To appear in the *Journal of Scientific Computing*, 2008, © Springer.

Nonlinear Dynamical Shape Priors for Level Set Segmentation

Daniel Cremers
Department of Computer Science
University of Bonn, Germany

Abstract

The introduction of statistical shape knowledge into level set based segmentation methods was shown to improve the segmentation of familiar structures in the presence of noise, clutter or partial occlusions. While most work has been focused on shape priors which are constant in time, it is clear that when tracking deformable shapes certain silhouettes may become more or less likely over time. In fact, the deformations of familiar objects such as the silhouettes of a walking person are often characterized by pronounced temporal correlations.

In this paper, we propose a nonlinear dynamical shape prior for level set based image segmentation. Specifically, we propose to approximate the temporal evolution of the eigenmodes of the level set function by means of a mixture of autoregressive models. We detail how such shape priors "with memory" can be integrated into a variational framework for level set segmentation. As an application, we experimentally validate that the *nonlinear* dynamical prior drastically improves the tracking of a person walking in different directions, despite large amounts of clutter and noise.

Keywords: Level sets, shape priors, dynamical systems, tracking.

1 Introduction

1.1 Level set methods

In this work, we are focused on the problems of segmentation and tracking: Given a sequence of images $I_1, \ldots, I_t$, where $I_i : \Omega \to \mathbb{R}$, we want to infer at any given time t the most likely shape $\mathcal{C}_t$ in the image plane $\Omega \subset \mathbb{R}$. Within the Bayesian framework, this is done by maximizing the posterior distribution $\mathcal{P}(\mathcal{C}_t \,|\, I_1, \ldots, I_t)$. This problem has been studied extensively, researchers have proposed dynamical models of shape and developed sophisticated frameworks to propagate the posterior distribution. Most of

this work is based on *explicit* contour representations (e.g. [2]).[1]

Yet, explicit boundary representations are known to suffer from several limitations when applied to shape learning and shape inference: Firstly, the matching of explicit contours requires to identify pairwise correspondences between points. In general this is a *combinatorial* problem – in particular if one wants to allow for local stretching or shrinking of the respective contours. While efficient matching algorithms have been developed based on dynamic programming, the integration of the resulting shape distances with statistical learning of shapes is still an open problem. Secondly, explicit boundary representations are typically constrained to a fixed topology. In practice, a shape of interest may undergo topological changes – it could be that a hole is torn into a 3D shape, or it could be that a single 3D object will induce 2D projections of varying topology. While the transition between two topological structures for explicit contours can be modeled based on sophisticated (and somewhat heuristic) decision processes (cf. [16]), the *matching* of explicit shapes with different topology for the sake of shape learning is not defined.

The level set method introduced by Osher and Sethian [19, 18] overcomes these drawbacks of explicit representations[2] as a means to implicitly propagate a boundary $\mathcal{C}(t)$ by evolving an appropriate embedding function $\phi : \Omega \times [0, T] \rightarrow \mathbb{R}$, where:

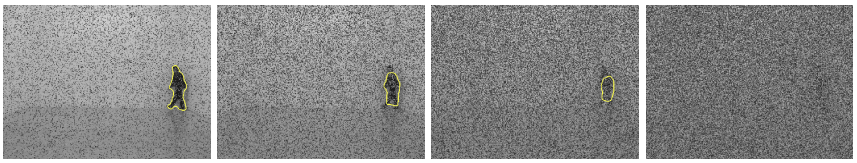
$$\mathcal{C}(t) = \{x \in \Omega \mid \phi(x, t) = 0\}. \tag{1}$$

In the context of shape learning and statistical shape inference, the level set method has several advantages:

- The implicit representation does not depend on a specific parameterization. Therefore shape matching does not require the computation of point-correspondences.
- Shape dissimilarity measures defined on the embedding functions can handle shapes of varying topology.
- The implicit representation (1) naturally generalizes to hypersurfaces in three or more dimensions, where the estimation of optimal point correspondences becomes a computationally cumbersome problem.

[1] While object-specific models representing human figures as kinematic chains of coupled geometric primitives allow for excellent results on tracking humans (cf. [23, 1, 24]), the geometric primitives and couplings are specified by a user. In contrast, our approach is based on a generic shape representation inferred from training data in an unsupervised manner.

[2] A precursor of the level set method was proposed by Dervieux and Thomasset [9].



25% noise 50% noise 75% noise 90% noise

Figure 1: Purely intensity-based image segmentation. With increasing amounts of noise, the intensity-based image segmentation using the method of Chan and Vese [5] gradually degrades.

1.2 Statistical shape priors for level set segmentation

The first applications of the level set method to image segmentation were pioneered in the early 90's by Malladi et al. [15], by Caselles et al. [3, 4], by Kichenassamy et al. [12] and by Paragios and Deriche [20].

Traditionally level set methods were applied to image segmentation by minimizing functionals which are based on local edge information [11, 4, 12] or based on regional homogeneity of the intensity function [17, 5, 26]. The results in Figure 1 show that purely intensity-based methods fail to provide the desired segmentation if the intensity information is degraded by increasing amounts of noise. In recent years, researchers have successfully introduced prior information about expected shapes into level set segmentation to cope with noise, background clutter and partial occlusions. Leventon et al. [13] modeled the embedding function by principal component analysis (PCA) of a set of training shapes and added appropriate driving terms to the level set evolution equation. Tsai et al. [25] suggested a more efficient formulation, where optimization is performed directly within the subspace of the first few eigenmodes. Rousson et al. [21] introduced shape information on the variational level. Cremers et al. introduced nonlinear statistical shape priors based on kernel density estimation [8].

By construction, these approaches were aimed to segment static images of an object of interest. Although they can be applied to tracking objects in image sequences (cf. [8]), they are not well-suited for this task, because they neglect the *temporal coherence of silhouettes* which characterizes deforming shapes. When tracking a deformable object, 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. The resulting set of silhouettes can be expected to contain strong temporal correlations. In [6], we recently proposed a simple linear dynamical shape model to capture such temporal correlations. Yet the use of linear models is limited to a single periodic motion.

Figure 2: Samples from a sequence of training silhouettes.

1.3 Contribution of this work

In this paper, we propose a more sophisticated dynamical shape model for level set segmentation which allows to simultaneously encode multiple dynamical modes. To this end, we approximate the temporal evolution of the level set embedding function by a *mixture of autoregressive models*. This leads to a *nonlinear* dynamical shape model for implicitly represented shapes. We detail the integration of nonlinear shape priors into the level set based segmentation process in a Bayesian framework. The resulting optimization problem is implemented using gradient descent inducing an evolution of the embedding function, driven both by the intensity information of the current image and by a time-dependent shape prior which relies on the segmentations obtained on the preceding frames. Experimental evaluation demonstrates that – in contrast to segmentation with static 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 over time. The *nonlinearity* of the dynamical shape model allows the shape deformation to undergo various dynamical modes. We demonstrate this by tracking a person walking in different directions. A preliminary version of this paper was presented at the International Conference on Computer Vision and Pattern Recognition [7].

2 Nonlinear Implicit Dynamical Shape Models

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 thus separate shape ϕ and transformation parameters θ, as one may want to use different models to represent and learn their temporal evolution.

Assume we are given a temporal sequence of training shapes such as the ones shown in Figure 2, represented by their embedding functions $\{\phi_1, \ldots, \phi_n\}$ and their transformation parameters $\{\theta_1, \ldots, \theta_n\}$. For uniqueness we require that all ϕ_i are signed distance functions. In the following, we will develop nonlinear dynamical models for implicit shape representations which allow to statistically model the above shape sequence.

2.1 A compact low-dimensional representation

It is well-known that statistical learning and inference can be performed more reliably and more efficiently in low-dimensional representations. For this reason, we revert to an approximation of the embedding functions associated with all training shapes by their principal components, i.e.

$$\phi_i(x) = \phi_0(x) \; + \; \sum_{j=1}^{n} \alpha_{ij}\, \psi_j(x), \tag{2}$$

where ϕ_0 denotes the mean embedding function and $\psi_1, \ldots, \psi_n$ the n largest eigenmodes with $n << N$. The expansion coefficients α_{ij} are given by the projection of each shape onto these eigenmodes:

$$\alpha_{ij} = \int (\phi_i - \phi_0)\, \psi_j \, dx, \tag{3}$$

Such PCA based representations of level set functions have been successfully applied for the construction of statistical shape priors in [13, 25, 21]. It should be pointed out that the application of PCA to the embedding function has certain limitations. The space of signed distance functions is not a linear space, such that a linear combination of eigenmodes will in general not be 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.

Let us denote the vector of the first n eigenmodes as

$$\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n). \tag{4}$$

Each sample shape ϕ_i is therefore approximated by the n-dimensional shape vector

$$\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{in}) = \int (\phi_i - \phi_0)\, \boldsymbol{\psi}\, dx. \tag{5}$$

Much theory has been developed for the statistical analysis of time series data. Overviews can be found in [14, 10]. Applications of dynamical systems to model deformable shapes were proposed among others in [2]. In our context, we intend to learn dynamical models for *implicitly* represented shapes. To allow for a more transparent presentation, we will gradually increase the model complexity from linear dynamical models of deformation, over joint models of deformation and transformation to nonlinear mixture models.

2.2 Linear implicit dynamical shape models

To learn a temporal model of the evolution of the level set function, one can approximate the shape vectors $\boldsymbol{\alpha}_t \equiv \boldsymbol{\alpha}_{\phi_t}$ representing sequence of level set functions by a Markov chain of order k [6]:

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu} + A_1\boldsymbol{\alpha}_{t-1} + A_2\boldsymbol{\alpha}_{t-2} + \ldots + A_k\boldsymbol{\alpha}_{t-k} + \boldsymbol{\eta}, \tag{6}$$

where $\boldsymbol{\eta}$ is zero-mean Gaussian noise with covariance Σ, $\boldsymbol{\mu}$ denotes the mean and A_i denote transition matrices. The probability of a shape conditioned on the shapes observed in previous time steps is therefore given by the corresponding autoregressive (AR) model of order k:

$$\mathcal{P}(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\alpha}_{1:t-1}) \propto \exp\left(-\frac{1}{2}\,\boldsymbol{v}^\top\,\Sigma^{-1}\,\boldsymbol{v}\right), \tag{7}$$

where

$$\boldsymbol{v} = \boldsymbol{\alpha}_t - \boldsymbol{\mu} - A_1\boldsymbol{\alpha}_{t-1} - A_2\boldsymbol{\alpha}_{t-2}\ldots - A_k\boldsymbol{\alpha}_{t-k} \tag{8}$$

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 $A_1, \ldots, A_k, \Sigma \in \mathbb{R}^{n\times n}$. We applied a maximum likelihood estimation using least squares. Different tests have been devised to quantify the accuracy of the model fit. Using dynamical models up to an order of 8, we found that according to Schwarz's Bayesian Criterion [22], our training sequences were best approximated by an autoregressive model of second order.

2.3 Models of deformation and transformation

In the previous section, we employed an autoregressive model to capture the temporal dynamics of implicitly represented shapes. To this end, we 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 3, for example, show a person walking "on the spot".

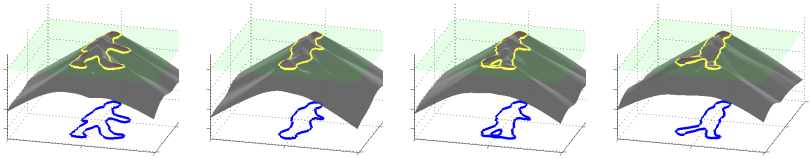


Figure 3: Synthesis of implicit dynamical shapes. Statistically generated embedding surfaces obtained by sampling from a second order autoregressive model, and the contours given by the zero level lines of the surfaces. The implicit formulation allows the embedded contour to change topology (third image).

In general, one can expect the deformation parameters $\boldsymbol{\alpha}_t$ 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 these transformations. At the same time, 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 $\triangle\theta_t$: $T_{\theta_t}x = T_{\triangle\theta_t}T_{\theta_{t-1}}x$. Instead of learning models of the absolute transformation θ_t, we can simply learn models of the update transformations $\triangle\theta_t$ (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 shape in the training sequence the deformation parameters $\boldsymbol{\alpha}_t$ and the transformation changes $\triangle\theta_t$, and fit the autoregressive models given in equations (7) and (8) to an ***extended shape vector***

$$\boldsymbol{\beta}_t \equiv \begin{pmatrix} \boldsymbol{\alpha}_t \\ \triangle\theta_t \end{pmatrix}. \tag{9}$$

Synthesizing from the autoregressive model allows to generate silhouettes of a walking person which are similar to the ones shown in Figure 3, but which move forward in space, starting from an arbitrary (user-specified) initial position.

2.4 Nonlinear implicit dynamical shape models

While linear dynamical models may be sufficient to model simple essentially periodical shape deformations, they are clearly insufficient when it comes

to modeling more complex dynamical processes. Much theory has been developed to model nonlinear dynamical systems. In the following, we will assume that the dynamics of our shape can be approximated using a collection of linear autoregressive models. Such mixtures models have been successfully applied to the tracking of human motion, based on user-specified shape representations by coupled geometric primitives [1]. The probability of an (extended) shape vector $\boldsymbol{\beta}_t$ conditioned on the shapes at previous time instances is approximated by a mixture of N autoregressive models of orders $\{k_i\}_{i=1..N}$ according to:

$$\mathcal{P}(\boldsymbol{\beta}_t \,|\, \boldsymbol{\beta}_{1:t-1}) \propto \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{|2\pi\Sigma_i|}} \exp\left(-\frac{1}{2}\, \boldsymbol{v}_i^\top\, \Sigma_i^{-1}\, \boldsymbol{v}_i\right), \tag{10}$$

where

$$\boldsymbol{v}_i = \boldsymbol{\beta}_t - \boldsymbol{\mu}_i - A_{i1}\boldsymbol{\beta}_{t-1} - A_{i2}\boldsymbol{\beta}_{t-2} \ldots - A_{ik_i}\boldsymbol{\beta}_{t-k_i}. \tag{11}$$

The fitting of a mixture of autoregressive models to a training sequence requires the estimation of the model parameters given by the number N of autoregressive models, the model orders $\{k_i\}$, the means $\{\boldsymbol{\mu}_i\}$ and transition matrices $\{A_{ij}\}_{j=1..k_i}$ associated with model i, where $i = 1, \ldots, N$. There exist sophisticated approaches to learn these parameters in an unsupervised manner. The key challenge is to solve the chicken-and-egg problem of simultaneously segmenting the sequence and estimating model parameters for each subsequence. This can be done using either iterative algorithms such as EM or direct approaches, for example by means of polynomial factorization [27].

Since the unsupervised learning of autoregressive mixture models is not the focus of this work, we will for simplicity pursue a semi-supervised learning process. Specifically we assume that our training sequence is already partitioned into subsequences each of which is fitted by a separate AR model. While we use the entire sequence to construct a PCA-based low-dimensional shape representation shared by all dynamical modes, we then learn separate autoregressive models to capture subsequences which are labeled, for example by a user marking them as "walking left", "walking right", "running left", etc. We will demonstrate that this approach allows to track objects undergoing different dynamics by using the same *nonlinear* dynamical shape prior.

3 Integration in a segmentation process

In the following, we will detail how the proposed nonlinear dynamical shape model can be imposed as a prior in variational image segmentation.

Assume we are given an image $I_t : \Omega \to \mathbb{R}$ from an image sequence and segmentations of the previous images in terms of shape vectors and transformations $\{\hat{\boldsymbol{\alpha}}_i, \hat{\theta}_i\}_{i=1,\ldots,t-1}$. The problem of segmenting the current frame I_t can then be addressed in the framework of Bayesian inference by computing the shape vector $\hat{\boldsymbol{\alpha}}_t$ and transformation $\hat{\theta}_t$ which maximize the conditional probability

$$\mathcal{P}(\boldsymbol{\alpha}_t, \theta_t \,|\, I_t, \hat{\boldsymbol{\alpha}}_i, \hat{\theta}_i) \propto \mathcal{P}(I_t \,|\, \boldsymbol{\beta}_t, \theta_t)\, \mathcal{P}(\boldsymbol{\alpha}_t, \theta_t \,|\, \hat{\boldsymbol{\alpha}}_i, \hat{\theta}_i).$$

Here we assumed that I_t only depends on the current segmentation, i.e. there is no further hidden dependence on the preceeding shape configurations. The expression $\mathcal{P}(\boldsymbol{\alpha}_t, \theta_t \,|\, \hat{\boldsymbol{\alpha}}_i, \hat{\theta}_i)$ is identical to the one in (10), with all past variables given by their optimal values $\hat{\boldsymbol{\alpha}}_i, \hat{\theta}_i$ for $i = 1, \ldots, t-1$.

Maximizing this conditional probability can be performed by minimizing its negative logarithm, which is – up to a constant – given by an energy of the form:

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

Assuming Gaussian-distributed intensities of object and background [28, 5], the data term is given by

$$\begin{aligned} E_{data}(\boldsymbol{\alpha}_t, \theta_t) = &\int \left(\frac{(I_t - \mu_1)^2}{2\sigma_1^2} + \log \sigma_1 \right) H\phi_{\boldsymbol{\beta}_t}\, dx \\ &+ \int \left(\frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \sigma_2 \right) \left(1 - H\phi_{\boldsymbol{\beta}_t} \right) dx, \end{aligned}$$

where, for notational simplicity, we have introduced the expression

$$\phi_{\boldsymbol{\beta}_t} \equiv \phi_0(T_{\theta_t} x) + \boldsymbol{\alpha}_t^\top \boldsymbol{\psi}(T_{\theta_t} x) \tag{13}$$

for the embedding function of a shape generated with parameters $\boldsymbol{\beta}_t$.

With the autoregressive mixture model (10), the dynamical shape energy is:

$$E_{shape}(\boldsymbol{\alpha}_t, \theta_t) = -\log \left[\sum_{i=1}^{N} \frac{1}{\sqrt{|2\pi\Sigma_i|}} \exp\!\left(-\frac{1}{2}\, \boldsymbol{v}_i^\top \Sigma_i^{-1}\, \boldsymbol{v}_i \right) \right],$$

with $\boldsymbol{v}_i$ defined in (11), replacing $\boldsymbol{\beta}_i$ by $\hat{\boldsymbol{\beta}}_i$ for $i < t$.

Tracking an object of interest over a sequence of images with a nonlinear dynamical shape prior can be done by minimizing energy (12). We pursue a gradient descent strategy. Due to space limitations, we will merely report the differential equations governing the evolution of the deformation component $\boldsymbol{\alpha}_t$ of the extended shape vector $\boldsymbol{\beta}_t$:

$$\frac{d\boldsymbol{\alpha}_t(\tau)}{d\tau} = -\frac{\partial E_{data}}{\partial \boldsymbol{\alpha}_t} - \nu\, \frac{\partial E_{shape}}{\partial \boldsymbol{\alpha}_t} \tag{14}$$

where τ denotes the artificial evolution time, as opposed to the physical time t.

The data term is given by:

$$\frac{\partial E_{data}}{\partial \boldsymbol{\alpha}_t} = \int_\Omega \left(\frac{(I_t - \mu_1)^2}{2\sigma_1^2} - \frac{(I_t - \mu_2)^2}{2\sigma_2^2} + \log \frac{\sigma_1}{\sigma_2} \right) \boldsymbol{\psi}(T_{\theta_t} x)\, \delta\big(\phi_{\boldsymbol{\beta}_t}\big)\, dx. \quad (15)$$

The gradient of the shape energy is given by:

$$\frac{\partial E_{shape}}{\partial \boldsymbol{\alpha}_t} = \sum_i \gamma_i \begin{pmatrix} 1_n & 0 \\ 0 & 0 \end{pmatrix} \Sigma_i^{-1}\, \boldsymbol{v}_i, \quad (16)$$

with $\boldsymbol{v}_i$ given in (11) and 1_n being the n-dim. unit matrix modeling the projection on the shape components of $\boldsymbol{v}_i$, where n is the number of shape modes. The normalized weights γ_i are given by:

$$\gamma_i = \frac{\tilde{\gamma}_i}{\sum_j \tilde{\gamma}_j}, \quad \tilde{\gamma}_i = \frac{1}{\sqrt{|2\pi\Sigma_i|}} \exp\left(-\frac{1}{2}\, \boldsymbol{v}_i^\top\, \Sigma_i^{-1}\, \boldsymbol{v}_i \right). \quad (17)$$

For every image I_t in the input sequence, the evolution of the shape $\boldsymbol{\alpha}_t$ is thus driven by the two terms shown in equations (15) and (16). These have the following very intuitive interpretations:

- The data term (15) draws the shape to separate the image intensities according to the estimated two Gaussian intensity models. Since the effect of variations in the shape vector $\boldsymbol{\alpha}_t$ onto the level set function $\phi_{\boldsymbol{\beta}_t}$ are given by the eigenmodes $\boldsymbol{\psi}$, the data term is a projection onto these eigenmodes.

- The shape term (16) induces a relaxation of the shape vector $\boldsymbol{\alpha}_t$ toward the most likely shape, as predicted by the nonlinear dynamical model based on the segmentations of previous time frames. This second term consists of a weighted sum of terms. Each term drives the current shape $\boldsymbol{\alpha}_t$ to the shape predicted by the i-th autoregressive model. The weights γ_i in (17) indicate how (relatively) well the respective dynamical models match the current dynamics. They (exponentially) suppress the influence of models which are not consistent with current and past estimates of shape and transformation. The weights γ_i thus indicates which dynamical models best represent the current observations. Plotting the weights γ_i over time allows to track an interpretation of the observed dynamics as a superposition of dynamical models.

Similar evolution equations can be derived for the transformation parameters.

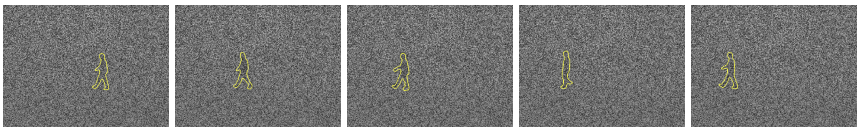


Figure 4: Segmentation with dynamical shape prior for 90% noise[3]. Compared to the segmentation results without shape prior shown in Figure 1, the dynamical shape prior provides reliable segmentations even if nine out of ten pixels are assigned a random intensity.

4 Experimental Results

For all experiments, we constructed a prior by hand-segmenting a sequence of a walking person. We additionally partitioned the training sequence into sections associated with different dynamical models. The subsequent computation of embedding functions, alignment, PCA and dynamical system parameters are done fully automatically. The weight ν of the prior was chosen constant for all experiments. While results are not too sensitive to the choice of ν, too large values of ν (too weak data term) tend to inhibit the detection of transitions between different dynamical modes.

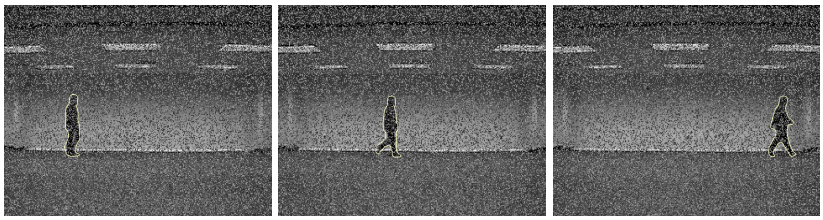
Coping with noise

Figure 4 shows segmentation results obtained with a dynamical shape prior for images corrupted by noise. While the segmentation without dynamical shape prior degrades even with moderate amounts of noise (see Figure 1), the same data term constrained with a dynamical shape prior at 90% noise[3] provides reliable segmentations where human observers fail.

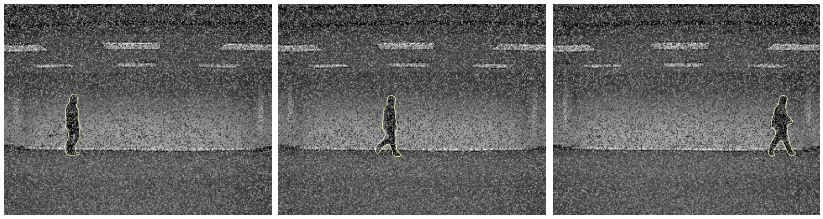
Linear versus nonlinear dynamical shape prior

The nonlinear dynamical shape prior (10) allows to integrate prior knowledge about *multiple* autoregressive models into the segmentation process. Figure 5 provides segmentation results obtained on a sequence showing a person walking in different directions with 50% noise superimposed. These indicate that in contrast to the linear prior (top row), the nonlinear dynamical prior (bottom row) can reliably enhance the segmentation of different dynamical shape modes, thereby allowing to track a person in different directions using a single shape prior. Since we merely imposed priors on the deformation (and not the transformation), the linear prior provides acceptable segmentations except that all generated silhouettes seem to be walking right. The close-ups in Figure 6 show that in contrast to the linear one, the nonlinear model selected the correct dynamical model in a data-driven manner.

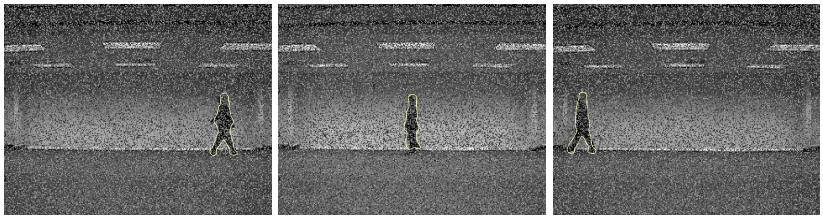
[3]90% noise means that 90% of pixel intensities were replaced by a random intensity.



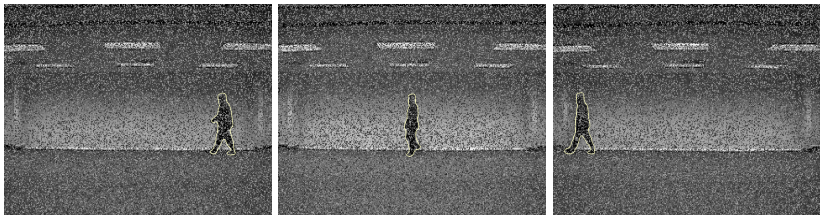
Tracking right with linear shape prior



Tracking right with nonlinear shape prior



Tracking left with linear shape prior



Tracking left with nonlinear shape prior

Figure 5: Linear versus nonlinear dynamical shape prior. While the **linear prior (first and third)** was built on people walking to the right, the **nonlinear prior (second and fourth)** simultaneously encodes both walking directions. Upon turning around (last three frames), the weights γ_i in 17 flip from 0 to 1 (and vice versa), indicating that the algorithm imposes the appropriate dynamical model in a data-driven manner. This leads to superior segmentation results in the second part of the sequence – see also the closeups in Figure 6.

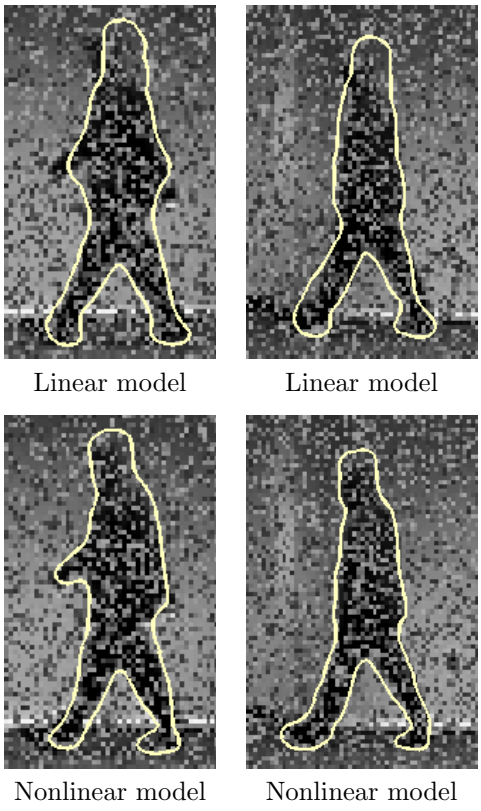


Figure 6: Zoom-in on Figure 5, last two columns. In contrast to the linear model (top) which incorrectly generates segmentations that look like a person walking right, the nonlinear dynamical prior (bottom) allows for the emergence of multiple walking modes, thereby providing segmentations which are consistent with people walking in either direction. This gives rise to superior segmentation results in the second part of the sequence.

Tracking through occlusions

Figure 7 demonstrates that one obtains accurate segmentations even when the walking person is fully occluded by an oncoming bar. This is due to the fact that the dynamical prior accumulates information over time and provides segmentations which are temporally consistent with the segmentations obtained on previous frames.

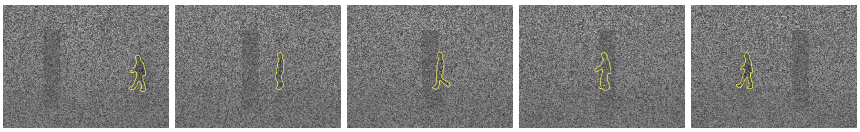


Figure 7: Dealing with noise and occlusion. The input sequence shows a person walking to the left occluded by a bar moving to the right, corrupted by 80% noise. Since the dynamical prior accumulates information over time, it allows for accurate segmentations even when the walking person is completely occluded – see the fourth frame.

5 Conclusion

In this work, we introduced a nonlinear dynamical shape model for implicitly represented shapes in order to cope with misleading low-level information in level set based image segmentation. Specifically, we proposed to approximate the temporal evolution of the eigenmodes of the level set function by a mixture of autoregressive models. In contrast to existing models for implicit shapes, the proposed approach allows to learn the temporal correlations characterizing deforming shapes in terms of multiple dynamical modes. The model can be integrated as a nonlinear dynamical shape prior in a Bayesian formulation of level set based image sequence segmentation. Experimental results confirm that the nonlinear dynamical shape prior outperforms the linear one when tracking a person walking in *different* directions through large amounts of noise and prominent occlusions.

References

[1] A. Agarwal and B. Triggs. Tracking articulated motion using a mixture of autoregressive models. In *Europ. Conf. on Computer Vision*, volume 3, pages 54–65, 2004.

[2] A. Blake and M. Isard. *Active Contours*. Springer, London, 1998.

[3] V. Caselles, F. Catté, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numer. Math.*, 66:1–31, 1993.

[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. A level set algorithm for minimizing the Mumford–Shah functional in image processing. In *IEEE Workshop on*

Variational and Level Set Methods, pages 161–168, Vancouver, CA, 2001.

[6] D. Cremers. Dynamical statistical shape priors for level set based tracking. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 28(8):1262–1273, August 2006.

[7] D. Cremers. Nonlinear dynamical shape priors for level set segmentation. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2007.

[8] 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*, 69(3):335–351, 2006.

[9] 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.

[10] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2003.

[11] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. of Computer Vision*, 1(4):321–331, 1988.

[12] 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.

[13] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Int. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 316–323, Hilton Head Island, SC, 2000.

[14] L. Ljung. *System Identification - Theory For the User*. Prentice Hall, Upper Saddle River, NJ, 1999.

[15] 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.

[16] T. McInerney and D. Terzopoulos. Topologically adaptable snakes. In *Proc. 5th Int. Conf. on Computer Vision*, pages 840–845, Los Alamitos, California, June 20–23 1995. IEEE Comp. Soc. Press.

[17] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[18] S. J. Osher and R. P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, New York, 2002.

[19] 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.

[20] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 22(3):266–280, 2000.

[21] M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3d segmentation in MRI imaging. In *MICCAI*, volume 2217 of *LNCS*, pages 209–216. Springer, 2004.

[22] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6:461–464, 1978.

[23] L. Sigal, S. Bhatia, S. Roth, M. Black, and M. Isard. Tracking loose-limbed people. In *Int. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 421–428, 2004.

[24] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas. Discriminative density propagation for 3d human motion estimation. In *Int. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 390–397, 2005.

[25] 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.

[26] 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.

[27] R. Vidal and Y. Hashambhoy. Recursive identification of switched arx models with unknown number of models and unknown orders. In *IEEE Conf. on Decision and Control*, Dec 2005.

[28] 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.