

# Dynamical Statistical Shape Priors for Level Set Based Sequence Segmentation

Daniel Cremers and Gareth Funka-Lea

Department of Imaging and Visualization  
Siemens Corporate Research, Princeton, NJ

**Abstract.** In recent years, researchers have proposed to introduce statistical shape knowledge into the level set method in order to cope with insufficient low-level information. While these priors were shown to drastically improve the segmentation of images or image sequences, so far the focus has been on statistical shape priors that are time-invariant. Yet, in the context of tracking deformable objects, it is clear that certain silhouettes 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 into a segmentation process in a Bayesian framework for image sequence segmentation. Experiments demonstrate that such shape priors with memory can drastically improve the segmentation of image sequences.

## 1 Level Set Based Image Segmentation

In 1988, Osher and Sethian [16] introduced the level set method<sup>1</sup> as a means to implicitly propagate boundaries  $C(t)$  in the image plane  $\Omega \subset \mathbb{R}^2$  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 contour 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. First, the implicit contour representation does not depend on a specific parameterization and during the propagation no control point regriding mechanisms need to be introduced. Second, evolving the embedding function allows topological changes such as splitting and merging of the embedded contour to be elegantly modeled. In the context of shape modeling and statistical learning of shapes, the latter property allows for the construction of shape dissimilarity measures defined on the embedding functions which can handle shapes of varying topology. Third, the implicit representation (1) naturally generalizes to hypersurfaces in three or more dimensions. To impose a unique correspondence between a contour and its embedding function one can constrain  $\phi$  to be a signed distance function, i.e.  $|\nabla\phi| = 1$  almost everywhere.

---

<sup>1</sup> A precursor of the level set method was proposed by Dervieux and Thomasset [8].

Starting in the early 90’s researchers proposed to apply the level set method to image segmentation (cf. [12, 3, 10, 17]). Level set implementations of the Mumford-Shah functional [14] were independently proposed in [4, 24].

In recent years, researchers have successfully introduced prior shape information into level set based segmentation schemes [11, 25, 21, 5, 19, 7, 22, 20]. Statistically learned shape information was shown to cope for missing or misleading information in the input images due to noise, clutter and occlusion. These shape priors were developed to segment objects of familiar shape in a given image. Although they can be applied to tracking objects in image sequences [6, 13, 7], they are not well-suited for this task, because they neglect the temporal coherence of silhouettes which characterizes the motion of 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 will develop statistical shape models for which the shape probability at a given time will depend 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 segmentation of image sequences as follows.

## 2 Level Set Based Tracking as Bayesian Inference

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

### 2.1 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  $T_\theta$  with a parameter vector  $\theta$ . 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  $\phi$  according to equation (1). Thus objects of interest will be given by  $\phi(T_\theta x)$ , where the transformation  $T_\theta$  acts on the grid, leading to corresponding transformations of the implicitly represented contour. We purposely separate shape  $\phi$  and transformation parameters  $\theta$  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. Assume we have already segmented the images at previous times in terms of

embedding functions  $\hat{\phi}_{1:t-1}$  and transformation parameters  $\hat{\theta}_{1:t-1}$ . The problem of segmenting the current frame  $I_t$  can then be addressed in the framework of Bayesian inference by maximizing the conditional probability

$$\mathcal{P}(\phi_t, \theta_t | I_{1:t}, \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1}) = \frac{\mathcal{P}(I_{1:t} | \phi_t, \theta_t, \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1}) \mathcal{P}(\phi_t, \theta_t | \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1})}{\mathcal{P}(I_{1:t} | \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1})},$$

with respect to the embedding function  $\phi_t$  and the transformation parameters  $\theta_t$ .<sup>2</sup> The denominator in the above expression does not depend on the estimated quantities and can therefore be neglected in the maximization.

In order to further reduce the complexity of the estimation problem, we will make the following assumptions:

- The images  $I_{1:t}$  are mutually independent and their probability only depends on the current shape and transformation. Therefore, the first term in the numerator reduces to:

$$\mathcal{P}(I_{1:t} | \phi_t, \theta_t, \hat{\phi}_{1:t-1}, \hat{\theta}_{1:t-1}) = \prod_{i=1}^t \mathcal{P}(I_i | \phi_i, \theta_i) = \mathcal{P}(I_t | \phi_t, \theta_t) \cdot \text{const.} \quad (2)$$

- We assume that the intensities of the shape of interest and of the background are independent samples from two Gaussian distributions  $K_{\mu, \sigma}(I) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I-\mu)^2}{2\sigma^2}\right)$  with unknown means  $\mu_1, \mu_2$  and variances  $\sigma_1, \sigma_2$ . As a consequence, the data term can be written as:

$$\begin{aligned} \mathcal{P}(I_t | \phi_t, \theta_t) &= \prod_{\substack{x \\ \phi(T_{\theta_t} x) \geq 0}} K_{\mu_1, \sigma_1}(I_t(x)) \prod_{\substack{x \\ \phi(T_{\theta_t} x) < 0}} K_{\mu_2, \sigma_2}(I_t(x)) \\ &\propto \exp\left(-\int_{\Omega} \left(\frac{(I_t - \mu_1)^2}{2\sigma_1^2} + \log \sigma_1\right) H\phi_t(T_{\theta_t} x) \right. \\ &\quad \left. + \left(\frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \sigma_2\right) (1 - H\phi_t(T_{\theta_t} x)) dx\right), \end{aligned} \quad (3)$$

where we have introduced the Heaviside step function  $H\phi \equiv H(\phi)$  to denote the areas where  $\phi$  is positive ( $H\phi = 1$ ) or negative ( $H\phi = 0$ ). Related intensity models for segmentation have been proposed among others in [26, 4]. The intensity model parameters  $\mu_i$  and  $\sigma_i$  are estimated jointly with the shape  $\phi_t$  and the transformation  $\theta_t$ . Their optimal values are simply given by the means and variances of the intensity  $I_t$  inside and outside the current shape. To keep the notation simple, we do not display them as part of the dynamic variables.

<sup>2</sup> Since the modeling of probability distributions on infinite-dimensional spaces is in general an open problem including issues of defining appropriate measures and of integrability, the functions  $\phi$  in this paper may be thought of as finite-dimensional approximations obtained by sampling the embedding functions on a regular grid.

- The prior probability of the current shape and transformation are mutually independent and only depend on their previous estimates. The second term in the numerator therefore simplifies as follows:

$$\mathcal{P}(\phi_t, \theta_t | \phi_{1:t-1}, \theta_{1:t-1}) = \mathcal{P}(\phi_t | \phi_{1:t-1}) \mathcal{P}(\theta_t | \theta_{1:t-1})$$

By this assumption, we therefore neglect couplings between shape and transformation. Since the focus of the present paper is on modeling temporally correlated shape deformations, we will simply assume a uniform prior on the transformation parameters, i.e.  $\mathcal{P}(\theta_t | \theta_{1:t-1}) = \text{const}$ . Rathi et al. [18] recently proposed a temporal model of these transformation parameters while not imposing any specific model on the shape. In this sense, our work is complementary to theirs. In the following, we will develop appropriate models for the conditional probability  $\mathcal{P}(\phi_t | \phi_{1:t-1})$ .

## 2.2 A Finite-dimensional Formulation

When estimating the conditional probability  $\mathcal{P}(\phi_t | \phi_{1:t-1})$  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 data is low. We will therefore recast the Bayesian inference in a low-dimensional formulation given within the subspace spanned by the largest principal eigenmodes of a set of sample shapes.

Let  $\{\phi_1, \dots, \phi_N\}$  be a temporal sequence of training shapes.<sup>3</sup> Let  $\phi_0$  denote the mean shape and  $\psi_1, \dots, \psi_n$  the  $n$  largest eigenmodes with  $n \ll N$ . We will then approximate each training shape as:

$$\phi_i(x) = \phi_0(x) + \sum_{j=1}^n \alpha_{ij} \psi_j(x), \quad (4)$$

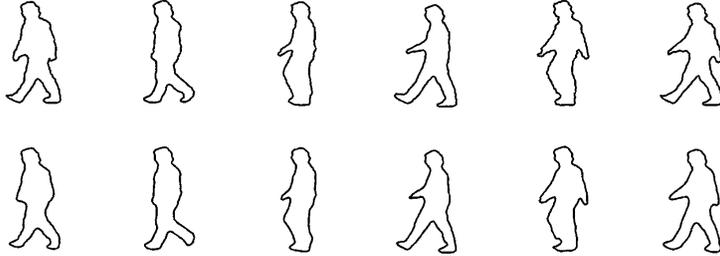
where

$$\alpha_{ij} = \langle \phi_i - \phi_0, \psi_j \rangle \equiv \int (\phi_i - \phi_0) \psi_j dx. \quad (5)$$

Such PCA based representations of level set functions have been successfully applied for the construction of statistical shape priors in [11, 24, 22]. In the following, we will denote the vector of the first  $n$  eigenmodes as  $\boldsymbol{\psi} = (\psi_1, \dots, \psi_n)$ .

<sup>3</sup> We assume that all training shapes  $\phi_i$  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 addition, it is quite possible that linear combinations result in empty shapes, i.e. the zero level set of a linear combination may be the empty set.

While level set based shape representations via harmonic embedding [9] do form a linear space, such representations are limited in practice, because not every shape can be represented by an appropriate harmonic function.



**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 (4).

Each sample shape  $\phi_i$  is therefore approximated by the  $n$ -dimensional shape vector  $\alpha_i = (\alpha_{i1}, \dots, \alpha_{in})$ . Similarly, an arbitrary shape  $\phi$  can be approximated by a shape vector of the form

$$\alpha_\phi = \langle \phi - \phi_0, \psi \rangle. \quad (6)$$

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.

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  $\theta_{1:t-1}$  obtained on the previous images  $I_{1:t-1}$ , we need to maximize the conditional probability

$$\mathcal{P}(\alpha_t, \theta_t | I_{1:t}, \{\hat{\alpha}, \hat{\theta}\}_{1:t-1}) \propto \mathcal{P}(I_{1:t} | \alpha_t, \theta_t, \{\hat{\alpha}, \hat{\theta}\}_{1:t-1}) \mathcal{P}(\alpha_t, \theta_t | \{\hat{\alpha}, \hat{\theta}\}_{1:t-1}), \quad (7)$$

with respect to the shape parameters  $\alpha_t$  and the transformation parameters  $\theta_t$ . One can introduce the same approximations as in the previous section. In all expressions the variables  $\phi_i$  are simply replaced by their shape vectors  $\alpha_i$ . Due to space limitations, we will not carry this out explicitly. The key contribution of this work, is to model the probability

$$\mathcal{P}(\alpha_t | \hat{\alpha}_{1:t-1}), \quad (8)$$

which constitutes the probability for observing a particular shape conditioned on the previously observed shapes.

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 [2]. In our context, we intend to learn dynamical models for the implicitly represented shapes.

### 3 Dynamical Statistical Shape Models

In the following, we propose to learn the temporal dynamics of a deforming shape by approximating the shape vectors  $\boldsymbol{\alpha}_t \equiv \boldsymbol{\alpha}_{\phi_t}$  of a sequence of silhouettes by a Markov chain (cf. [2, 15]) of order  $k$ , i.e.:

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu} + A_1\boldsymbol{\alpha}_{t-1} + A_2\boldsymbol{\alpha}_{t-2} + \dots + A_k\boldsymbol{\alpha}_{t-k} + \boldsymbol{\eta}. \quad (9)$$

The state at time  $t$  is therefore given by a linear combination of previous states, modeled by a mean  $\boldsymbol{\mu} \in \mathbb{R}^n$  and transition matrices  $A_1, \dots, A_k \in \mathbb{R}^{n \times n}$ , and zero-mean Gaussian noise  $\boldsymbol{\eta} \in \mathbb{R}^n$  with covariance  $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$  superimposed. The probability of a shape conditioned on the shapes observed in previous time steps is therefore given by the corresponding autoregressive model of order  $k$ :

$$\mathcal{P}(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{1:t-1}) \propto \exp\left(-\frac{1}{2} \mathbf{v}^\top \boldsymbol{\Sigma}^{-1} \mathbf{v}\right), \quad (10)$$

where

$$\mathbf{v} = \boldsymbol{\alpha}_t - \boldsymbol{\mu} - A_1\boldsymbol{\alpha}_{t-1} - A_2\boldsymbol{\alpha}_{t-2} \dots - A_k\boldsymbol{\alpha}_{t-k} \quad (11)$$

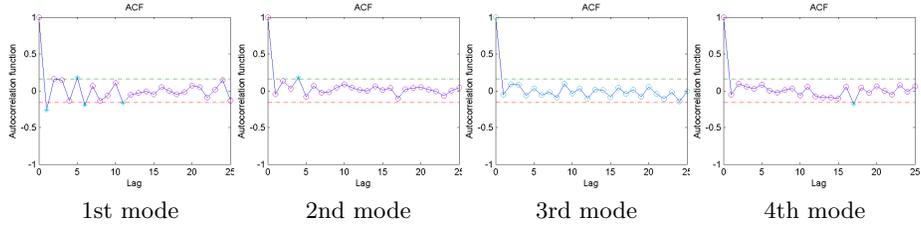
Various methods have been proposed in the literature to estimate the model parameters given by the mean  $\boldsymbol{\mu} \in \mathbb{R}^n$  and the matrices  $A_1, \dots, A_k, \boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ . We applied a stepwise least squares algorithm proposed in [15]. Different 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 [23]. 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 sequence of 151 consecutive silhouettes, we estimated the parameters of a second order autoregressive model. We subsequently validated this model by plotting the autocorrelation functions of the residuals associated with each of the modeled eigenmodes – see Figure 2. These show that the residuals are essentially uncorrelated.

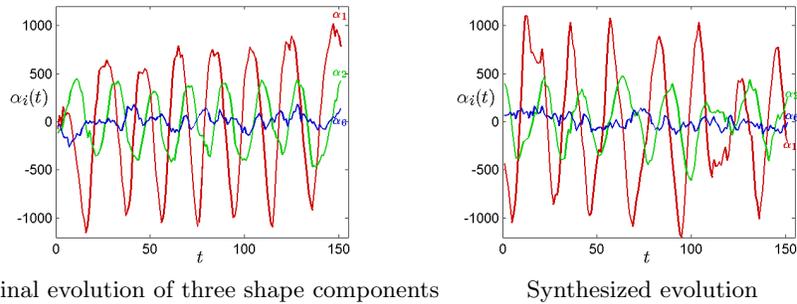
In addition, the estimated model parameters allow us to synthesize a walking sequence according to (9).<sup>4</sup> Figure 3 shows the temporal evolution of the first, second and sixth eigenmode in the input sequence (left) and in the synthesized sequence. Clearly, the second order model captures some of the key elements of the oscillatory behaviour.

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  $256 \times 256$  grid, the model merely retains a mean shape  $\phi_0$ , 6 eigenmodes  $\boldsymbol{\psi}$  and the autoregressive model parameters given by a 6-dimensional mean and three  $6 \times 6$  matrices. This amounts to 458851 instead of 9895936 parameters, corresponding to a compression to 4.6% of the original size.

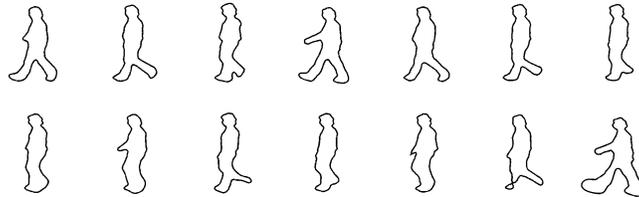
<sup>4</sup> In order to remove the dependency on the initial conditions, the first several hundred samples were discarded from the synthesized sequence.



**Fig. 2. Autocorrelation functions.** To validate the accuracy of the fitted autoregressive model, we plotted the autocorrelation functions of the residuals associated with the first four shape modes. Except for the first mode, more than 95% of autocorrelations (for lag > 0) lie within the confidence limits of an IID process.



**Fig. 3. Model comparison.** The original shape sequence (**top**) and the sequence synthesized by a statistically learned second order Markov chain (**bottom**) exhibit similar oscillatory behaviour and amplitude modulation. The plots show the temporal evolution of the first, second and sixth shape eigenmode.



**Fig. 4. Synthetically generated walking sequence.** Sample silhouettes generated by a statistically learned second order Markov model on the embedding functions – see equation (9). While the Markov model captures much of the typical oscillatory behaviour of a walking person, not all generated samples correspond to permissible shapes – cf. the last two silhouettes on the bottom right. Yet, as we shall see in Section 5, the model is sufficiently accurate to constrain the segmentation process in a meaningful way.

## 4 Dynamical Shape Priors in Variational Segmentation

Maximizing the conditional probability (7) under the assumptions introduced in Section 2 can be done by minimizing the negative logarithm of (7). Up to a constant, the latter is given by:

$$E(\boldsymbol{\alpha}_t, \theta_t) = E_{data}(\boldsymbol{\alpha}_t, \theta_t) + \nu E_{dynamics}(\boldsymbol{\alpha}_t). \quad (12)$$

According to equation (3), the data term is given by:

$$E_{data} = \int_{\Omega} \left( \frac{(I_t - \mu_1)^2}{2\sigma_1^2} + \log \sigma_1 \right) H\phi_{\boldsymbol{\alpha}_t, \theta_t} + \left( \frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \sigma_2 \right) (1 - H\phi_{\boldsymbol{\alpha}_t, \theta_t}) dx,$$

where, for notational simplicity, we have introduced the expression  $\phi_{\boldsymbol{\alpha}_t, \theta_t} \equiv \phi_0(T_{\theta_t}x) + \boldsymbol{\alpha}_t^\top \boldsymbol{\psi}(T_{\theta_t}x)$  to denote the embedding function of a shape generated with deformation parameters  $\boldsymbol{\alpha}_t$  and transformed with parameters  $\theta_t$ .

Using the autoregressive model (10), the dynamical shape energy is given by:

$$E_{dynamics}(\boldsymbol{\alpha}_t) = \frac{1}{2} \mathbf{v}^\top \Sigma^{-1} \mathbf{v} \quad (13)$$

with  $\mathbf{v}$  defined in (11).

Tracking an object of interest over a sequence of images  $I_{1:t}$  with a statistically learnt dynamical shape prior can be done by minimizing energy (12). In this work, we pursue a gradient descent strategy leading to the following differential equations to estimate the shape vector  $\boldsymbol{\alpha}_t$  and  $\theta_t$ :

$$\frac{d\boldsymbol{\alpha}_t(\tau)}{d\tau} = - \frac{\partial E_{data}(\boldsymbol{\alpha}_t, \theta_t)}{\partial \boldsymbol{\alpha}_t} - \nu \frac{dE_{dynamics}(\boldsymbol{\alpha}_t)}{d\boldsymbol{\alpha}_t} \quad (14)$$

where  $\tau$  denotes the artificial evolution time, as opposed to the physical time  $t$ . The first term is given by:

$$\frac{\partial E_{data}}{\partial \boldsymbol{\alpha}_t} = \left\langle \boldsymbol{\psi}, \delta(\phi_{\boldsymbol{\alpha}_t}) \left( \frac{(I_t - \mu_1)^2}{2\sigma_1^2} - \frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \frac{\sigma_1}{\sigma_2} \right) \right\rangle,$$

and the second one is given by:

$$\frac{dE_{dynamics}}{d\boldsymbol{\alpha}_t} = \Sigma^{-1} \mathbf{v}, \quad (15)$$

with  $\mathbf{v}$  given in (11). 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 the effect of variations in the shape vector  $\boldsymbol{\alpha}_t$  are given by the eigenmodes  $\boldsymbol{\psi}$ , the data term is a projection onto these eigenmodes. The second term induces a relaxation of the shape vector  $\boldsymbol{\alpha}_t$  toward the most likely shape, given the shapes obtained on previous time frames.

Minimization with respect to the transformation parameters  $\theta_t$  is obtained by evolving the respective gradient descent equation given by:

$$\frac{d\theta_t(\tau)}{d\tau} = - \frac{\partial E_{data}}{\partial \theta_t} = - \left\langle \nabla \boldsymbol{\psi} \frac{d(T_{\theta_t}x)}{d\theta_t}, \delta(\phi_{\boldsymbol{\alpha}_t}) \left( \frac{(I_t - \mu_1)^2}{2\sigma_1^2} - \frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \frac{\sigma_1}{\sigma_2} \right) \right\rangle.$$

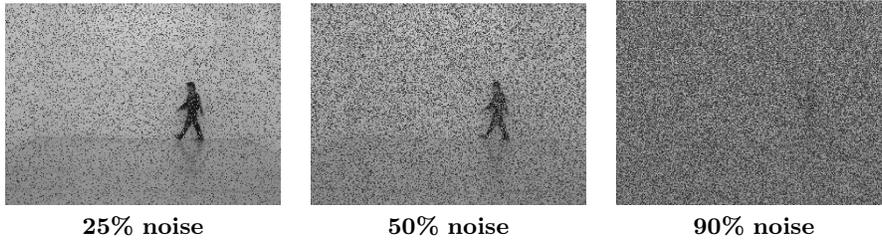


Fig. 5. Images from a sequence with increasing amounts of noise.<sup>5</sup>



Fig. 6. Sample segmentations with a static shape prior on a walking sequence with 25% noise. Constraining the level set evolution to a low-dimensional subspace allows to cope with a certain amount of noise.



Fig. 7. Sample segmentations with a static shape prior on a walking sequence with 50% noise. Using merely a static shape prior, the segmentation scheme cannot cope with larger amounts of noise.

## 5 Segmentation and Tracking Results

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

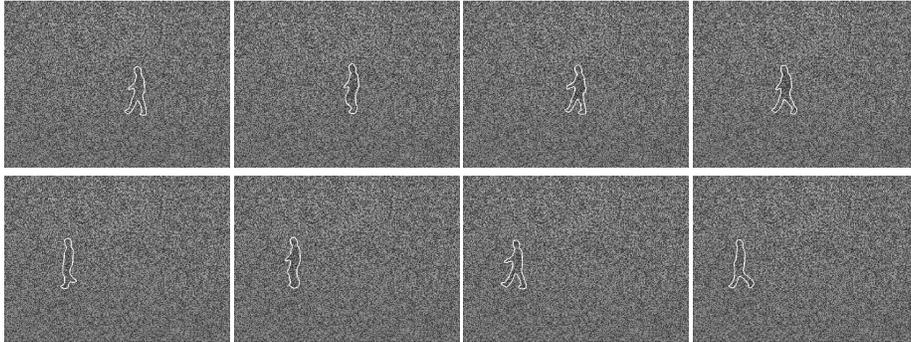
To construct the shape prior, we hand-segmented a sequence of a walking person, centered and binarized each shape. Subsequently, we determined the set of signed distance functions  $\{\phi_i\}_{i=1..N}$  associated with each shape and computed the dominant 6 eigenmodes. Projecting each training shape on 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  $\mu$ , the transition matrices  $A_1, A_2$  and the noise covariance  $\Sigma$  shown in equation (10). Subsequently, we compared segmentations of noisy sequences obtained by segmentation in the 6-dimensional subspace without and with the dynamical statistical prior.

Figure 5 shows a sample input frame from a sequence with 25%, 50%, and 90% noise.<sup>5</sup> Figure 6 shows a set of segmentations obtained without dynamical

<sup>5</sup> 90% noise means that 90% of all pixels were replaced by a random intensity sampled from a uniform distribution.



**Fig. 8. Segmentation using a dynamical statistical shape prior based on a second order autoregressive model.** In contrast to the segmentation in Figure 7, the prior imposes statistically learned information about the *temporal dynamics* of the shape evolution to cope with misleading low-level information.



**Fig. 9. Tracking with dynamical statistical shape prior to cope with larger amounts of noise.** The input images were corrupted with 90% of noise. Yet, the statistically learned dynamical shape model allows to disambiguate the low-level information. These experiments confirm that our tracking schemes can indeed compete with the capacities of human observers.

shape prior on a sequence with 25% noise. While the segmentation without dynamical prior is successful with little noise, Figure 7 shows that it eventually breaks down when the noise level is increased.

Figure 8 shows segmentations of the same sequence as in 7 obtained with a *dynamical* statistical shape prior derived from a second order autoregressive model. Figure 9 shows that the dynamical statistical shape prior provides for good segmentations, even with 90% noise. Clearly, exploiting the temporal statistics of dynamical shapes allows to make the segmentation process very robust to missing and misleading information.

## 6 Conclusion

In this work, we introduced *dynamical* statistical shape models for implicitly represented shapes. In contrast to existing statistical shape models for implicit shapes, these models capture the temporal correlations which characterize deforming shapes such as the consecutive silhouettes of a walking person or the 2D projections of a rotating 3D object. Therefore they 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.

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 implicit shape models therefore support shapes of varying topology and are easily extended to higher-dimensional shapes (i.e. surfaces).

With the estimated dynamical models one can synthesize shape sequences of arbitrary length. In the context of a sequence 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 the synthesized sequence for various shape eigenmodes. In addition, we validated 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 process in a meaningful way.

To this end, we developed a Bayesian formulation for level set based image sequence segmentation, which allows to impose the statistically learnt dynamical models as shape priors in the segmentation process. 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 stochastic optimization methods).

Experimental results confirm that the resulting shape priors make it possible to reliably track familiar deformable objects despite large amounts of noise. Future work is focused on further quantitative performance analysis, on the development of statistical models which capture the joint dynamics of deformation and transformation modes, and on the optimization with stochastic methods.

## Acknowledgments

We thank Alessandro Bissacco and Payam Saisan for providing the image sequence data. We thank Gianfranco Doretto and Paolo Favaro for helpful discussions on autoregressive models.

## References

1. H. Akaike. Autoregressive model fitting for control. *Ann. Inst. Statist. Math.*, 23:163–180, 1971.
2. A. Blake and M. Isard. *Active Contours*. Springer, London, 1998.
3. V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 694–699, Boston, USA, 1995.
4. T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
5. 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.
6. 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 *LNCS*, pages 93–108, Copenhagen, May 2002. Springer.

7. D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: Teaching level sets to walk. In *Pattern Recognition*, volume 3175 of *LNCS*, pages 36–44. Springer, 2004.
8. 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.
9. A. Duci, A. Yezzi, S. Mitter, and S. Soatto. Shape representation via harmonic embedding. In *ICCV*, pages 656–662, 2003.
10. S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *IEEE Intl. Conf. on Comp. Vis.*, pages 810–815, 1995.
11. 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.
12. R. Malladi, J. A. Sethian, and B. C. Vemuri. A topology independent shape modeling scheme. In *SPIE Conf. on Geometric Methods in Comp. Vision II*, volume 2031, pages 246–258, 1994.
13. M. Moelich and T. Chan. Tracking objects with the Chan-Vese algorithm. Technical Report 03-14, Computational Applied Mathematics, UCLA, Los Angeles, 2003.
14. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
15. A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM T. on Mathematical Software*, 27(1):27–57, 2001.
16. 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.
17. 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.
18. 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 Comp. Vision and Patt. Recognition*, 2005. To appear.
19. 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 *LNCS*, pages 50–61, Prague, 2004. Springer.
20. M. Rousson and D. Cremers. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In *MICCAI*, 2005. To appear.
21. M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Proc. of the Europ. Conf. on Comp. Vis.*, volume 2351 of *LNCS*, pages 78–92, Copenhagen, May 2002. Springer, Berlin.
22. M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3d segmentation in MRI imaging. In *MICCAI*, pages 209–216, 2004.
23. G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6:461–464, 1978.
24. 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.
25. 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.
26. S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900, 1996.