HAL
open science

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

Charles Kervrann, Fabrice Heitz

To cite this version:

Charles Kervrann, Fabrice Heitz. A Hierarchical statistical framework for the segmentation of deformable objects in image sequences. [Research Report] RR-2133, INRIA. 1993. inria-00074539

HAL Id: inria-00074539

https://inria.hal.science/inria-00074539

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

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

Charles Kervrann, Fabrice Heitz

N° 2133

Décembre 1993

PROGRAMME 4

Robotique,
image
et vision

Rapport de recherche

1993

ISSN 0249-6399

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

Charles Kervrann, Fabrice Heitz

Programme 4 — Robotique, image et vision
Projet TEMIS

Rapport de recherche n° 2133 — Décembre 1993 — 26 pages

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 computed template. Global deformations are modeled using a Karhunen Loeve expansion of the distorsions observed on a representative population. Local deformations are modeled using (first-order) Markov processes. The statistical hierarchical model is used to represent the *a priori* structure of the shapes to be extracted from the image sequence. The optimal bayesian estimate of the global and local deformations is obtained by minimizing a global objective function depending on the global deformation parameters and on the local random deformation process. A partial optimal solution is estimated by stochastic and deterministic relaxation procedures. The use of global optimization algorithms yields robust and reliable segmentations in adverse situations such as low signal-to-noise ratio, non-gaussian noise or occlusions. This procedure also leads to solutions which do not depend on the initial configuration of the model. The technique is demonstrated on synthetic as well as on real-world image sequences showing moving hands with partial occlusions.

Key-words: deformable models, bayesian estimation, global optimization, motion segmentation

(Résumé : tsvp)

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : (33) 99 84 71 00 – Télécopie : (33) 99 38 38 32

Un modèle statistique hiérarchique pour la segmentation d'objets déformables dans une séquence d'images

Résumé : Dans cet article, nous proposons un cadre statistique général pour la modélisation et la segmentation d'objets déformables bidimensionnels en mouvement dans des séquences d'images. La représentation de l'objet s'appuie sur une description hiérarchique des déformations appliquées à une forme prototype. Les déformations globales sont identifiées au moyen d'une décomposition de Karhunen Loeve des distorsions observées sur une population représentative. Les déformations locales sont modélisées par des processus markoviens. Le modèle statistique hiérarchique est utilisé pour représenter la structure *a priori* des formes à extraire dans la séquence d'images. L'estimée bayesienne optimale des déformations globales et locales est obtenu par minimisation d'une fonction d'énergie globale dépendant des paramètres globaux de déformation et du processus des déformations locales. Une solution optimale partielle est estimée par des procédures de relaxation stochastique et déterministe. L'utilisation d'algorithmes d'optimisation globale produit des segmentations fiables et robustes dans des situations diverses telles qu'un faible rapport signal-à-bruit, la présence d'un bruit non gausssien ou d'occlusions. En outre, cette procédure conduit à des solutions indépendantes de la configuration initiale du modèle. Cette technique a été validée sur des séquences d'images synthétiques et réelles présentant des mains en mouvement avec des occlusions partielles.

Mots-clé : modèles déformables, estimation bayesienne, optimisation globale, segmentation au sens du mouvement

1 Introduction

Up to the middle of the eighties, the models developed in computer vision have essentially been specified for the description and the analysis of rigid objects undergoing rigid movements. The representation and processing of deformations has recently gained considerable popularity (especially in the last five years). In an increasing number of applications fields (remote sensing, meteorology, oceanography, biological or biomedical images, analysis of human motion, turbulence analysis), the shapes and dynamic phenomena to be modeled may undergo deformations which have to be analysed and characterized.

The introduction of deformable models to represent such phenomena is motivated by different points :

- at first, the model enables to take into account complex *a priori* knowledge on the structure of the objects to be handled and their variability. This *a priori* knowledge leads to more robust segmentation or tracking schemes ;
- secondly, deformation analysis yields promising future prospects as far as the characterization and the interpretation of the dynamic behavior of complex objects is concerned.

In this paper, we study the representation of the outline of 2D deformable shapes in the image plane. Our approach relies on the description of the shape of interest by a deformable template which incorporates (statistical) knowledge about the shape, including its variability. The deformations of the original template are represented using a hierarchical model. Global deformations parameters include transformations from the group of similarity (translation, rotation, scale) and parameters which control the main variation modes of the original template. The group of similarity transformations enables global registration of the shape on the input data. To represent global deformations of the shape, a technique described recently by Cootes *el al.* [11] has been adopted. Cootes *et al.* [11] propose to retain the first modes of variation in the Karhunen-Loeve (KL) expansion of the deformations observed on a representative population. Local deformations are modeled, at a second level of the hierarchical representation, as local random perturbations on the outline of the globally deformed pattern. This local deformation process is assumed to be a first-order markov process, which takes into account dependencies between the local deformations ap-

plied to neighboring points. This local process can be seen as a refinement of the global deformations applied to the input pattern.

The general mathematical framework under concern to perform segmentation is based on global bayesian estimation. The joint distribution of the deformable template is derived and a Maximum A Posteriori (MAP) estimate of its 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 similarity transformations and the global deformation modes are considered as parameters of the probabilistic model. These parameters are estimated using a Maximum Likelihood procedure. The global optimization of the joint-likelihood of the model is performed with a fast stochastic relaxation procedure. The low dimensionality of the problem enables to compute quasi-optimal estimates in reasonable cpu time. Only a few (typically between 5 and 10) relevant modes are indeed used to represent the global deformations. Local deformations are estimated in turn, using a deterministic relaxation procedure.

The introduction of a statistical model and global optimization procedures has several advantages :

- It leads to segmentation schemes which are robust to noise. The method shows indeed an excellent robustness to high noise levels and to non-gaussian noise statistics.

- The final configuration of the deformable template does not depend on the initialization. This point is of course important in non-supervised segmentation schemes.

The remainder of this paper is organized as following : in Section 2, we briefly describe the different approaches proposed in the literature to model deformable shapes. In Section 3, we describe the hierarchical deformable model considered in this paper. The joint likelihood of the global statistical model is derived and the different optimization procedures used in the bayesian estimation of deformations are presented. Experimental results are reported in Section 4. Synthetic images with very low signal-to-noise ratio, non-gaussian noise and occlusions are first processed. The method is also demonstrated on real image sequences showing hands moving against a background with partial occlusions.

2 Deformable models in image analysis

Deformable models are mathematical models which incorporate knowledge about shapes and their variations, [18]. These models have been used in still as well as in the analysis of dynamic images to extract, track or characterize deformable objects.

Different classes of deformable phenomena may be encountered ;

- articulated objects (composed of rigid parts),
- structured shapes undergoing elastic deformations (hands, biomedical or biological shapes,...)
- fluid deformable phenomena (clouds,...).

The models we consider in this paper are suited to the representation of objects belonging to the two first classes.

Since the early work of Kass, Witkin and Terzopoulos on active contour models (or snakes) [21], deformable models have gained increasing popularity in computer vision, [9, 19]. First considered in static image segmentation, these models are now used with success in image restoration [18], to track deformable structures in image sequences [8, 20, 25] or to characterize objects with the help of deformation analysis, [2, 19].

Computer vision methods relying on deformable models are often expressed as the minimization of (global) energy functions describing the interactions between the observed data and the parameters of the model. Deterministic or stochastic optimization methods are used to this end. Deterministic approaches converge to model configurations corresponding to local minima of the energy function. Initialization, close to the desired solution are therefore required [9, 11, 21]. Stochastic schemes may be used to obtain optimal solutions but are often cpu expensive [1, 19, 18].

General purpose closed contours ("snakes" and variants) controlled by elastic forces based on local curvature, inflating forces and image based potentials (created for instance by local edges) have been used to extract continuous contour lines [3, 10, 9, 15, 21, 23, 36]. Their limits and drawbacks are now well-known, [3, 9, 23]. The optimization of the energy function associated to these models is performed using variational principles and finite differences techniques, [3, 21] which need an appropriate initialization to converge to a relevant solution. Automatic procedures have been proposed

in this context [29]. Snakes are not well adapted to the modeling of shapes with discontinuities although some techniques have been proposed recently to take them into account, [3]. These models have been generalized to the representation of deformable surfaces by Cohen *et al* [9] for 3D segmentation in medical imaging. Finite element methods have been introduced in this context.

Application-tailored parameterized templates have also been proposed in cases where strong a priori knowledge about the shape to analyse is available. The parameterized templates described by Bouthemy [5] and Yuille, [24, 37] rely on a specific description of the structure of the shape to be represented. These models have been used to detect atmospheric disturbances in meteorological pictures [5] and to extract and track deformable features such as eyes or lips in human faces [37]. These models are hand-built using simple parameterized 2D geometric representations and adjusted by minimizing an energy function.

3D physical models based on rigid and deformable parts, primarily used in Computer Graphics, have been considered in image analysis to segment and track deformable objects [26, 28, 27, 34, 35, 33]. Deformable structures are modeled using parametric models such as superquadrics [33] or other polynomial shape models [27]. The evolution of the shape is governed by the laws of rigid and non rigid dynamics expressed by a set of Lagrangian equation of motion. Modal analysis methods (stemmed from mechanics) have been introduced in this context. These methods allow to generate different shapes using the free vibration modes of parametric models [26, 27, 33].

A dual point of view on elastic matching focuses on models of random deformations for a given initial shape (deformable template). Grenander *et al* [2, 18] and Mardia *et al.* [25] have obtained very promising results in image restoration and segmentation by considering statistical deformable models which describe the statistics of *local* deformations (transformations) applied to an original template. Markov models have been introduced, along with bayesian estimation methods in order to derive the optimal random deformations [18]. 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 computationally demanding [18], when no initial guess, close to the optimal solution can be provided.

In this paper we propose an alternate approach which relies on a hierarchical statistical description of deformations in which both global and local deformations are represented. Global deformations are described with

a reduced number of parameters that are optimized with a fast stochastic optimization technique. Local deformations are introduced to refine the globally deformed template. The global description of deformations we consider relies on a (statistical) modal decomposition based on the Karhunen-Loeve transform. Staib and Duncan, [32, 31] have proposed to use a decomposition on a Fourier basis, associated to iterative minimization techniques to analyse deformable objects. Cootes *et al*, [12, 11] consider the Karhunen-Loeve (KL) decomposition 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 [12, 11], 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 [11].

The method we present here combines the advantages of fast global optimization techniques (which are insensitive to initialization) with a compact hierarchical statistical description of deformations. This yields fast model adjustment and robust segmentation. The proposed model is described in the next section.

3 A hierarchical statistical deformable model

Our approach relies on the description of the object class of interest using a "deformable template" which incorporates *a priori* knowledge on the structure of the object and its variability. A particular configuration of the deformable shape is derived from the original template by specifying two kind of transformations (figure 1) :

- global deformations correspond to transformations from the group of similarity (rotation, translation, scale) and to the main variation modes associated to a KL transform of the deformations observed on a training set, as proposed by Cootes [12].
- local deformations are modeled as a random process which modifies locally the location of the points belonging to the globally deformed template.

This hierarchical decomposition of deformations enables an accurate approximation of the outline of the deformable shape.

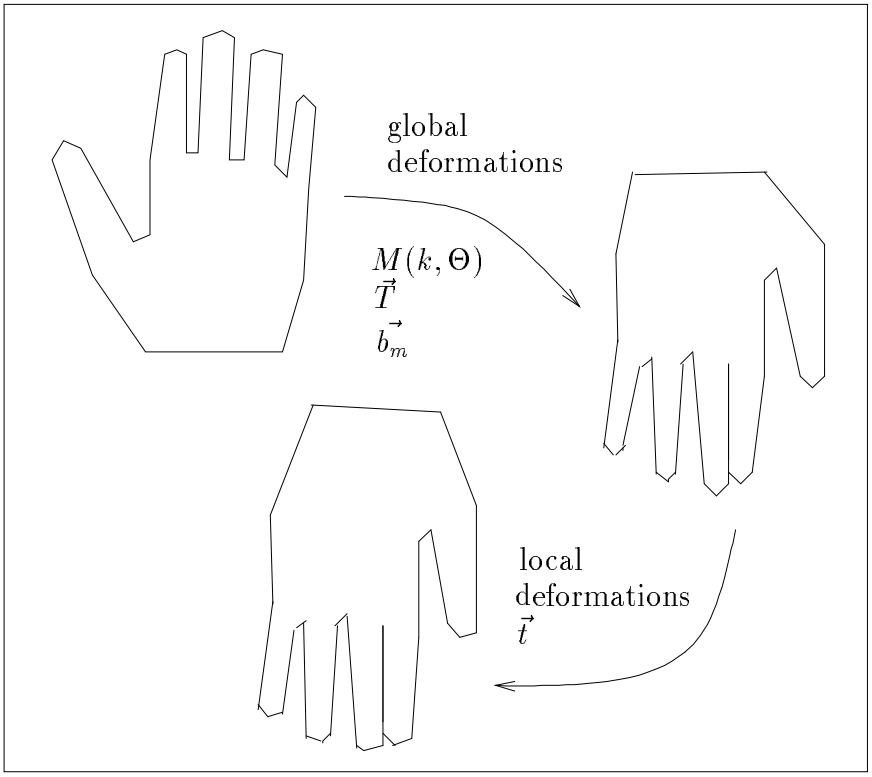

Figure 1: *Hierarchical description of deformations*

3.1 Description of global deformations

To identify the main variation modes of the shape of interest, an (off-line) training procedure is performed on a representative population.

The shapes from the training set are approximated by their polygonal outlines, obtained by standard edge detection and edge linking operators. The different shapes are aligned in the same way with respect to a set of axes, using the method described by Cootes *et al.* in [12]. The most salient points of the outline are labeled in a consistent way and the original template is defined as the "mean shape" of the set of aligned shapes (figure 2).

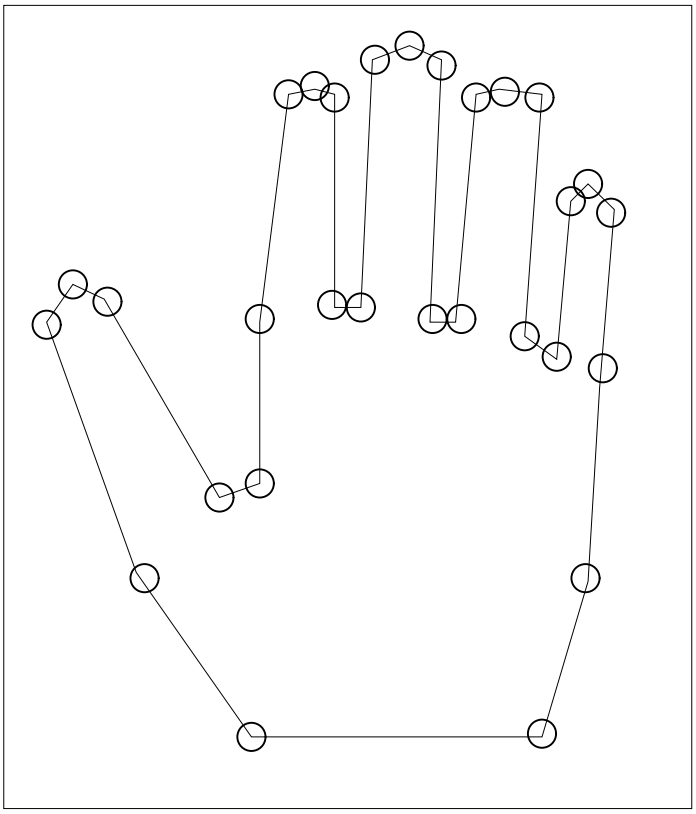

Figure 2: *A 29 point model of a hand*

Let $\vec{x_i}$ be the vector representing the n labeled points of the ith shape in the training set :

$$\vec{x_i} = (x_{i1}, y_{i1}, x_{i2}, y_{i2}, ..., x_{ik}, y_{ik}, ..., x_{in}, y_{in})^T \tag{1}$$

the mean shape $\vec{x}^\star$ is defined as :

$$\vec{x}^\star = \frac{1}{N_f} \sum_{i=1}^{N_f} \vec{x_i} \tag{2}$$

where N_f is the number of shapes in the training set.

The variabilities within the training set are characterized by the displacements $\vec{d_{x_i}}$ of the different labeled points with respect to the mean model :

$$\vec{d_{x_i}} = \vec{x_i} - \vec{x}^\star \tag{3}$$

The variation modes of the model are described by the unit eigenvectors $\vec{p_i}$ ($i = 1$ à $2n$) of the covariance matrix S, defined by :

$$S = \frac{1}{N_f} \sum_{i=1}^{N_f} \vec{d_{x_i}}^T \vec{d_{x_i}} \tag{4}$$

and :

$$S\vec{p_i} = \lambda_i \vec{p_i} \tag{5}$$

where λ_i is the ith eigenvalue of S. Eigenvectors associated to the largest eigenvalues correspond to the most significant variation modes. The new basis associated to this KL transform is :

$$\vec{b} \;=\; P^T \vec{dx} \;=\; P^T(\vec{x} - \vec{x}^\star) \tag{6}$$

where $P = (\vec{p_1}, \vec{p_2}, \ldots, \vec{p_{2n}})$ is the eigenvector matrix and $\vec{b}$ corresponds to the parameters of the deformations in the new basis. An accurate description of the main variation modes may be obtained by retaining only the m eigenvectors associated to the m largest eigenvalues [12] :

$$\vec{x} \;=\; \vec{x}^\star + P_m \vec{b}_m \tag{7}$$

where $P_m = (\vec{p_1}, \vec{p_2}, \ldots, \vec{p_m})$ and $\vec{b}_m = (b_1, b_2, \ldots, b_m)^T$.

The global deformation of the statistical model will thus be characterized by the control parameters $b_i,\ i = 1, ..., m$.

Eigenvalue	$\lambda_i / \sum_{i=1}^{2n} \lambda_i$	total %
λ_1	44%	44%
λ_2	21%	65%
λ_3	14%	79%
λ_4	6%	85%
λ_5	5%	90%
λ_6	4%	94%
λ_7	3%	97%
λ_8	2%	99%

Table 1: *Normalized eigenvalues computed on a training set of 27 point models of hands*

Table 1 presents the eigenvalues computed on a 27 point model of a hand, using a training set of 11 shapes. The percentage of the total variation that is accounted for by each eigenvalue is given in this table. We can notice that a few modes only (from five to seven) integrate more than 90% of the total variability.

Besides the global deformation parameters, we also consider global transformations from the similarity group which lead to the following model $\vec{X}^\star$ for global deformations :

$$\vec{X}^\star \;=\; M(k,\theta)\left[\vec{x}^\star + P_m \vec{b}_m\right] \;+\; \vec{T} \tag{8}$$

The transformations from the similarity group include a translation $\vec{T} = (t_x, t_y, t_x, t_y, ...)$, a rotation (of angle θ) and a scale change by a factor k :

$$M(k,\theta) \;=\; \begin{pmatrix} kcos\theta & -ksin\theta & 0 & 0 & 0 & ... \\ ksin\theta & kcos\theta & 0 & 0 & 0 & ... \\ 0 & 0 & kcos\theta & -ksin\theta & 0 & ... \\ 0 & 0 & ksin\theta & kcos\theta & 0 & ... \\ 0 & 0 & 0 & 0 & ... & ... \end{pmatrix} \tag{9}$$

3.2 Modeling local deformations

Due to the KL transform and other approximations, model (8) only accounts for the global deformations of the shape of interest. 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 $\vec{t}$ is assumed to be a first-order markov process, which takes into account dependencies between the local deformations applied to neighboring points. The global statistical model, taking into account global and local deformations becomes :

$$\vec{X} \;=\; M(k,\theta)\left[\vec{x}^{\star} + P_m \vec{b}_m\right] \;+\; \vec{T} \;+\; \vec{t} \tag{10}$$

where :

$\vec{t} = (\vec{t_1}, \vec{t_2}, \ldots, \vec{t_{2n}})$ is the random local deformation process applied on the n labeled points. $\vec{t}$ is assumed to be a zero-mean Gauss-Markov random process :

$$p(\vec{t}) \;=\; \frac{1}{C}\, exp \;-\; \frac{1}{2}\, \vec{t}^{\,T}\, R^{-1}\, \vec{t} \tag{11}$$

where R is the covariance matrix of $\vec{t}$ and C is a the partition function. Under first-order assumptions on this Markov random process, the joint distribution of vector $\vec{t}$ may be rewritten as :

$$p(\vec{t}) \;=\; \frac{1}{C}\, exp \;-\; \frac{1}{2} \sum_{i=1}^{2n} \left[\, \frac{1}{\epsilon_i^2}\,(t_i \;-\; t_{i-1})^2 \;+\; \frac{1}{\sigma_i^2}\, t_i^2 \,\right] \tag{12}$$

where σ_i^2 and ϵ_i^2 are the parameters of the model. ϵ_i^2 weights the interactions between neighboring points. As far as σ_i^2 is concerned, low values for σ_i^2 draw the shape towards the globally deformed model $\vec{X}^{\star}$. These parameters have been considered constant and adjusted to the same values in all experiments.

3.3 Application to image segmentation

Unsupervised image segmentation remains a fundamental issue in low-level image analysis and computer vision since most high-level processing schemes must rely on reliable and robust segmentations. The purpose of segmentation is to partition the image into meaningful components. Each component corresponds to a region of similar attribute (luminance, color, texture, depth, motion, etc.) [6, 13, 17]. The extracted components may be used in a second step in medium or high-level identification or interpretation schemes.

Significant improvements have been obtained in segmentation problems by introducing global statistical models such as Markov Random Field models or deformable models [16, 19, 18]. In this paper we consider the pro-

blem of extracting (and tracking) moving deformable objects in an image sequence.

3.3.1 Bayesian estimation of deformations

The hierarchical statistical deformable model defined in the previous section is used as an a priori model within a global bayesian estimation scheme. To estimate the deformations, the method proceeds as follows :
• 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.
• Observations are combined, within a global statistical model, with a priori knowledge on the structure (and variability) of the shape, in order to derive estimates of the deformations.

This defines a consistent mathematical framework for determining the global and local deformation parameters of the template. The joint distribution of the local deformations is derived and a Maximum A Posteriori (MAP) estimate of the deformable template $\vec{X}$ 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.

It is easy to verify that the prior distribution of $\vec{X}$ is a first order Gauss-Markov distribution (according to the assumption on the statistics of $\vec{t}$, see Eq. 12) :

$$p(\vec{X}) \;=\; \frac{1}{C}\, exp \,-\, \frac{1}{2}\,(\vec{X} - \vec{X}^{\star})^{\,T}\, R^{-1}\,(\vec{X} - \vec{X}^{\star}) \tag{13}$$

with :

$$\vec{X} \;=\; \vec{X}^{\star} \;+\; \vec{t} \tag{14}$$

and

$$\vec{X}^{\star} \;=\; M(k,\theta)\left[\vec{x}^{\star} + P_m \vec{b}_m\right] \;+\; \vec{T} \tag{15}$$

This prior distribution controls the local and global deformations of the original template. The global similarity transformations $M(k,\theta)$, $\vec{T}$ and the global deformation modes $\vec{b}_m$ are considered as deterministic parameters of

the probabilistic model (Eq. 13) and are estimated separately using a Maximum Likelihood (ML) method.

Let now $O = (O_s,\ s \in S)$ designate an observation field defined on a rectangular lattice S (in our image sequence analysis problem, observations are related to the spatiotemporal variations of the intensity function).

The Maximum A Posteriori (MAP) estimate of the deformable template is defined by :

$$\begin{aligned} \vec{X}_{opt} &\triangleq \arg\max_{\vec{X}}\ p(\vec{X}|O) \\ &= \arg\max_{\vec{X}}\ p(O|\vec{X})p(\vec{X}) \end{aligned} \tag{16}$$

The distribution $p(O|\vec{X})$ describes the interaction between the observations and the deformations to estimate. This distribution is described in the next section.

3.3.2 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. In order to extract moving objects from the image sequence, temporal variations are computed using two complementary methods. The first one measures variations $O_1(s)$ on three successive images ; the second one estimates the changes $O_2(s)$ between the current image and a reference image which is created and updated on line :

$$\begin{aligned} O_1(s) &= min(|I_t(s) - I_{t-1}(s)|, |I_{t+1}(s) - I_t(s)|) \\ O_2(s) &= |I_{ref}(s) - I_t(s)| \end{aligned} \tag{17}$$

The reference image $I_{ref}(s)$ is obtained using a linear estimator of the background described in [14] . Observations $O_1(s)$ present high values for points belonging to a moving object and low values for background points. Temporal gradients such as $O_1(s)$ are known to yield poor observations in homogeneous regions. The second observation $O_2(s)$ based on a reference

image is less sensitive to this problem and is used as a complementary information.

Distribution $p(O|\vec{X})$ is specified by describing the interactions between local observations $O_1(s)$ and $O_2(s)$ and the configuration of the deformable model. For a given configuration of the template, the image can be partitioned into two regions : the inside of the template $\vec{X}$ (denoted R^-) and the outside of $\vec{X}$ (denoted R^+) corresponding to the background. The following (Gibbs) distribution is adopted for $p(O|\vec{X})$:

$$p(O|\vec{X}) \;=\; \frac{1}{C'}\, exp\, -\, [\, \sum_{s\,\in\,R^-} |O(s)-1\,| \;+\; \sum_{s\,\in\,R^+} |O(s)-0\,| \;\,] \tag{18}$$

where :

$$O(s) \;=\; max(\Gamma_\alpha(O_1(s)), \Gamma_\beta(O_2(s))) \tag{19}$$

$$\begin{aligned} \Gamma_\eta(y) &= 1 \quad if \quad y > \eta \\ \Gamma_\eta(y) &= 0 \quad else \end{aligned} \tag{20}$$

and C' is a normalization constant.

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.

3.3.3 Global optimization

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

$$\vec{X}_{opt} \;=\; \arg\max_{\vec{X}}\; p(O|\vec{X})p_\Theta(\vec{X}) \tag{21}$$

where Θ denotes the global parameters of the model :

$$\Theta \;=\; (\, M(k,\theta), \vec{T}, \vec{b_m}\,) \tag{22}$$

The joint distribution may thus be rewritten as a Gibbs distribution :

$$p(O|\vec{X})p_\Theta(\vec{X}) \;=\; \frac{1}{Z} exp\; -U_\Theta(O,\vec{X}), \tag{23}$$

with the following energy function :

$$U_\Theta(O,\vec{X}) = \sum_{s \in R^-} |O(s)-1| + \sum_{s \in R^+} |O(s)-0| + \frac{1}{2}(\vec{X}-\vec{X}^\star)^T R^{-1}(\vec{X}-\vec{X}^\star). \quad (24)$$

Let us notice that the partition function Z does not depend on the global parameter Θ. To estimate the global parameters Θ we have adapted a method recently described by Lakshmanan *et al.* [22] for estimating global model parameters in Markov Random Field-based unsupervised segmentation. Model parameters are estimated alternately with the configuration of the model, yielding a partial optimal solution [22]. In our case, the global parameters Θ are estimated using a Maximum Likelihood procedure. A global optimization of the joint-likelihood (Eq. 23) with respect to parameter Θ is performed with a fast stochastic relaxation procedure. The MAP estimate of the deformable template $\vec{X}$ is then approximated using a deterministic relaxation scheme.

This two optimization steps have to be performed alternately according to the following procedure [22] :

Initialization : $\vec{X}_{opt} = \vec{x}^\star$
Do alternately :

- perform a stochastic optimization to determine the global model parameters Θ.

$$\Theta_{opt} = Arg\ Min_\Theta\ U_\Theta(O,\vec{X}_{opt}) \quad (25)$$

- perform a deterministic local relaxation using a modified version of the ICM algorithm [4] to determine the local deformation process $\vec{t}$.

$$\vec{X}_{opt} = Arg\ Min_{\{\vec{X} = \vec{X}^\star + \vec{t}\}}\ U_{\Theta_{opt}}(O,\vec{X}) \quad (26)$$

End do

In practice only one iteration of this optimization loop is performed. This procedure has shown to provide good and stable results and additional iterations have only produced slight improvements on the final estimates.

In most methods involving deformable templates (apart from the work of Grenander [18]), deterministic minimization algorithm are used. They are known to be very sensitive to local minima of the objective function to minimize. 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, as will be seen.

Contrary to the model of Grenander [18] which only describes *local* deformations, the hierarchical model proposed here may be optimized efficiently thanks to the reduced number of global parameters. The local deformations are identified using a deterministic relaxation algorithm (ICM) which has fast convergence properties. Local minima are not a problem since the first step (global optimization) provides a good initialization for the deterministic refinement procedure.

The global optimization which has been implemented relies on a standard simulated annealing algorithm based on the Gibbs sampler dynamics [16] (the Metropolis algorithm has been tested with similar results). It is easily shown that the global energy variations ΔU, for each new configuration to be tested, only depends on regions $A \cup C$ depicted on figure 3.

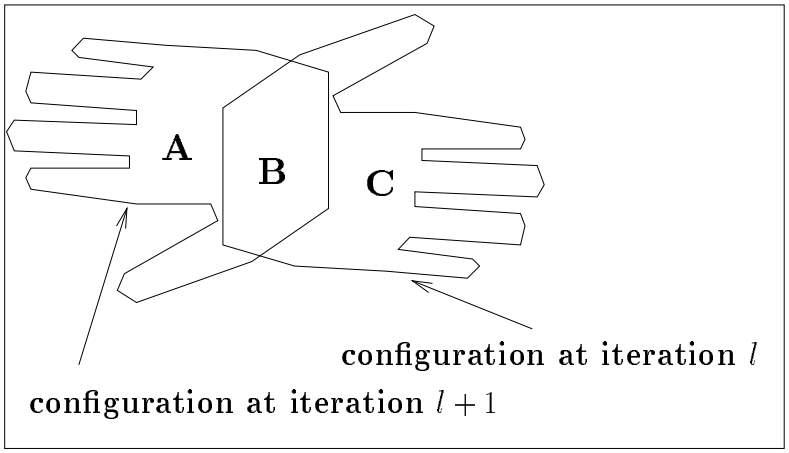

Figure 3: *Regions concerned by the computation of global energy variations*

To determine these different regions, the model is delineated using the standard Bresenham [7] algorithm and regions R^+ and R^+ are obtained using the Smith [30] coloring algorithm.

4 Experimental results

In our experiments, we have considered the segmentation of shapes corresponding to hands moving against a background. The case of hand patterns, studied by Grenander [18], Cootes [12, 11] and other authors (on static images) provides a good example of a complex deformable structure. Synthetic observations corresponding to motion detection maps have first been generated from real hands outline in order to study the effect of (non gaussian) noise and partial occlusions. Results on real-world image sequences are also presented.

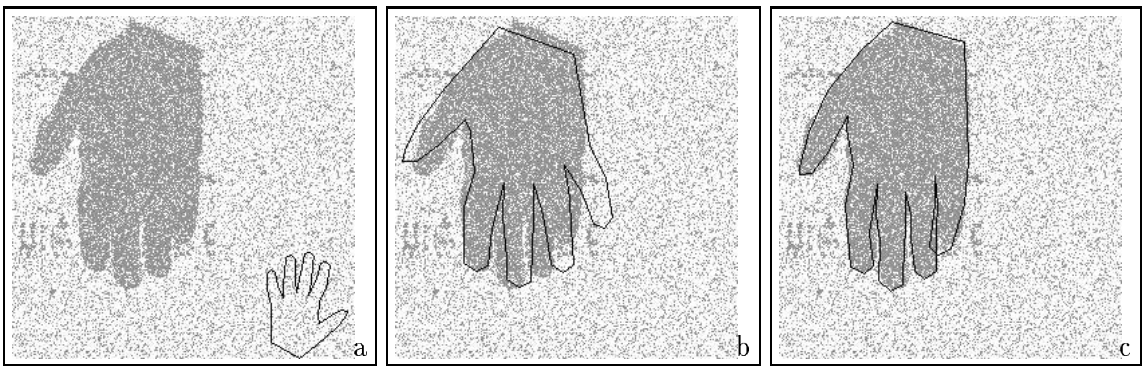

Figure 4: *Intermediate steps in the segmentation process (see text)*

In the first experiments on synthetic images, the detection map $\{O(s), s \in S\}$ has been simulated. More precisely, we have considered an image in which the inside of the hand is constant and equal to 1 (moving object) and the outside of the hand is equal to 0 (background). The outline of the hand is the boundary of a real hand (which does not belong to the training set)[1]. This "synthetic" image is corrupted by changing, with a given probability p, the values of the binary observations $O(s)$. Very low signal-to-noise have been simulated as well as occlusions. On these synthetic examples, a 40 point models has been specified to represent the hand. 11 different (representative)

[1]The training set has been built from discretized outlines of scanned images of hand.

hand configurations have been used in the training set. In the KL transform the 9 first most significant variation modes b_i, corresponding to 95 % of the observed variability have been retained. In the optimization process, parameters b_i are bounded within a three standard variation interval :

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i} \tag{27}$$

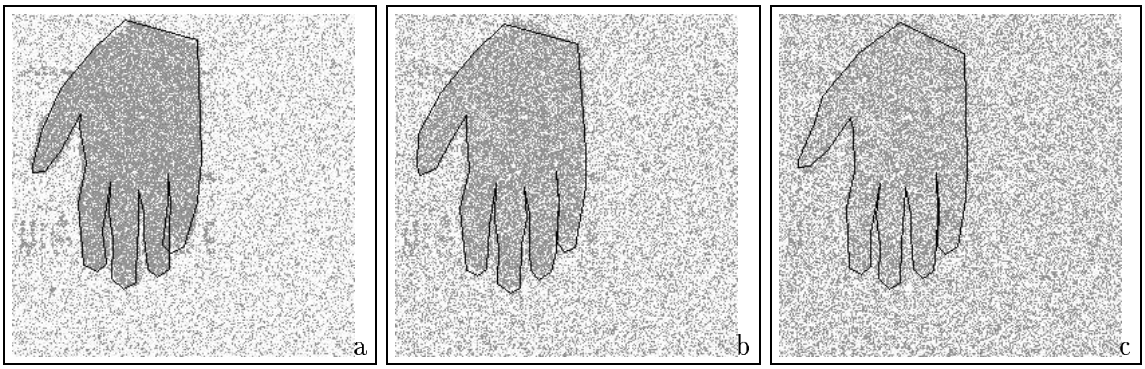

Figure 5: *Segmentations with different signal-to-noise ratio*

Figure 4 presents the result of the MAP estimation with a signal-to-noise ratio corresponding to $p = 0.2$. The noisy observations $O(s)$ are shown. The initial template configuration is generated at random, and can be seen in the lower right part of the image (figure 4a). Figure 4b shows the result of global optimization on the similarity transformation parameters only (translation, rotation and scale parameters). The optimization of the different deformation modes (figure 4c) enables to refine the template and to segment the shape accurately.

Figure 5 shows segmentations obtained with the same scheme on images with different signal-to-noise ratio (corresponding respectively to $p = 0.2$, 0.3 and 0.4). We notice that the hierarchical statistical model is able to recover the underlying shape, even with very low signal-to-noise ratios (see for instance figure 5c).

In the case of figure 6 occlusions (in addition to noise) have been simulated. In figure 6b and 6c, fingers have been suppressed by "digital surgery". These different results tend to prove that the approach is robust to occlusions and to non gaussian noise. The method is able to find a relevant template configuration although some important observations have been lost.

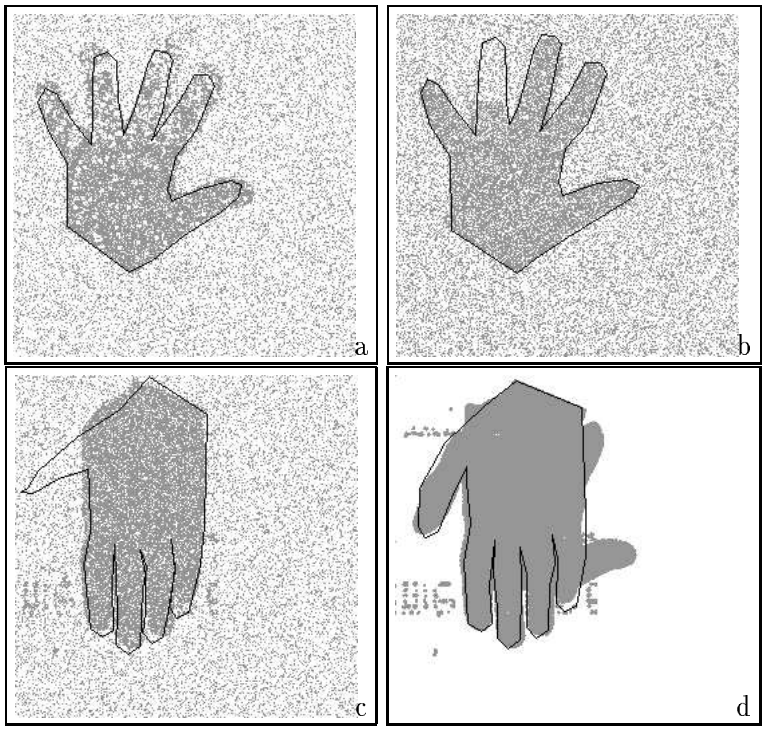

Figure 6: *Segmentations with different signal-to-noise ratio and partial occlusions*

Figure 7 presents the processing of a real-world image sequence composed of six successive frames. The original sequence was composed of more than one hundred frames. The algorithm was able to extract and track the hand reliably over the whole sequence. The tracking was performed by initializing the deformable template in the current frame by the final configuration obtained on the previous frame. This initialization is close to the new configuration if the temporal sampling of the sequence is tight enough. A good initial configuration enables to use a faster annealing schedule in the global optimization process and thus to save cpu time. A slower annealing schedule was used on the first frame for which a random initialization was provided. Let us notice, however, that a pure deterministic optimization procedure was unable to track the structure reliably over the sequence (due to many local minima in the energy function).

In this first real-world example, the hand is moving against a homogeneous background (a textured background is used in the next example). Figures 7a to 7c and 7g to 7i show the result of the global optimization step on the translation-rotation-scale transformations and on the global deformation parameters. A quite accurate result is already obtained in this first preliminary step, which defines the initialization for the local refinement. The results of the deterministic relaxation process (local refinement) are depicted in figures 7d to 7f and 7j to 7k.

The model adopted in this case was a 30 point model built from a training set of 22 hands which belonged to different persons[2]. We have considered in each case the seven most significant deformation modes corresponding to 90% of the total observed variability.

In all experiments the parameter values $\sigma_i^2 = 4$ and $\epsilon_i^2 = 1$ were adopted. The results turned out to be largely insensitive to the adjustment of these parameters (due to the unsupervised global optimization step). The total cpu time, on a Sun-4 station (sun IPX) was about 8 minutes for the complete processing of one 256x256 frame. This algorithm is of course slower than the deterministic scheme proposed by Cootes [11] (about 52.5s on a workstation) but it does not require an initialization close the optimal configuration and is less sensitive to local minima. Contrary to the method described in [11], which is edge-based, the approach presented here is region-based which leads also more robust results.

The robustness of the approach is best illustrated on a second sequence showing a hand moving against a textured background (figure 8). The se-

[2]The processed images did not belong to the training set.

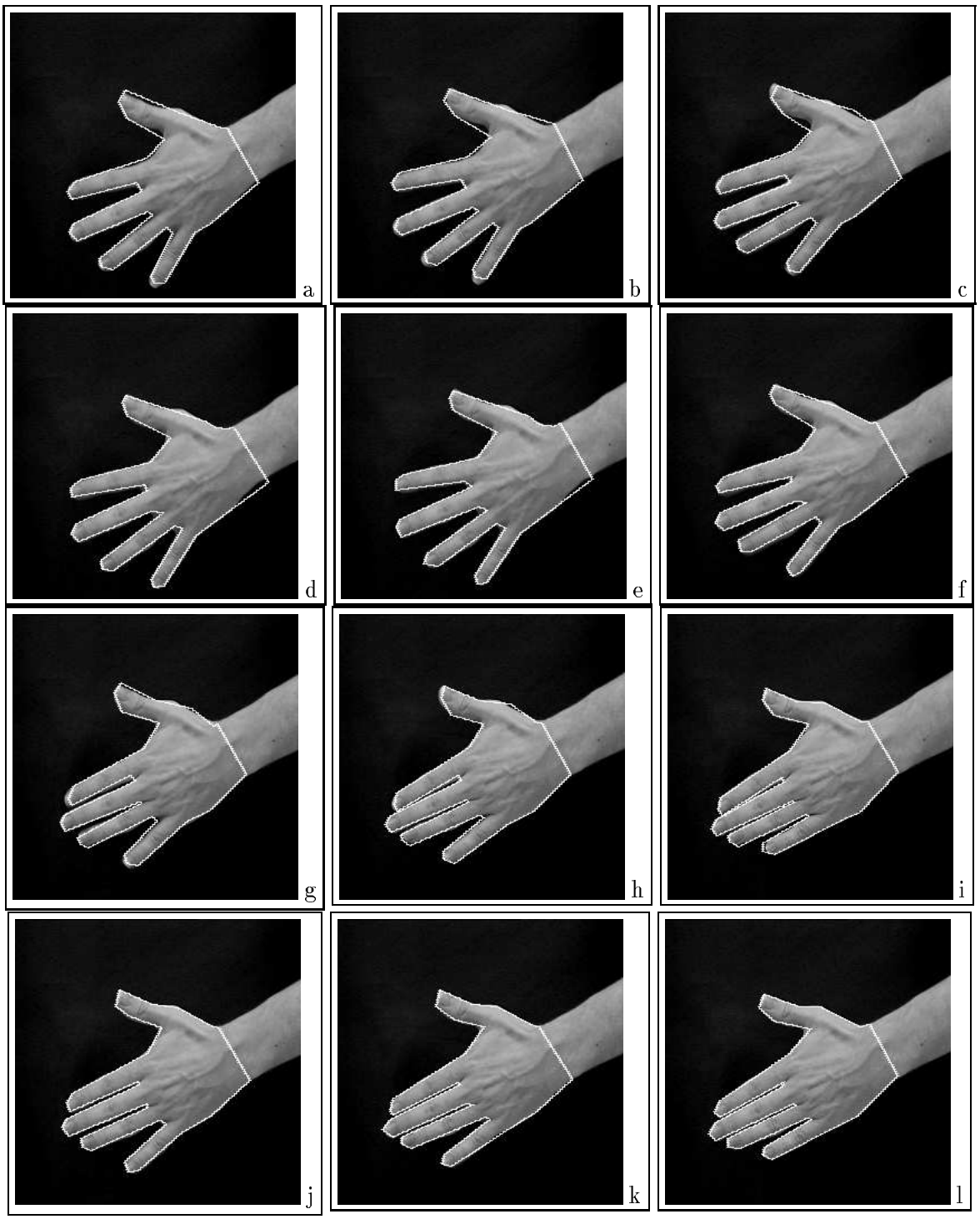

Figure 7: *Segmentation and tracking of a moving hand using the hierarchical deformable model (homogeneous background). a-c and g-i: result of the optimization on the global parameters* Θ *of the model on six successive frames. d-f and j-l: final* MAP *estimate (including the local deformation process).*

quence also contains a partial occlusion situation (figures 8g to 8l) in which two fingers disappear behind an object which is located in the foreground of the scene. Six frames are presented. The deformable model is the same as in the previous example. Again, figures 8a to 8c and 8g to 8i show the result of the first global optimization step on the translation-rotation-scale transformations and on the global deformation parameters. The results of the deterministic relaxation process are depicted in figures 8d to 8f and 8j to 8k. The hierarchical model and the statistical approach are able to segment the shape, to track it and to maintain it even in the partial occlusion area. The observations $O(s)$ (Eq. 19), corresponding to the different frames of this sequence are presented in figure 9. As can be seen the observations are noisy and partial (in the occlusion area).

5 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.

This approach has been demonstrated on synthetic as well as on real-world images sequences for the segmentation of moving deformable objects. It is shown to be robust even in adverse situations (very low signal-to-noise ratio, non gaussian noise and occlusions).

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.

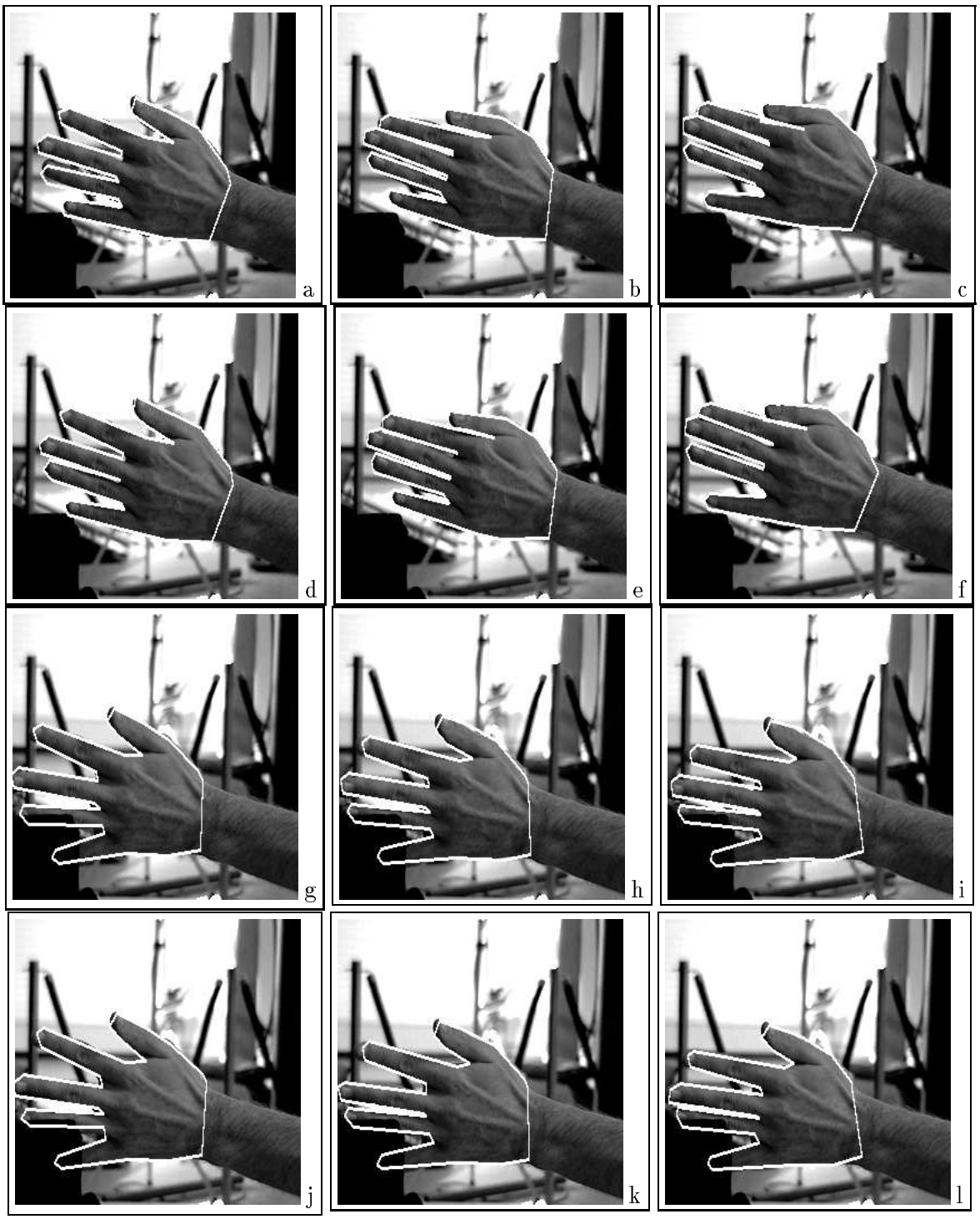

Figure 8: *Segmentation and tracking of a moving hand using the hierarchical deformable model (textured background and occlusions). a-c and g-i: result of the optimization on the global parameters* Θ *of the model on six successive frames. d-f and j-l: final* MAP *estimate (including the local deformation process).*

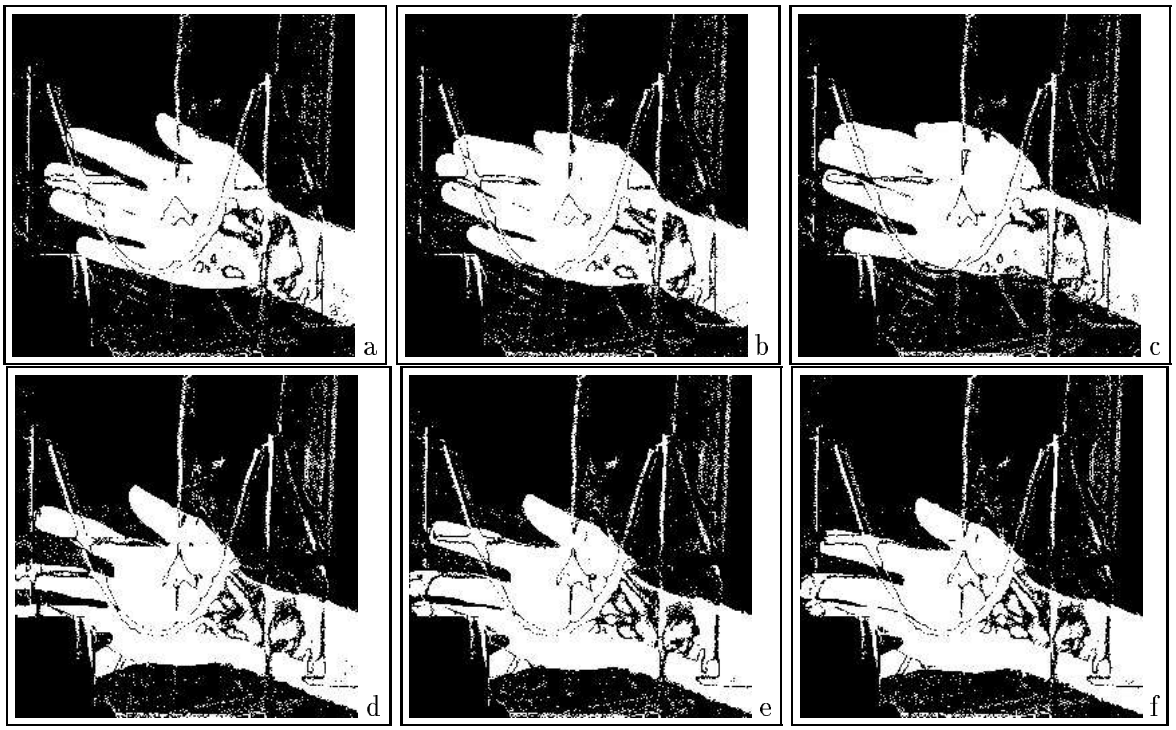

Figure 9: *Observations maps $O(s)$ computed on the image sequence of the moving hand (textured background and occlusions). The original sequence is presented figure 8.*

References

[1] E.H.L. AARTS and P.J.M. van LAARHOVEN. – *Simulated Annealing: Theory and Applications.* – D. Reidel Publishing Company, 1987.

[2] Y. AMIT, U. GRENANDER, and M. PICCIONI. – Structural image restoration through deformable templates. – *Jour. Americ. Statist. Assoc.*, Vol. 86, No 414, Theory and Methods: pp. 376–387, June 1991.

[3] M. O. BERGER. – Tracking rigid and no polyhedral objects in an image sequence. – In *8th Scandinavian Conf. Image Analysis*, pages pp. 945–957, Tromso, Norway, May 1993.

[4] J. BESAG. – On the statistical analysis of dirty pictures. – *J. Royal Statist. Soc.*, Vol. 48, Serie B, No 3: pages 259–302, 1986.

[5] P. BOUTHEMY and A. BENVENISTE. – Modeling of atmospheric disturbances in meteorological pictures. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No 5: pp. 587–600, Sept 1984.

[6] P. BOUTHEMY and E. FRANÇOIS. – Motion segmentation and qualitative dynamic scene analysis from an image sequence. – *Int. J. Computer Vision*, Vol 10, No 2: pages 157–182, 1993.

[7] J. E. BRESENHAM. – Algorithm for computer control of a digital plotter. – *IBM System Journal*, Vol. 4, No 1: pp. 25–30, 1965.

[8] I. COHEN, N. AYACHE, and P. SULGER. – Tracking points on deformable objects using curvature information. – In *ECCV 92*, pages 458–466, Santa Margherita Ligure, Italy, May 1992.

[9] I. COHEN, L.D. COHEN, and N. AYACHE. – Using deformable surfaces to segment 3D images and infer differential structures. – *CVGIP : Image Understanding*, Vol. 56, No 2: pp. 242–263, Sept. 1992.

[10] L. COHEN. – On active contour models and balloons. – *CVGIP : Image Understanding*, Vol. 53, No 2: pp. 211–218, 1991.

[11] T.F. COOTES and C.J. TAYLOR. – Active shape models : Smart snakes. – In *British, Machine Vision Conf.*, pages 266–275, Leeds, UK, Sept. 1992.

[12] 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.

[13] H. DERIN and H. ELLIOT. – Modeling and segmentation of noisy and textured images using Gibbs random fields. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 9, No 1: pages 39–55, Jan. 1987.

[14] G.W. DONOHOE, D.R. HUSH, and N. AHMED. – Change detection for target detection and classification in video sequences. – In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 1084–1087, New-York, 1988.

[15] P. FUA and Y. LECLERC. – Model driven edge detection. – *Machine Vision and Applications*, Vol. 3: pp. 45–56, 1990.

[16] S. GEMAN and D. GEMAN. – Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No 6: pages 721–741, November 1984.

[17] C. GRAFFIGNE. – *Experiments in Texture Analysis and Segmentation.* – PhD thesis, Brown University, 1986.

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

[19] U. GRENANDER and D. M. KEENAN. – Towards automated image understanding. – *J. Applied Statistics*, Vol. 16,No 2: pp. 207–221, 1989.

[20] I.L. HERLIN and N. AYACHE. – Features extraction and analysis methods for sequences of ultrasound images. – *Image and Vision Computing*, Vol. 10: pp. 673–682, Dec. 1992.

[21] 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.

[22] S. LAKSHMANAN and H. DERIN. – Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No 8: pages 799–813, 1989.

[23] F. LEITNER, I. MARQUE, S. LAVALLEE, and P. CINQUIN. – Dynamic segmentation. Finding the edge with snake splines. – In *Curves and Surfaces*, Chamonix, France, 1990. Academic Press.

[24] P. LIPSON, A.L. YUILLE, D.O'KEEFE, J. CAVANAUGH, J. TAAFLE, and D. ROSENTHAL. – Deformable templates for feature extraction from medical images. – In *Proc. First European Conf. Comp. Vision*, pages 413–417, Antibes,France, April 1990.

[25] K.V. MARDIA and T.J. HAINSWORTH. – Deformable templates in image sequences. – In *Proc. Int. Conf. Pattern Recognition*, pages 132–135, La Haye, The Netherlands, Sept 1992.

[26] C. NASTAR and N. AYACHE. – Fast segmentation, tracking and analysis of deformable objects. – In *Proceedings of the Fourth International Conference on Computer Vision (ICCV '93)*, pages 275–279, Berlin, Germany, May 1993.

[27] A.P. PENTLAND and B. HOROWITZ. – Recovery of non-rigid motion and structure. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No 7: pp. 730–742, July 1991.

[28] A.P. PENTLAND and S. SCLAROFF. – Closed-form solutions for physically based shape modeling and recognition. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No 7: pp. 715–729, July 1991.

[29] N. ROUGON and F. PRETEUX. – Directional adaptive deformable models for segmentation with applications to 2D and 3D medical images. – In *SPIE Conference on Medical Imaging VII, Image Processing*, volume 1898, Newport Beach, California , USA, February 1993.

[30] A. R. SMITH. – Tint filling. – In *SIGGRAPH-ACM*, pages 276–283, August 1979.

[31] 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, November 1992.

[32] L.H. STAIB and J.S. DUNCAN. – Parametrically deformable contour models. – In *IEEE Proc. Conf. Comp. Vision Pattern Recognition*, pages 98–103, San Diego, USA, June 1989.

[33] D. TERZOPOULOS and D. METAXAS. – Dynamic 3D models with local and global deformations : Deformable superquadrics. – *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No 7: pp. 703–714, July 1991.

[34] D. TERZOPOULOS and A. WITKIN. – Deformable models. physically based models with rigid and deformable components. – *IEEE Computer Graphics and Applications*, pages 41–51, Nov. 1988.

[35] D. TERZOPOULOS, A. WITKIN, and M. KASS. – Constraints on deformable models : Recovering 3D shape and nonrigid motion. – *Artificial Intelligence*, Vol. 36: pp. 91–123, 1988.

[36] D.J. WILLIAMS and M. SHAH. – A fast algorithm for active contours and curvature estimation. – *CVGIP : Image Understanding*, Vol. 55, No 1: pp. 14–26, Jan. 1992.

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

Unité de recherche INRIA Lorraine, Technôpole de Nancy-Brabois, Campus scientifique,
615 rue de Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS
Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
(France)
ISSN 0249-6399