A Hierarchical Markov Modeling Approach for the Segmentation and Tracking of Deformable Shapes

Charles Kervrann

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

and

Fabrice Heitz

ENSPS/LSIIT URA CNRS 1871, Boulevard Sébastien Brant, 67400 Illkirch, France
E-mail: Fabrice.Heitz@ensps.u-strasbg.fr

In many applications of dynamic scene analysis, the objects or structures to be analyzed undergo deformations that have to be modeled. In this paper, we develop a hierarchical statistical modeling framework for the representation, segmentation, and tracking of 2D deformable structures in image sequences. The model relies on the specification of a template, on which global as well as local deformations are defined. Global deformations are modeled using a statistical modal analysis of the deformations observed on a representative population. Local deformations are represented by a (first-order) Markov random process. A model-based segmentation of the scene is obtained by a joint bayesian estimation of global deformation parameters and local deformation variables. Spatial or spatio-temporal observations are considered in this estimation procedure, yielding an edge-based or a motion-based segmentation of the scene. The segmentation procedure is combined with a temporal tracking of the deformable structure over long image sequences, using a Kalman filtering approach. This combined segmentation-tracking procedure has produced reliable extraction of deformable parts from long image sequences in adverse situations such as low signal-to-noise ratio, nongaussian noise, partial occlusions, or random initialization. The approach is demonstrated on a variety of synthetic as well as real-world image sequences featuring different classes of deformable objects. © 1998 Academic Press

Key Words: deformable models; Markov models; image sequence analysis; segmentation; tracking; Kalman filtering.

1. INTRODUCTION

Until the middle of the 1980s, the representations developed in model-based image processing essentially aimed at the description and analysis of 2D and 3D rigid structures undergoing rigid movements. However, 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 undergo deformations which have to be analyzed and characterized. The representation and processing of deformations has thus recently gained considerable popularity, especially in the past ten years.

The modeling of deformations, however, remains an intricate problem, due to the wide range of shapes and distorsions which may be encountered: articulated structures (composed of rigid parts) [53], shapes undergoing elastic deformations (i.e., hands, biomedical, or biological shapes) [46, 49, 59], or fluid flows [41, 60]. This calls for the development of appropriate mathematical models, adapted to each particular class of deformations.

1.1. A Hierarchical Markov Modeling Approach

The statistical model introduced in this paper aims at representing 2D moving structures undergoing elastic deformations. Our approach relies on the description of the shape of interest by a deformable template which incorporates statistical knowledge about the shape and its variability [29]. The representation of deformations is based on a Markov modeling [22, 23, 25, 32, 51] of local deformations, associated to a modal representation of global shape distorsions [15]. This hierarchical statistical representation of deformations has shown itself to be very flexible. It has been used with success to extract and track a large variety of deformable shapes (showing high variability) over long image sequences (typically several hundred of frames).

In our representation, the parameters describing global shape deformations 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 a first crude registration of the shape on the input data. In addition, to control the main variation modes of the original template, a Karhunen–Loeve (KL) expansion of the deformations observed on a representative population is used. Following [14, 15], a

modal approximation of global deformations is obtained by retaining the first eigenvectors of this KL expansion. As a consequence, only a relatively small number of parameters enters into the specification of a particular configuration of the model.

Local deformations are modeled, at a second level of the hierarchical representation, as local random perturbations of the shape and are assumed to follow a first-order gaussian Markov process. This local process can be considered as a refinement of the global deformations applied to the original shape, since the main deformation modes have been captured by the preliminary global shape deformation modeling step. The use of a statistical local deformation process has been inspired by the work of Grenander *et al.* [23, 24] on stochastic pattern representation.

The use of the Markov modeling framework enables the derivation of—in a bayesian sense—optimal estimates of deformations as well as the development of well-founded techniques to estimate the parameters of the model. In the experiments that have been conducted, the proposed approach has shown itself to be robust to nongaussian noise as well as to partial occlusions. (Figures 4 and 5 present examples of image segmentations obtained in such adverse conditions.)

1.2. Statistical Image Segmentation

The segmentation of a deformable shape is based on a global bayesian estimation scheme, in which the previously described statistical representation is used as an *a priori* model. A maximum *a priori* (MAP) estimate of the deformation process is obtained by maximizing a highly nonlinear joint probability distribution [22, 23] describing the interactions between data (spatial or temporal gradients extracted from the image sequence) and the deformation process. The global parameters of the model (controlling the global deformations modes obtained from the KL expansion) are obtained simultaneously with the segmentation using a marginalized maximum likelihood (MML) estimator [43]. In motion-based segmentation, global optimization techniques are used to obtain estimates which do not depend on the initial configuration of the model (see Section 4.3.2). This saves the operator the bother of providing manual initializations for the model (even for the first frame). A completely data-driven segmentation is then obtained.

1.3. Temporal Tracking of the Deformable Model

The statistical segmentation procedure outlined in the previous section is combined with a Kalman filter-based temporal tracking of the global deformation parameters of the template. The tracking procedure is reinitialized when an abrupt kinematical change in the deformable movement is detected [30]. Tracking the deformable structure significantly reduces the computational cost of the segmentation method over a long image sequence by propagating good initial segmentations between two successive frames. For the same reason, tracking also enables us to process large movements more reliably.

The combined segmentation-tracking procedure is summarized on the flow diagram depicted in Fig. 1, which presents the way the two procedures interact. The segmentation procedure is composed of three steps:

- An initialization step that provides the initial template configuration for the segmentation procedure.
- A second step in which the global deformation parameters that roughly describe the configuration of the template in the current frame are estimated. This step also provides the measurements used by the Kalman filter to update its current state estimate.
- A final step in which local deformations that refine the description of the deformable shape are obtained. This step provides the final segmentation at time t.

The initial segmentation at time $t \neq 0$ is defined by the template configuration predicted from the previous frame $t - \Delta t$ by the Kalman filter (unless an abrupt change is detected). At time $t = 0$ or when an abrupt change in the movement has been detected, the Kalman filter is reinitialized. In this case, the initial segmentation is provided manually, at random ($t = 0$), or by using the updated estimate obtained at $t - \Delta t$.

The tracking procedure interacts with the segmentation procedure through the filtering of the global deformation parameters provided by the segmentation:

- At first an abrupt change test compares the global deformation parameters obtained at time t by the segmentation procedure to the parameters predicted at time $t - \Delta t$ and eventually decides to reinitialize the tracking procedure.
- Then, the global deformation parameters estimated by the segmentation procedure are used as measurements to update the Kalman filter state.
- Finally the Kalman filter predicts the global deformations of the template at time $t + \Delta t$.

The segmentation and the tracking procedures are detailed in Sections 4 and 5 of this paper.

1.4. Paper Organization

The remainder of this paper is organized as follows. Background and related studies are presented in Section 2. The statistical hierarchical deformable template, combining global deformation modes and a local deformation (Markov) process, is described in Section 3. Image segmentation through bayesian estimation of global and local deformations is considered in Section 4. The segmentation is experimentally shown to be robust to model initialization and nongaussian noise as well as partial occlusions. Section 5 describes the tracking procedure, based on a recursive temporal filtering of the model parameters. The statistical procedure for the detection of abrupt changes in the kinematic behavior of the deformable shape is presented and evaluated. Experimental results obtained by the deformable model-based segmentation and tracking procedure on long, real-world image sequences, are commented on in Section 6. Four case studies, featuring a variety of deformable shapes, are considered: the segmentation and tracking of hands and lips, the
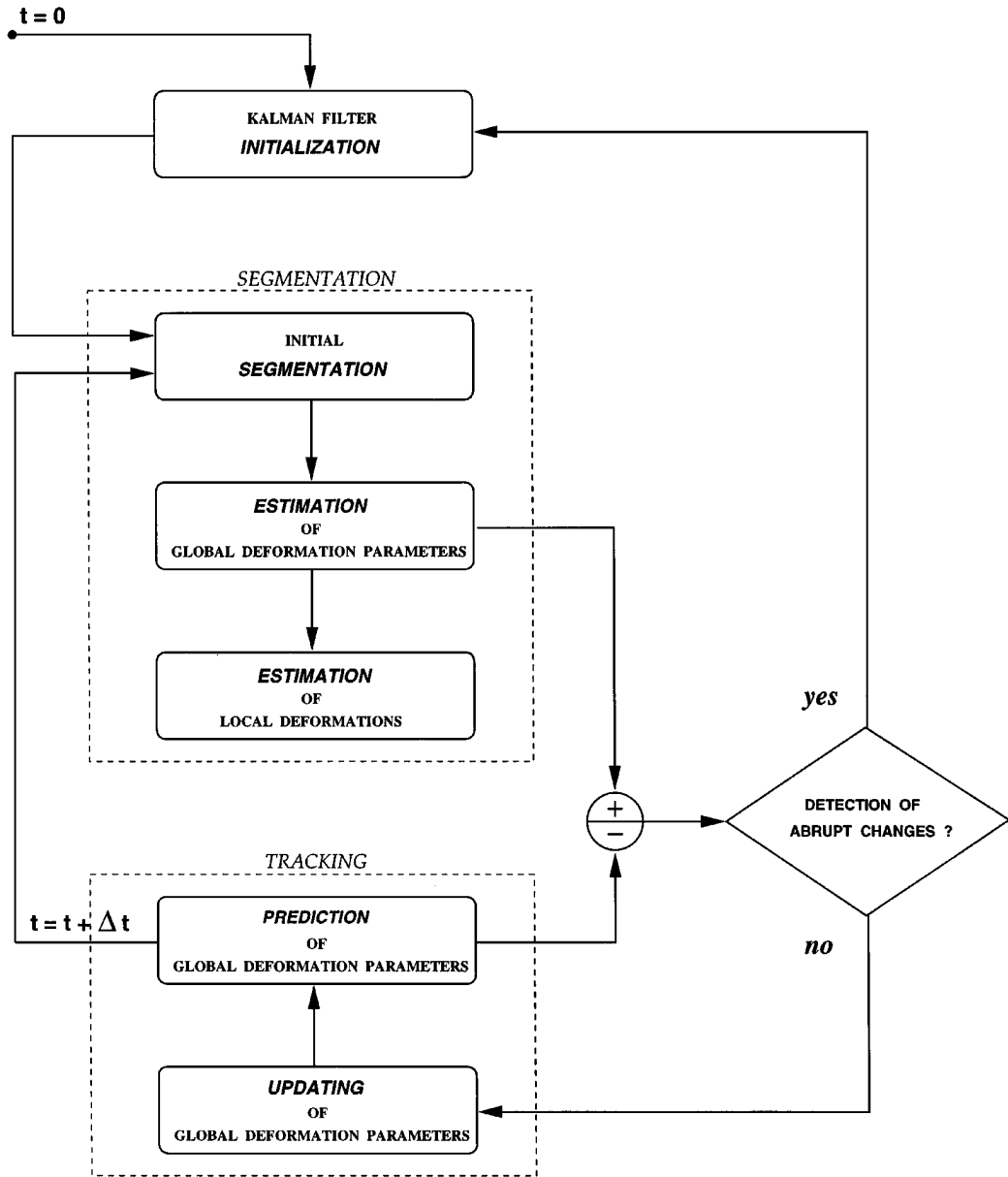
FIG. 1. The deformable model-based segmentation and tracking scheme.

tracking of the coiling of a cinema film for servoing purposes, and the extraction of a beating heart in echocardiographic medical imaging (four additional case studies may be found in [34]). Limitations of the method are also discussed and illustrated.

2. BACKGROUND AND RELATED WORK

Deformable models are mathematical models which incorporate knowledge about shapes and their variations [23]. Since the early work of Kass, Witkin, and Terzopoulos on active contour models (or snakes) [28], deformable models have gained increasing popularity in computer vision, [13, 24, 49]. First considered in static image segmentation, these models are now used with

success in image restoration [23] to track deformable structures in image sequences [30, 39, 46] or to characterize objects with the help of deformation analysis [1, 24, 41]. To this end, deterministic as well as stochastic approaches have been devised.

Among the deterministic approaches, 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 [13, 28]. Their limits and drawbacks are now well known. The optimization of the energy function associated to active contours models is generally performed using variational principles and finite differences techniques [28], which need an appropriate initialization to converge

to a relevant solution. Snakes are not well adapted to the modeling of shapes with discontinuities or multiple shapes, although some techniques proposed recently take them into account [44]. These models have been generalized to the representation of deformable surfaces by Cohen *et al.* [13] for 3D segmentation in medical imaging. Finite element methods have been introduced in this context. More sophisticated concepts have been introduced recently, based on the theory of surfaces evolution and geometric flows [35, 38]. These evolution models generally do not require manual initialization and automatically handle different object topologies, allowing the detection of an arbitrary number of structures in the image.

Deterministic 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 [46, 49, 50, 56]. Deformable structures are modeled using parametric models such as superquadrics [56] or other polynomial shape models [49]. The evolution of the shape is governed by the laws of rigid and nonrigid dynamics expressed by a set of Lagrangian equations of motion. Modal analysis methods (stemming from mechanics) have been introduced in this context. These methods allow us to generate different shapes using the free vibration modes of parametric models [46, 49, 56].

Application-tailored parameterized templates have been proposed in cases where strong *a priori* knowledge about the shape being analyzed is available. The parameterized templates described by Bouthemy *et al.* [10] and Yuille *et al.* [59] 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 [10] and to extract and track deformable features such as eyes or lips in human faces [59]. In [59], the elastic model is adjusted by minimizing an energy function. These models are hand-built using simple parameterized 2D geometric representations. In this approach, building a new model can thus become a quite slow and tedious task.

A more flexible approach was devised by Cootes *et al.* in [14, 15]. In their approach, the shape structure and the parameters describing its deformations are learned from a training set of representative shapes. Modal approximation techniques, based on the (orthogonal) KL transform, allow us to approximate the deformations of the shapes belonging to the learning set on a low-dimension eigenspace. KL-based modal approximation, also known as principal component analysis, is a standard technique in pattern recognition [48]. This technique has been used by, among others, Turk and Pentland [58], for the retrieval and recognition of human faces in large databases, Cootes *et al.* [15], and Martin *et al.* [40] for the description of deformable shapes. Recently, Nayar and Murase [45] took advantage of this compact representation for the development of real-time recognition systems of 3D objects from gray level appearance images. In the case of 2D deformation models, five to ten parameters are usually sufficient to obtain an accurate modal approximation. In [14, 15], the deformation parameters were 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 parameters [15]. A global fitting algorithm based on genetic algorithms was also proposed in [26]. Other orthogonal transforms have been suggested to represent deformations on low dimensional spaces. Staib and Duncan [55] used, for instance, a standard decomposition on a Fourier basis, associated to iterative minimization techniques to analyze deformable objects. Chuang *et al.* introduced wavelet representations in [12] to decompose a curve into components at different scales: coarse scale components are related to global shape features while the finer scale components contain local detail information. Let us however notice that the KL transform yields the more compact representation in every case.

A second point of view on elastic matching focuses on models of *random deformations* for a given initial shape (deformable template). Grenander *et al.* [1, 23] and Mardia *et al.* [39] obtained 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 in this context, along with bayesian estimation methods, in order to derive optimal random deformations estimates [23]. Monte Carlo techniques are necessary to compute optimal MAP estimates. Unfortunately, due to the very large size of the space of configuration, the computation of the MAP estimate is computationally demanding [23], when no initial guess close to the optimal solution can be provided.

The statistical modeling approach we present here combines the advantages of a compact description of global deformations along with an accurate description of local deformations. These two levels of description are embedded within a single statistical (Markov) model, yielding several advantages, with respect to other related statistical [1, 23, 39] or deterministic [15] approaches:

- The use of a Markov modeling framework allows the derivation of (in a bayesian sense) optimal estimates of deformations.
- Model parameters are estimated simultaneously with the deformations using well-founded statistical techniques.
- Contrary to [23], global deformations are described with a reduced number of parameters. The optimal shape configuration may thus be easily obtained with fast stochastic optimization techniques. Local deformations are also estimated with low computational cost, since they simply correspond to a refinement of the globally deformed shape.
- Contrary to the approach of Cootes *et al.* [14, 15], the segmentation scheme proposed herein has shown itself to be robust to nongaussian noise and small occlusions.
- The computation of optimal statistical estimates is based on optimization schemes that do not necessarily require initial template configurations that are close to the desired solution. Initializations may be defined at random, leading to segmentation procedures that are completely data driven.
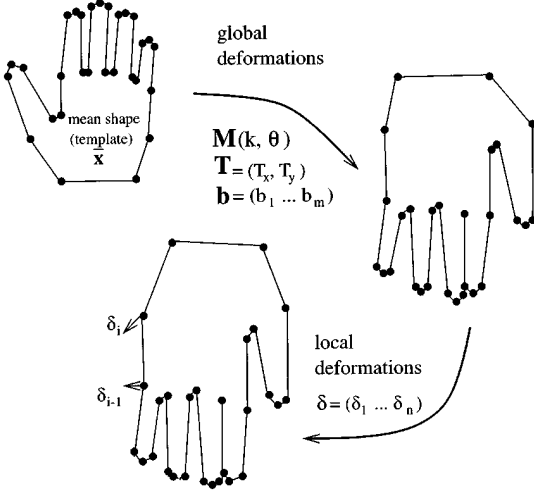
FIG. 2. Hierarchical description of deformations.

The hierarchical Markov model is described in the next section.

3. A HIERARCHICAL STATISTICAL DEFORMABLE MODEL

Modal analysis methods have proven highly efficient for describing deformations using a reduced number of significant parameters [15, 40, 46, 50, 56]. In our hierarchical model, the description of global deformations relies on a *statistical* modal decomposition based on the KL transform [15, 48, 58]. The KL expansion allows us to approximate the global deformations observed on a training set of representative shapes on a low-dimensional eigenspace. Only the first most significant deformation modes are retained in this transform, yielding a shape-tailored representation of global deformations [15, 40]. A local deformation process, described by a Markov model, is introduced at a second level of the hierarchy to refine the global shape representation (Fig. 2). This two-level hierarchical description of deformations (Fig. 2) has shown itself to be very flexible for representing accurately a wide variety of deformable shapes (see Section 6 in which a selection of shape models is presented).

3.1. Description of Global Deformations

To obtain a compact application-tailored description of the object of interest, the shape template and its main deformation modes are characterized using an *off-line*[1] training procedure. This training procedure relies on the KL expansion of the deformations observed on a representative population. This procedure, first proposed by Cootes *et al.*, is described in detail in [14, 15]. It is briefly recalled here.

Following [14, 15], a particular shape $\mathbf{x}_k$ belonging to the training population is represented by a set of n labeled points which approximates its outline (Fig. 2):

$$\mathbf{x}_k = (x_{k1}, y_{k1}, x_{k2}, y_{k2}, \ldots, x_{kn}, y_{kn})^T.$$

The n labeled points correspond to the most salient points of the shape outline; these "landmarks" are extracted manually on the learning population. Following [15], the shapes belonging to the learning population are normalized in scale and aligned with respect to a common reference frame. The mean shape $\bar{\mathbf{x}}$ and the covariance matrix $\mathbf{C}$ of shapes $\{\mathbf{x}_k\}$ are computed from this set of normalized shapes. The main deformation modes of the template model $\mathbf{X}$ are then described by the eigenvectors of $\mathbf{\Phi}$ of $\mathbf{C}$, with the largest eigenvalues [14, 15]. The globally deformed template is defined by (Fig. 2)

$$\mathbf{X} = \mathbf{M}(k, \theta)[\bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}] + \mathbf{T}, \qquad (1)$$

where

- $\mathbf{T}$ and $\mathbf{M}(k, \theta)$ account for rigid transformations of the template in the image plane ($\mathbf{T}$ is a global translation vector, and $\mathbf{M}(k, \theta)$ performs a rotation by θ and a scaling by k),
- $\mathbf{\Phi} = (\phi_1, \phi_2, \ldots, \phi_m)$ is the matrix of the first $m(m < 2n)$ eigenvectors associated to the m largest eigenvalues and $\mathbf{b} = (b_1, \ldots, b_m)^T$ is a vector containing the weights for these m deformation modes.

A global configuration of the deformable template is thus described by $4 + m$ parameters corresponding to rigid transformation (four parameters) and m modal weights b_j, $j = 1, \ldots, m$. In practice, only five to seven modes are necessary to stand for more than 90% of the variability observed on the training population [15]. This first crude representation is refined using a local deformation process, as described in the next section.

3.2. Description of Local Deformations

A local deformation process $\boldsymbol{\delta}$, applied to the n labeled points, is introduced to refine the global description presented in the previous section. These local deformations are considered as *random perturbations* (represented by local random translations) that are superimposed on the globally deformed shape (Fig. 2). The local deformation vector $\boldsymbol{\delta}$ is described by a Gauss–Markov process defined on the graph corresponding to the outline of the deformable template. The Gauss–Markov distribution models the statistical interactions between the local random deformations applied to neighboring points of the template [29]. The complete model is expressed as (Fig. 2)

$$\mathbf{Y} = \mathbf{X} + \boldsymbol{\delta} = \mathbf{M}(k, \theta)[\bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}] + \mathbf{T} + \boldsymbol{\delta}, \qquad (2)$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)^T$ and $\delta_i = (\delta_{x_i}, \delta_{y_i})^T$. The probability distribution of $\boldsymbol{\delta}$ is defined by

$$P(\boldsymbol{\delta}) = \frac{1}{Z_p} \exp -\frac{1}{2}\boldsymbol{\delta}^T \mathbf{R}^{-1} \boldsymbol{\delta}, \qquad (3)$$

---

[1] A partial on-line training procedure enabling a simultaneous image segmentation and updating of deformation modes has been described by the authors in [31].

where $\mathbf{R}$ is the covariance matrix of $\boldsymbol{\delta}$ and Z_p designates the partition function. Assuming a first-order Gauss–Markov model (i.e., a first-order neighborhood structure on the graph), the joint distribution of $\boldsymbol{\delta}$ becomes

$$P(\boldsymbol{\delta}) = \frac{1}{Z_p} \exp -\frac{1}{2} \sum_{i=1}^{n} \left[\frac{1}{\varepsilon_i^2} \|\delta_i - \delta_{i-1}\|^2 + \frac{1}{\sigma_i^2} \|\delta_i\|^2 \right]. \quad (4)$$

Parameters σ_i^2 and ε_i^2 are interpreted as variance parameters. Parameters ε_i^2 weight the interactions between neighboring points and control the smoothness of local deformations. Parameters σ_i^2 control the amplitude of the local deformation vectors. Low values for σ_i^2 will draw the hierarchical model $\mathbf{Y}$ toward the globally deformed shape $\mathbf{X}$. In our experiments, these parameters were assumed to be constant: $\sigma_i^2 = \sigma^2$ and $\varepsilon_i^2 = \varepsilon^2$, $\forall i$. This is of course an approximation since σ_i^2 and ε_i^2 should depend on the distance between the points of index $i - 1$ and i, unless the feature points are equally spaced. Adopting constant values for these parameters in our implementation has, however, proved satisfactory in practice: the goal was to favor smooth shapes, in particular in the presence of missing data (occlusions, noise, etc.). To this end the values $\sigma = 2$ and $\varepsilon = 1$ have been adopted and kept constant in all experiments.

4. DEFORMABLE MODEL-BASED IMAGE SEGMENTATION

Significant improvements have been obtained in image segmentation problems by introducing global statistical models such as Markov random field models (MRF) [22, 25, 32, 51] or statistical deformable models [23, 24] that *constrain* the segmentation process. In the following we consider the problem of extracting and tracking moving deformable objects in an image sequence.

Our approach for the segmentation of deformable shapes relies on a bayesian formulation of the problem. The hierarchical model defined in the previous section (Eq. (2)) is considered as an *a priori* statistical model describing the configurations of the shape of interest. Besides, one or more specialized modules extract from the image sequence low-level features (spatial or spatio-temporal gradients) that will be used as observations in the bayesian estimation process.

4.1. Bayesian Estimation of Deformations

Let $\mathbf{d} = \{d_s, s \in S\}$ designate an observation field defined on a rectangular lattice S. The observation field $\mathbf{d}$, extracted from the image sequence, is related to the spatio-temporal variations of the intensity function. The segmentation problem is formulated as the MAP estimation of the (hidden) random process $\mathbf{Y}$ from the observation field $\mathbf{d}$:

$$\mathbf{Y}^* = \arg\max_{\mathbf{Y}} P(\mathbf{d} \mid \mathbf{Y}) P(\mathbf{Y}) = \arg\max_{\mathbf{Y}} P(\mathbf{Y}, \mathbf{d}). \quad (5)$$

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

$$P(\mathbf{Y}) = \frac{1}{Z_p} \exp -\frac{1}{2}(\mathbf{Y} - \mathbf{X}(\Theta))^T \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}(\Theta)), \quad (6)$$

where $\mathbf{R}$ is the covariance matrix of process $\boldsymbol{\delta}$ (see Eqs. (3) and (4)) and the mean $\mathbf{X}(\Theta)$ corresponds to (see Eq. (1))

$$\mathbf{X}(\Theta) = \mathbf{M}(k, \theta)[\bar{\mathbf{x}} + \Phi\mathbf{b}] + \mathbf{T}. \quad (7)$$

$\Theta = (\mathbf{M}(k, \theta), \mathbf{T}, \mathbf{b})$ denotes here the (deterministic) set of *hyperparameters* of this probabilistic model [17].

The distribution $P(\mathbf{d} \mid \mathbf{Y})$ is the likelihood of the observation field given the deformation process. This distribution depends on the image attributes used in the segmentation and on the observations at hand. In our case, this likelihood is specified as a Gibbs distribution [22] which incorporates specific knowledge on the application

$$P(\mathbf{d} \mid \mathbf{Y}) = \frac{1}{Z_d} \exp -E_d(\mathbf{Y}, \mathbf{d}), \quad (8)$$

where $E_d(\mathbf{Y}, \mathbf{d})$ is an energy function and Z_d is the partition function ($E_d(\mathbf{Y}, \mathbf{d})$ is specified in the next section, in the cases of motion-based and edge-based image segmentation).

The joint distribution appearing in Eq. (5) is thus also a Gibbs distribution:

$$P(\mathbf{Y}, \mathbf{d}) = \frac{1}{Z_p Z_d} \exp -E_\Theta(\mathbf{Y}, \mathbf{d}), \quad (9)$$

where

$$E_\Theta(\mathbf{Y}, \mathbf{d}) = E_d(\mathbf{Y}, \mathbf{d}) + \frac{1}{2}(\mathbf{Y} - \mathbf{X}(\Theta))^T \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}(\Theta)). \quad (10)$$

Let us notice that $Z = Z_p Z_d$ does not depend on Θ since Θ only appears in the mean of the Gaussian distribution (see Eq. (6)). On the other hand, the normalizing constant $Z_d = \int \exp -E_d(\mathbf{Y}, \mathbf{d}) \, d\mathbf{d}$ and hence Z generally depends on $\mathbf{Y}$. However, we show (see Appendix A or [34]) that for the segmentation models considered in Section 4.2, Z_d may approximately be considered as constant (the result is exact for the motion-based segmentation model and approximate in the case of the edge-based segmentation model). In the following we consider that $Z = Z_p Z_d$ does not depend on $\mathbf{Y}$. As a consequence, the MAP estimation of the deformable template comes to the minimization of global energy function $E_\Theta(\mathbf{Y}, \mathbf{d})$. The final configuration of the template is thus a compromise between prior information on the deformable structure and image-derived information (see Eq. (10)). Strong prior information about the template deformations is embedded in the model, thanks to the modal expansion described in Section 3.1. This helps obtaining robust segmentations in adverse situations such as noise or occlusions.

The modeling of $E_d(\mathbf{Y}, \mathbf{d})$ is considered in the next section in two cases: motion-based image segmentation and edge-based image segmentation.

FIG. 3. Observation maps for the segmentation. (a) Thresholded temporal gradients; (b) spatial gradients.

4.2. Segmentation Models

In the segmentation process, the global energy function $E_d(\mathbf{Y}, \mathbf{d})$ stands for the interactions between data $\mathbf{d}$ and the deformable template $\mathbf{Y}$. The modeling of $E_d(\mathbf{Y}, \mathbf{d})$ is clearly problem dependent. In this section we present two different models for $E_d(\mathbf{Y}, \mathbf{d})$. Both aim at extracting objects of interest from an image sequence. The first model yields a motion-based segmentation of the scene: objects are characterized by their motion with respect to the background. The background may be static or may itself be in motion. This first model relies on temporal gradient data. The second model is a more standard edge-based segmentation model, relying on spatial gradient data. We present in Section 6 several case studies corresponding to these two models, applied on real-world image sequences.

4.2.1. A Motion-Based Segmentation Model

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. We first assume that the camera is static. In order to extract moving objects from the image sequence, temporal variations (Fig. 3a) are estimated using two complementary methods. The first one measures variations $d_1(s)$ on three successive images; the second one estimates the changes $d_2(s)$ between the current image and a reference image which is created and updated on line,

$$d_1(s) = \min(|I_t(s) - I_{t-\Delta t}(s)|, |I_{t+\Delta t}(s) - I_t(s)|),$$
$$d_2(s) = |I_{\text{ref}}(s) - I_t(s)|, \tag{11}$$

where Δt designates the time step between two successive frames. As can be seen, three frames are necessary to obtain a segmentation: $I_{t-\Delta t}$, I_t, and $I_{t+\Delta t}$. The tracking procedure is applied once the first segmentation has been obtained, i.e., after the third frame. The reference image $I_{\text{ref}}(s)$ is constructed using a linear estimator of the background described in [20]. Observations $d_1(s)$ present high values for points belonging to a moving object and low values for background points. Temporal gradients such as $d_1(s)$ are known to yield poor observations in homogeneous (i.e., nontextured) regions or in the presence of self-overlapping of an object mask during the displacement. The second observation $d_2(s)$ based on a reference image of the background is

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

The observation field (data) is thus defined as

$$d(s) = \max(s_\alpha(d_1(s)), s_\beta(d_2(s))), \tag{12}$$

where

$$s_\eta(y) = 1 \quad \text{if } y > \eta$$
$$s_\eta(y) = 0 \quad \text{otherwise,} \tag{13}$$

where α and β are two thresholds for the detection of significant motion. Energy $E_d(\mathbf{Y}, \mathbf{d})$ then describes the statistical interaction between the thresholded temporal gradients $d(s)$ 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 $\Gamma_\mathbf{Y}^I$ corresponding to the object of interest and the outside of the template $\Gamma_\mathbf{Y}^O$ corresponding to the background. Energy $E_d(\mathbf{Y}, \mathbf{d})$ tends to enclose moving points inside the deformable model and to reject static points, belonging to the background, outside the outline of the model:

$$E_d(\mathbf{Y}, \mathbf{d}) = \sum_{s \in \Gamma_\mathbf{Y}^I} |d(s) - 1| + \sum_{s \in \Gamma_\mathbf{Y}^O} |d(s) - 0|. \tag{14}$$

This model can also be generalized to situations in which the camera is itself moving (inducing a global motion on the background) by using a preprocessing step to compensate for the apparent motion of the background [34, 47].

4.2.2. An Edge-Based Segmentation Model

In many applications, the only relevant clue available for performing the segmentation corresponds to spatial gradient information, related to photometric edges (Fig. 3b). This information allows an accurate segmentation of the deformable structure, provided that the template be initialized close enough to the desired solution. This initialization may be done manually or using preprocessing steps based for instance on mathematical morphology [54] or on the Hough transform [36].

In this case, we adopt the following standard form for the energy term related to observations [59]:

$$E_d(\mathbf{Y}, \mathbf{d}) \propto - \sum_{s \in \Gamma_\mathbf{Y}} \|\nabla I(s)\|. \tag{15}$$

$\nabla I(s)$ designates here the spatial gradient vector at site s and

Γ_Y is the boundary of the deformable structure. The energy function $E_d(\mathbf{Y}, \mathbf{d})$ is here simply defined as the integral of the spatial gradient along the boundary of the deformable model. This energy was first introduced in [28] for a snake model; it was afterward adapted to deformable templates in [59]. Variants of this energy function are now commonly used in many edge-based segmentation approaches.

4.3. Estimation of the Model Hyperparameters and Optimization

4.3.1. *Marginalized Maximum Likelihood Estimation*

Thanks to an approximation presented in the following, it is possible to design a simple, noniterative procedure for estimating the hyperparameters of the stochastic deformable model. For notation conveniences, the joint distribution of $\mathbf{Y}$ and $\mathbf{d}$ is redefined as

$$\mathsf{P}(\mathbf{Y}, \mathbf{d} \mid \Theta) = \frac{1}{Z} \exp -E_\Theta(\mathbf{Y}, \mathbf{d}), \qquad (16)$$

where, as already noticed, Z does not depend on Θ or on $\mathbf{Y}$. The segmentation problem is formulated as the joint estimation of $\mathbf{Y}$ and of the (unknown) set of hyperparameters Θ [17, 43]. Since Θ is unknown, a standard criterion for estimating Θ is the MML criterion [43]:

$$\Theta^* = \arg \max_\Theta \int_{\mathbf{Y}} \mathsf{P}(\mathbf{Y}, \mathbf{d} \mid \Theta) \, d\mathbf{Y}$$

$$= \arg \max_\Theta \int_{\mathbf{Y}} \mathsf{P}(\mathbf{d} \mid \mathbf{Y}, \Theta) \, \mathsf{P}(\mathbf{Y} \mid \Theta) \, d\mathbf{Y}. \qquad (17)$$

The estimate of $\mathbf{Y}$ is computed in turn using the MML estimate Θ^* and the already considered MAP criterion,

$$\mathbf{Y}^* = \arg \max_\mathbf{Y} \mathsf{P}(\mathbf{Y}, \mathbf{d} \mid \Theta^*). \qquad (18)$$

Estimation of Θ. Θ is estimated according to Eq. (17). Note that the local deformation process δ can usually be considered as a *local* random refinement of the globally deformed shape. This simplifying assumption expresses the fact that $\mathbf{Y}$ remains concentrated around the globally deformed shape $\mathbf{X}(\Theta)$. Under this assumption, the variance of stochastic process δ is small compared to the size of the deformable shape and the gaussian distribution of $\mathbf{Y}$ may be approximated by a Dirac distribution:

$$\mathsf{P}(\mathbf{Y} \mid \Theta) = \frac{1}{Z_p} \exp -\frac{1}{2}(\mathbf{Y} - \mathbf{X}(\Theta))^T \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}(\Theta))$$

$$\approx \delta(\mathbf{Y} - \mathbf{X}(\Theta)). \qquad (19)$$

It follows

$$\Theta^* \approx \arg \max_\Theta \int_{\mathbf{Y}} \mathsf{P}(\mathbf{d} \mid \mathbf{Y}, \Theta) \, \delta(\mathbf{Y} - \mathbf{X}(\Theta)) \, d\mathbf{Y}$$

$$= \arg \max_\Theta \mathsf{P}(\mathbf{d} \mid \mathbf{Y} = \mathbf{X}(\Theta))$$

$$= \arg \min_\Theta E_d(\mathbf{X}(\Theta), \mathbf{d}). \qquad (20)$$

Thus the MML criterion reduces to the minimization of the data-related energy term $E_d(\mathbf{Y}, \mathbf{d})$, for $\mathbf{Y} = \mathbf{X}(\Theta)$.

The simplifying assumption (19) is, in general, approximately true if the shape to be analyzed is close to the shapes that have been used to train the KL-based modal decomposition. This assumption may however not hold when the observed shape differs significantly from the shapes belonging to the training population or when the observed shape cannot be obtained linearly from these training shapes. In such a case the local deformations may contribute to the configuration of the template in a *nonlocal manner* (see for instance Figs. 11 and 12) and the estimate of δ should be used to iteratively refine the estimation of Θ. This feedback has been tested in a first implementation of the segmentation procedure, but has not been retained since it produces no noticeable enhancement of the visual quality of the segmentation (but yields a significant increase of the computational load).

Estimation of $\mathbf{Y}$. The MAP estimate of $\mathbf{Y}$ is easily derived given the estimate Θ^*. The MAP criterion (Eq. (18)) comes to the minimization of the global energy function

$$\mathbf{Y}^* = \arg \min_\mathbf{Y} E_{\Theta^*}(\mathbf{Y}, \mathbf{d}), \qquad (21)$$

where $E_{\Theta^*}(\mathbf{Y}, \mathbf{d})$ is defined by Eq. (10).

To summarize, the segmentation of the structure of interest requires the estimation of the hyperparameters according to Eq. (20) and the minimization of the global energy function according to Eq. (21). These (nonlinear) optimization steps are performed in different ways, depending on the clues used in the segmentation process. This is explained and illustrated on synthetic examples in the following sections.

4.3.2. *Stochastic and Deterministic Optimization*

Motion-Based Segmentation. In order to be insensitive to the initial configuration of the deformable model, global optimization techniques are performed on the *first* frame of the image, to determine Θ^*, according to Eq. (20). This global optimization step relies on a *simulated annealing* algorithm based on the Gibbs sampler [22]. Although stochastic, this algorithm leads to a rather fast adjustment of the deformable shape, thanks to the reduced number of parameters to estimate. This procedure yields robust segmentations and avoids a manual initialization of the model. Local deformations (corresponding to the estimation of $\mathbf{Y}^*$) are obtained in a second step, according to Eq. (21), using a fast deterministic relaxation algorithm known as ICM [5]. This relaxation algorithm is initialized by $\mathbf{Y} = \mathbf{X}(\Theta^*)$. The total cpu time for the segmentation of the first frame is about 3 or 4 min for a 256×256 image on a standard SUN/SPARC 10 workstation.

The subsequent frames are then analyzed using only the deterministic ICM algorithm coupled with a tracking procedure of the deformable model, to be described in Section 5. The tracking procedure provides good initializations from one frame to the next, avoiding resorting to the relatively time consuming simulated annealing algorithm. As a consequence, the cpu time
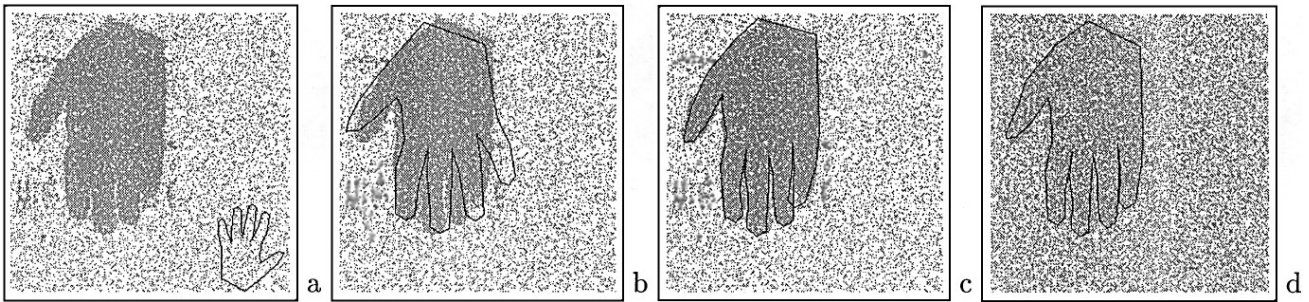
FIG. 4. (a, b, c,) Initialization and convergence of the deformable model. (d) Segmentation with low signal-to-noise ratio.

falls down to 45 s for one image. In [23], Grenander *et al.* reported several hours cpu time to perform the segmentation of a hand, with a partial manual initialization[2] on a 128×70 image. Grenander's experiments were performed 10 years ago and thus it is difficult to compare cpu times. However, our approach is intrinsically faster, thanks to the modal decomposition of shape deformations and shape tracking. The local deterministic approach proposed by Cootes *et al.* in [15] leads to cpu times comparable to ours (about 50 s are announced in [15] on a standard workstation) but requires a manual initialization of the model [15]. A *genetic algorithm* was proposed in [26], along with the model of Cootes *et al.*, to cope with the model initialization problem, but this algorithm is known to have a complexity at least equivalent to simulated annealing (cpu times are not given in [26]).

The robustness of our approach to initial conditions, noise, and missing data (partial occlusions) is illustrated Figs. 4 and 5 on a hand template. In these experiments, observation maps $d(s)$, obtained from real hand outlines, have been corrupted by nongaussian noise. Observation maps with different signal-to-noise ratios and partial occlusions have been simulated. In any case, the prototype shape is initialized at random in the image (here in the lower right part of the image—see Fig. 4a).

Figure 4b shows the result of a global optimization with respect to Θ, considering only the similarity transform parameters (i.e., global rotations, translations, and scale changes). The complete estimation of Θ (including the deformation hyperparameters from the KL expansion) yields to the segmentation depicted in Fig. 4c which is considered satisfactory. A similar result is obtained, in Fig. 4d, with a very low signal-to-noise ratio and the same random initialization. Figure 5 shows similar segmentation results, in the presence of partial occlusions (missing fingers— Figs. 5a, 5b) or when two structures are superimposed (Fig. 5c). Figure 5d presents a case in which the algorithm gets confused by two superimposed hands. This illustrates the limitation of the approach, which does not specifically handle large occlusions or widespread cluttered areas (robust estimation methods might be devised to address this issue [6]). These results however demon-

strate the ability of the approach to provide robust segmentations in adverse situations such as small occlusions or missing data (an example of the segmentation of a real-world sequence of a moving hand with a partial occlusion is presented in Section 6).

Edge-Based Segmentation. The optimization scheme is similar to that for the second segmentation model presented in Section 4.2.2. However, a global minimization by a stochastic algorithm is not appropriate here, since the segmentation criterion is generally valid only within a *region of interest*, close to the expected solution [15, 28, 46, 55]. A standard deterministic optimization procedure (steepest gradient descent) is therefore used in this context [55] to determine Θ^*. Local deformations (corresponding to the estimation of $\mathbf{Y}^*$) are obtained as before, using the fast deterministic ICM relaxation algorithm [5]. The deformable template is here initialized manually on the first frame of the sequence. Subsequent frames are initialized by temporal predictions of the template obtained by the tracking procedure described in Section 5. This method was able to extract and track reliably complex deformable structure over long image sequences, with only one manual initialization (on the first image), as is demonstrated in Section 6.

The tracking procedure, based on a Kalman filtering of the model parameters, is described in the following section.

5. TRACKING THE DEFORMABLE MODEL OVER A LONG IMAGE SEQUENCE

The segmentation procedure described in the previous section is combined with a temporal tracking of the hyperparameters of the model, as explained in the Introduction of this paper. Tracking the deformable structure over a long image sequence [8, 16, 30, 46] reduces significantly the computational cost of the segmentation method (by propagating good initializations from one frame to the next) and enables us to process large movements more reliably. The temporal prediction and filtering of the model parameters also provide valuable information about the global dynamic behavior of the deformable structure over time, which might for instance be used for interpretation purposes. A review of existing methods in curve tracking may be found in [7].

Contrary to standard approaches, based on the tracking of discrete low-level or intermediate-level image primitives such

---

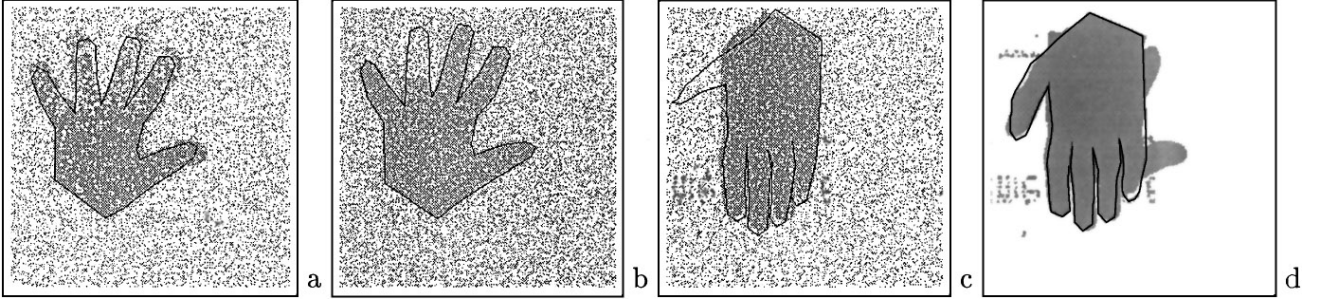[2] A point of the template was constrained to take up a specified location in the image.

FIG. 5. Segmentations with missing data and partial occlusions.

as characteristic points [2], segments [19], or regions [42], the tracking procedure proposed here considers the structure of interest and its possible deformations as a whole by tracking the hyperparameters $\boldsymbol{\Theta} = (\mathbf{M}(k, \theta), \mathbf{T}, \mathbf{b})$ describing the configuration of the deformable template. The tracking of *rigid* motion parameters, describing the movement of homogeneous regions, was proposed in [42]. This approach has recently been extended to *rigid* motions associated to active contour models in [3]. In [57], Terzopolous and Szelisky use a Kalman filter to guide the dynamic evolution of an active contour model. In their approach (called "Kalman snakes") the deformable model is governed by the laws of nonrigid dynamics. An approach akin to ours [30] was proposed independently by Blake *et al.* in [9] for tracking piece-wise smooth curves in the image plane. The tracker consists in an estimator of the parametrization of the motion of a template defined by B-splines. The method includes a learning process in order to tune the tracker to movements observed on a training set. This method enables us to reach real time tracking on simple templates and movements. In [14], Cootes *et al.* outlined a temporal prediction method for tracking the parameters of their deformable model using a heuristic first order linear prediction equation. Such a model does, however, only stand for very simple movements. The tracking scheme presented here, contrary to the one proposed in [3, 42], takes into account the deformations of the moving structure. Contrary to [14], it incorporates different kinematical and noise models, and an *a posteriori* filtering of the model parameters, based on the Kalman filter recursive estimation. Moreover we have introduced a procedure for the statistical detection of abrupt changes in the tracking model [27], which enables us to reinitialize the Kalman filter when a significant change in the kinematical behavior of the deformable structure is detected. In the following, we mainly concentrate on the dynamical evolution model and on the observation model, both used in the formulation of the problem. The Kalman filter equations are recalled in Appendix B.

5.1. Temporal Evolution of the Deformable Template

Considering the tracking of the model over a long sequence, the model is redefined at time t as

$$\mathbf{X}_t = \mathbf{M}_t[\bar{\mathbf{x}} + \boldsymbol{\Phi}\mathbf{b}_t] + \mathbf{T}_t. \tag{22}$$

The tracking procedure is based on a linear recursive estimation of the hyperparameters $\boldsymbol{\Theta}_t$ of the model. Let us recall that $\boldsymbol{\Theta}_t = \{\mathbf{M}_t, \mathbf{T}_t, \mathbf{b}_t\}$ contains the rigid transformation parameters and modal deformation parameters for the template. The local deformations $\boldsymbol{\delta}$ are not considered in the temporal filtering equations because it is difficult to predict the statistical behavior of the local deformation process over time. Tracking $\boldsymbol{\delta}$ is in fact not necessary, since adopting $\boldsymbol{\delta} = 0$ as an initialization in our segmentation scheme yields fast convergence to the desired solution, as explained in Section 4.3.2. On the other hand, the temporal tracking of the global deformation parameters $\boldsymbol{\Theta}_t$ is highly advisable since a good prediction (from one frame to the next) avoids the need for expensive global optimization procedures or manual initialization of the template.

In our implementation, the state vector $\mathbf{s}_t$ of the Kalman filter contains parameters $\mathbf{T}_t$, $\mathbf{M}_t$, and $\mathbf{b}_t$ and their derivatives. Measurements $\mathbf{w}_t$ used in the recursive filter are provided by the segmentation procedure and correspond to the MML estimation of the global parameters $\boldsymbol{\Theta}_t^* = \{\mathbf{M}_t^*, \mathbf{T}_t^*, \mathbf{b}_t^*\}$, as described in Section 4.3.1.

The dynamic evolution of the hyperparameters is obtained from a second order Taylor expansion of template $\mathbf{X}_t$:

$$\begin{cases} \mathbf{X}_{t+\Delta t} = \mathbf{X}_t + \Delta t\dot{\mathbf{X}}_t + \dfrac{\Delta t^2}{2}\ddot{\mathbf{X}}_t, \\ \dot{\mathbf{X}}_{t+\Delta t} = \dot{\mathbf{X}}_t + \Delta t\ddot{\mathbf{X}}_t. \end{cases} \tag{23}$$

By substituting Eq. (22) into Eq. (23), we obtain the dynamic equations for the global deformation parameters (see Appendix C):

$$\begin{cases} \mathbf{M}_{t+\Delta t} = \mathbf{M}_t + \Delta t\dot{\mathbf{M}}_t + \dfrac{\Delta t^2}{2}\ddot{\mathbf{M}}_t, \\ \dot{\mathbf{M}}_{t+\Delta t} = \dot{\mathbf{M}}_t + \Delta t\ddot{\mathbf{M}}_t, \\ \mathbf{T}_{t+\Delta t} = \mathbf{T}_t + \Delta t\dot{\mathbf{T}}_t + \dfrac{\Delta t^2}{2}\ddot{\mathbf{T}}_t, \\ \dot{\mathbf{T}}_{t+\Delta t} = \dot{\mathbf{T}}_t + \Delta t\ddot{\mathbf{T}}_t, \\ \mathbf{b}_{t+\Delta t} = \mathbf{b}_t + \Delta t\dot{\mathbf{b}}_t + \dfrac{\Delta t^2}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t, \\ \dot{\mathbf{b}}_{t+\Delta t} = \left[\mathbf{I} + \dfrac{3\Delta t}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}(\dot{\mathbf{M}}_t\boldsymbol{\Phi} - \dot{\mathbf{M}}_{t+\Delta t}\boldsymbol{\Phi})\right]\dot{\mathbf{b}}_t \\ \qquad + \left[\Delta t\mathbf{I} - \dfrac{\Delta t^2}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\dot{\mathbf{M}}_{t+\Delta t}\boldsymbol{\Phi}\right]\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t \end{cases} \tag{24}$$

with

$$\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1} = [(\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi})^T(\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi})]^{-1}(\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi})^T. \qquad (25)$$

Based on these dynamics and on a kinematical model described in the next section, we derive the standard Kalman filter equations associated to the global parameters $\mathbf{M}_t$, $\mathbf{T}_t$, and $\mathbf{b}_t$ (see Appendix B).

5.2. The Constant Velocity Kalman Filter

Several kinematic models were considered by the authors in [30, 34] in order to track deformable shape over long sequences. A "constant velocity" filter with white noise, a "constant velocity" filter with correlated noise, and an "instantaneous velocity" filter have been implemented. All filters give qualitatively similar final results, even though some intermediate state estimates may be quite different. We limit here our presentation to the constant velocity filter which has given satisfactory results in our experiments. The other filters are detailed in [34].

The constant velocity filter assumes approximate constant velocities for the model hyperparameters $\mathbf{M}_t$ and $\mathbf{T}_t$. Neglecting higher order derivatives we get

$$\begin{cases} \dot{\mathbf{M}}_{t+\Delta t} \simeq \dot{\mathbf{M}}_t \\ \dot{\mathbf{T}}_{t+\Delta t} \simeq \dot{\mathbf{T}}_t. \end{cases} \qquad (26)$$

Let us notice that by substituting Eq. (26) into Eq. (24) one also gets an approximate constant velocity for $\mathbf{b}_t$:

$$\dot{\mathbf{b}}_{t+\Delta t} \simeq \left[\mathbf{I} + \frac{3\Delta t}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}(\dot{\mathbf{M}}_t - \dot{\mathbf{M}}_{t+\Delta t})\boldsymbol{\Phi}\right]\dot{\mathbf{b}}_t \simeq \dot{\mathbf{b}}_t. \qquad (27)$$

In the constant velocity filter, the second order derivatives of the hyperparameters (i.e., their acceleration) are considered as (small) random accelerations, modeled as white noise. These assumptions correspond to a standard kinematic model that has, for instance, been used to track isolated points or line segments [19].

State variable model. If $\mathbf{z}_t$ stands for $\mathbf{T}_t$, $\mathbf{b}_t$, or $\mathbf{M}_t$, the state vector is defined as $\mathbf{s}_t = (\mathbf{z}_t \dot{\mathbf{z}}_t)^T$. The dynamic evolution of the system is described by

$$\mathbf{s}_{t+\Delta t} = \mathbf{A}_t\mathbf{s}_t + \boldsymbol{\xi}_t, \qquad (28)$$

where $\mathbf{A}_t$ is the state transition matrix and $\boldsymbol{\xi}_t$ is a zero-mean white gaussian noise with covariance matrix $\mathbf{Q}_t = \mathbb{E}[\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T]$. For our kinematic model, the transition matrices $\mathbf{A}_t$ for parameters $\mathbf{M}_t$, $\mathbf{T}_t$, and $\mathbf{b}_t$ are easily derived from Eqs. (24) and (26). The transition matrices for $\mathbf{M}_t$ and $\mathbf{T}_t$ have the same expression:

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{I} & \Delta t\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}. \qquad (29)$$

Matrix $\mathbf{A}_t$ for the deformation parameter $\mathbf{b}_t$ is expressed as

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{I} & \Delta t\mathbf{I} \\ \mathbf{0} & \mathbf{I} + \frac{3\Delta t}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}[\dot{\mathbf{M}}_t - \dot{\mathbf{M}}_{t+\Delta t}]\boldsymbol{\Phi} \end{pmatrix}. \qquad (30)$$

The expression of the noise term $\boldsymbol{\xi}_t$ on the state vector is given, on the one hand, for $\mathbf{M}_t$ and $\mathbf{T}_t$ and on the other hand for $\mathbf{b}_t$ by

$$\boldsymbol{\xi}_t = \begin{pmatrix} \frac{\Delta t^2}{2}\ddot{\mathbf{z}}_t \\ \Delta t\ddot{\mathbf{z}}_t \end{pmatrix},$$

$$\boldsymbol{\xi}_t = \begin{pmatrix} \frac{\Delta t^2}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{z}}_t \\ \left[\Delta t\mathbf{I} - \frac{\Delta t^2}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\dot{\mathbf{M}}_{t+\Delta t}\boldsymbol{\Phi}\right]\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{z}}_t \end{pmatrix}, \qquad (31)$$

where the acceleration term $\ddot{\mathbf{z}}_t$ is, as already explained, modeled as a zero mean white noise with variance σ_{acc}^2.

As can be noted, there is a nonlinear coupling between parameters $\dot{\mathbf{b}}$ and $\dot{\mathbf{M}}$ since the prediction and updating of $\dot{\mathbf{b}}$ depends on $\dot{\mathbf{M}}$. In order to obtain an approximate linear formulation of the recursive estimation, we have uncoupled the filters on parameters $\mathbf{M}$, $\mathbf{T}$, and $\mathbf{b}$ by approximating $\dot{\mathbf{M}}_t$ by $\dot{\mathbf{M}}_{t|t}$ and $\dot{\mathbf{M}}_{t+\Delta t}$ by $\dot{\mathbf{M}}_{t+\Delta t|t}$, where $\dot{\mathbf{M}}_{t|t}$ is the state estimate at time t and $\dot{\mathbf{M}}_{t+\Delta t|t}$ is the predicted state at time $t + \Delta t$ (see Appendix B). This approximation yields good experimental results in practice.

Observation model. The observations $\mathbf{w}_t$ are the result of the noisy measurement of the hyperparameters $\boldsymbol{\Theta}_t = \{\mathbf{M}_t, \mathbf{T}_t, \mathbf{b}_t\}$. These measurements are obtained using the MML estimation procedure described in Section 4.3.1 (see Fig. 1). This yields the standard observation model [19]

$$\mathbf{w}_t = \mathbf{H}_t\mathbf{s}_t + \boldsymbol{\eta}_t, \qquad (32)$$

where

$$\mathbf{H}_t = (\mathbf{I}, \mathbf{0}).$$

$\mathbf{I}$ is the identity matrix. The statistical properties of the noise term $\boldsymbol{\eta}_t$ could theoretically be derived from the properties of the MML estimator that provides the measurements. It is approximated here by a zero-mean white gaussian noise with covariance matrix $\mathbf{V}_t = \mathbb{E}[\boldsymbol{\eta}_t\boldsymbol{\eta}_t^T] = \sigma_\eta^2\mathbf{I}$. This approximation has provided satisfactory results.

5.3. Tracking Procedure and Detection of Abrupt Changes

The tracking procedure may be described as the following: at time t the current state estimate (corresponding to the hyperparameters and their derivatives) is $\hat{\mathbf{s}}_{t|t}$. The prediction step of the Kalman filter (Eqs. (45) and (46) in Appendix B) defines the predicted location $\hat{\mathbf{s}}_{t+\Delta t|t}$ of the deformable model in the next frame (at time $t + \Delta t$). The predicted state $\hat{\mathbf{s}}_{t+\Delta t|t}$ is generally close to
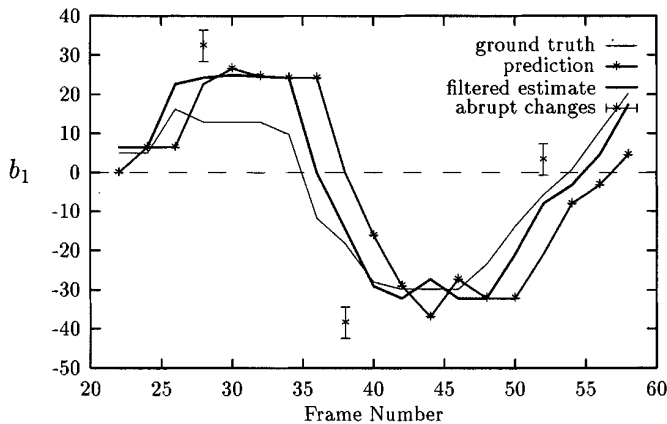
FIG. 6. Tracking of the 1st modal amplitude b_1 (constant velocity filter).

the optimal configuration corresponding to a relevant segmentation (unless an abrupt change occurs). Measurements $\mathbf{w}_{t+\Delta t}$ at time $t + \Delta t$ are defined as the MML estimates of the model hyperparameters, as described in Section 4.3. These estimates are computed by a fast deterministic optimization procedure (the ICM algorithm [5]) that tunes the hyperparameters toward the spatial or temporal gradients extracted at time $t + \Delta t$, according to Eq. (20). The updated state $\hat{\mathbf{s}}_{t+\Delta t|t+\Delta t}$ is finally derived from the classical Kalman filter equations (Eqs. (42) and (43) in Appendix B).

The prediction and filtering of the hyperparameters of the deformable template is completed by a statistical detection of abrupt changes which allows us to reinitialize the Kalman filter when the observed movement does not correspond to the underlying kinematic model. Hinkley's cumulative sum test [27] that we have used for this purpose is detailed in [30, 34]. The Kalman filter is reinitialized when an abrupt change is detected. In the case of motion-based segmentation, a global optimization step is performed in order to obtain reliable estimates for the model hyperparameters. In the case of edge-based segmentation, global optimization is not desirable (see Section 4.3.2.), and the tracker is reinitialized with the filtered estimate obtained on the previous frame.

The number of abrupt changes detected by the cumulative sum test essentially depends on the complexity of the deformable

motion. Although only a few abrupt changes were observed in practice on most image sequences presented in Section 6, the number of significant changes may noticeably increase when the underlying kinematic behavior becomes highly nonstationary. A second filter that is based on an on-line computation of the instantaneous velocities of the parameters is described in [30, 34]. Although these instantaneous velocities are quite noisy, this filter leads to less abrupt change detections in the case of complex kinematic behaviors. The final estimates and computational loads are, however, similar for these two kinematic models. The instantaneous velocity filter is not presented here due to space limitation, and we refer to [30, 34] for additional details.

The performances of the constant velocity Kalman filters are illustrated in Fig. 6 which shows plots of the predicted and filtered estimates obtained, for the first modal amplitude b_1, on a synthetic hand image sequence. Since the ground truth is known in this case, one can evaluate the performance of the tracking. The simulated deformable movement has been generated from Eq. (22), using eigenvectors obtained from a training sequence and specifying adequately the other parameters. Detections of abrupt changes are represented Fig. 6 by vertical bars. As already mentioned, in this case, when an abrupt change is detected, the Kalman filtering is reinitialized and a global optimization step is performed in order to estimate reliable model hyperparameters. As can be seen in Fig. 6, the Kalman filter enables a reliable tracking of the deformation parameter b_1 (similar qualitative results are observed for the other parameters). Abrupt changes are mainly detected in regions of fast evolution of the considered parameter.

The contribution of the Kalman filter-based tracking procedure is also demonstrated, with the same hand template, on a real world sequence in Figs. 7 and 8. Figure 7 shows a segmentation result obtained without the tracking procedure described in this section. The optimal MML estimate, obtained at time $t - \Delta t$ by a stochastic relaxation algorithm, is presented Fig. 7a. Figure 7b depicts the initialization considered in the next image, at time t, when no tracking procedure is used. This initialization corresponds to a simple projection (without any temporal prediction) of the MML estimate obtained at time $t - \Delta t$. Figure 7c shows what happens in this case when a fast deterministic optimization algorithm is used to determine the model parameters at time t.
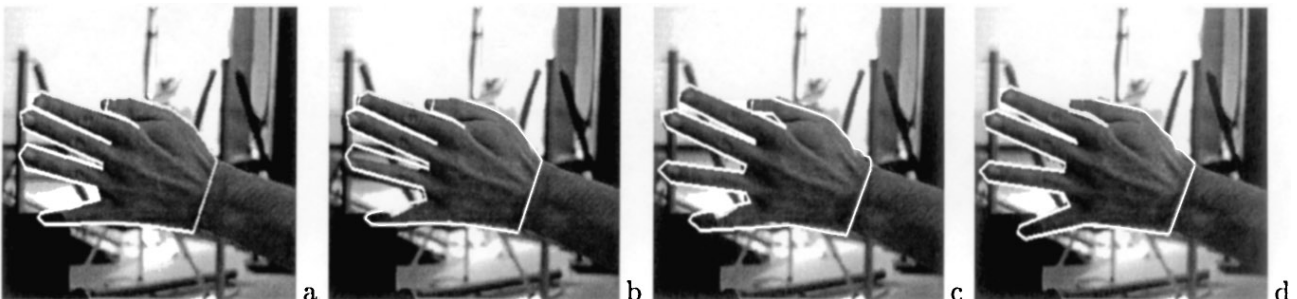

FIG. 7. Motion-based segmentation of a hand without the tracking procedure (see text).
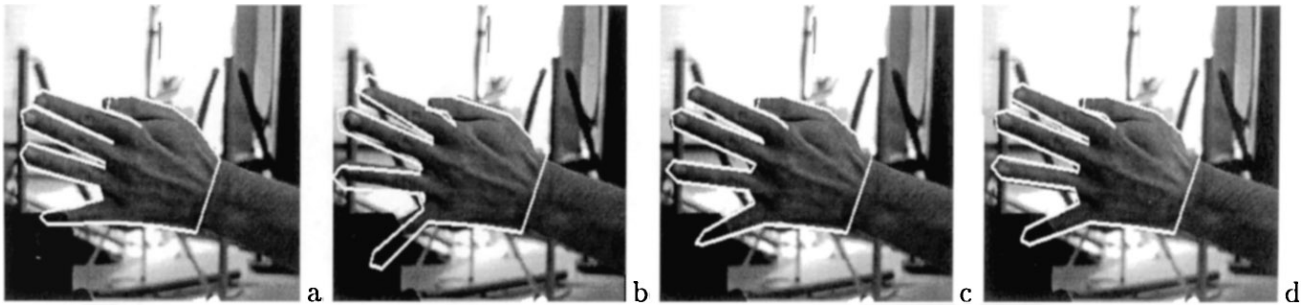
FIG. 8. Motion-based segmentation of a hand with the tracking procedure (see text).

As can be seen, the segmentation is not satisfactory, especially on the little finger (Fig. 7c) because the deterministic optimization algorithm remains trapped in a local minimum of the energy function, corresponding here to a wrong segmentation, because the initialization is not close enough to the desired solution. One has thus to resort to time-consuming stochastic optimization algorithms to obtain a relevant segmentation (see the result in Fig. 7d—4 min cpu time on a SUN/SPARC10 workstation). A better solution consists in introducing the tracking procedure and coupling it with a fast deterministic optimization procedure as explained previously. The computational load of the tracking procedure is negligible (about 1 s cpu time per frame), and tracking provides better initializations from one frame to the next. This enables us to reduce the cpu time to about 45 s, with the same qualitative results as for global optimization. Figure 8 shows an example of a tracking with the constant velocity Kalman filter (similar qualitative results are obtained with the instantaneous velocity filter). Figure 8a presents the result of the estimate at time $t - \Delta t$. The predicted configuration at time t is presented in Fig. 8b. As can be seen in Fig. 8c, a deterministic optimization of the model parameters yields a satisfactory segmentation in this case (Fig. 8c also corresponds to the measurements used in the Kalman filter at time t). Finally, Fig. 8d depicts the final configuration of the deformable model corresponding to the Kalman filtered estimate along with the estimation of the local Markovian

deformation process δ. Although this final estimate (Fig. 8d) is close to the measurement in Fig. 8c, a local adjustment of the model is noticeable on the little finger.

Experimental results on real-world sequences processed by the complete segmentation-tracking scheme are presented in Section 6.

6. EXPERIMENTAL RESULTS

The Markov modeling approach for the segmentation and tracking of deformable shapes is demonstrated here on a variety of applications corresponding to different image classes and deformable models.

A deformable hand model. In our first experiments, we have considered the segmentation and tracking of hands moving against a textured background, with partial occlusions. The hand was considered here as a 2D deformable structure. The image sequence presented in Fig. 9 is composed of more than 100 256×256 frames. As can be observed in Figs. 9c and 10c, the little finger is partially occluded by a box in the foreground. This is conspicuous in the observation field $d(s)$ (temporal gradients) represented in Fig. 10c, on which large regions of missing or noisy data can be seen.

The model of the deformable hand structure is a 30-point model computed from a training set of 22 hands which belonged

FIG. 9. Segmentation and tracking of a moving hand on a cluttered background (image size: 256×256).
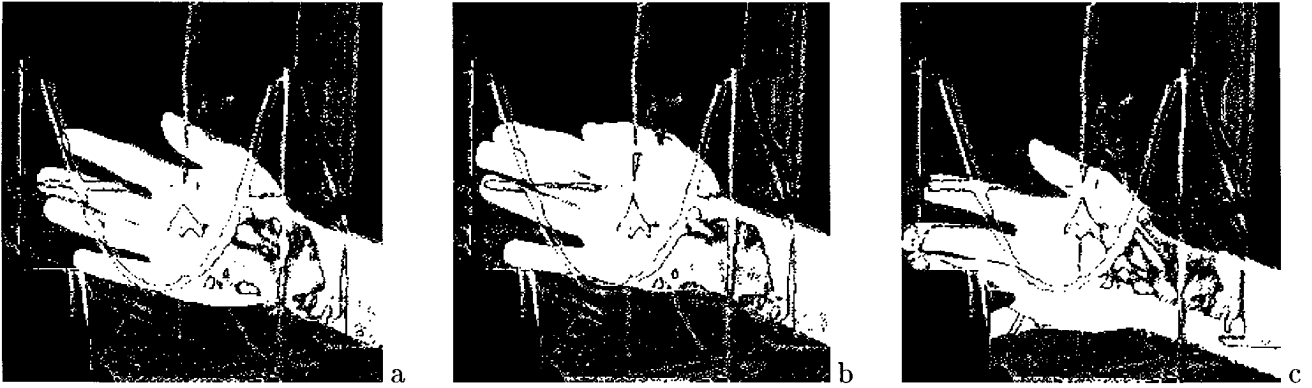
FIG. 10. Observation fields (temporal gradients) $d(s)$ used for the segmentation of the hand.

to different persons (the processed images did not belong to the training set). The initial configuration of the model is defined at random on the first frame of the sequence. A global optimization algorithm is used on this first image. The next images are processed using fast deterministic optimization techniques coupled with a Kalman filtering of the hyperparameters of the model, as explained in Section 5. Using this procedure, a reliable tracking and segmentation of the deformable structure were obtained in all cases, over the whole image sequence. Figure 9 shows three (nonconsecutive) images extracted from the sequence. Let us notice that a relevant segmentation has been obtained on the occluded finger, although the corresponding observation field is missing (Figs. 9c and 10c). The average cpu time on this sequence is about 45 s per frame on a standard SUN/SPARC10 workstation.

The contribution of the local deformation process $\boldsymbol{\delta}$ is demonstrated by comparing the segmentations obtained in Figs. 11 and 12. Figure 11 presents the optimal segmentation obtained

on two successive frames, considering the global deformation parameters only (i.e., $\mathbf{M}(k, \theta)$, $\mathbf{T}$, and $\mathbf{b}$) in the estimation procedure. The same frames have been processed (Fig. 12) considering both global deformation parameters and the local deformation process $\boldsymbol{\delta}$. As expected, the local deformation process captures important details that have been overlooked by the global deformation parameters (which are constrained by the learning sequence).

A deformable mouth model. In a second experiment, we have considered the standard problem [59] of tracking a mouth in face image sequences (Figs. 13 and 14). Spatial gradients are the only clues used here to perform the segmentation. The mouth template is defined as a 29 point model interpolated by a cubic B-spline. Its structure and deformation modes have been learned from a sequence of 10 relevant images. In Fig. 13, the deformable model is initialized manually on the first frame. In this case, the "instantaneous velocity" Kalman filter proved more
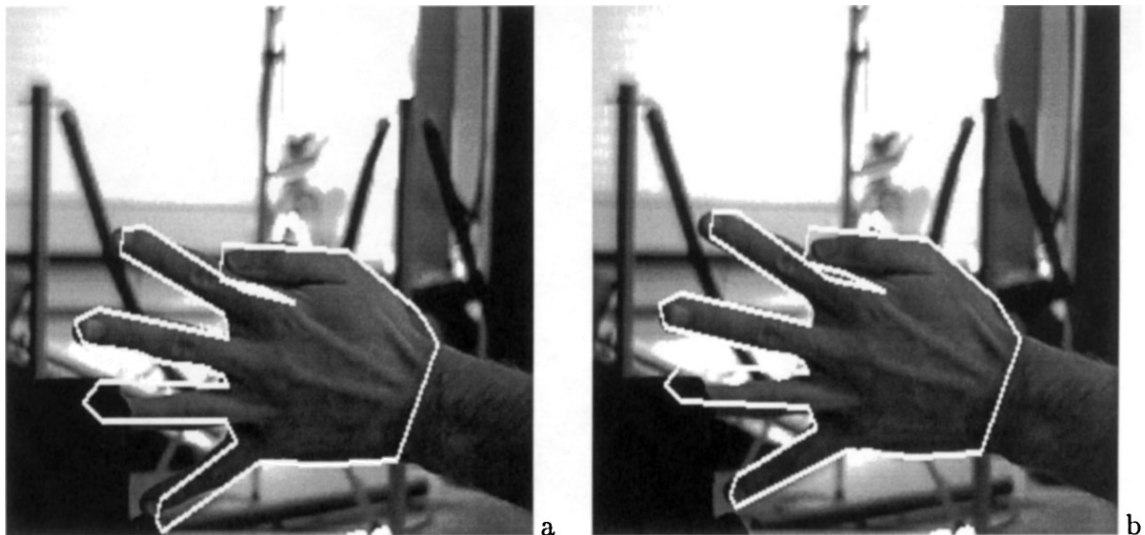

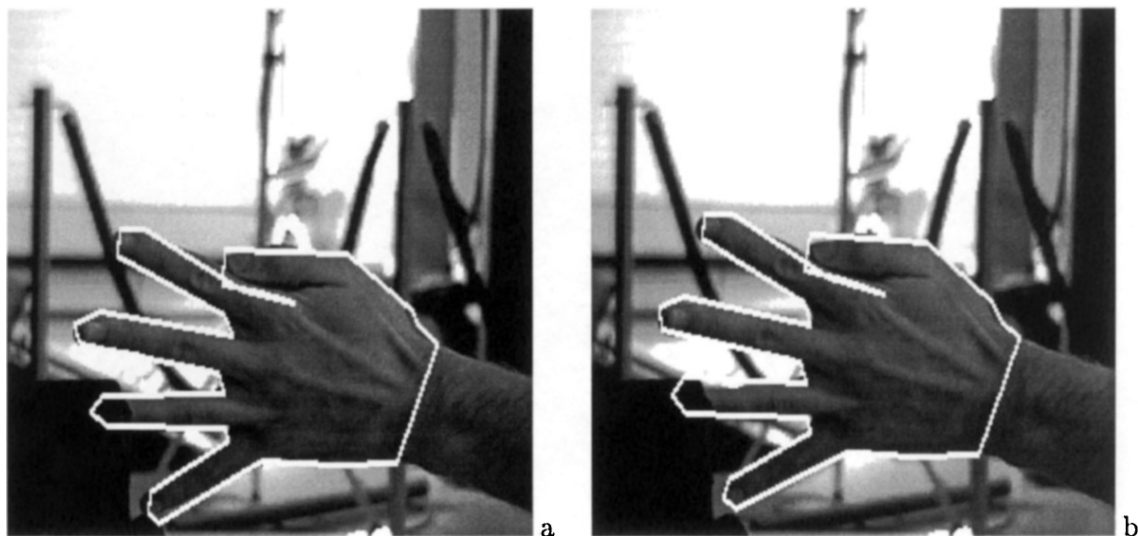**FIG. 11.** Motion-based segmentation of a hand—estimation of global deformations only.

FIG. 12. Motion-based segmentation of a hand—estimation of global and local deformations.

robust than the constant velocity filter due to abrupt changes in the kinematic behavior of the structure. The segmentation and tracking results are very satisfactory, especially if one takes into account the low quality of the image data available in this case (see Fig. 3b). These results may be compared with the results, presented in a similar case study, in [59]. The cpu time is about 15 s for an (256×256) image in this case. The segmentation and tracking of the mouth of a different person with the *same* model is presented in Fig. 14 in the standard "Claire" sequence. Similar results are obtained in this case, showing that the model is able to represent a variety of realizations of a mouth structure.
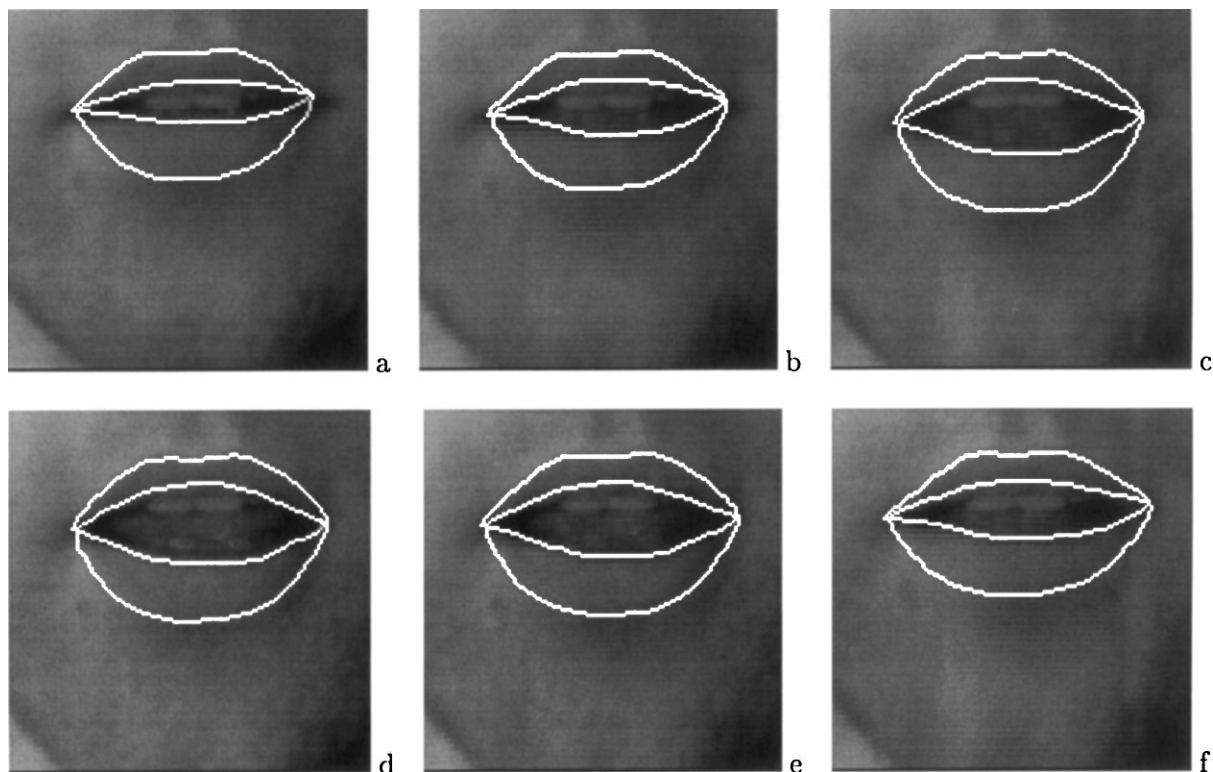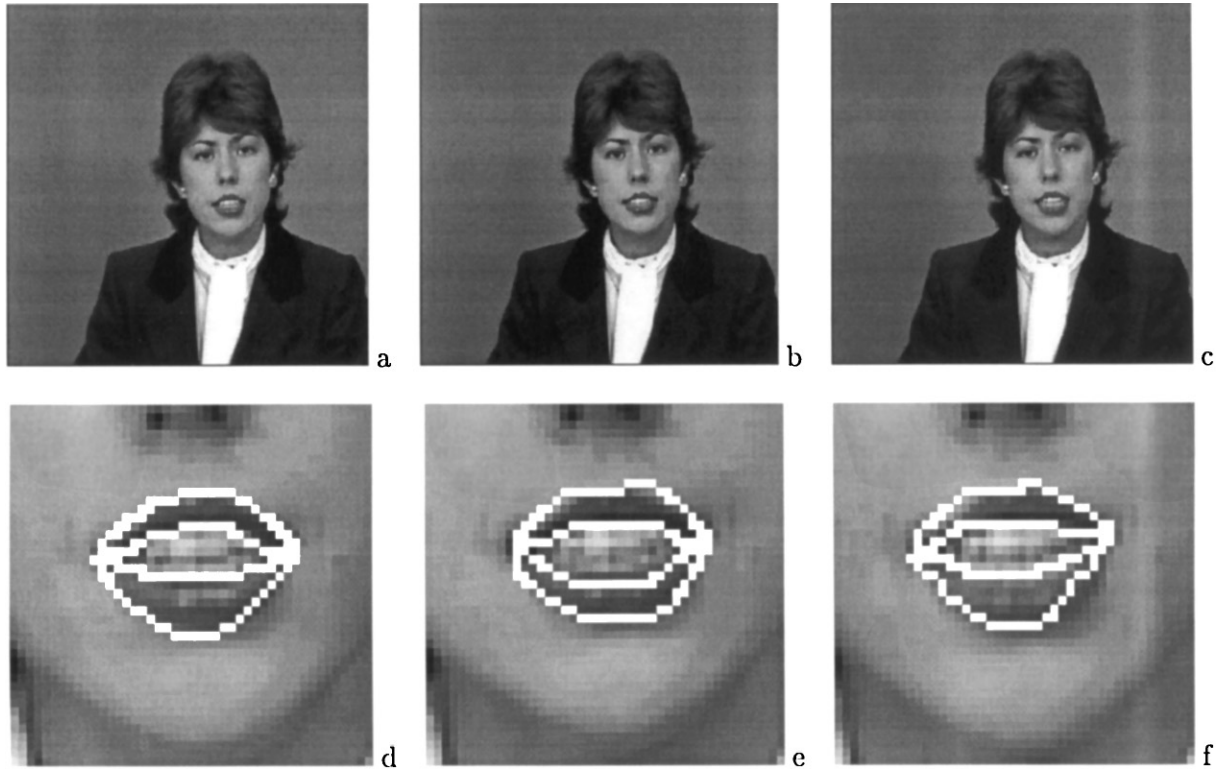

FIG. 13. Tracking of a mouth ("Mouth" sequence, 128×128).

FIG. 14. Tracking of a mouth ("Claire" sequence, 256×256). (a, b, c) Original sequence (frames 20, 23, 27, respectively). (d, e, f) Segmentation and tracking of the lips.

Modeling the uncoiling bow of a film reel. A third case study, which has not been considered previously in the literature, is presented in Fig. 15 to illustrate the variety of application fields which may be addressed with this approach. The problem is here to track the uncoiling bow of a film reel. The issue is to control the high uncoiling speed of the film and to keep up a constant film tensile strength on the reel, using standard servoing procedures [61]. To this end, we have designed a 16-point model interpolated by a B-spline curve to perform the *visual servoing*. The learning procedure has been conducted with 16 relevant examples of the uncoiling bow deformations. The initialization of the model on the first frame is done at hand (nonsupervised initialization procedures could, however, be devised in this case).

Promising results have been obtained with the segmentation and tracking procedure described in this paper (successive frames are presented in Fig. 15; the frame rate is 10 Hz, yielding quite large displacements). Because of the simplicity of the model, the cpu times reduce here to about 5 s per image on the same workstation. The final estimate of the stochastic deformable template is superimposed (in black) on each image (it is located close to the center of the reel).

Modeling deformable structures in medical images. The statistical segmentation and tracking procedure also proved efficient and reliable in a standard segmentation problem featuring deformable anatomical structures in medical images. Figure 16 presents, for instance, the extraction and tracking of the left ven-

tricle during a cardiac cycle, in an ultrasound image sequence. The observations correspond again to spatial gradient maps and the template is defined by 20 characteristic points designated on the silhouette of the ventricle. The computation cost is about 25 s for one 128×128 image in this case. These segmentation results have been estimated to be satisfactory in practice by expert physicians.

Other experimental results featuring other deformable models are reported in [34, 33].

Limits of the proposed algorithm. To assess the robustness of the combined segmentation/tracking procedure, we conducted the following complementary experiment: we simulated two synthetic sequences composed of 43 (256×256) frames showing a moving hand, at first without occlusion and then by introducing increasing occlusions on the fingers (see Fig. 17 for a selection of frames). The simulated deformable movement was the same for the sequences with and without occlusion and the ground truths, i.e., the true modal amplitudes, were known. Evaluation was accomplished by comparing the estimated modal amplitudes to the known mode values associated to the sequence containing no occlusion. Figure 17 depicts four typical segmentation results for each synthetic sequence. The top row of Fig. 17 illustrates the expected results and the bottom row presents the segmentations obtained when the fingers are partially occluded.

To compare two estimated shapes at time t, we compute the dot product (correlation measurement) of the two corresponding
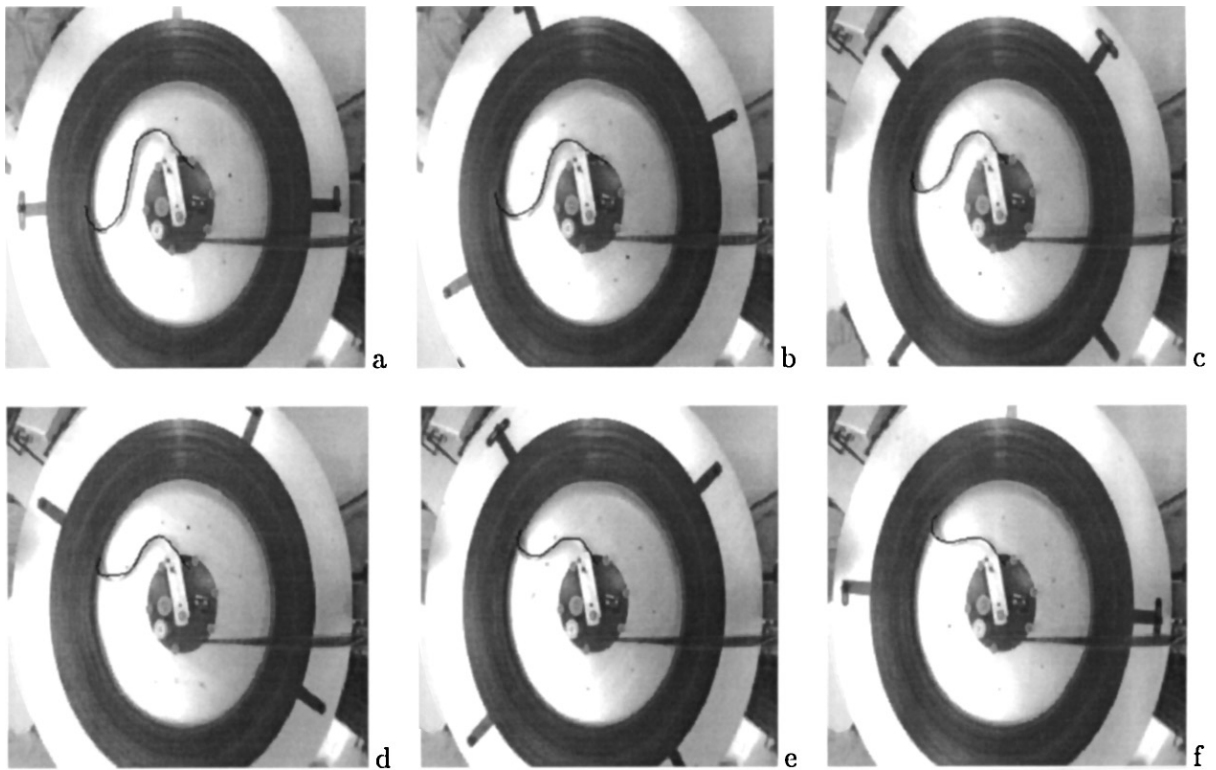
FIG. 15. Tracking of the uncoiling bow of a film reel (image sequence: 256×256; courtesy of Laboratoire d'Automatique des Arts et Métiers).
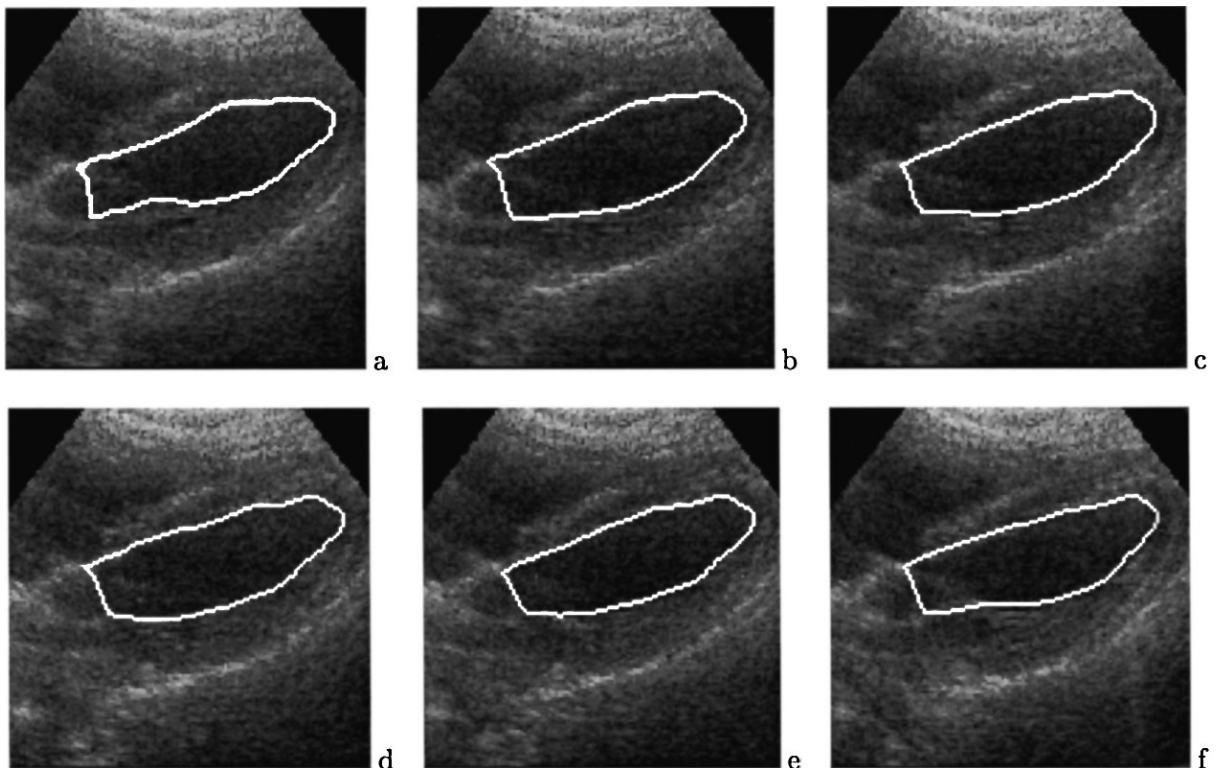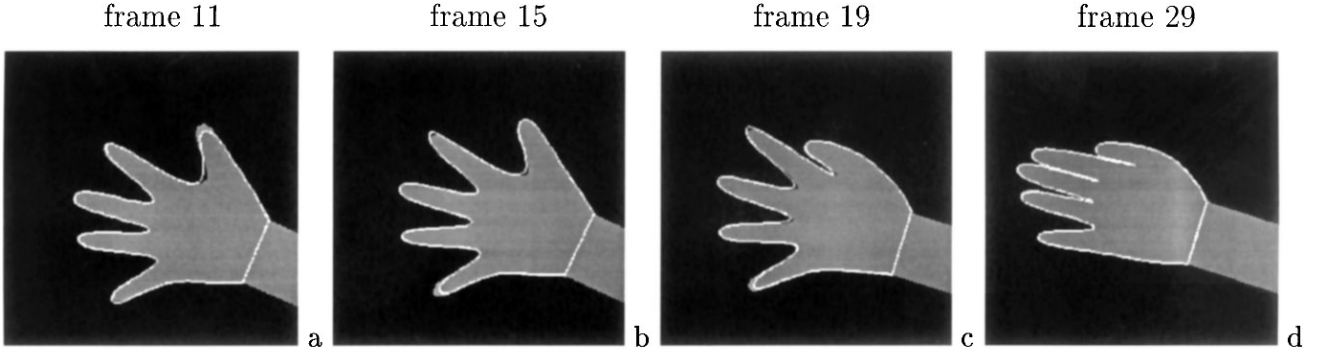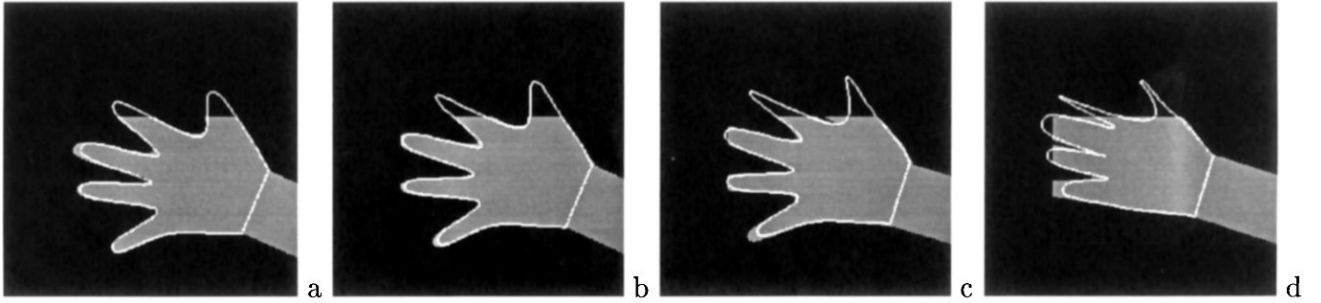

FIG. 16. Tracking of the left ventricle in ultrasound imagery (image size: 128×128).

frame 11   frame 15   frame 19   frame 29

Motion-based segmentation of a hand in presence of no occlusion

Motion-based segmentation of a hand in presence of partial occlusions

FIG. 17.   Motion-based segmentation of a moving hand with increasing occlusion.

vectors of modal values $\mathbf{b}_t^{\text{ref}}$ and $\mathbf{b}_t$ as proposed in [50]:

$$\epsilon = \frac{\mathbf{b}_t^{\text{ref}} \cdot \mathbf{b}_t}{\|\mathbf{b}_t^{\text{ref}}\| \|\mathbf{b}_t\|}. \tag{33}$$

This measurement ϵ reflects the distance between the two shapes. For two similar shapes, ϵ approaches 1. Figure 18 shows plots of the correlation measurement and of the proportion of occlusion in each frame. As can be seen, the mode values are

reliably recovered if the proportion of occlusions does not exceed about 10% in this case. Beyond 10% occlusion the correlation measure decreases quickly. For instance, in Fig. 17, frames 11 and 15 show satisfactory segmentations, whereas the shape is not correctly extracted in frame 19 and 29. The robustness of the segmentation/tracking procedure was also evaluated on real-world sequences, exhibiting hands with occlusions, with similar qualitative results.

7. CONCLUSION

In this paper, we have presented a statistical modeling framework for the representation, segmentation, and tracking of deformable objects in image sequences. The approach relies on the definition of a stochastic deformable template on which hierarchical deformations are applied (transformations from the similarity group and global and local deformations). The hierarchical decomposition of deformations enables a shape-tailored and compact although accurate description of deformations. Bayesian estimates of the model parameters and of the deformation processes are computed using stochastic and/or deterministic optimization techniques. Besides, a recursive temporal filtering of the model parameters is introduced to track the deformable structure over time. Tracking the deformable structure over a long image sequence reduces significantly the computation cost of the segmentation method (by ensuring the
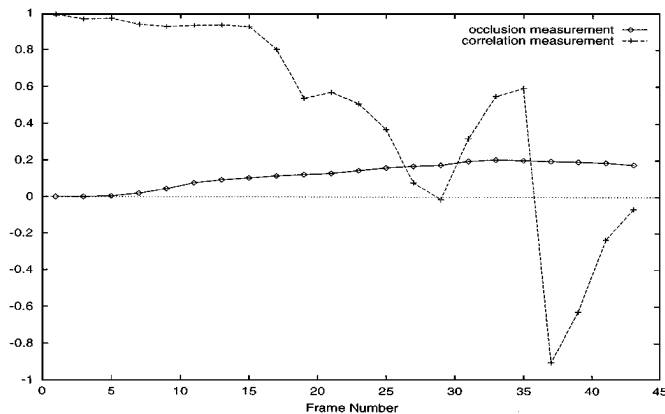
FIG. 18.   Performance analysis of the segmentation/tracking procedure.

propagation of relevant initializations over time) and enables us to process large movements reliably.

The contribution of this approach has been illustrated on synthetic as well as on real-world images sequences for the segmentation of a wide range of deformable objects. The method was shown to be robust even in adverse situations (such as low signal-to-noise ratio, nongaussian noise, and partial occlusions or missing data).

Several promising directions may be explored for continued research, starting from the approach proposed in this paper.

Let us first note that the proposed modeling and algorithmic framework is comprehensive and suited to the representation of a large class of deformable objects. It may also be adapted to segmentation problems based on other image attributes (color, texture, depth, etc.; see for instance [33]).

The manual training step necessary to compute the deformation modes of the model may become cumbersome, especially when 3D structures are to be handled. We currently investigate unsupervised training methods in order to alleviate the learning procedure (preliminary results have been presented in [31]). Unsupervised training methods define a first step toward a completely data-driven segmentation algorithm for deformable structures.

Handling large occlusions in image sequences is also a challenging issue for the deformable model-based approach. Robust estimation or outlier rejection techniques might be introduced with profit in the estimation of the local and global deformation parameters of the template, in order to cope with large occlusion areas [6]. alternatively, a robust Kalman filter [52] could also be considered in the tracking procedure.

Finally, the filtering and tracking of the parameters of the deformable templates yield promising future prospects as far as the characterization and the interpretation of the dynamic behavior of complex objects is concerned. Interpretation methods for nonrigid motions, similar to those presented in [11] for rigid movements, could, for instance, be devised to provide a qualitative interpretation and classification of deformable movements. This issue is of great importance in many application fields, for instance in medical imaging, when pathological deformations have to be detected.

APPENDIX A

Derivation of the Partition Function for the Statistical Model

In this Appendix we show that the partition function $Z_d = \int \exp -E_d(\mathbf{Y}, \mathbf{d}) \, d\mathbf{d}$ of the observation model does not depend on the template configuration $\mathbf{Y}$. We refer to Sections 4.1 and 4.2 for notations and show the result for the two observation models considered in this paper: motion-based segmentation model and edge-based segmentation model. In the case of the edge-based segmentation model, this property results from an approximation.

Motion-Based Segmentation Model

The normalizing constant Z_d related to the motion-based segmentation model described on Section 4.2 is easily derived as

$$Z_d = \int \exp -\left\{ \sum_{s \in \Gamma_{\mathbf{Y}}^I} |d(s) - 1| + \sum_{s \in \Gamma_{\mathbf{Y}}^O} |d(s) - 0| \right\} d\mathbf{d}$$

$$= \int \prod_{s \in \Gamma_{\mathbf{Y}}^I} \exp -|d(s) - 1| \prod_{s \in \Gamma_{\mathbf{Y}}^O} \exp -|d(s) - 0| \, d\mathbf{d}, \quad (34)$$

where $d(s) \in \{0, 1\}$ is the binary thresholded temporal gradient observed at site s. Since $d(s)$ is a binary variable, the expression of Z_d becomes

$$Z_d = \int (e^{-1})^{\alpha_d} (e^{-1})^{\beta_d} \, d\mathbf{d}, \quad (35)$$

where

- α_d designates the number of sites $s \in \Gamma_{\mathbf{Y}}^I$ for which $d(s) = 0$ (given a realization $\mathbf{d}$);
- β_d designates the number of sites $s \in \Gamma_{\mathbf{Y}}^O$ for which $d(s) = 1$ (given $\mathbf{d}$).

Let $\eta_d = \alpha_d + \beta_d$ ($0 \leq \eta_d \leq M$ where M is the total number of sites in the image). The number of possible binary data configurations $\mathbf{d}$ leading to a given value of η_d is simply $\binom{M}{\eta_d}$. The total number of possible data configurations is

$$2^M = \sum_{\eta_d=0}^{M} \binom{M}{\eta_d}. \quad (36)$$

From standard binomial series, we get

$$Z_d = \int e^{-\eta_d} \, d\mathbf{d} = \sum_{\eta_d=0}^{M} \binom{M}{\eta_d} \cdot e^{-\eta_d} = (1 + e^{-1})^M. \quad (37)$$

As can be seen Z_d does not depend on the template configuration $\mathbf{Y}$.

Edge-Based Segmentation Model

The edge-based segmentation model used in our implementation was described in [55].

The data field $\mathbf{d}$, defined on lattice S ($|S| = M$), is assumed to be a noise-corrupted version of an idealized gradient map depicting the boundary $\Gamma_{\mathbf{Y}}$ of the object of interest

$$\mathbf{d} = \mathbf{f} + \boldsymbol{\gamma},$$

where $\boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma_\gamma^2 \mathbf{I})$ is a white gaussian noise and

$$\forall s, \quad f(s) = \begin{cases} \mu & \text{if } s \in \Gamma_{\mathbf{Y}} \\ 0 & \text{else} \end{cases}.$$

The gradient map is idealized; i.e., one assumes that μ is constant through the image. Under this assumption, the likelihood function of the observed data may be expressed as

$$P(\mathbf{d} \mid \mathbf{Y}) = \frac{1}{(2\pi\sigma_\gamma^2)^{\frac{M}{2}}} \exp -\frac{1}{2\sigma_\gamma^2} \left\{ \sum_{s \in S} (d(s) - f(s))^2 \right\}$$

$$= \frac{1}{(2\pi\sigma_\gamma^2)^{\frac{M}{2}}} \exp -\frac{1}{2\sigma_\gamma^2} \left\{ \sum_{s \in S} d(s)^2 + \mu^2 \mid \Gamma_\mathbf{Y} \mid \right.$$

$$\left. - 2\mu \sum_{s \in \Gamma_\mathbf{Y}} d(s) \right\}.$$

If one assumes that the length of the boundary $\mid\Gamma_\mathbf{Y}\mid$ is approximately constant for all configurations of the deformable template, i.e., $\mid\Gamma_\mathbf{Y}\mid \simeq \mid\Gamma\mid$ does not depend on $\mathbf{Y}$, it follows that

$$P(\mathbf{d} \mid \mathbf{Y}) = \frac{1}{(2\pi\sigma_\gamma^2)^{\frac{M}{2}}} \exp -\frac{\mu^2 \mid\Gamma\mid}{2\sigma_\gamma^2} \exp -\frac{1}{2\sigma_\gamma^2} \sum_{s \in S} d(s)^2$$

$$\times \exp -\frac{\mu}{\sigma_\gamma^2} \sum_{s \in \Gamma_\mathbf{Y}} d(s). \tag{38}$$

This approximation is well verified once the scale factor of the template has been determined, in particular near convergence.

In Eq. (38), the term $\exp -\frac{1}{2\sigma_\gamma^2} \sum_{s \in S} d(s)^2$ does not depend on $\mathbf{Y}$ and can thus be discarded for the MAP estimation. Finally, one gets the model

$$P(\mathbf{d} \mid \mathbf{Y}) = \frac{1}{Z_d} \exp -\frac{\mu}{\sigma_\gamma^2} \sum_{s \in \Gamma_\mathbf{Y}} d(s), \tag{39}$$

where, as already stated, $d(s)$ is an "idealized" gradient map, $d(s) = \|\nabla I(s)\|$, and Z_d does not depend on $\mathbf{Y}$. This model expresses the correlation between the boundary template with the boundary strength in the data image [55].

APPENDIX B

The Kalman Filter Equations

In this Appendix we recall the standard Kalman filter equations used in our tracker.

The dynamic evolution of the system is described by

$$\mathbf{s}_{t+\Delta t} = \mathbf{A}_t \mathbf{s}_t + \boldsymbol{\xi}_t, \tag{40}$$

where $\mathbf{s}_t$ is the state vector, $\mathbf{A}_t$ is the state transition matrix, and $\boldsymbol{\xi}_t$ is a zero-mean white gaussian noise with covariance matrix $\mathbf{Q}_t = \mathbb{E}[\boldsymbol{\xi}_t \boldsymbol{\xi}_t^T]$. The measurement $\mathbf{w}_t$ is a linear function of the state vector

$$\mathbf{w}_t = \mathbf{H}_t \mathbf{s}_t + \boldsymbol{\eta}_t, \tag{41}$$

where $\boldsymbol{\eta}_t$ corresponds to a zero-mean white gaussian noise with covariance matrix $\mathbf{V}_t = \mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\eta}_t^T]$.

If $\hat{\mathbf{s}}_{t_2|t_1}$ denotes the minimum mean square error estimate of $\mathbf{s}_{t_2}$ given the measurements up to $t_1(t_1 \leq t_2)$ and $\mathbf{P}_{t_2|t_1}$ is the associated error covariance matrix, the Kalman filter updating equations are

$$\hat{\mathbf{s}}_{t|t} = \hat{\mathbf{s}}_{t|t-\Delta t} + \mathbf{K}_t[\mathbf{w}_t - \mathbf{H}_t \hat{\mathbf{s}}_{t|t-\Delta t}], \tag{42}$$

$$\mathbf{P}_{t|t} = [\mathbf{I} - \mathbf{K}_t \mathbf{H}_t]\mathbf{P}_{t|t-\Delta t}, \tag{43}$$

where the Kalman gain $\mathbf{K}_t$ is defined as

$$\mathbf{K}_t = \mathbf{P}_{t|t-\Delta t} \mathbf{H}_t^T \left[\mathbf{H}_t \mathbf{P}_{t|t-\Delta t} \mathbf{H}_t^T + \mathbf{V}_t \right]^{-1}. \tag{44}$$

The prediction equations are

$$\hat{\mathbf{s}}_{t+\Delta t|t} = \mathbf{A}_t \hat{\mathbf{s}}_{t|t}, \tag{45}$$

$$\mathbf{P}_{t+\Delta t|t} = \mathbf{A}_t \mathbf{P}_{t|t} \mathbf{A}_t^T + \mathbf{Q}_t. \tag{46}$$

APPENDIX C

Dynamic Evolution of the Deformable Template Parametrization

In this Appendix we derive the equations describing the temporal evolution of the model parametrization.

Consider the tracking of the model (redefined at time t) over a long sequence:

$$\mathbf{X}_t = \mathbf{M}_t[\bar{\mathbf{x}} + \boldsymbol{\Phi}\mathbf{b}_t] + \mathbf{T}_t. \tag{47}$$

The global deformable model $\mathbf{X}_t$ and its first-order and second-order temporal derivatives may easily be expressed at time t by

$$\begin{cases} \mathbf{X}_t = \mathbf{M}_t[\bar{\mathbf{x}} + \boldsymbol{\Phi}\mathbf{b}_t] + \mathbf{T}_t, \\ \dot{\mathbf{X}}_t = \dot{\mathbf{M}}_t[\bar{\mathbf{x}} + \boldsymbol{\Phi}\mathbf{b}_t] + \mathbf{M}_t\boldsymbol{\Phi}\dot{\mathbf{b}}_t + \dot{\mathbf{T}}_t, \\ \ddot{\mathbf{X}}_t = \ddot{\mathbf{M}}_t[\bar{\mathbf{x}} + \boldsymbol{\Phi}\mathbf{b}_t] + \mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t + 2\dot{\mathbf{M}}_t\boldsymbol{\Phi}\dot{\mathbf{b}}_t + \ddot{\mathbf{T}}_t. \end{cases} \tag{48}$$

In this modeling, we assume that $\boldsymbol{\Phi}$ and $\bar{\mathbf{x}}$ do not depend on time t, i.e., that the deformation eigenvectors of the deformable template are stationary. The temporal evolution of the model is described by a second-order Taylor expansion of vector $\mathbf{X}_t$ and $\dot{\mathbf{X}}_t$:

$$\begin{cases} \mathbf{X}_{t+\Delta t} = \mathbf{X}_t + \Delta t \dot{\mathbf{X}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{X}}_t, \\ \dot{\mathbf{X}}_{t+\Delta t} = \dot{\mathbf{X}}_t + \Delta t \ddot{\mathbf{X}}_t. \end{cases} \tag{49}$$

By substituting Eq. (47) into Eq. (49), we obtain

$$\mathbf{M}_{t+\Delta t}\bar{\mathbf{x}} + \mathbf{M}_{t+\Delta t}\boldsymbol{\Phi}\mathbf{b}_{t+\Delta t} + \mathbf{T}_{t+\Delta t} = \left[\mathbf{M}_t + \Delta t\dot{\mathbf{M}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{M}}_t\right]\bar{\mathbf{x}}$$

$$+ \left[\mathbf{M}_t + \Delta t\dot{\mathbf{M}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{M}}_t\right]\boldsymbol{\Phi}\mathbf{b}_t + \left[\Delta t\mathbf{M}_t + \Delta t^2\dot{\mathbf{M}}_t\right]\boldsymbol{\Phi}\dot{\mathbf{b}}_t$$

$$+ \left[\frac{\Delta t^2}{2}\mathbf{M}_t\right]\boldsymbol{\Phi}\ddot{\mathbf{b}}_t + \left[\mathbf{T}_t + \Delta t\dot{\mathbf{T}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{T}}_t\right]. \quad (50)$$

The temporal evolution of hyperparameters $\mathbf{M}_t$ and $\mathbf{T}_t$ are obtained directly from (50) by inspection and identification:

$$\mathbf{M}_{t+\Delta t} = \mathbf{M}_t + \Delta t\dot{\mathbf{M}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{M}}_t, \quad (51)$$

$$\mathbf{T}_{t+\Delta t} = \mathbf{T}_t + \Delta t\dot{\mathbf{T}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{T}}_t. \quad (52)$$

The dynamic evolution of the modal amplitudes is obtained by identifying the remaining terms in (50):

$$\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi}\mathbf{b}_{t+\Delta t}$$

$$= \left[\mathbf{M}_t + \Delta t\dot{\mathbf{M}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{M}}_t\right]\boldsymbol{\Phi}\mathbf{b}_t + [\Delta t\mathbf{M}_{t+\Delta t} + \Delta t^2\dot{\mathbf{M}}_t]\boldsymbol{\Phi}\dot{\mathbf{b}}_t$$

$$+ \frac{\Delta t^2}{2}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t$$

$$= \mathbf{M}_{t+\Delta t}\boldsymbol{\Phi}\mathbf{b}_t + \left[\Delta t\mathbf{M}_{t+\Delta t} - \frac{\Delta t^3}{2}\ddot{\mathbf{M}}_t\right]\boldsymbol{\Phi}\dot{\mathbf{b}}_t$$

$$+ \frac{\Delta t^2}{2}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t$$

$$= \mathbf{M}_{t+\Delta t}\boldsymbol{\Phi}[\mathbf{b}_t + \Delta t\dot{\mathbf{b}}_t] - \frac{\Delta t^3}{2}\ddot{\mathbf{M}}_t\boldsymbol{\Phi}\dot{\mathbf{b}}_t + \frac{\Delta t^2}{2}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t. \quad (53)$$

By neglecting the higher order terms, it follows that

$$\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi}\mathbf{b}_{t+\Delta t} = \mathbf{M}_{t+\Delta t}\boldsymbol{\Phi}[\mathbf{b}_t + \Delta t\dot{\mathbf{b}}_t] + \frac{\Delta t^2}{2}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t. \quad (54)$$

Due to the reduction of dimension stemming from the KL transform, Eq. (54) is an *overdetermined system* of $2n$ equations with $m < 2n$ unknowns corresponding to the modal weights $\mathbf{b}_{t+\Delta t}$. This system is solved by considering the pseudoinverse — $\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}$ of matrix $\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi}$:

$$\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1} = [(\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi})^T(\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi})]^{-1}(\mathbf{M}_{t+\Delta t}\boldsymbol{\Phi})^T. \quad (55)$$

The final equation for the evolution of $\mathbf{b}_t$ becomes

$$\mathbf{b}_{t+\Delta t} = \mathbf{b}_t + \Delta t\dot{\mathbf{b}}_t + \frac{\Delta t^2}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t. \quad (56)$$

A similar derivation (see [34] for details) leads to the equations describing the evolution of $\dot{\mathbf{M}}_t$, $\dot{\mathbf{T}}_t$, and $\dot{\mathbf{b}}_t$. The dynamic equations for the global deformation parameters may finally be summarized as the following:

$$\begin{cases}
\mathbf{M}_{t+\Delta t} = \mathbf{M}_t + \Delta t\dot{\mathbf{M}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{M}}_t, \\[2mm]
\dot{\mathbf{M}}_{t+\Delta t} = \dot{\mathbf{M}}_t + \Delta t\ddot{\mathbf{M}}_t, \\[2mm]
\mathbf{T}_{t+\Delta t} = \mathbf{T}_t + \Delta t\dot{\mathbf{T}}_t + \frac{\Delta t^2}{2}\ddot{\mathbf{T}}_t, \\[2mm]
\dot{\mathbf{T}}_{t+\Delta t} = \dot{\mathbf{T}}_t + \Delta t\ddot{\mathbf{T}}_t, \\[2mm]
\mathbf{b}_{t+\Delta t} = \mathbf{b}_t + \Delta t\dot{\mathbf{b}}_t + \frac{\Delta t^2}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t, \\[2mm]
\dot{\mathbf{b}}_{t+\Delta t} = \left[\mathbf{I} + \frac{3\Delta t}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}(\dot{\mathbf{M}}_t\boldsymbol{\Phi} - \dot{\mathbf{M}}_{t+\Delta t}\boldsymbol{\Phi})\right]\dot{\mathbf{b}}_t \\[2mm]
\qquad + \left[\Delta t\mathbf{I} - \frac{\Delta t^2}{2}\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\dot{\mathbf{M}}_{t+\Delta t}\boldsymbol{\Phi}\right]\tilde{\boldsymbol{\Phi}}_{t+\Delta t}^{-1}\mathbf{M}_t\boldsymbol{\Phi}\ddot{\mathbf{b}}_t
\end{cases}$$

REFERENCES

1. Y. Amit, U. Grenander, and M. Piccioni, Structural image restoration through deformable templates, *J. Am. Statist. Assoc.* **86**, 1991, 376–387.

2. A. Azarbayejani, B. Horowitz, and A. Pentland, Recursive estimation of structure and motion using relative orientation constraints, in *Proc. Conf. Comp. Vision Pattern Rec., New York, 1993*, pp. 294–299.

3. B. Bascle, P. Bouthemy, N. Deriche, and F. Meyer, Tracking complex primitives in an image sequence, in *Proc. Int. Conf. Pattern Recognition, Jerusalem, 1994*, pp. 426–431.

4. A. Baumberg and D. Hogg, Learning flexible models from image sequences, in *Proc. European Conf. Computer Vision, Stockholm, 1994*, pp. 299–308.

5. J. Besag, On the statistical analysis of dirty pictures, *J. Royal Statist. Soc. B* **48**, 1986, 259–302.

6. M. A. Black and A. D. Jepson, Eigentracking: robust matching and tracking of articulated objects using a view-based representation, in *Proc. European Conf. Computer Vision, Cambridge, UK, April 1996*, pp. 329–342.

7. A. Blake and A. Yuille, *Active Vision*, MIT Press, 1992.

8. A. Blake, R. Curwen, and A. Zisserman, A framework for spatiotemporal control in the tracking of visual contours, *Int. J. Comput. Vision* **11**, 1993, 127–145.

9. A. Blake, M. Isard, and D. Reynard, Learning to track the visual motion of contours, *Artificial Intelligence* **78**, 1995, 179–212.

10. P. Bouthemy and A. Benveniste, Modeling of atmospheric disturbances in meteorological pictures, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 1984, 587–600.

11. P. Bouthemy and E. Francois, Motion segmentation and qualitative dynamic scene analysis from an image sequence, *Int. J. Comput. Vision* **10**, 1993, 157–182.

12. G. C.-H. Chuang and C.-C. Jay Kuo, Wavelet descriptor of planar curves: theory and applications, *IEEE Trans. Image Process.* **5**, 1996, 56–70.

13. I. Cohen, L. D. Cohen, and N. Ayache, Using deformable surfaces to segment 3D images and infer differential structures, *CVGIP: Image Understanding* **56**, 1992, 242–263.

14. T. F. Cootes, A. H. Hill, C. J. Taylor, and J. Haslam, The use of active models for locating structures in medical images, *Image and Vision Comput.* **12**, 1994, 355–365.

15. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, Active shape models—their training and application, *CVGIP: Image Understanding* **61**, 1994, 38–59.

16. R. Curwen and A. Blake, Dynamic contours: real-time active splines, in *Active Vision* (A. Blake and A. Yuille, Ed.), Chap. 3, pp. 39–57, MIT Press, Cambridge, MA, 1992.

17. G. Demoment, Image reconstruction and restoration: overview of common estimation structures and problems, *IEEE Trans. Acoust., Speech, Signal Process.* **37**, 1989, 2024–2036.

18. A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Statist. Soc. B* **39**, 1977, 1–38.

19. R. Deriche and O. Faugeras, Tracking line segments, in *Proc. European Conf. Comput. Vision, Antibes, 1990*, pp. 259–268.

20. 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, New York, 1988*, pp. 1084–1087.

21. N. Friedland and D. Adam, Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing, *IEEE Trans. Medical Imaging* **8**, 1989, 344–353.

22. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 1984, 721–741.

23. U. Grenander, Y. Chow, and D. M. Keenan, *Hands. A Pattern Theoretic Study of Biological Shapes*, Springer-Verlag, Berlin/Heidelberg/New York, 1991.

24. U. Grenander and D. M. Keenan, Towards automated image understanding, *J. Appl. Statist.* **16**, 1989, 207–221.

25. F. Heitz and P. Bouthemy, Multimodal estimation of discontinuous optical flow using Markov random fields, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1993, 1217–1232.

26. A. H. Hill and C. J. Taylor, Model based image interpretation using genetic algorithms, *Image and Vision Comput.* **10**, 1992, 295–300.

27. D. V. Hinkley, Inference about the change-point from cumulative sum-tests, *Biometrika* **58**, 1971, 509–523.

28. M. Kass, A. Witkin, and D. Terzopolous, Snakes: Active contour models, in *Proc. Int. Conf. Computer Vision, London, 1987*, pp. 259–268.

29. C. Kervrann and F. Heitz, A hierarchical statistical framework for the segmentation of deformable objects in image sequences, in *Proc. Conf. Comp. Vision Pattern Rec., Seattle, 1994*, pp. 724–728.

30. C. Kervrann and F. Heitz, Robust tracking of stochastic deformable models in image sequences, in *Proc. Int. Conf. Image Processing, Austin, 1994*, Vol. III, pp. 88–92.

31. C. Kervrann and F. Heitz, Learning structure and deformation modes of nonrigid objects in long image sequences, in *Proc. Int. Workshop on Automatic Face and Gesture Recognition, Zurich, 1995*, pp. 104–109.

32. C. Kervrann and F. Heitz, A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics, *IEEE Trans. Image Process.* **4**, 1995, 856–862.

33. C. Kervrann and F. Heitz, Statistical model-based segmentation of deformable motion, in *Proc. Int. Conf. Image Processing, Lausanne, 1996*, Vol. I, pp. 937–940.

34. C. Kervrann, *Statistical Models for the Segmentation and Tracking of 2D Deformable Structures in Image Sequences*, Ph.D. thesis, Université de Rennes I, France, 1995. [In French]

35. B. B. Kimia, A. R. Tannenbaum, and S. W. Zucker, Shapes, shocks, and deformations I: the components of two-dimensional shape and the reaction-diffusion space, *Int. J. Comput. Vision* **19**, 1995, 189–224.

36. K. L. Lai and R. T. Chin, Deformable contours: modeling and extraction, in *Proc. Conf. Comp. Vision Pattern Rec., Seattle, 1994*, pp. 601–608.

37. S. Lakshmanan and H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1989, 799–813.

38. R. Malladi, J. A. Sethian, and B. C. Vemuri, Shape modeling with front propagation: a level set approach, *IEEE Trans. Pattern Anal. Machine Intell.* **17**, 1995, 158–175.

39. K. V. Mardia and T. F. Hainsworth, Deformable templates in image sequences, in *Proc Int. Conf. Pattern Recognition, La Haye, 1992*, pp. 132–135.

40. J. Martin, A. Pentland, and R. Kikinis, Shape analysis of brain structures using physical and experimental modes, in *Proc. Conf. Comp. Vision Pattern Rec., Seattle, 1994*, pp. 752–755.

41. M. Maurizot, P. Bouthemy, B. Delyon, A. Iouditsky, and J. M. Odobez, Locating singular points and characterizing deformable flow fields in image sequences, in *Proc. Int. Conf. Image Processing, Washington, 1995*, pp. 88–92.

42. F. Meyer and P. Bouthemy, Region-based tracking using affine motion models in long image sequences, *CVGIP: Image Understanding* **60**, 1994, 119–140.

43. A. Mohammad Djafari, On the estimation of hyperparameters in bayesian approach of solving inverse problems, in *Proc. Int. Conf. Acoust., Speech, Signal Processing, Minneapolis, 1993*, pp. 495–498.

44. T. McInerney and D. Terzopoulos, Topologically adaptable snakes, in *Int. Conf. Computer Vision, Cambridge, MA, 1995*, pp. 840–845.

45. H. Murase and S. K. Nayar, Visual learning and recognition of 3D objects from appearance, *Int. J. Comput. Vision* **14**, 1995, 5–24.

46. C. Nastar and N. Ayache, Fast segmentation, tracking and analysis of deformable objects, in *Proc. Int. Conf. Computer Vision, Berlin, 1993*, pp. 275–279.

47. J. M. Odobez and P. Bouthemy, Detection of multiple objects using multiscale Markov random fields, with camera compensation, in *Proc Int. Conf. Image Processing, Austin, 1994*, pp. 257–261.

48. E. Oja, *Subspace Methods of Pattern Recognition*, Research Studies Press, Hertfordshire, 1983.

49. A. Pentland and B. Horowitz, Recovery of non-rigid motion and structure, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 1991, 730–742.

50. A. Pentland and S. Sclaroff, Closed-form solutions for physically based shape modeling and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 1991, 715–729.

51. P. Perez and F. Heitz, Restriction of a Markov random field on a graph and multiresolution statistical image modeling, *IEEE Trans. Information Theory* **42**, 1996, 180–190.

52. R. Rao, *Robust Kalman Filters for Prediction, Recognition and Learning*, Technical Report 645, The University of Rochester, Computer Science Dept., Dec. 1996.

53. K. Rohr, Incremental recognition of pedestrians from image sequences, in *Proc. Conf. Comp. Vision Pattern Rec., New York, 1993*, pp. 8–13.

54. N. Rougon and F. Preteux, Marqueurs déformables: segmentation par contour actif et morphologie mathématique, in *Actes du 8e congrès AFCET/INRIA Reconnaisance des Formes et Intelligence Artificielle, Lyon, 1991*, pp. 955–966. [In French]

55. L. H. Staib and J. S. Duncan, Boundary finding with parametrically deformable models, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 1992, 1061–1075.

56. D. Terzopoulos and D. Metaxas, Dynamic 3D models with local and global deformations: deformable superquadrics, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 1991, 703–714.

57. D. Terzopoulos and R. Szelisky, Tracking with Kalman Snakes, in *Active Vision* (A. Blake and A. Yuille, Ed.), Chap. 1, pp. 3–20, MIT Press, Cambridge, MA, 1992.

58. M. Turk and A. Pentland, Eigenfaces for recognition, *J. Cognitive Neurosci.* **3**, 1991, 71–86.

59. A. Yuille, P. W. Hallinan, and D. S. Cohen, Feature extraction from faces using deformable templates, *Int. J. Comput, Vision* **8**, 1992, 99–111.

60. J. Zhong, T. S. Huang, and R. J. Adrian, Salient structures analysis of fluid flow, in *Proc. Conf. Comp. Vision Pattern Rec., Seattle, 1994*, pp. 310–315.

61. D. Zugaj and V. Lattuati, Contrôle en temps réel de la forme d'une boucle de déroulement par vision artificielle, in *QCAV '95, Le Creusot, France, 1995*. [In French]