A hierarchical statistical framework for the segmentation of deformable objects in image sequences

Charles Kervrann and Fabrice Heitz

IRISA/INRIA, Campus Universitaire de Beaulieu,
35042 Rennes Cedex, France
E-mail: kervrann@irisa.fr, heitz@irisa.fr

IEEE Computer Vision Pattern Recognition
June 1994, Seattle, USA

A Hierarchical Statistical Framework for the Segmentation of Deformable Objects in Image Sequences

Charles Kervrann and Fabrice Heitz
IRISA / INRIA - Rennes
Campus Universitaire de Beaulieu
F-35042 Rennes Cédex, France

Abstract

In this paper, we propose a new statistical framework for modeling and extracting 2D moving deformable objects from image sequences. The object representation relies on a hierarchical description of the deformations applied to a template. Global deformations are modeled using a Karhunen Loeve expansion of the distorsions observed on a representative population. Local deformations are modeled by a (first-order) Markov process. The optimal bayesian estimate of the global and local deformations is obtained by maximizing a non-linear joint probability distribution using stochastic and deterministic optimization techniques. The use of global optimization techniques yields robust and reliable segmentations in adverse situations such as low signal-to-noise ratio, non-gaussian noise or occlusions. Moreover, no human interaction is required to initialize the model. The approach is demonstrated on synthetic as well as on real-world image sequences showing moving hands with partial occlusions.

1 Introduction

In an increasing number of application fields the objects to be modeled undergo deformations which have to be analysed and characterized. Deformable models [3, 4, 7] are mathematical models which incorporate knowledge about shapes and their variations. These models have been used with success in the analysis of still as well as dynamic images, to extract, track or characterize deformable objects.

In this paper we introduce a modeling framework which relies on a hierarchical statistical description of shape deformations, in which both global and local deformations are represented. The global description of deformations relies on a (statistical) modal decomposition introduced recently by Cootes *et al.* in [2]. Local deformations are modeled as local random perturbations and are assumed to follow a first-order markov process; they can be seen as a refinement of the global deformations applied to the original shape. The joint distribution of the deformable template is derived and a Maximum A Posteriori (MAP) estimate of the deformations is obtained by minimizing a global energy (objective) function describing the interactions between observations (spatial or temporal gradients extracted from the image) and the deformation process. The method combines the advantages of fast global optimization techniques with a compact hierarchical statistical description of deformations. This yields fast model adjustment and robust segmentation.

Computer vision methods relying on deformable templates are often expressed as the minimization of (global) energy functions describing the interactions between the observed data and the variables of the model [4, 8, 9]. In most methods [1, 5, 7, 8, 9] (apart from the work of Grenander *et al.* [4]), deterministic optimization algorithms are used to solve for the model configuration. These approaches require human interaction to initialize the model and are also known to be very sensitive to local minima of the objective function. The global optimization process which is considered here does not have this drawback: it is robust to noise, occlusions and does not require an initial configuration of the model close to the optimal solution.

Staib and Duncan [8] have proposed to use Fourier descriptors, associated to bayesian estimation methods to analyse deformable objects. Human interaction is however needed to initialize the model. Cootes *et al*, [2] consider the Karhunen Loeve (KL) transform of the global deformations observed on a training set of representative shapes. KL analysis allows to approximate the global deformations of the original template by superimposing the main variations modes extracted from the shapes belonging to the learning set. Five to ten parameters are usually sufficient to obtain an accurate description. In [3], these global deformation parameters are adjusted to fit the model on edges

extracted from the image. A deterministic relaxation scheme, which requires an initialization close to the optimal configuration, is used to find the deformation modes [3]. In [4] Grenander *et al.* has obtained very promising results in image restoration and segmentation by modeling local deformations using Markov processes. Monte-Carlo techniques are used to compute the Maximum A Posteriori (MAP) estimate of these local deformations. Due to the large size of the space of configuration, the computation of the MAP estimate is generally computationally demanding [4].

The hierarchical model proposed here can be optimized efficiently thanks to its reduced number of global deformation parameters. Local deformations are identified using a deterministic relaxation algorithm which has fast convergence properties. Local minima are not a problem since optimization on the global deformation parameters provides a good initialization for the deterministic refinement procedure. This paper is a short version of technical report [6] in which additional detail about the method as well as other experimental results can be found.

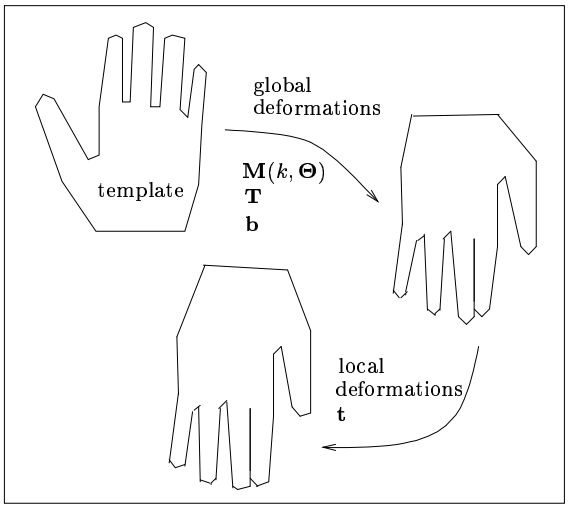

Figure 1: *Hierarchical description of deformations*

2 A hierarchical statistical deformable model

Following Cootes *et al.* [2], we represent the shape of interest by the point distribution of a "deformable template" which incorporates *a priori* knowledge on the structure of the object and its variability. A particular shape $\mathbf{X}$ is represented by a set of n labeled points which approximate its outline [2] (see Fig. 1).

The global deformations of shape $\mathbf{X}$ are characterized by a displacement vector $\mathbf{dX}$ with respect to a pre-computed mean shape (the "template") $\mathbf{X}^\star$ [2]. A KL expansion of the displacement vectors observed on a representative population allows to obtain a good approximation for the actual deformations, on a low dimension space [2]. If $\mathbf{P}$ designates the matrix of the m unit eigenvectors corresponding to the m largest eigenvalues, and if $\mathbf{b}$ denotes the vector $(m \times 1)$ corresponding to the m most significant deformation modes, the deformable template is represented by the following model [6]:

$$\mathbf{X} \;=\; \mathbf{M}(k,\theta)\,[\mathbf{X}^\star + \mathbf{P}\mathbf{b}] \;+\; \mathbf{T} \qquad (1)$$

Global transformations from the similarity group (rotation of angle θ, scale change by a factor k and translation $\mathbf{T}$) are taken into account in this model (Fig. 1).

A local deformation process is introduced to refine this first (eventually crude) description. Local deformations are modeled as random perturbations on the location of the points belonging to the globally deformed pattern. This local deformation process $\mathbf{t}$ is assumed to follow a first-order Markov random process, which takes into account interactions between neighboring points. The global statistical model becomes (Fig. 1):

$$\mathbf{Y} \;=\; \mathbf{M}(k,\theta)\,[\mathbf{X}^\star + \mathbf{P}\mathbf{b}] \;+\; \mathbf{T} \;+\; \mathbf{t} \qquad (2)$$

where: $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n)$ and $\mathbf{t}_i = (\mathbf{t}_{x_i}, \mathbf{t}_{y_i})$, is the random local deformation process applied on the n labeled points. $\mathbf{t}$ is assumed to be a zero-mean first-order Gauss-Markov random process:

$$p(\mathbf{t}) \;=\; \frac{1}{C}\, exp\, -\frac{1}{2}\, \mathbf{t}^{\,T}\, \mathbf{R}^{-1}\, \mathbf{t} \qquad (3)$$

where $\mathbf{R}$ is the covariance matrix of $\mathbf{t}$ and C is the partition function. The joint distribution of vector $\mathbf{t}$ may be written as:

$$p(\mathbf{t}) = \frac{1}{C}\, exp{-}\frac{1}{2} \sum_{i=1}^{n} \left[\frac{1}{\epsilon_i^2} \left\| t_i - t_{i-1} \right\|^2 + \frac{1}{\sigma_i^2} \left\| t_i \right\|^2 \right] \quad (4)$$

where σ_i^2 and ϵ_i^2 are the parameters of the model. ϵ_i^2 weights the interactions between neighboring points. Low values for σ_i^2 draw the shape towards the globally deformed model $\mathbf{X}$. In all experiments, these parameters have been kept constant ($\sigma_i^2 = 4$ and $\epsilon_i^2 = 1$).

3 Bayesian estimation of deformations

The hierarchical statistical deformable model defined in the previous section is used as an a priori model wi-

thin a global bayesian estimation scheme. One or more specialized modules extract from the image sequence, low-level features (spatio-temporal gradients) that will be used as observations in the estimation process.

Let $\mathbf{O} = (O_s,\ s \in S)$ designate an observation field defined on a rectangular lattice S related to the spatiotemporal variations of the intensity function. The Maximum A Posteriori (MAP) estimate of the deformable template is defined by:

$$\mathbf{Y}_{opt} = \arg\max_{\mathbf{Y}}\ p(\mathbf{O}|\mathbf{Y})\, p(\mathbf{Y}) \qquad (5)$$

The distribution $p(\mathbf{O}|\mathbf{Y})$ (presented in the next section) describes the interactions between the observations and the deformations to estimate.

According to the assumption on the statistics of $\mathbf{t}$ (see Eq. 3), $\mathbf{Y}$ follows a first order Gauss-Markov process:

$$p(\mathbf{Y}) = \frac{1}{C}\, exp\, -\frac{1}{2}\, (\mathbf{Y}-\mathbf{X})^{\,T}\, \mathbf{R}^{-1}\, (\mathbf{Y}-\mathbf{X}) \qquad (6)$$

This prior distribution controls the local and global deformations of the original template. The similarity transformations $\mathbf{M}(k,\theta)$, $\mathbf{T}$ and the global deformation modes $\mathbf{b}$ are considered as deterministic parameters of this probabilistic model and are estimated using a Maximum Likelihood (ML) method.

4 Segmentation model

We have considered here the particular problem of the extraction of moving objects from images sequences in the case where the camera is static. In this context, we rely on observations related to temporal gradients extracted from the image sequence.

Let $I_t(s),\ s \in S$ denote the intensity function, where $s = (x, y)$ designates the 2-D spatial image coordinates and t the time axis. A first set of observations corresponds to temporal changes $O_1(s)$ evaluated on three successive images; a second observation set is related to the changes $O_2(s)$ between the current image and a reference image $I_{ref}(s)$ which is created and updated on line [6]:

$$\begin{aligned} O_1(s) &= \min(|I_t(s) - I_{t-1}(s)|, |I_{t+1}(s) - I_t(s)|) \\ O_2(s) &= |I_{ref}(s) - I_t(s)| \end{aligned} \qquad (7)$$

Temporal gradients such as $O_1(s)$ are known to yield poor observations in homogeneous regions. The second observations $O_2(s)$ are less sensitive to this problem and are used as complementary information.

For a given configuration of the template, the image can be partitioned into two regions: the inside of the template $\mathbf{X}$ (R^-) and the outside of $\mathbf{X}$ (R^+) corresponding to the background. The following (Gibbs) distribution $p(\mathbf{O}|\mathbf{X})$ is specified to describe the interactions between local observations $\mathbf{O}(s)$ and the configuration of the deformable model:

$$p(\mathbf{O}|\mathbf{Y}) = \frac{1}{C'}\, exp -[\sum_{s\,\in\, R^-} |O(s)-1\,| + \sum_{s\,\in\, R^+} |O(s)-0\,|\ ] \qquad (8)$$

where C' is a normalization constant and:

$$O(s) = \max(\Gamma_\alpha(O_1(s)), \Gamma_\beta(O_2(s))) \qquad (9)$$

$$\Gamma_\eta(y) = 1 \ \ if\ y > \eta \quad and \quad \Gamma_\eta(y) = 0 \ \ else \qquad (10)$$

This distribution tends to enclose moving points inside the deformable model and to reject static points belonging to the background outside the outline of the model.

5 Global optimization

The Maximum A Posteriori (MAP) estimate of the deformable template may be expressed as:

$$\mathbf{Y}_{opt} = \arg\max_{\mathbf{Y}}\ p(\mathbf{O}|\mathbf{Y}) p_{\mathbf{\Theta}}(\mathbf{Y}) \qquad (11)$$

where $\mathbf{\Theta}$ is the global parameter vector of the model:

$$\mathbf{\Theta} = (\mathbf{M}(k,\theta),\ \mathbf{T},\ \mathbf{b}) \qquad (12)$$

The joint distribution may be rewritten as a Gibbs distribution where Z is the partition (Z does not depend on $\mathbf{\Theta}$):

$$p(\mathbf{O}|\mathbf{Y})\, p_{\mathbf{\Theta}}(\mathbf{Y}) = \frac{1}{Z}\, exp\ -U_{\mathbf{\Theta}}(\mathbf{O},\mathbf{Y}), \qquad (13)$$

with the following energy function:

$$\begin{aligned} U_{\mathbf{\Theta}}(\mathbf{O},\mathbf{Y}) &= \sum_{s\,\in\, R^-} |O(s)-1\,| + \sum_{s\,\in\, R^+} |O(s)-0\,| \\ &+ \frac{1}{2}\, (\mathbf{Y}-\mathbf{X}(\mathbf{\Theta}))^{\,T}\, \mathbf{R}^{-1}\, (\mathbf{Y}-\mathbf{X}(\mathbf{\Theta})). \end{aligned} \qquad (14)$$

Starting from the initial model configuration $\mathbf{Y} = \mathbf{X}^\star$, the model global parameters $\mathbf{\Theta}$ are estimated alternately with the configuration of the model, yielding a partial optimal solution [6]. In our case, the global parameter vector $\mathbf{\Theta}$ is estimated using a Maximum Likelihood procedure. A global optimization of the joint-likelihood (Eq. 13) with respect to parameter vector $\mathbf{\Theta}$ is first performed with a fast stochastic relaxation procedure (relying on a standard simulated annealing

algorithm based on the Gibbs sampler dynamics [6]). This optimization step yields:

$$\mathbf{\Theta}_{opt} \;=\; Arg\,Min_{\mathbf{\Theta}}\; U_{\mathbf{\Theta}}(\mathbf{O},\mathbf{Y}) \qquad (15)$$

The MAP estimate of the deformable template $\mathbf{Y}$ is then approximated using a fast deterministic relaxation scheme corresponding to a modified version of the ICM algorithm [6]. This step determines the local deformation process $\mathbf{t}$:

$$\mathbf{Y}_{opt} \;=\; Arg\,Min_{\{\mathbf{Y}\,=\,\mathbf{X}(\mathbf{\Theta}_{opt})\,+\,\mathbf{t}\}}\; U_{\mathbf{\Theta}_{opt}}(\mathbf{O},\mathbf{Y}). \quad (16)$$

6 Experimental results

In our experiments, we have considered the segmentation of complex deformable structures [1, 2, 4] corresponding to hands moving against a background.

In the first experiments (Fig. 2), observation maps $O(s)$, obtained from real hand outlines, have been corrupted by non gaussian noise. Occlusions have also been simulated (Fig. 3). The initial model configuration (a 40 point model here), corresponding to $\mathbf{Y} = \mathbf{X}^{\star}$, can be seen in the lower right part of Fig. 2a. Fig. 2b shows the result of global optimization on the similarity transformation parameters only. The result of the global optimization on the nine first most significant variation modes (vector $\mathbf{b}$) is shown Fig. 2c. As can be seen, the boundary is localized accurately, even with very low signal-to-noise ratios (Fig. 2d), for corrupted (Fig. 3a, 3d) or occluded observations (Fig. 3b, 3c).

The robustness of the approach is best illustrated on a part of a real sequence (of more than one hundred frames) showing a hand moving against a textured background (Fig. 4a to 4c). The algorithm was able to extract and track the hand reliably over the whole sequence by initializing the deformable template (a 30 point model here) in the current frame by the final configuration obtained on the previous frame and by considering the seven first modes. The noisy and partial observations $O(s)$ (Eq. 9), corresponding to the different frames of this sequence are presented in Fig. 5 (please note the occlusions on the fingers).

The total cpu time, on a Sun-4 station (sun IPX) was about 6 minutes for the complete processing of one 256x256 frame.

7 Conclusion

In this paper, we have presented a global bayesian framework for modeling and processing deformable shapes. The technique relies on the definition of a deformable template on which hierarchical deformations are applied. The deformations are described using global statistical models and the optimal bayesian estimate of these deformations is computed using stochastic and deterministic optimization techniques.

The proposed modeling and algorithmic framework is comprehensive and suited to the representation of a large class of deformable objects. It may be adapted to segmentation problem based on other image attributes (luminance, color, texture, depth, etc.). The use of a hierarchical deformable model also yields promising future prospects as far as the characterization and the interpretation of the dynamic behavior of complex objects is concerned.

References

[1] A. BLAKE, R. CURWEN, and A. ZISSERMAN. – A framework for spatiotemporal control in the tracking of visual contours. – *Int. J. Computer Vision*, Vol. 11, No 2: pp. 127–145, oct. 1993.

[2] T.F. COOTES, C.J. TAYLOR, D.H. COOPER, and J. GRAHAM. – Training models of shape from sets of examples. – In *British, Machine Vision Conf.*, pages 9–18, Leeds, UK, Sept. 1992.

[3] T.F. COOTES, C.J. TAYLOR, A. LANITIS, D.H. COOPER, and J. GRAHAM. – Building and using flexible models incorporating grey-level information. – In *Proc. 4th Int. Conf. Comp. Vis.*, pages 242–246, Berlin, Germany, May 1993.

[4] U. GRENANDER, Y. CHOW, and D.M. KEENAN. – *Hands. A Pattern Theoretic Study of Biological Shapes.* – Springer, 1991.

[5] M. KASS, A. WITKIN, and D. TERZOPOULOS. – Snakes : Active contour models. – In *Proc. First Int. Conf. Comp. Vis.*, pages 259–268, London, UK, June 1987.

[6] C. KERVRANN and F. HEITZ. – A hierarchical statistical framework for the segmentation of deformable objects in image sequences. – Technical Report No 2133, INRIA, Dec. 1993.

[7] C. NASTAR and N. AYACHE. – Fast segmentation, tracking and analysis of deformable objects. – In *Proc. 4th Int. Conf. Comp. Vis.*, pages 275–279, Berlin, Germany, May 1993.

[8] L. H. STAIB and J. S. DUNCAN. – Boundary finding with parametrically deformable models. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 14, No 11: pp. 1061–1075, Nov. 1992.

[9] A.L. YUILLE. – Feature extraction from faces using deformable templates. – *Int. J. Computer Vision*, Vol. 8, No 2: pp. 99–111, 1992.

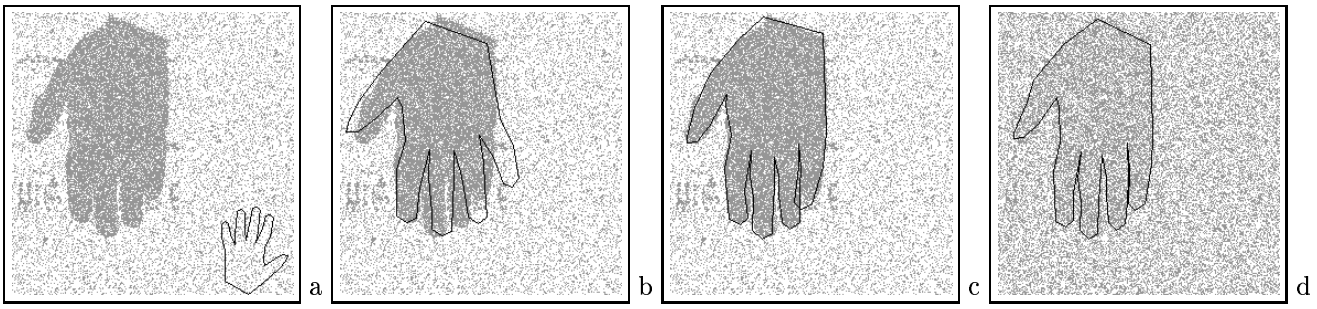

Figure 2a-b-c: *Intermediate steps in the segmentation process.* - d: *Result with low signal-to-noise ratio.*

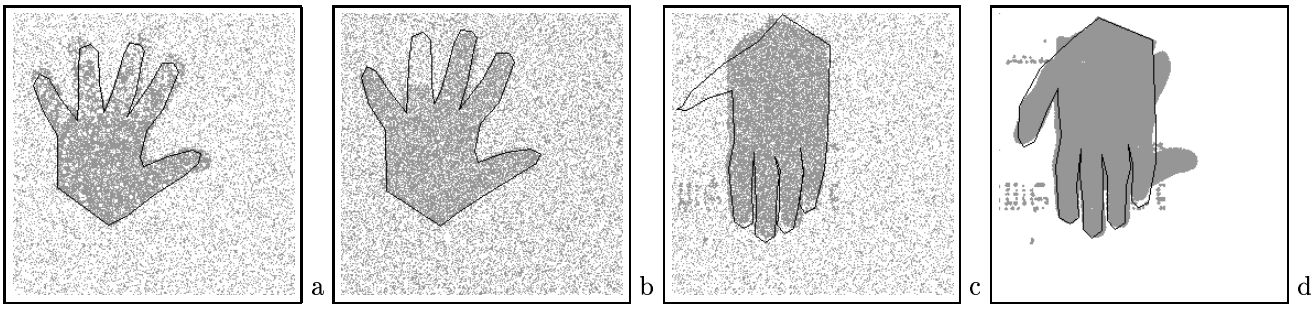

Figure 3: *Segmentations with different signal-to-noise ratio and partial occlusions.*

Figure 4: *Segmentation and tracking of a moving hand against a textured background.*

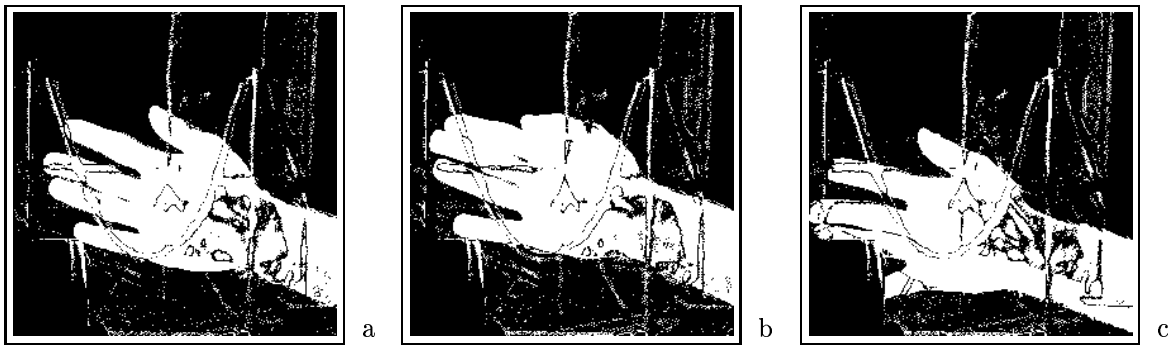

Figure 5: *Observations maps $O(s)$ used for the segmentation of the moving hand.*