

Dynamical Statistical Shape Priors for Level Set Based Tracking

Daniel Cremers, *Member, IEEE*

Abstract—In recent years, researchers have proposed to introduce statistical shape knowledge into level set based segmentation methods in order to cope with insufficient low-level information. While these priors were shown to drastically improve the segmentation of familiar objects, so far the focus has been on statistical shape priors which are *static in time*. Yet, in the context of tracking deformable objects, it is clear that certain silhouettes (such as those of a walking person) may become more or less likely over time. In this paper, we tackle the challenge of learning *dynamical* statistical models for implicitly represented shapes. We show how these can be integrated as dynamical shape priors in a Bayesian framework for level set based image sequence segmentation. We assess the effect of such shape priors “with memory” on the tracking of familiar deformable objects in the presence of noise and occlusion. We show comparisons between dynamical and static shape priors, between models of pure deformation and joint models of deformation and transformation, and we quantitatively evaluate the segmentation accuracy as a function of the noise level and of the camera frame rate. Our experiments demonstrate, that level set based segmentation and tracking can be strongly improved by exploiting the temporal correlations among consecutive silhouettes which characterize deforming shapes.

I. INTRODUCTION

In 1988, Osher and Sethian [21] introduced the level set method¹ as a means to implicitly propagate hypersurfaces $C(t)$ in a domain $\Omega \subset \mathbb{R}^n$ by evolving an appropriate embedding function $\phi : \Omega \times [0, T] \rightarrow \mathbb{R}$, where:

$$C(t) = \{x \in \Omega \mid \phi(x, t) = 0\}. \quad (1)$$

The ordinary differential equation propagating explicit boundary points is thus replaced by a partial differential equation modeling the evolution of a higher-dimensional embedding function. The key advantages of this approach are well-known: Firstly,

the implicit boundary representation does not depend on a specific parameterization, during the propagation no control point regriding mechanisms need to be introduced. Secondly, evolving the embedding function allows to elegantly model topological changes of the boundary such as splitting and merging. In the context of statistical shape learning, this allows to construct shape dissimilarity measures defined on the embedding function which can handle shapes of varying topology. Thirdly, the implicit representation naturally generalizes to hypersurfaces in three or more dimensions. To impose a unique correspondence between a shape and its embedding function one can constrain ϕ to be a signed distance function, i.e. $|\nabla\phi| = 1$ almost everywhere, with $\phi > 0$ inside and $\phi < 0$ outside the shape.

The first applications of level set methods to image segmentation were pioneered in the early 90’s by Malladi et al. [17], by Caselles et al. [4], by Kichenassamy et al. [13] and by Paragios and Deriche [23]. Level set implementations of the Mumford-Shah functional [19] were proposed by Chan and Vese [5] and by Tsai et al. [30].

In recent years, researchers have successfully introduced prior shape information into level set based segmentation schemes. Leventon et al. [15] proposed to model the embedding function by principal component analysis (PCA) of a set of training shapes and to add appropriate driving terms to the level set evolution equation, Tsai et al. [31] performed optimization directly within the subspace of the first few eigenmodes. Rousson et al. [27], [28] suggested to introduce shape information on the variational level, while Chen et al. [6] imposed shape constraints directly on the contour given by the zero level of the embedding function. More recently, Riklin-Raviv et al. [25] proposed to introduce projective invariance by slicing the signed distance function at various angles. Models to simultaneously impose shape knowledge about multiple objects were proposed by Cremers et al. [10].

Manuscript received July 20, 2005; revised December 6, 2005. Recommended for acceptance by G. Sapiro.

Daniel Cremers is with the Department of Computer Science, University of Bonn, Germany. E-mail: dcremers@cs.uni-bonn.de

¹A precursor of the level set method was proposed in [11].

In the above works, statistically learned shape information was shown to cope for missing or misleading information in the input images due to noise, clutter and occlusion. The shape priors were developed to segment objects of familiar shape in a given image. However, although they can be applied to tracking objects in image sequences [8], [18], [9], they are not well-suited for this task, because they neglect the *temporal coherence of silhouettes* which characterizes many deforming shapes.

When tracking a three-dimensional deformable object over time, clearly not all shapes are equally likely at a given time instance. Regularly sampled images of a walking person, for example, exhibit a typical pattern of consecutive silhouettes. Similarly, the projections of a rigid 3D object rotating at constant speed are generally not independent samples from a statistical shape distribution. Instead, the resulting set of silhouettes can be expected to contain strong temporal correlations.

In this paper, we develop temporal statistical shape models for implicitly represented shapes. In particular, the shape probability at a given time depends on the shapes observed at previous time steps. The integration of such dynamical shape models into the segmentation process can be elegantly formulated within a Bayesian framework for level set based image sequence segmentation. The resulting optimization problem can be implemented by a partial differential equation for the level set function. It models an evolution of the interface which is driven both by the intensity information of the current image as well as by a dynamical shape prior which relies on the segmentations obtained on the preceding frames. Experimental evaluation demonstrates that – in contrast to existing approaches to segmentation with statistical shape priors – the resulting segmentations are not only similar to previously learned shapes, but they are also consistent with the temporal correlations estimated from sample sequences. The resulting segmentation process can cope with large amounts of noise and occlusion because it exploits prior knowledge about *temporal* shape consistency and because it aggregates information from the input images over time (rather than treating each image independently).

The development of dynamical models for implicitly represented shapes and their integration into image sequence segmentation on the basis of the

Bayesian framework draws on much prior work in various fields. The theory of dynamical systems and time series analysis has a long tradition in the literature (see for example [22], [16]). Autoregressive models were developed for *explicit* shape representations among others by Blake, Isard and coworkers [2], [3]. In these works, successful tracking results were obtained by particle filtering based on edge-information extracted from the intensity images. Although our dynamical shape representations were inspired by the above works, our method differs from these in three ways:

- We propose dynamical models for *implicitly* represented shapes. As a consequence, our dynamical shape model can automatically handle shapes of varying topology. Our model trivially extends to higher dimensions (e.g. 3D shapes), since we do not need to deal with the combinatorial problem of determining point correspondences and issues of control point regridding associated with explicit shape representations.
- Our method integrates the intensity information of the input images in a statistical formulation inspired by [19], [32], [5]. This leads to a region-based tracking scheme rather than an edge-based one. The statistical formulation implies that – with respect to the assumed intensity models – our method optimally exploits the input information. It does not rely on a precomputation of heuristically defined image edge features. Yet, the assumed probabilistic intensity models are quite simple (namely Gaussian distributions). More sophisticated models for intensity, color or texture of objects and background could be employed. But this is not the focus of the present paper.
- The Bayesian a posteriori optimization is solved in a variational setting by gradient descent rather than by stochastic sampling techniques. While this limits our algorithm to only track the most likely hypothesis (rather than multiple hypotheses), it facilitates an extension to higher-dimensional shape representations without the drastic increase in computational complexity inherent to sampling methods. While there exist algorithms to efficiently compute global minima of a certain class of cost functionals [14], the functional derived in this work does not fall within this class.

Recently, Goldenberg et al. [12] successfully applied PCA to an aligned shape sequence to classify the behavior of periodic shape motion. Though this work is also focused on characterizing moving implicitly represented shapes, it differs from ours in that shapes are not represented by the level set embedding function (but rather by a binary mask), it does not make use of autoregressive models, and it is focused on behavior classification of presegmented shape sequences rather than segmentation or tracking with dynamical shape priors.

The paper is structured as follows. In Section II, we introduce a Bayesian formulation for level set based image sequence segmentation and specify which assumptions we make in order to end up with a computationally feasible problem. In Section III, we introduce dynamical models which allow to learn dynamical statistical priors for implicitly represented shapes. In Section IV, we show how the Bayesian inference can be computed by energy minimization and derive appropriate partial differential equations. In Section V, we provide experimental results aimed at evaluating several properties of our method: We show that the dynamical prior can cope with large amounts of noise, while a static shape prior – even with moderate amounts of noise – gets stuck in a local minimum after the first few frames. We quantify the segmentation accuracy for a dynamical shape prior, trained on a specific walking sequence, when applied to sequences of different walking speed. And finally, we show how dynamical priors which capture the joint distribution of deformations and transformations outperform purely deformation-based dynamical priors, when dealing with occlusions.

A preliminary version of this work was published in [7].

II. LEVEL SET BASED TRACKING AS BAYESIAN INFERENCE

In this section, we will introduce a Bayesian formulation for level set based image sequence segmentation. We first treat the general formulation in the space of embedding functions and subsequently propose a computationally efficient formulation in a low-dimensional subspace.

A. General Formulation

In the following, we define as *shape* a set of closed 2D contours modulo a certain transformation

group, the elements of which are denoted by T_θ with a parameter vector θ . Depending on the application, these may be rigid-body transformations, similarity or affine transformations or larger transformation groups. The shape is represented implicitly by an embedding function ϕ according to equation (1). Thus objects of interest will be given by $\phi(T_\theta x)$, where the transformation T_θ acts on the grid, leading to corresponding transformations of the implicitly represented contour. We purposely separate shape ϕ and transformation parameters θ since one may want to use different models to represent and learn their respective temporal evolution.

Assume we are given consecutive images $I_t : \Omega \rightarrow \mathbb{R}$ from an image sequence, where $I_{1:t}$ denotes the set of images $\{I_1, I_1, \dots, I_t\}$ at different time instances. Using the Bayesian formula (with all expressions conditioned on $I_{1:t-1}$), the problem of segmenting the current frame I_t can then be addressed by maximizing the conditional probability

$$\mathcal{P}(\phi_t, \theta_t | I_{1:t}) = \frac{\mathcal{P}(I_t | \phi_t, \theta_t, I_{1:t-1}) \mathcal{P}(\phi_t, \theta_t | I_{1:t-1})}{\mathcal{P}(I_t | I_{1:t-1})}, \quad (2)$$

with respect to the embedding function ϕ_t and the transformation parameters θ_t .² For the sake of brevity, we will not delve into the philosophical interpretation of the Bayesian approach. We merely point out that the Bayesian framework can be seen as an inversion of the image formation process in a probabilistic setting.

The denominator in (2) does not depend on the estimated quantities and can therefore be neglected in the maximization. Moreover, the second term in the numerator can be rewritten using the Chapman-Kolmogorov equation [22]:

$$\begin{aligned} \mathcal{P}(\phi_t, \theta_t | I_{1:t-1}) &= \int \mathcal{P}(\phi_t, \theta_t | \phi_{1:t-1}, \theta_{1:t-1}) \\ &\cdot \mathcal{P}(\phi_{1:t-1}, \theta_{1:t-1} | I_{1:t-1}) d\phi_{1:t-1} d\theta_{1:t-1} \end{aligned} \quad (3)$$

In the following, we will make several assumptions which are aimed at simplifying expression (2), leading to a computationally more feasible estimation problem:

- We assume that the images $I_{1:t}$ are mutually independent:

$$\mathcal{P}(I_t | \phi_t, \theta_t, I_{1:t-1}) = \mathcal{P}(I_t | \phi_t, \theta_t). \quad (4)$$

²Modeling probability distributions on infinite-dimensional spaces is in general an open problem, including issues of defining appropriate measures and of integrability. Therefore the functions ϕ may be thought of as finite-dimensional approximations obtained by sampling the embedding functions on a regular grid.

- We assume that the intensities of the shape of interest and of the background are independent samples from two Gaussian distributions with unknown means μ_1, μ_2 and variances σ_1, σ_2 . As a consequence, the data term above can be written as:

$$\mathcal{P}(I_t | \phi_t, \theta_t) = \prod_{\substack{x \\ \phi_t(T_{\theta_t}x) \geq 0}} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(I_t(x)-\mu_1)^2}{2\sigma_1^2}} \cdot \prod_{\substack{x \\ \phi_t(T_{\theta_t}x) < 0}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(I_t(x)-\mu_2)^2}{2\sigma_2^2}} \quad (5)$$

where we have introduced the Heaviside step function $H\phi \equiv H(\phi)$ to denote the areas where ϕ is positive ($H\phi = 1$) or negative ($H\phi = 0$). Related intensity models have been proposed among others in [19], [32], [5]. The model parameters μ_i and σ_i are estimated jointly with the shape ϕ_t and the transformation θ_t . Their optimal values are given by the means and variances of the intensity I_t inside and outside the current shape:

$$\mu_1 = \frac{\int I_t H\phi_t dx}{\int H\phi_t dx}, \quad \sigma_1^2 = \frac{\int (I_t - \mu_1)^2 H\phi_t dx}{\int H\phi_t dx}, \quad (6)$$

and similarly for μ_2 and σ_2 with $H\phi_t$ replaced by $(1 - H\phi_t)$. To keep the notation simple, we do not display these parameters as part of the dynamic variables.

- To avoid the computational burden of considering all possible intermediate shapes $\phi_{1:t-1}$ and transformations $\theta_{1:t-1}$ in equation (3), we assume the distributions of previous states to be strongly peaked around the maxima of the respective distributions:

$$\mathcal{P}(\phi_{1:t-1}, \theta_{1:t-1} | I_{1:t-1}) \approx \delta(\phi_{1:t-1} - \hat{\phi}_{1:t-1}) \delta(\theta_{1:t-1} - \hat{\theta}_{1:t-1}), \quad (7)$$

where $(\hat{\phi}_i, \hat{\theta}_i) = \arg \max \mathcal{P}(\phi_i, \theta_i | I_{1:i-1})$ are the estimates of shape and transformation obtained for the past frames, and $\delta(\cdot)$ denotes the Dirac delta function. An alternative justification for this approximation is the following: Assume that due to memory limitations, the tracking system cannot store the acquired images, but that it merely stores the past estimates of shape and transformation. Then the inference

problem at time t reduces to that of maximizing the conditional distribution

$$\mathcal{P}(\phi_t, \theta_t | I_t, \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1}) \propto \mathcal{P}(I_t | \phi_t, \theta_t) \mathcal{P}(\phi_t, \theta_t | \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1}) \quad (8)$$

with respect to the embedding function ϕ_t and the transformation parameters θ_t . This is equivalent to the original inference problem (2) subject to the approximation (7).

- A central contribution of this paper is to model the joint prior on shape ϕ_t and transformation θ_t conditioned on previous shapes and transformations. To this end, we will consider two approximations:

In a first step, we will assume that shape and transformation are mutually independent, i.e. $\mathcal{P}(\phi_t, \theta_t | \phi_{1:t-1}, \theta_{1:t-1}) = \mathcal{P}(\phi_t | \phi_{1:t-1}) \mathcal{P}(\theta_t | \theta_{1:t-1})$, and assume a uniform prior on the transformation parameters, i.e. $\mathcal{P}(\theta_t | \theta_{1:t-1}) = \text{const}$. This is complimentary to the recent work of Rathi et al. [24] who proposed a temporal model for these transformation parameters, while not imposing any specific model on the shape.

In a second step, we will then consider the more general case of a joint distribution $\mathcal{P}(\phi_t, \theta_t | \phi_{1:t-1}, \theta_{1:t-1})$ of shape and transformation parameters, taking into account couplings between shape and transformation. Experimental results demonstrate that this leads to superior performance in dealing with occlusions.

B. Finite-dimensional Formulation

When estimating the conditional probability $\mathcal{P}(\phi_t, \theta_t | \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1})$ in (8) from sample data, one needs to revert to finite-dimensional approximations of the embedding function. It is well-known that statistical models can be estimated more reliably if the dimensionality of the model and the data are low. We will therefore recast the Bayesian inference in a low-dimensional formulation within the subspace spanned by the largest principal eigenmodes of a set of sample shapes. We therefore exploit the training sequence in a twofold way: Firstly, it serves to define a low-dimensional subspace in which to perform estimation. And secondly, within this subspace we use it to learn dynamical models for implicit shapes.



Fig. 1. Low-dimensional approximation of a set of training silhouettes. The silhouettes (above) are approximated by the first 6 principal components of their embedding functions (below) – see equation (9).

Let $\{\phi_1, \dots, \phi_N\}$ be a temporal sequence of training shapes.³ Let ϕ_0 denote the mean shape and ψ_1, \dots, ψ_n the n largest eigenmodes with $n \ll N$. We will then approximate each training shape as:

$$\phi_i(x) = \phi_0(x) + \sum_{j=1}^n \alpha_{ij} \psi_j(x), \quad (9)$$

where

$$\alpha_{ij} = \langle \phi_i - \phi_0, \psi_j \rangle \equiv \int (\phi_i - \phi_0) \psi_j dx. \quad (10)$$

Such PCA based representations of level set functions have been successfully applied for the construction of statistical shape priors in [15], [30], [28], [26]. In the following, we will denote the vector of the first n eigenmodes as $\psi = (\psi_1, \dots, \psi_n)$. Each sample shape ϕ_i is therefore approximated by the n -dimensional shape vector $\alpha_i = (\alpha_{i1}, \dots, \alpha_{in})$. Similarly, an arbitrary shape ϕ can be approximated by a shape vector of the form

$$\alpha_\phi = \langle \phi - \phi_0, \psi \rangle. \quad (11)$$

Figure 1 shows a set of silhouettes from a sequence of a walking person and their approximation by the first 6 eigenmodes. While this approximation is certainly a rough approximation lacking some of the details of the shape, we found it sufficiently accurate for our purpose.

³We assume that all training shapes ϕ_i are signed distance functions. Yet an arbitrary linear combination of eigenmodes will in general not generate a signed distance function. While the proposed statistical shape models favor shapes which are close to the training shapes (and therefore close to the set of signed distance functions), not all shapes sampled in the considered subspace will correspond to signed distance functions.

In analogy to the derivation presented in the previous section, the goal of image sequence segmentation within this subspace can then be stated as follows: Given consecutive images $I_t : \Omega \rightarrow \mathbb{R}$ from an image sequence, and given the segmentations $\hat{\alpha}_{1:t-1}$ and transformations $\hat{\theta}_{1:t-1}$ obtained on the previous images $I_{1:t-1}$, we need to maximize the conditional probability

$$\begin{aligned} & \mathcal{P}(\alpha_t, \theta_t | I_t, \hat{\alpha}_{1:t-1}, \hat{\theta}_{1:t-1}) \\ &= \frac{\mathcal{P}(I_t | \alpha_t, \theta_t) \mathcal{P}(\alpha_t, \theta_t | \hat{\alpha}_{1:t-1}, \hat{\theta}_{1:t-1})}{\mathcal{P}(I_t | \hat{\alpha}_{1:t-1}, \hat{\theta}_{1:t-1})}, \end{aligned} \quad (12)$$

with respect to the shape parameters α_t and the transformation parameters θ_t . The key contribution of this work, is to model the conditional probability

$$\mathcal{P}(\alpha_t, \theta_t | \hat{\alpha}_{1:t-1}, \hat{\theta}_{1:t-1}), \quad (13)$$

which constitutes the probability for observing a particular shape α_t and a particular transformation θ_t at time t , conditioned on the parameter estimates for shape and transformation obtained on previous images.

III. DYNAMICAL STATISTICAL SHAPE MODELS

Abundant theory has been developed to model temporally correlated time series data. Applications of dynamical systems to model deformable shapes were proposed among others in [3]. In our context, we intend to learn dynamical models for implicitly represented shapes. To simplify the presentation, we will first focus on dynamical models of shape *deformation*. In other words, we will assume a uniform distribution on the transformation parameters and merely model the conditional distribution $\mathcal{P}(\alpha_t | \hat{\alpha}_{1:t-1})$.

A. Dynamical Models of Deformation

In the following, we propose to learn the temporal dynamics of a deforming shape by approximating the shape vectors $\alpha_t \equiv \alpha_{\phi_t}$ of a sequence of level set functions by a Markov chain (cf. [20]) of order k , i.e.:

$$\alpha_t = \mu + A_1 \hat{\alpha}_{t-1} + A_2 \hat{\alpha}_{t-2} + \dots + A_k \hat{\alpha}_{t-k} + \eta, \quad (14)$$

where η is zero-mean Gaussian noise with covariance Σ . The probability of a shape conditioned on the shapes observed in previous time steps is

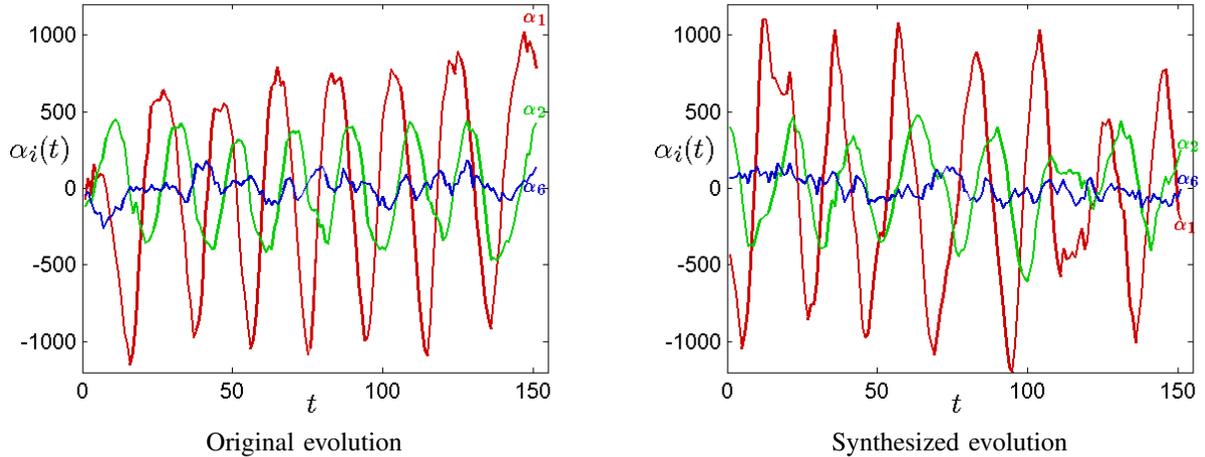


Fig. 2. Synthesized Shape Modes. The plots show the temporal evolution of the first, second and sixth shape eigenmode. The original shape sequence (**left**) and the sequence synthesized by a statistically learned second order Markov chain (**right**) exhibit similar oscillatory behavior and amplitude modulation.

therefore given by the corresponding autoregressive model of order k :

$$\mathcal{P}(\alpha_t | \hat{\alpha}_{1:t-1}) \propto \exp\left(-\frac{1}{2} \mathbf{v}^\top \Sigma^{-1} \mathbf{v}\right), \quad (15)$$

where

$$\mathbf{v} \equiv \alpha_t - \boldsymbol{\mu} - A_1 \hat{\alpha}_{t-1} - A_2 \hat{\alpha}_{t-2} \dots - A_k \hat{\alpha}_{t-k} \quad (16)$$

Various methods have been proposed in the literature to estimate the model parameters given by the mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and the transition and noise matrices $A_1, \dots, A_k, \Sigma \in \mathbb{R}^{n \times n}$. We applied a stepwise least squares algorithm proposed in [20]. Different tests have been devised to quantify the accuracy of the model fit. Two established criteria for model accuracy are Akaike's Final Prediction Error [1] and Schwarz's Bayesian Criterion [29]. Using dynamical models up to an order of 8, we found that according to Schwarz's Bayesian Criterion, our training sequences were best approximated by an autoregressive model of second order.

From a training sequence of 151 consecutive silhouettes, we estimated the parameters of a second order autoregressive model. We subsequently validated that the residuals of all shape modes are essentially uncorrelated by computing their autocorrelation functions. For the first two shape modes, the autocorrelation functions are plotted in Figure 3.

In addition, the estimated model parameters allow us to synthesize a walking sequence according to

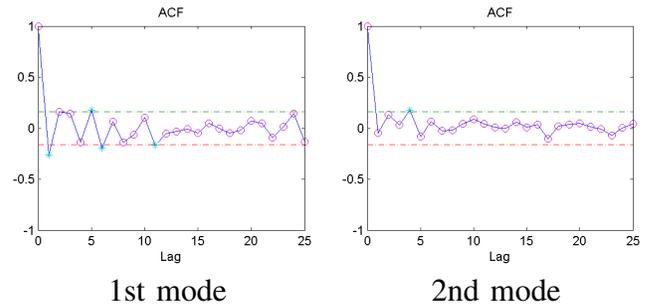


Fig. 3. Autocorrelation functions for the first two shape modes.

(14).⁴ Figure 2 shows the temporal evolution of the first, second and sixth eigenmode in the input sequence (left) and in the synthesized sequence (right). Clearly, the second order model captures some of the key elements of the oscillatory behavior.

While the synthesized sequence does capture the characteristic motion of a walking person, Figure 4 shows that the individual synthesized silhouettes do not in all instances mimic valid shapes. We believe that such limitations can be expected from a model which strongly compresses the represented input sequence: Instead of 151 shapes defined on a 256×256 grid, the model merely retains a mean shape ϕ_0 , 6 eigenmodes $\boldsymbol{\psi}$ and the autoregressive model parameters given by a 6-dimensional mean and three 6×6 matrices. This amounts to 458851 instead of 9895936 parameters, corresponding to a compression to 4.6% of the original size.

⁴To remove the dependency on the initial conditions, the first several hundred samples were discarded.

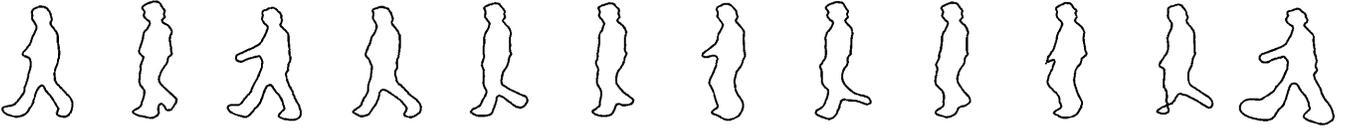


Fig. 4. Synthetically generated walking sequence. Sample silhouettes generated by a statistically learned second order Markov model on the embedding functions – see equation (14) and Figure 5. While the Markov model captures much of the typical oscillatory behavior of a walking person, not all generated samples correspond to permissible shapes – cf. the last two silhouettes. Yet, as we shall see in Section V, the model is sufficiently accurate to appropriately constrain a segmentation process.

While the synthesis of dynamical shape models using autoregressive models has been studied before (cf. [3]), we want to stress the fact that in this work we are synthesizing shapes based on an *implicit* representation. To further clarify this key idea of our paper, we show in Figure 5 a sequence of statistically synthesized embedding functions and the induced contours given by the zero level line of the respective surfaces. In particular, this implicit representation allows to synthesize shapes of varying topology. The silhouette on the bottom left of Figure 5, for example, consists of two contours.

B. Dynamics of Deformation and Transformation

In the previous section, we introduced autoregressive models to capture the temporal dynamics of implicitly represented shapes. To this end, we had removed the degrees of freedom corresponding to transformations such as translation and rotation before performing the learning of dynamical models. As a consequence, the learning only incorporates deformation modes, neglecting all information about pose and location. The synthesized shapes in Figure 4, for example, show a walking person which is walking “on the spot”.

In general, one can expect the deformation parameters α_t and the transformation parameters θ_t to be tightly coupled. A model which captures the joint dynamics of shape and transformation would clearly be more powerful than one which neglects the transformations. Yet, we want to learn dynamical shape models which are invariant to translation, rotation and other transformations. To this end, we can make use of the fact that the transformations form a group which implies that the transformation θ_t at time t can be obtained from the previous transformation θ_{t-1} by applying an incremental transformation $\Delta\theta_t$: $T_{\theta_t}x = T_{\Delta\theta_t}T_{\theta_{t-1}}x$. Instead of learning models of the absolute transformation θ_t , we can simply learn models of the update transformations $\Delta\theta_t$

(e.g. the changes in translation and rotation). By construction, such models are invariant with respect to the global pose or location of the modeled shape.

To jointly model transformation and deformation, we simply obtain for each training shape in the learning sequence the deformation parameters α_i and the transformation changes $\Delta\theta_i$, and fit the autoregressive models given in equations (15) and (16) to the combined vector

$$\tilde{\alpha}_t = \begin{pmatrix} \alpha_t \\ \Delta\theta_t \end{pmatrix}. \quad (17)$$

In the case of the walking person, we found that – as in the stationary case – a second order autoregressive model gives the best model fit. Synthesizing from this model allows to generate silhouettes of a walking person which are similar to the ones shown in Figure 4, but which move forward in space, starting from an arbitrary (user-specified) initial position.

IV. DYNAMICAL SHAPE PRIORS IN VARIATIONAL SEQUENCE SEGMENTATION

Given an image I_t from an image sequence and given a set of previously segmented shapes with shape parameters $\alpha_{1:t-1}$ and transformation parameters $\theta_{1:t-1}$, the goal of tracking is to maximize the conditional probability (12) with respect to shape α_t and transformation θ_t . This can be performed by minimizing its negative logarithm, which is – up to a constant – given by an energy of the form:

$$E(\alpha_t, \theta_t) = E_{data}(\alpha_t, \theta_t) + \nu E_{shape}(\alpha_t, \theta_t). \quad (18)$$

The additional weight ν was introduced to allow a relative weighting between prior and data term. In particular if the intensity information is not consistent with the assumptions (Gaussian intensity distributions of object and background), a larger

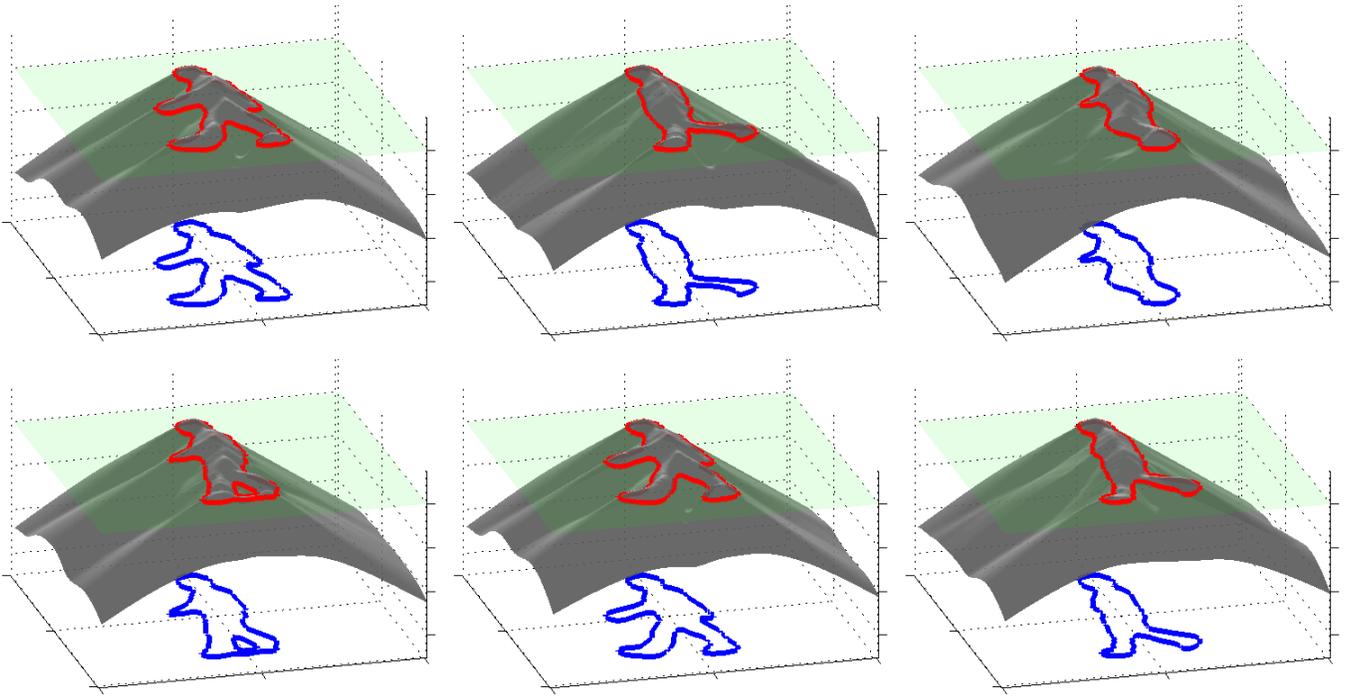


Fig. 5. Synthesis of implicit dynamical shapes. Statistically generated embedding surfaces obtained by sampling from a second order autoregressive model, and the contours given by the zero level lines of the synthesized surfaces. The implicit formulation allows the embedded contour to change topology (bottom left image).

weight of ν is preferable. Following (5), the data term is given by:

$$E_{data}(\alpha_t, \theta_t) = \int \left(\frac{(I_t - \mu_1)^2}{2\sigma_1^2} + \log \sigma_1 \right) H \phi_{\alpha_t, \theta_t} dx + \int \left(\frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \sigma_2 \right) (1 - H \phi_{\alpha_t, \theta_t}) dx, \quad (19)$$

where, for notational simplicity, we have introduced the expression

$$\phi_{\alpha_t, \theta_t} \equiv \phi_0(T_{\theta_t} x) + \alpha_t^\top \psi(T_{\theta_t} x), \quad (20)$$

to denote the embedding function of a shape generated with deformation parameters α_t and transformed with parameters θ_t .

Using the autoregressive model (15), the shape energy is given by:

$$E_{shape}(\alpha_t, \theta_t) = \frac{1}{2} \mathbf{v}^\top \Sigma^{-1} \mathbf{v} \quad (21)$$

with \mathbf{v} defined in (16). To incorporate the joint model of deformation and transformation introduced in Section III-B, the above expression for \mathbf{v} needs to be enhanced by the relative transformations $\Delta\theta$:

$$\mathbf{v} \equiv \begin{pmatrix} \alpha_t \\ \Delta\theta_t \end{pmatrix} - \boldsymbol{\mu} - \sum_{i=1}^k A_i \begin{pmatrix} \hat{\alpha}_{t-i} \\ \Delta\hat{\theta}_{t-i} \end{pmatrix}, \quad (22)$$

where $\boldsymbol{\mu}$ and A_i denote the statistically learned mean and transition matrices for the joint space of deformations and transformations, and k is the model order. In our experiments, we chose a model order of $k = 2$. One can easily show that a second order autoregressive model can be interpreted as a stochastic version of a time-discrete damped harmonic oscillator. As a consequence, it is well-suited to model essentially oscillatory shape deformations. However, we found that higher-order autoregressive models provide qualitatively similar results.

Tracking an object of interest over a sequence of images $I_{1:t}$ with a dynamical shape prior can be done by minimizing energy (18). In this work, we pursue a gradient descent strategy, leading to the following differential equations to estimate the shape vector α_t :

$$\frac{d\alpha_t(\tau)}{d\tau} = - \frac{\partial E_{data}(\alpha_t, \theta_t)}{\partial \alpha_t} - \nu \frac{\partial E_{shape}(\alpha_t, \theta_t)}{\partial \alpha_t}$$

where τ denotes the artificial evolution time, as opposed to the physical time t . The data term is given by:

$$\frac{\partial E_{data}}{\partial \alpha_t} = \int \psi(x) \eta(x) dx,$$

with the abbreviation

$$\eta(x) \equiv \delta(\phi_{\alpha_t}) \left(\frac{(I_t - \mu_1)^2}{2\sigma_1^2} - \frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \frac{\sigma_1}{\sigma_2} \right),$$

and the shape term is given by:

$$\frac{\partial E_{shape}}{\partial \alpha_t} = \frac{\partial \mathbf{v}}{\partial \alpha_t} \Sigma^{-1} \mathbf{v} = \begin{pmatrix} 1_n & 0 \\ 0 & 0 \end{pmatrix} \Sigma^{-1} \mathbf{v}, \quad (23)$$

with \mathbf{v} given in (22) and 1_n being the n -dim. unit matrix modeling the projection on the shape components of \mathbf{v} , where n is the number of shape modes. These two terms affect the shape evolution in the following manner: The first term draws the shape to separate the image intensities according to the two Gaussian intensity models. Since variations in the shape vector α_t affect the shape through the eigenmodes ψ , the data term is a projection onto these eigenmodes. The second term induces a relaxation of the shape vector α_t toward the most likely shape, as predicted by the dynamical model based on the shape vectors and transformation parameters obtained for previous time frames.

Similarly, minimization with respect to the transformation parameters θ_t is obtained by evolving the respective gradient descent equation given by:

$$\frac{d\theta_t(\tau)}{d\tau} = - \frac{\partial E_{data}(\alpha_t, \theta_t)}{\partial \theta_t} - \nu \frac{\partial E_{shape}(\alpha_t, \theta_t)}{\partial \theta_t} \quad (24)$$

where the data term is given by

$$\frac{\partial E_{data}(\alpha_t, \theta_t)}{\partial \theta_t} = \int \nabla \psi(x) \frac{d(T_{\theta_t} x)}{d\theta_t} \eta(x) dx \quad (25)$$

and the driving term from the prior is given by:

$$\frac{\partial E_{shape}}{\partial \theta_t} = \frac{\partial \mathbf{v}}{\partial \theta_t} \Sigma^{-1} \mathbf{v} = \frac{d(\Delta \theta_t)}{d\theta_t} \begin{pmatrix} 0 & 0 \\ 0 & 1_s \end{pmatrix} \Sigma^{-1} \mathbf{v}, \quad (26)$$

where, as above, the shape prior contributes a driving force toward the most likely transformation predicted by the dynamical model. The block diagonal matrix in (26) simply models the projection onto the s transformation components of the joint vector \mathbf{v} defined in (22).

V. EXPERIMENTAL RESULTS

A. Dynamical versus Static Statistical Shape Priors

In the following, we will apply the dynamical statistical shape prior introduced above for the purpose of level set based tracking.

To construct the shape prior, we hand-segmented a sequence of a walking person, centered and binarized each shape. Subsequently, we determined the set of signed distance functions $\{\phi_i\}_{i=1..N}$ associated with each shape and computed the dominant 6 eigenmodes. Projecting each training shape onto these eigenmodes, we obtained a sequence of shape vectors $\{\alpha_i \in \mathbb{R}^6\}_{i=1..N}$. We fitted a second order multivariate autoregressive model to this sequence by computing the mean vector $\boldsymbol{\mu}$, the transition matrices $A_1, A_2 \in \mathbb{R}^{6 \times 6}$ and the noise covariance $\Sigma \in \mathbb{R}^{6 \times 6}$ shown in equation (15). Subsequently, we compared segmentations of noisy sequences obtained by segmentation in the 6-dimensional subspace without and with the dynamical statistical shape prior. The segmentation without dynamical prior corresponds to that obtained with a uniform prior in the subspace of the first few eigenmodes, as proposed in [30]. While there exist alternative models for static shape priors (for example the Gaussian model [15] or non-parametric statistical models [9], [26]), we found that all of these exhibit a qualitatively similar limitation when applied to image sequence segmentation (see Figure 8): they tend to get stuck in local minima because they do not exploit temporal shape correlations.

Figure 6 shows a sample input frame from a sequence with 25%, 50%, 75%, and 90% noise.⁵ Figure 7 shows a set of segmentations obtained with a uniform static shape prior on a sequence with 25% noise. While this segmentation without dynamical prior is successful in the presence of moderate noise, Figure 8 shows that it eventually breaks down when the noise level is increased. Since static shape priors do not provide for predictions in time, they have a tendency of getting stuck to the shape estimate obtained on the previous image.

Figure 9 shows segmentations of the same sequence as in 8 obtained with a dynamical statistical shape prior derived from a second order autoregressive model. Figures 10 and 11 show that the statistical shape prior provides for good segmentations, even with 75% or 90% noise. Clearly, exploiting the temporal statistics of dynamical shapes allows to make the segmentation process very robust to missing and misleading information.

⁵90% noise means that 90% of all pixels were replaced by a random intensity sampled from a uniform distribution. Note that our algorithm easily handles uniform noise, although its probabilistic formulation is based on the assumption of Gaussian noise.

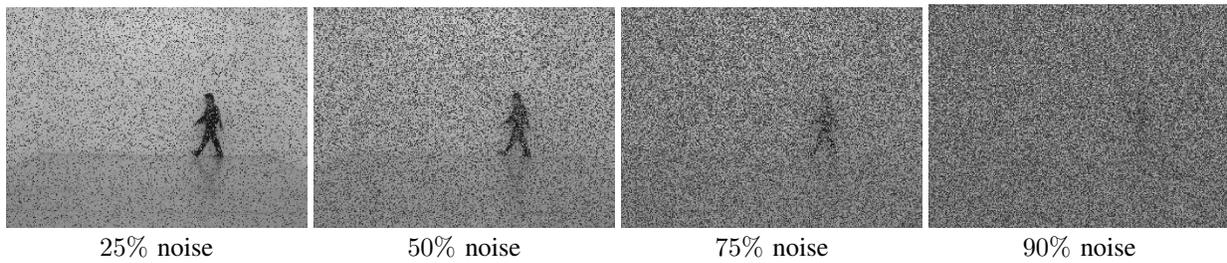


Fig. 6. Images from a sequence with increasing amount of noise.⁵

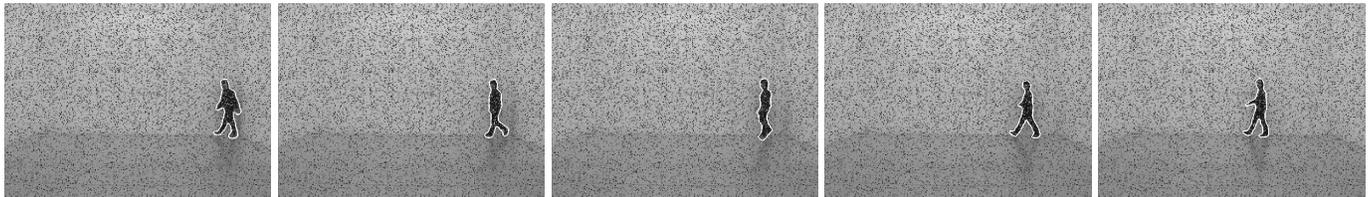


Fig. 7. Static shape prior at 25% noise: Constraining the level set evolution to a low-dimensional subspace allows to cope with some noise.

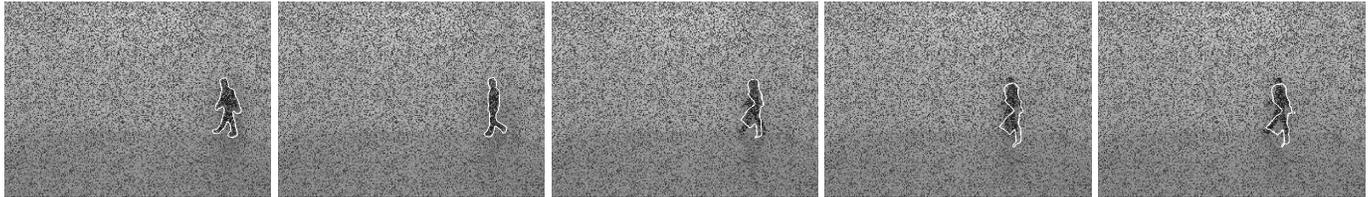


Fig. 8. Static shape prior at 50% noise: The segmentation with a static prior gets stuck in a local minimum after the first few frames.



Fig. 9. Dynamical shape prior at 50% noise: In contrast to the segmentation with a static prior shown in Figure 8, the dynamical prior (using a second-order autoregressive model) imposes statistically learned information about the temporal dynamics of the shape evolution to cope with missing or misleading low-level information.

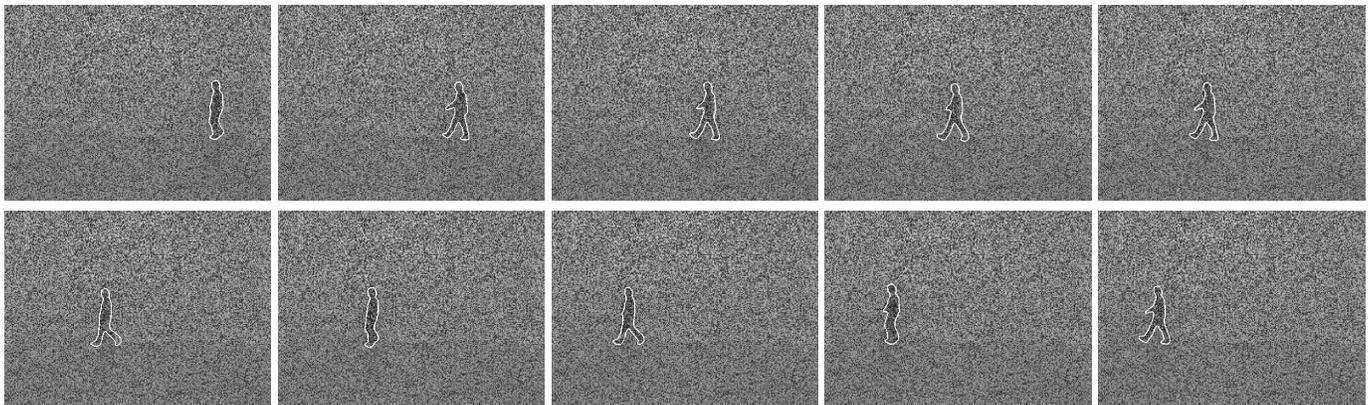


Fig. 10. Dynamical shape prior at 75% of noise: The statistically learned dynamical model allows to disambiguate the low-level information.

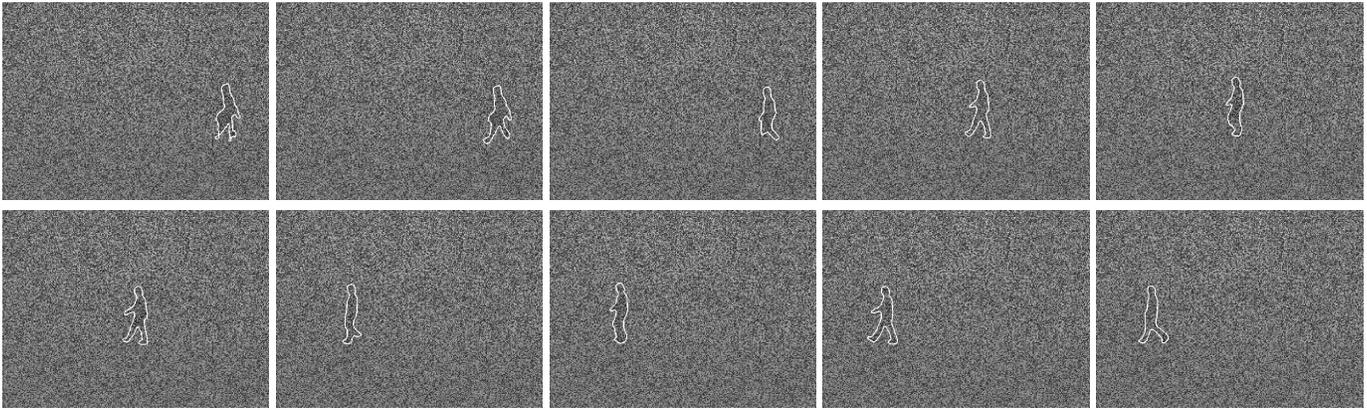


Fig. 11. Dynamical shape prior at 90% noise. Quantitative comparison with the ground truth, shown in Figure 12, left side, indicates that our tracking scheme can compete with the capacities of human observers, providing reliable segmentations where human observers fail. The segmentation of the first three or four frames is inaccurate, since the segmentation process accumulates image information over time.

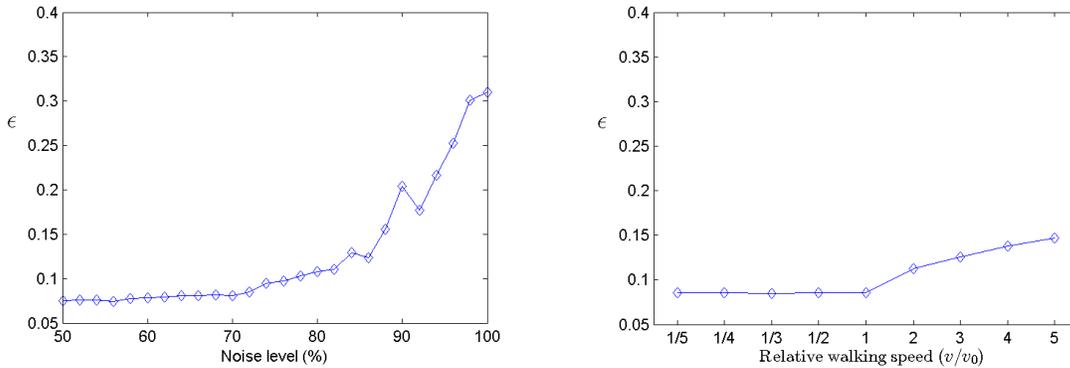


Fig. 12. Quantitative evaluation of the segmentation accuracy. The relative segmentation error is plotted for increasing amounts of noise (**left**) and for varying walking speed (**right**). Even for 100% noise the segmentation error remains below 1 because the process integrates a good estimate of the initial position and a model of the translational motion. The plot on the right shows that for walking speeds v slower than the learned one v_0 , the segmentation error (with 70% noise) remains low, whereas for faster walking sequences, the accuracy slowly degrades. Yet, even for sequences of 5 times the learned speed, the dynamical shape prior outperforms the static one.

B. Quantitative Evaluation of the Noise Robustness

In order to quantify the accuracy of segmentation, we hand segmented the original test sequence. Subsequently, we defined the following error measure:

$$\epsilon = \frac{\int (H\phi(x) - H\phi_0(x))^2 dx}{\int H\phi(x) dx + \int H\phi_0(x) dx}, \quad (27)$$

where H is again the Heaviside step function, ϕ_0 is the true segmentation and ϕ the estimated one. This error corresponds to the relative area of the set symmetric difference, i.e. the union of both shapes minus its intersection, divided by the sum of the areas. We decided for this measure, because it takes on values within the range $0 \leq \epsilon \leq 1$, where $\epsilon = 0$ corresponds to the perfect segmentation.

Figure 12, left side, shows the segmentation error averaged over a test sequence as a function of

the noise level. We used a dynamical shape prior of deformation and transformation (Section III-B), initializing the segmentation process with a an estimate of the initial location. The plot shows several things: Firstly, the error remains fairly constant for noise levels below 60%. The residual error of around 5% can be ascribed to the discrepancy between the estimated dynamical model and the true sequence, accumulating errors introduced by the principal component approximation and the approximation by autoregressive models. Secondly, as can be expected, the error increases for larger values of noise. The deviation from monotonicity (in particular at 90% noise) is probably an effect of statistical fluctuation.

C. Robustness to Frequency Variation

Assume that we have learned a dynamical model of a walking person from a sequence of a fixed walking speed v_0 . Clearly, the estimated model will be tuned to this specific walking speed. Yet, we cannot guarantee that the person in the test sequence will be walking at exactly the same speed. Equivalently, we may not be sure – even if the walking speed is identical – that the camera frame rate is the same. In order to be practically useful, the proposed prior must be robust to variations in the walking frequency and frame rate.

To validate this robustness, we synthetically generated test sequences of different walking speed by either dropping certain frames (in order to speed up the gait) or by replicating frames (thereby slowing down the gait). Figure 12, right side, shows the segmentation error ϵ , defined in (27), averaged over test sequences with 70% noise and speeds which vary from $1/5$ the speed of the training sequence to 5 times the original speed. While the accuracy is not affected by slowing down the sequence, it degrades gradually once the speed is increased. Yet, the segmentation process is quite robust to such drastic changes in speed. The reason for this robustness is twofold: Firstly, the Bayesian formulation allows to combine model prediction and input data in a way that the segmentation process constantly adapts to the incoming input data. Secondly, the autoregressive model only relies on the last few estimated silhouettes to generate a shape probability for the current frame. It does not assume long range temporal consistency and can thus handle sequences with varying walking speed. Our experiments show that even for sequences of 5 times the original walking sequence segmentation with a dynamical model is superior to segmentation with a static model. This is not surprising: In contrast to the static model, the dynamical model does provide a prediction of the temporal shape evolution. Even if this prediction is suboptimal for strongly differing walking speeds, it still allows to enhance the segmentation process.

D. Dynamics of Deformation and Transformation

In Section III-B, we introduced dynamical models to capture the joint evolution of deformation and transformation parameters. On the tasks we have shown so far, we found pure deformation models and joint models of deformation and transformation

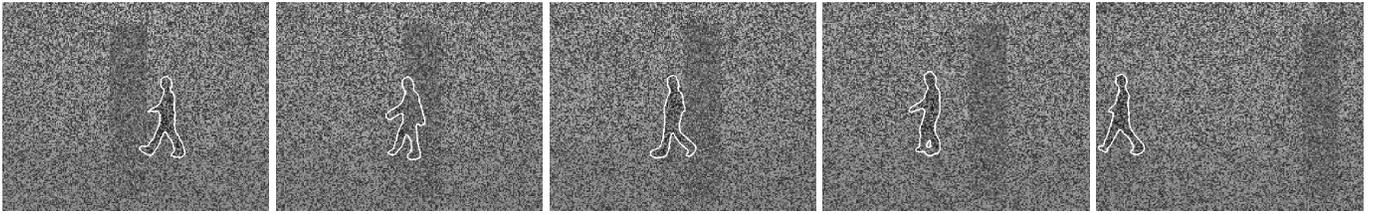
to provide similar segmentation results. While the joint model provides a prior about the transformation parameters which are most likely at a given time instance, the pure-deformation model requires these parameters to be estimated solely from the data.

As a final example, we generated a segmentation task where the transformation parameters cannot be reliably estimated from the data due to a prominent occlusion. The test sequence shows a person walking from right to left and an occluding bar moving from left to right, corrupted by 80% noise. Figure 13, top row, shows segmentations obtained with a dynamical shape prior capturing both deformation and transformation. Even when the walking silhouette is completely occluded, the model is capable of generating silhouettes walking to the left and adapts to the image data, once the figure reappears.

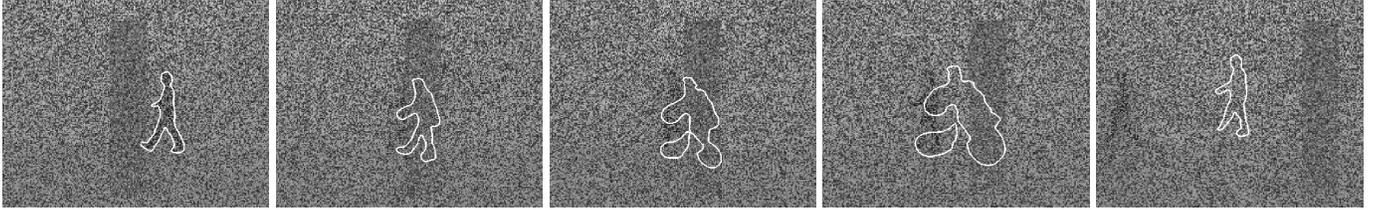
The bottom row of Figure 13, on the other hand, shows the segmentation of the same frames with a dynamical model which only incorporates the shape deformation. Since no knowledge about translation is assumed, the segmentation process needs to rely entirely on the image information in order to estimate the transformation parameters. As a consequence, the segmentation process is misled by the prominent occlusion. When the figure reappears from behind the bar, the process integrates contradictory information about translation provided by the person walking to the left and by the bar moving to the right. Once the figure of interest is lost, the prior simply “hallucinates” silhouettes of a person walking “on the spot” — see the last image on the bottom right. Although a “failed” experiment, we believe that this result illuminates best how the dynamical model and the image information are fused within the Bayesian formulation for image sequence segmentation.

VI. CONCLUSION

In this work, we introduced dynamical statistical shape models for implicitly represented shapes. In contrast to existing shape models for implicit shapes, these models capture the temporal correlations which characterize deforming shapes such as the consecutive silhouettes of a walking person. Such *dynamical* shape models account for the fact that the probability of observing a particular shape at a given time instance may depend on the shapes observed at previous time instances.



Segmentation with a dynamical prior of joint deformation and transformation.



Segmentation with a dynamical prior on the deformation only.

Fig. 13. Tracking in the presence of occlusion. The input sequence shows a person walking to the left occluded by a bar moving to the right. While the top row is generated with a dynamical prior integrating both deformation and transformation, the bottom row uses a dynamical prior which merely captures the deformation component. Since the latter does not provide predictions of the translational motion, the estimation of translation is purely based on the image data. It is misled by the occlusion and cannot recover, once the person reappears from behind the bar.

For the construction of statistical shape models, we extended the concepts of Markov chains and autoregressive models to the domain of implicitly represented shapes. The resulting dynamical shape models therefore allow to handle shapes of varying topology. Moreover, they are easily extended to higher-dimensional shapes (i.e. surfaces).

The estimated dynamical models allow to synthesize shape sequences of arbitrary length. For the case of a walking person, we validated the accuracy of the estimated dynamical models, comparing the dynamical shape evolution of the input sequence to that of synthesized sequences for various shape eigenmodes, and verifying that the residuals are statistically uncorrelated. Although the synthesized shapes do not in all instances correspond to valid shapes, one can nevertheless use the dynamical model to constrain a segmentation and tracking process in such a way that it favors familiar shape evolutions.

To this end, we developed a Bayesian formulation for level set based image sequence segmentation, which allows to impose the statistically learned dynamical models as a shape prior for segmentation processes. In contrast to most existing approaches to tracking, autoregressive models are integrated as statistical priors in a *variational approach* which can be minimized by local gradient descent (rather than by stochastic optimization methods).

Experimental results confirm that the dynamical

shape priors outperform static shape priors when tracking a walking person in the presence of large amounts of noise. We provided quantitative evaluation of the segmentation accuracy as a function of noise. Moreover, we validated that the model-based segmentation process is quite robust to large (up to a factor of 5) variations in frame rate and walking speed. Finally, we showed that a dynamical prior in the *joint* space of deformation and transformation outperforms a purely deformation-based prior, when tracking a walking person through prominent occlusions. Future research is focused on developing nonlinear dynamical models for implicit shapes in order to account for more complex shape deformations.

ACKNOWLEDGMENTS

We thank Gareth Funka-Lea, Paolo Favaro, Gianfranco Doretto, René Vidal, and Mikaël Rousson for fruitful discussions. We thank Alessandro Bissacco and Payam Saisan for providing the image sequence data used in the experiments. This research was supported by the German National Science Foundation (DFG), grant #CR-250/1-1.

REFERENCES

- [1] H. Akaike. Autoregressive model fitting for control. *Ann. Inst. Statist. Math.*, 23:163–180, 1971.
- [2] A. Blake, B. Bascle, M. Isard, and J. MacCormick. Statistical models of visual shape and motion. *Philos. Trans. Roy. Soc. London, A*, 356:1283–1302, 1998.

- [3] A. Blake and M. Isard. *Active Contours*. Springer, London, 1998.
- [4] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 694–699, Boston, USA, 1995.
- [5] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
- [6] Y. Chen, H. Tagare, S. Thiruvankadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. Geiser. Using shape priors in geometric active contours in a variational framework. *Int. J. of Computer Vision*, 50(3):315–328, 2002.
- [7] D. Cremers and G. Funka-Lea. Dynamical statistical shape priors for level set based tracking. In N. Paragios et al., editors, *Intl. Workshop on Variational and Level Set Methods*, volume 3752 of *Lect. Not. Comp. Sci.*, pages 210–221. Springer, 2005.
- [8] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford–Shah based segmentation. In A. Heyden et al., editors, *Europ. Conf. on Comp. Vis.*, volume 2351 of *Lect. Not. Comp. Sci.*, pages 93–108, Copenhagen, May 2002. Springer.
- [9] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int. J. of Computer Vision*, 2006. To appear.
- [10] D. Cremers, N. Sochen, and C. Schnörr. A multiphase dynamic labeling model for variational recognition-driven image segmentation. *Int. J. of Computer Vision*, 66(1):67–81, 2006.
- [11] A. Dervieux and F. Thomasset. A finite element method for the simulation of Raleigh–Taylor instability. *Springer Lect. Notes in Math.*, 771:145–158, 1979.
- [12] R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky. Behavior classification by eigendecomposition of periodic motion. *Pattern Recognition*, 38:1033–1043, July 2005.
- [13] S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *IEEE Int. Conf. on Computer Vision*, pages 810–815, 1995.
- [14] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 24(5):657–673, 2004.
- [15] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR*, volume 1, pages 316–323, Hilton Head Island, SC, 2000.
- [16] L. Ljung. *System Identification - Theory For the User*. Prentice Hall, Upper Saddle River, NJ, 1999.
- [17] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 17(2):158–175, 1995.
- [18] M. Moelich and T. Chan. Tracking objects with the Chan-Vese algorithm. Technical Report 03-14, Computational Applied Mathematics, UCLA, Los Angeles, 2003.
- [19] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
- [20] A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM T. on Mathematical Software*, 27(1):27–57, 2001.
- [21] S. J. Osher and J. A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49, 1988.
- [22] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1984.
- [23] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *Int. J. of Computer Vision*, 46(3):223–247, 2002.
- [24] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi. Particle filtering for geometric active contours and application to tracking deforming objects. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 2–9, 2005.
- [25] T. Riklin-Raviv, N. Kiryati, and N. Sochen. Unlevel sets: Geometry and prior-based segmentation. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3024 of *Lect. Not. Comp. Sci.*, pages 50–61, Prague, 2004. Springer.
- [26] M. Rousson and D. Cremers. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In *MICCAI*, volume 1, pages 757–764, 2005.
- [27] M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Europ. Conf. on Comp. Vis.*, volume 2351 of *Lect. Not. Comp. Sci.*, pages 78–92. Springer, 2002.
- [28] M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3d segmentation in MRI imaging. In *MICCAI*, pages 209–216, 2004.
- [29] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6:461–464, 1978.
- [30] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. Model-based curve evolution technique for image segmentation. In *Comp. Vision Patt. Recog.*, pages 463–468, Kauai, Hawaii, 2001.
- [31] A. Tsai, A. J. Yezzi, and A. S. Willsky. Curve evolution implementation of the Mumford–Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Trans. on Image Processing*, 10(8):1169–1186, 2001.
- [32] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996.



Daniel Cremers received the BS in Mathematics (1994) and Physics (1994), and an MS (Diplom) in Theoretical Physics (1997) from the University of Heidelberg. In 2002 he obtained a PhD in Computer Science from the University of Mannheim, Germany. Subsequently he spent two years as a postdoctoral researcher at the University of California at Los Angeles and one year as a permanent researcher at Siemens Corporate Research in Princeton, NJ. Since October 2005 he heads the Research Group for Computer Vision, Image Processing and Pattern Recognition at the University of Bonn, Germany. His research is focused on statistical and variational methods in computer vision. He received several awards, in particular the *Best Paper of the Year 2003* by the Pattern Recognition Society, the *Olympus Award 2004*, and the *UCLA Chancellor's Award for Postdoctoral Research 2005*.