# Generative Model and consistent estimation algorithms for non-rigid deformation model

S. Allassonnière[1], Y. Amit[2], E. Kuhn[1], A. Trouvé[3]

[1]LAGA - Université Paris 13, [2]University of Chicago, [3]CMLA - ENS Cachan

http://www.cmla.ens-cachan.fr/∼allasson Stephanie.Allassonniere@cmla.ens-cachan.fr

**Abstract:** The link between Bayesian and variational approaches is well known in the image analysis community in particular in the context of deformable models. However, true generative models and consistent estimation procedures are usually not available and the current trend is the computation of statistics mainly based on PCA analysis. We advocate in this paper a careful statistical modeling of deformable structures and we propose an effective and consistent estimation algorithm for the various parameters (geometric and photometric) appearing in the models.

## 1 Introduction

One primary difficulty in the context of deformable template models is the initial choice of the template and of various parameters in the energies underlying the registration process. This problem is of utmost important in the context of medical imaging and computational anatomy where people try to provide statisticals models for anatomical and functional variability, but also in many problems of object detection and scene interpretation. Building real generative model, that handle pose variability and yield effective likelihood ratio tests for various discriminative purposes, is a fundamental issue mainly unsolved in the context of non-rigid objects.

A first step toward a statistical approach for the estimation of templates has been proposed by C.A. Glasbey and K.V. Mardia in 2001.

Our goal here is to propose a coherent statistical framework for dense deformable templates both in terms of the probability model, and in terms of the effective estimation procedure of the template *and* of the deformation covariance structure.

## 2 The Observation Model

Let $(y_i)_{1\leq i\leq n}$ be the gray level observed data. Each $y_i$ is defined on a grid of pixels $\Lambda \hookrightarrow \mathbb{R}^2$ where for each $s \in \Lambda$, $x_s$ is the location of pixel $s$ in a specified domain $D \subset \mathbb{R}^2$. The template is a function from $\mathbb{R}^2$ to $\mathbb{R}$and we consider the small deformation framework to caracterise the observations: we assume the existence of an unobserved deformation field $z : \mathbb{R}^2 \to \mathbb{R}^2$ such that $y(s) = I_0(x_s - z(x_s)) + \sigma\epsilon(s) = zI_0(s) + \sigma\epsilon(s)$ where $\epsilon(s)$ are i.i.d $\mathcal{N}(0,1)$, independent of all other variables.

## 3 The Template and Deformation Model

The template $I_0$ and the deformation $z$ belong to $V_p$ and $V_g$, 2 RKHS with respective kernels $K_p$ and $K_g$: Given $(p_k)_{1\leq k\leq k_p}$ and $(g_k)_{1\leq k\leq k_g})$ $\exists\alpha \in \mathbb{R}^{k_p}$ and $(\beta^{(1)},\beta^{(2)}) \in \mathbb{R}^{k_g}\times\mathbb{R}^{k_g}$ such as:

$$I_0(x) = \mathbf{K_p}\alpha(x), = \sum_{k=1}^{k_p} K_p(x,p_k)\alpha(k)\ ,$$
$$z_\beta(x) = (\mathbf{K_g}\beta)(x) = \sum_{k=1}^{k_g} K_g(x,g_k)(\beta^{(1)}(k),\beta^{(2)}(k)).$$

## 4 Parameters and Likelihood

*General model (includes mixtures of deformable templates)*:
Model parameters: $\theta = (\theta^\tau = (\alpha_\tau, \sigma^2_\tau, \Gamma^\tau_g))_{1\leq\tau\leq T}$ where $T$ = *#components*,
Weight of the different mixtures: $\rho = (\rho_\tau)_{1\leq\tau\leq T}$.
Let $\theta^\tau_g = \Gamma^\tau_g$ and $\theta^\tau_p = (\alpha_\tau, \sigma^2_\tau)$ and $\theta \in \Theta$ an open set.
For each observation $y_i$ we consider the pair of unobserved variables $\xi_i = (\beta_i, \tau_i)$. The likelihood of the observed data is:

$$q(y|\theta,\rho) = \sum_{\tau=1}^{T}\int q(y|\beta_\tau,\theta_p,\rho)q(\beta_\tau,\theta_g,\rho)\rho(\tau)d\beta_\tau$$

where the density functions are given by a Bayesian model.

## 5 The Bayesian Model

The generative probabilistic model is given by:

$$\begin{cases} \rho \sim \nu_\rho \\ \theta = (\theta^\gamma_g, \theta^\gamma_p)_{1\leq\gamma\leq} \sim \otimes_{\gamma=1}(\nu_g\otimes\nu_p) \mid \rho \\ \gamma_1^n \sim \otimes_{i=1}^n \rho \mid \eta = (\theta,\rho) \\ \beta_1^n \sim \otimes_{i=1}^n \mathcal{N}(0,\Gamma_g^{\gamma_i})\mid \eta,\ \gamma_1^n \\ y_1^n \sim \otimes_{i=1}^n \mathcal{N}(z_{\beta_i} I_{\alpha_i}, \sigma^2_{\gamma_i} Id_\Lambda) \mid \beta_1^n,\ \eta,\ \gamma_1^n \end{cases}$$

with the following prior distributions:

$$\begin{cases} \nu_g(d\Gamma_g) \propto \left(\exp(-\langle\Gamma_g^{-1},\Gamma_g^0\rangle/2)\dfrac{1}{\sqrt{|\Gamma_g|}}\right)^{a_g} d\Gamma_g,\ a_g > 2k_g+1 \\ \nu_p(d\sigma^2, d\alpha) \propto \left(\exp\left(-\dfrac{\sigma_0^2}{2\sigma^2}\right)\dfrac{1}{\sqrt{\sigma^2}}\right)^{a_p} \exp\left((\alpha-\mu_p)^t(\Gamma_p)^{-1}(\alpha-\mu_p)\right)d\sigma^2 d\alpha \\ \nu_\rho(\rho) = \left(\prod_{\tau=1}^{T}\rho(\tau)\right)^{a_\rho}. \end{cases}$$

## 6 Estimation: Theoretical results in the 1 component case

**Theorem 1 (Existence of the MAP estimator)** *For any sample $y_1^n$, there exists $\hat\theta_n \in \Theta$ such that*

$$q(\hat\theta_n|y_1^n) = \sup_{\theta\in\Theta} q(\theta|y_1^n)\,.$$

**Theorem 2 (Consistency)** *Assume that $\Theta_*$ is non empty. Then, for any compact set $K \subset \Theta$,*

$$\lim_{n\to+\infty} P(\ \delta(\hat\theta_n,\Theta_*) \geq \epsilon\ \wedge \hat\theta_n \in K\ ) = 0\,,$$

*where $\delta$ is any metric compatible with the usual topology on $\Theta$.*

Moreover, if we introduce a baseline image $I_b : \mathbb{R}^2 \to \mathbb{R}$ set the template as $I_\alpha = \mathbf{K_p}\alpha + I_b$ , and denote for any $R > 0$:

$$(1)\quad \begin{cases} \Theta^R = \{\ \theta = (\alpha,\sigma^2,\Gamma) \mid \alpha\in\mathbb{R}^{k_p},\ |\alpha|\leq R,\ \sigma^2\in\mathbb{R}^*_+,\ \Gamma\in\Sigma^+_{2k_g}(\mathbb{R})\ \} \\ \Theta^R_* = \{\ \theta\in\Theta^R \mid E_P(\log q(y|\theta)) = \sup_{\theta\in\Theta^R} E_P(\log q(y|\theta))\ \} \end{cases}$$

**Theorem 3 (Consistency on bounded prototypes)** *Assume that $dim_\beta < dim_y$, that $P(dy) = p(y)dy$ where the density $p$ is bounded with exponentially decaying tails and that the observations $y_1^n$ are i.i.d under $P$. Assume also that the baseline $I_b$ satisfies $|I_b(x)| > a|x| + b$ for some positive constant $a$. Then $\Theta^R_* \neq \emptyset$ and for any $\epsilon > 0$*

$$\lim_{n\to\infty} P(\delta(\hat\theta^R_n,\Theta^R_*)\geq\epsilon) = 0\,,$$

*where $\delta$ is any metric compatible with the topology on $\Theta^R$.*

## 7 Estimation with the EM algorithm

A natural approach with unobserved variables: The Em algorithm: We compute $\hat\eta = \arg\max_\eta q(\eta|y_1^N)$. This can be rewritten as:

$$\max_{\eta,\nu}\left[\int \log q(y,u|\eta)\nu(u)\mu(du) - \int \nu(u)\log\nu(u)\mu(du)\right],$$

which yields to 2 maximisation steps. We iterate the following 2 steps:

**E Step:** Compute the posterior law on $(\beta_i,\gamma_i), i = 1,\dots,n$ as a product of the following distributions:

$$\nu_{l,i}(\beta,\gamma) = \frac{q(y_i|\beta,\alpha_{\gamma,l})q(\beta|\Gamma^\gamma_{g,l})\rho_l(\gamma)}{\sum_{\gamma'}\int q(y_i|\beta',\alpha_{\gamma',l})q(\beta'|\Gamma^{\gamma'}_{g,l})\rho_l(\gamma')d\beta'}$$

**M Step:** $\eta_{l+1} = \arg\max_\eta E_{\nu_l(d\xi_1^n)}(\log q(\eta,\beta_1^n,\gamma_1^n|y_1^n))$.

## 1 Fast approximation with modes

The M step require the computaion of expectations with respect to $\nu_{i,l}(\beta,\tau)$ which has no simple form.
*Solution proposed:* Approximation with modes:
$\nu_{i,l}(d\beta_{i,\tau},\tau) \simeq \delta_{\beta^*_{i,\tau},\tau}$ where $\forall\tau$:

$$\beta^*_{i,\tau} = \arg\max_\beta \log q(\beta|\alpha_{\tau,l},\sigma^2_{\tau,l},\Gamma^\tau_{g,l},y_i) = \arg\min_\beta\left\{\frac{1}{2}\beta^t(\Gamma^\tau_{g,l})^{-1}\beta + \frac{1}{2\sigma^2_{l,\tau}}|y_i - K_p^\beta\alpha_{\tau,l}|^2\right\}\ .$$

And the joint posterior distribution on $(\beta_i,\tau_i)$ is approximated by a discrete distribution concentrated at the $T$ points $\beta^*_{i,\tau}$ with weights:

$$w_l(\tau) = \frac{q(y_i|\beta^*_{i,\tau},\alpha_{\tau,l})q(\beta^*_{i,\tau}|\Gamma^\tau_{g,l})\rho_l(\tau)}{\sum_{\tau'} q(y_i|\beta^*_{i,\tau'},\alpha_{\tau',l})q(\beta^*_{i,\tau'}|\Gamma^{\tau'}_{g,l})\rho_l(\tau')}.$$

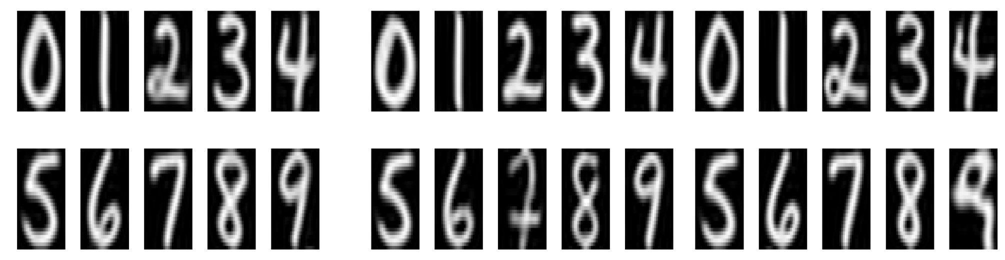
## 2 Using a stochastic version of the EM algorithm

*Second solution:* coupling SAEM with MCMC procedure:
This yields to the 3 following steps:

- Draw the missing data using a transition probability of a convergent Markov Chain having the posterior distribution as stationary distribution: Simulation step: $\beta^{l+1} \sim \Pi_{\theta_l}(\beta^l,.)$.
- Approximate the complete likelihood using the previous simulations: Stochastic approximation: $Q_{l+1}(\theta) = Q_l(\theta) + \Delta_l[\log q(y,\beta^{l+1}|\theta) - Q_l(\theta)]$ where $(\Delta_l)$ is a non increasing sequence with limit 0 of positive step-size.
- Parameter update in a M-step: Maximisation step: $\theta_{l+1} = \arg\max Q_{l+1}(\theta)$.

## 8 Experiments: Estimated templates

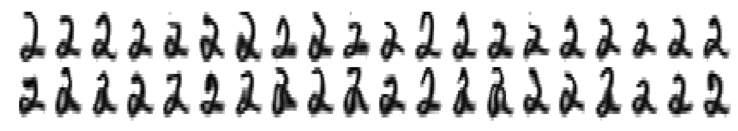
Training set: 20 images per class for 1 component and 40 for 2 components.
Results after 20 EM iterations.



*Left: one component prototype. Right: 2 components prototypes.*

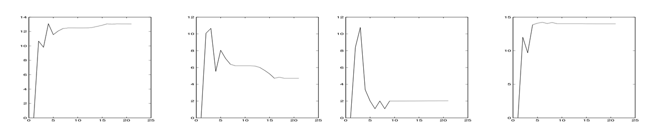
## 9 The estimated geometric distribution

To be able to notice the geometrical effets learned through the covariance matrix, we compare the effects of one learned deformation on the corresponding template and on other elements either in the same class or for an other digit.



*Top: Synthetized 2's with template from second component of the previous results and proper covariance. Bottom: Same template using covariance matrix of other 2 component.*

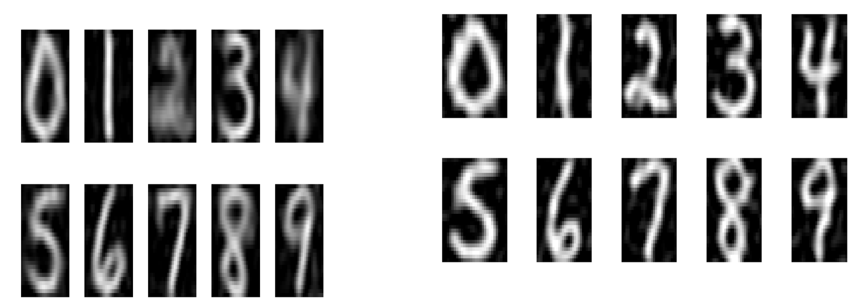
*Top: Synthetized 3's with the corresponding template and covaiance matrix. Bottom: Same template using covariance matrix of other one of the component of the 2's*



*Evolution of the symetric Kullback distance between the current value of $\Gamma_g$ and the prior center $\Gamma_0$. Left: 2 components of class 0. Right: 2 components of class 7*

## 10 In the presence of noise

The stochastic procedure has shown more robustness and accuracy in the presence of noise. The mode approximation is biaised because of the high number of local maxima of the likelihood.



*Left: prototypes in noisy framework learned with the mode approximation. Right: Prototypes in noisy framework learned with the stochastic EM algorithm*

## 11 Multi-component case in the stochastic EM algorithm: Some problems encountered

In this particular framework, the theoretical convergence of the Markov Chain cannot be numerically reached. To generate the new simulation of each missing data, we use a Gibbs sampler procedure. The first iteration of the EM algorithm affect each image in a class with probablility 1/2 then no change of class occures ; the probability for an image to be affected to an other class is too small and generally under the computer precision.
*Solutions currently studied:*

- Concider a model of mixture of the pevious model and other missing variables: the deformations of an image and the weight of an image for each class $(\beta^\tau, p^\tau)_\tau$.
  Problem: this model is no more exponential, no concergence has been yet proved and the implementation is more complex.
- Consider an other simulation method based on the Gibbs sampler for the deformation and on an other law for the class of a given image. (Theory and algorithm in progress...)

## Références

[1] S. Allassonnière, Y. Amit, A. Trouvé, *Toward a coherent statistical framework for dense deformable template estimation*, Submitted to JRSS.
[2] S. Allassonnière, Y. Amit, E. Kuhn, A. Trouvé, *Generative model and consistent estimation algorithms for non-rigid deformation models.*, ICASSP 2006.