9, 10, 11] but in principle all kind of information can be combined e.g. texture or colour. Let the observation model be Gibbs distributed::

$$P(\boldsymbol{y}|\boldsymbol{v}) = \frac{1}{Z_o(\boldsymbol{\beta})} exp\{-\frac{1}{\lambda} I(\boldsymbol{y},\boldsymbol{v},\boldsymbol{\beta})\} \quad (4)$$

where $Z_o(\boldsymbol{\beta}) = \sum_{\boldsymbol{v}\in\boldsymbol{\Omega}} \exp\{-\frac{1}{\lambda} I(\boldsymbol{y},\boldsymbol{v},\boldsymbol{\beta})\}$ is the normalizing constant, $\lambda \in \boldsymbol{\Gamma_\lambda}$ is the regularization parameter that determines the relative influence of the prior and the observation model, $I(\boldsymbol{y},\boldsymbol{v},\boldsymbol{\beta}) : \boldsymbol{\Psi} \mapsto \boldsymbol{R}$ is the energy function representing the interaction between $\boldsymbol{y}$ and $\boldsymbol{v}$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_l) \in \boldsymbol{\Gamma_\beta}$ are the observation parameters which gives the relative influence of the different types of information.

3. Supervised model parameter estimation

Assume a training set of q template parameter sets $\boldsymbol{V}_1, .., \boldsymbol{V}_q$ corresponding to some structure within an image $\boldsymbol{y}$ are known.

3.1. The prior model

In the case where the assumption in theorem 1 is fulfilled the prior parameters can be estimated by using theorem 1 and the maximum likelihood (ML) estimator, which is defined as:

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}, \boldsymbol{V}_1, ..., \boldsymbol{V}_q) \quad (5)$$

where it is assumed that $\hat{\boldsymbol{\alpha}} \in \boldsymbol{\Gamma_\alpha}$. If $\boldsymbol{V}_1, .., \boldsymbol{V}_q$ are stochastic independent and the covariance is rank sufficient then is the likelihood function directly given as:

$$L(\hat{\boldsymbol{\alpha}}, \boldsymbol{V}_1, ..., \boldsymbol{V}_q) = P(\boldsymbol{V}_1, ..., \boldsymbol{V}_q) = \prod_{k=1}^{q} P(\boldsymbol{V}_k) \quad (6)$$

where $P(\boldsymbol{V}_1, ..., \boldsymbol{V}_q)$ is the simultaneous density function for the prior distribution $P(\boldsymbol{V}_k)$ given by (3). If the covariance is rank deficient, i.e. $rank(\boldsymbol{\Sigma}) = p - r$ for $r > 0$, it is necessary to condition on at least r variables to obtain full rank. In this case e.g. the pseudo-likelihood [1] can be used as an approximation to the true likelihood:

$$\begin{array}{rcl} PL(\hat{\boldsymbol{\alpha}}, \boldsymbol{V}_1, ..., \boldsymbol{V}_q) & = & P(\boldsymbol{V}_1, ..., \boldsymbol{V}_q) \\ & = & \prod_{k=1}^{q} \prod_{i=1}^{m} P(\boldsymbol{V}_{k,i}|\boldsymbol{V}_k \setminus \boldsymbol{V}_{k,i}) \end{array} \quad (7)$$

If $\boldsymbol{V}_1, .., \boldsymbol{V}_q$ are not stochastic independent then $P(\boldsymbol{V}_1, ..., \boldsymbol{V}_q)$ must be rewritten using conditional probabilities and Bayes theorem. When the prior parameters are estimated it is possible to simulate the prior model by sampling in the unconditional or conditional prior distribution.

3.2. The posterior model

Due to the image information in $P(\boldsymbol{y}|\boldsymbol{v})$ is it impossible to make any distribution assumption about $P(\boldsymbol{v}|\boldsymbol{y})$. This make it impossible to use the ML estimator for estimation

of the observation and regularization parameters, because it is infeasible to calculate the normalizing constant even for small $\boldsymbol{\Omega}$. A likelihood approximation with the normalizing constant removed from $P(\boldsymbol{v}|\boldsymbol{y})$ or a criterion equivalent to the minimax criterion proposed by [9] can't be used either, because $P(\boldsymbol{v}|\boldsymbol{y})$ without the normalizing constant in many case would not be convex within $(\boldsymbol{\Gamma_\lambda}, \boldsymbol{\Gamma_\beta})$. The observation and regularization parameters are instead estimated as the parameters which minimize the least squares error (LSE) between the training set parameters $\boldsymbol{V}_1, .., \boldsymbol{V}_q$ and the estimated parameters $\hat{\boldsymbol{v}}_1, ..., \hat{\boldsymbol{v}}_q$:

$$(\hat{\lambda}, \hat{\boldsymbol{\beta}}) = \arg\min_{\lambda, \boldsymbol{\beta}} \frac{1}{pq} \sum_{i=1}^{q} (\boldsymbol{V}_i - \hat{\boldsymbol{v}}_i)' \boldsymbol{C} (\boldsymbol{V}_i - \hat{\boldsymbol{v}}_i) \qquad (8)$$

where $\boldsymbol{C}$ is diagonal matrix where each parameter can be assigned an estimation weight (in most cases $\boldsymbol{C}$ is chosen equal to the identity matrix $\boldsymbol{I}$) and $\hat{\boldsymbol{v}}_i$ is the maximum a posteriori (MAP) estimate of parameter set i defined as $\hat{\boldsymbol{v}}_i = \arg\max_{\boldsymbol{v}_i} P(\boldsymbol{v}_i|\boldsymbol{y})$. Many different technics have been applied for MAP estimation as deterministic [3, 7, 8, 9, 11], stochastic [2, 6] and heuristic optimization algorithms [5].

If the assumption in theorem 1 is not fulfilled the model parameters $(\boldsymbol{\alpha}, \lambda, \boldsymbol{\beta})$ can still be estimated by expanding (8) to:

$$(\hat{\boldsymbol{\alpha}}, \hat{\lambda}, \hat{\boldsymbol{\beta}}) = \arg\min_{\boldsymbol{\alpha}, \lambda, \boldsymbol{\beta}} \frac{1}{pq} \sum_{i=1}^{q} (\boldsymbol{V}_i - \hat{\boldsymbol{v}}_i)' \boldsymbol{C} (\boldsymbol{V}_i - \hat{\boldsymbol{v}}_i) \quad (9)$$

If $log\ \ P(\boldsymbol{v}|\boldsymbol{y})$ is a linear function of $(\boldsymbol{\alpha}, \lambda, \boldsymbol{\beta})$ then (9) have to be solved with respect to the constraint: $\lambda + \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{l} \beta_i = constant$ to limit the number of solutions to one.

4. Unsupervised model parameter estimation

The unsupervised model parameter estimation is based on a modified version of the Expectation-Maximization (EM) algorithm [4]:

1. Start with the observations $\boldsymbol{y}$, an initial estimate $\boldsymbol{v}^0$ of $\boldsymbol{v}$ and a guess $(\boldsymbol{\alpha}^0, \lambda^0, \boldsymbol{\beta}^0)$ for $(\boldsymbol{\alpha}, \lambda, \boldsymbol{\beta})$.

2. Estimate $\hat{\boldsymbol{v}}^{t+1}$ by $\boldsymbol{v}^{t+1} = \arg\max_{\boldsymbol{v}}\ P(\boldsymbol{v}|\boldsymbol{y})$.

3. Use the algorithm for supervised model parameter estimation (Section 3) to estimate $(\hat{\boldsymbol{\alpha}}^{t+1}, \hat{\lambda}^{t+1}, \hat{\boldsymbol{\beta}}^{t+1})$ taking $\hat{\boldsymbol{v}}^t$ as training set.

4. Go to 2 for a number of iterations or until $(\boldsymbol{\alpha}^t, \lambda^t, \boldsymbol{\beta}^t)$ has approximately converged.

5. Experimental results

A deformable template is used for localization of the horizontal yarns in a system for automated visual inspection of textile [5]. The system should be used for inspection of above 35 different types of textile and new types are developed frequently, so it is necessary that a non-expert can train the system, i.e. select the model parameters $(\boldsymbol{\alpha}, \lambda, \boldsymbol{\beta})$.

The horizontal yarn k is modelled as $\boldsymbol{v}_k = (v_{k,1}, v_{k,1}, v_{k,2}, ..., v_{k,p})'$, where $v_{k,j}$ is the vertical position of yarn k in the vertical space j between the vertical yarns. The horizontal position of the vertical space j is assumed known, because it is easily located as a local maxima in a vertical projection of the image [5]. The distance between each vertical space j is assumed to be constant.

The posterior probability of the deformable model used for textile inspection is given by [5]:

$$\begin{array}{l} P(\boldsymbol{v}_k|\boldsymbol{v}_{k-1}, \boldsymbol{y}) = \frac{1}{Z_o(\boldsymbol{\alpha}) Z_p(\boldsymbol{\alpha})} \times \\ exp\{-\alpha_1 \sum_{j=2}^{p} (v_{k,j} - v_{k,j-1})^2 \\ -\alpha_2 \sum_{j=1}^{p} (v_{k,j} - v_{k-1,j} - d)^2 \\ -\frac{1}{\lambda} \sum_{j=1}^{p} I(v_{k,j}, j)\} \end{array} \qquad (10)$$

where the two first terms $U(\boldsymbol{v}, \boldsymbol{\alpha}) = \alpha_1 \sum_{j=2}^{p} (v_{k,j} - v_{k,j-1})^2 + \alpha_2 \sum_{j=1}^{p} (v_{k,j} - v_{k-1,j} - d)^2$ correspond to the prior model. The first term favours strictly horizontal threads. The second term describes that the thread $\boldsymbol{v}_k$ should be placed in a predefined distance $d \in \boldsymbol{R}_+$ from $\boldsymbol{v}_{k-1}$. The third term in $P(\boldsymbol{v}_k|\boldsymbol{v}_{k-1}, \boldsymbol{y})$ is the observation model where $I(v_{k,j}, j)$ is the negative horizontal mean at the vertical position $v_{k,j}$ of the pixels in the vertical space j. $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \boldsymbol{R}_+^2$ are the prior parameters and $\lambda \in \boldsymbol{R}_+$ is the regularization parameter. Let $U(\boldsymbol{v}, \boldsymbol{\alpha}) = \alpha_1 (\boldsymbol{N}\boldsymbol{v}_k)'(\boldsymbol{N}\boldsymbol{v}_k) + \alpha_2 (\boldsymbol{v}_k - (\boldsymbol{v}_{k-1} + \boldsymbol{d}))'(\boldsymbol{v}_k - (\boldsymbol{v}_{k-1} + \boldsymbol{d})) = \alpha_1 \boldsymbol{v}_k' \boldsymbol{N}' \boldsymbol{N} \boldsymbol{v}_k + \alpha_2 (\boldsymbol{v}_k \boldsymbol{I} \boldsymbol{v}_k - 2(\boldsymbol{v}_{k-1} + \boldsymbol{d}) \boldsymbol{v}_k + (\boldsymbol{v}_{k-1} + \boldsymbol{d})'(\boldsymbol{v}_{k-1} + \boldsymbol{d}) = \boldsymbol{v}_k' (\alpha_1 \boldsymbol{N}' \boldsymbol{N} + \alpha_2 \boldsymbol{I}) \boldsymbol{v}_k - 2\alpha_2 (\boldsymbol{v}_{k-1} + \boldsymbol{d}) \boldsymbol{v}_k + (\boldsymbol{v}_{k-1} + \boldsymbol{d})'(\boldsymbol{v}_{k-1} + \boldsymbol{d})$ where

$$N = \begin{bmatrix} 0 & 0 & ... & 0 \\ -1 & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & 1 \end{bmatrix} \qquad (11)$$

$$N'N = \begin{bmatrix} 1 & -1 & ... & 0 \\ -1 & 2 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & 1 \end{bmatrix} \qquad (12)$$

From above it's seen that the prior model fulfill the assumption in theorem 1 (which was obvious because the prior only consists of quadratic terms). By theorem 1 $P(\boldsymbol{v}_k|\boldsymbol{v}_{k-1}) \in N(\boldsymbol{\mu}_k(\boldsymbol{\alpha}), \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$ where:

$$\boldsymbol{\mu}_k(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}(\boldsymbol{\alpha}) 2\alpha_2(\boldsymbol{v}_{k-1} + \boldsymbol{d})$$
$$\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1} =$$
$$\begin{bmatrix} 2\alpha_1 + 2\alpha_2 & -2\alpha_1 & \dots & 0 \\ -2\alpha_1 & 4\alpha_1 + 2\alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2\alpha_1 + 2\alpha_2 \end{bmatrix} \quad (13)$$

5.1. Estimation of model parameters

Assume a training set of 150 yarns $\boldsymbol{V}_1, ..., \boldsymbol{V}_{150}$ is manual marked in a image by an operator, see subset in figure 2 and 1. It can be shown that the rank of $\boldsymbol{\Sigma}^{-1}$ is full $\forall(\alpha_1, \alpha_2) \in \boldsymbol{R}_+^2$, but the prior probabilities are not independent. Because the prior probability is a Markov Random field and by using Bayes theorem the likelihood is given by:

$$\begin{aligned} & L(\boldsymbol{\alpha}, \boldsymbol{V}_1, ..., \boldsymbol{V}_{150}) \\ & = P(\boldsymbol{V}_1, ..., \boldsymbol{V}_{150}) \\ & = P(\boldsymbol{V}_1) P(\boldsymbol{V}_2, ..., \boldsymbol{V}_{150} | \boldsymbol{V}_1) \\ & = P(\boldsymbol{V}_1) P(\boldsymbol{V}_2 | \boldsymbol{V}_1) P(\boldsymbol{V}_3, ..., \boldsymbol{V}_{150} | \boldsymbol{V}_2) \\ & = P(\boldsymbol{V}_1) \prod_{k=2}^{150} P(\boldsymbol{V}_k | \boldsymbol{V}_{k-1}) \end{aligned} \quad (14)$$

where $P(\boldsymbol{V}_k | \boldsymbol{V}_{k-1}) = \frac{1}{\sqrt{2\pi}^p} \frac{1}{\sqrt{det \boldsymbol{\Sigma}(\boldsymbol{\alpha})}} exp\{-0.5(\boldsymbol{V}_k - \boldsymbol{\Sigma}(\boldsymbol{\alpha}) 2\alpha_2 (\boldsymbol{V}_{k-1} + \boldsymbol{D}))' \boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1} (\boldsymbol{V}_k - \boldsymbol{\Sigma}(\boldsymbol{\alpha}) 2\alpha_2 (\boldsymbol{V}_{k-1} + \boldsymbol{D}))\}$ and $P(\boldsymbol{V}_1)$ is constant because $\boldsymbol{v}_1$ is estimated using an ad hoc procedure.

Prior parameters $\alpha_1 = 1.0476$ and $\alpha_2 = 0.0172$ are then estimated for the known yarns corresponding to figure 1. Simulations of the prior model are then performed by sampling in the Gaussian distribution $N(\boldsymbol{\mu}_k(\boldsymbol{\alpha}), \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$, see figure 1. If the simulated horizontal yarns are compared to the real yarns this seems to verify that the prior model with the estimated parameters is a good model of the yarns. The regularization parameter $\lambda = 0.3652$ is then estimated using the LSE (8) with $\boldsymbol{C} = \boldsymbol{I}$ and a heuristic MAP-estimation algorithm [5]. The yarns in the image in figure 2 is located using the estimated parameters with very good results, see figure 2. To examine the variation of the estimated parameters within the same textile sample, two parameter sets have been estimated on two different pieces of one sample. This was done for two different samples and the mean variation on α_1, α_2 and λ was 1.9%, 5.2% and 9.5%, respectively. The parameter variation between samples depends on the amount of variations and defects within the samples used for estimation. This is so significant that the estimated parameters often can be used to discriminate between god and bad samples. Using 5 representative samples were the parameters estimated with the mean $\alpha_1 = 1.0214, \alpha_2 = 0.0202$ and $\lambda = 0.4145$ and the standard deviation 0.2009, 0.0066 and 0.0717, respectively. The parameters were then used to locate the yarns in the 5 samples with very good results leading to an average increase on only 6% in the LSE (8) compared to the LSE obtained with the individual estimated parameters. The method for supervised parameter estimation have also been tested on other types of textile with good results.

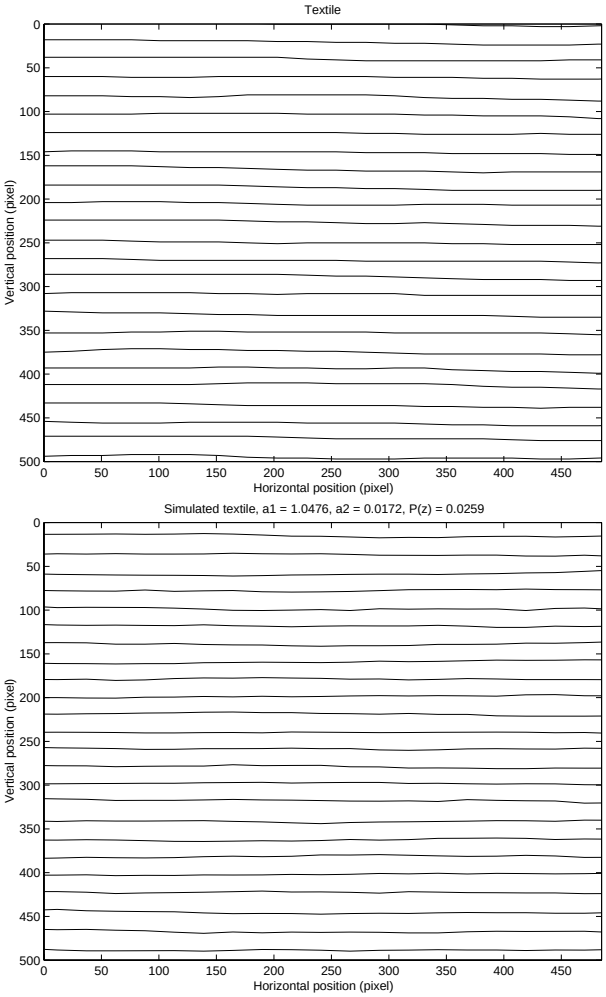

Figure 1. Manually marked horizontal yarns (top) and simulated horizontal yarns (bottom)

The unsupervised method performs well, but it is a little bit sensitive to the initial guess $(\boldsymbol{\alpha}^0, \lambda^0, \boldsymbol{\beta}^0)$ because it tends to get caught in local maxima if the regularization parameter is chosen very large or small. This sensitivity is a well known problem for the EM algorithm. For 6 randomly chosen $(\boldsymbol{\alpha}^0, \frac{1}{\lambda^0}, \boldsymbol{\beta}^0) \in [0.2; 2]$ under the constrain $\alpha_1^0 < \frac{1}{\lambda^0}$ and $\alpha_2^0 < \frac{1}{\lambda^0}$ were the parameters corre-

sponding to the yarns in figure 2 estimated with the mean $\alpha_1 = 1.3429, \alpha_2 = 0.0059$ and $\lambda = 0.5050$ and the standard deviation 0.1708, 0.0001 and 0.1005. The algorithm converged within 10 iterations. If the estimated parameters are compared with the previous estimated parameter there exists some differences, but still the LSE is only increased by 14 % compared to the LSE for the individual estimated parameters, and the yarns are still located very well, see figure 2. Similar results for the unsupervised parameter estimation are obtained for other textile samples and types.

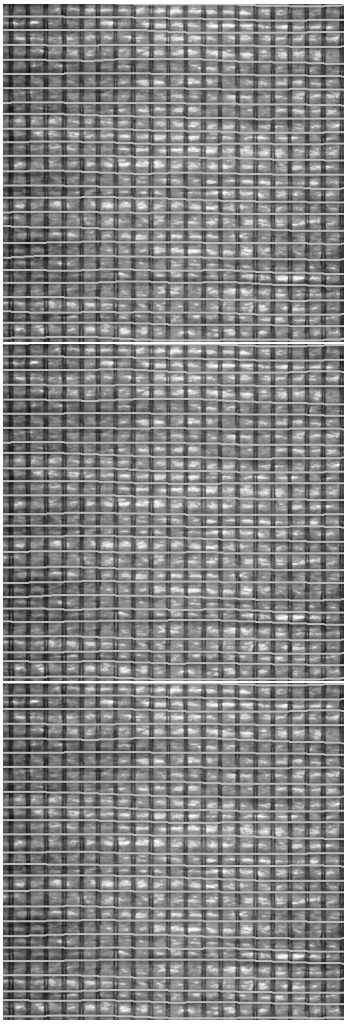

Figure 2. Textile with manually marked horizontal yarns (top), yarns located using supervised estimated parameters (center) and yarns located using unsupervised estimated parameters (bottom).

6. Conclusion

A method for supervised and unsupervised estimation of model parameters in deformable templates have been presented. Experimental results are successful and indicates that the methods are robust, though the unsupervised method is a little sensitive to the initial parameter configurations. The opportunity for simulations of the prior model seems to be a good tool for verification of the model. The presented methods also contains interesting perspectives regarding using the estimated parameters as features for description of the located structure and regarding automated model selection using an information criterion.

References

[1] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.

[2] J. M. Carstensen. An active lattice model in a bayesian framework. *Computer Vision and Image Understanding*, 63(2):380–387, 1996.

[3] L. D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53(2):211–218, 1991.

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society Series B*, 39:1–38, 1977.

[5] R. Fisker and J. M. Carstensen. Automated visual inspection of textile. In *Proc. The 10th Scandinavian Conf. on Image Analysis*, pages 173–179, Lappeenranta, 1997.

[6] U. Grenander, Y. Chow, and D. M. Keenan. *Hands: A Pattern Theoretic Study of Biological Shapes*. Springer-Verlag, 1991.

[7] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):267–278, 1991.

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 8(2):321–331, 1988.

[9] K. F. Lai and R. T. Chin. Deformable contours: Modeling and extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(11):1084–1090, 1995.

[10] O. V. Larsen, P. Radeva, and E. Marti. Guidelines for choosing optimal parameters of elasticity for snakes. In *Proc. 6th Comp. Anal. of Image and Patterns*, pages 106–113, 1995.

[11] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.