PERGAMON

Pattern Recognition 36 (2003) 1929 – 1943

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Shape statistics in kernel space for variational image segmentation

Daniel Cremers[a,*] , Timo Kohlberger[b], Christoph Schnörr[b]

[a] *Department of Computer Science, University of California, Los Angeles, CA 90095, USA*
[b] *Computer Vision, Graphics, and Pattern Recognition Group, Department of Mathematics and Computer Science, University of Mannheim, D-68131 Mannheim, Germany*

Received 15 January 2003; accepted 15 January 2003

Abstract

We present a variational integration of nonlinear shape statistics into a Mumford–Shah based segmentation process. The nonlinear statistics are derived from a set of training silhouettes by a novel method of density estimation which can be considered as an extension of kernel PCA to a probabilistic framework.

We assume that the training data forms a Gaussian distribution after a nonlinear mapping to a higher-dimensional feature space. Due to the strong nonlinearity, the corresponding density estimate in the original space is highly non-Gaussian.

Applications of the nonlinear shape statistics in segmentation and tracking of 2D and 3D objects demonstrate that the segmentation process can incorporate knowledge on a large variety of complex real-world shapes. It makes the segmentation process robust against misleading information due to noise, clutter and occlusion.

Keywords: Probabilistic kernel PCA; Nonlinear shape statistics; Density estimation; Image segmentation; Variational methods; Diffusion snakes

1. Introduction

One of the challenges in the field of image segmentation is the incorporation of prior knowledge on the shape of the segmenting contour. A common approach is to learn the shape of an object statistically from a set of training shapes, and to then restrict the segmenting contour to a submanifold of familiar shapes during the segmentation process. For the problem of segmenting a specific known object this approach was shown to drastically improve segmentation results (cf. [1,2]).

Although the shape prior can be quite powerful in compensating for misleading information due to noise, clutter and occlusion in the input image, most approaches are limited in their applicability to more complicated shape variations of real-world objects. Commonly, the permissible shapes are assumed to form a multivariate Gaussian distribution, which essentially means that all possible shape deformations correspond to linear combinations of a set of eigenmodes, such as those given by principal component analysis (cf. [1,3–6]). In particular, this means that for any two permissible shapes, the entire sequence of shapes obtained by a linear morphing of the two shapes is permissible as well. Once the set of training shapes exhibits highly nonlinear shape deformations—such as different 2D views of a 3D object—one finds distinct clusters in shape space corresponding to the stable views of an object. Moreover, each of the clusters may by itself be quite non-Gaussian. The Gaussian hypothesis will then result in a mixing of the different

[*] Corresponding author.

E-mail addresses: cremers@cs.ucla.edu (D. Cremers), tiko@uni-mannheim.de (T. Kohlberger), schnoerr@uni-mannheim.de (C. Schnörr).

URLs: http://www.cs.ucla.edu/∼cremers, http://www.cvgpr.uni-mannheim.de, http://www.cvgpr.uni-mannheim.de

0031-3203/03/$30.00 © 2003 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0031-3203(03)00056-6

views, and the space of accepted shapes will be far too large for the prior to sensibly restrict the contour deformation.

A number of models have been proposed to deal with nonlinear shape variation. However, they often suffer from certain drawbacks. Some involve a complicated model construction procedure [7]. Some are supervised in the sense that they assume prior knowledge on the structure of the nonlinearity [8]. Others require prior classification with the number of classes to be estimated or specified beforehand and each class being assumed Gaussian [9,10]. And some cannot be easily extended to shape spaces of higher dimension [11].

In the present paper, we present a density estimation approach which is based on Mercer kernels [12,13] and which does not suffer from any of the mentioned drawbacks. Our work has been inspired by recent developments in the machine learning community [14]. It comprises and extends results which were presented on two conferences [15,16]. In Section 2, we review the variational integration of a linear shape prior into Mumford–Shah based segmentation. In Section 3, we give an intuitive example for the limitations of the linear shape model. In Section 4, we present the nonlinear density estimate which was first introduced in Ref. [15]. We compare it to related approaches and give estimates of the involved parameters. In Section 5, we illustrate its application to artificial 2D data and to silhouettes of real objects. In Section 6, this nonlinear shape prior is integrated into segmentation. We propose a variational integration of similarity invariance. In Section 7, numerous examples of segmentation with and without shape prior on static images and tracking sequences finally confirm the properties of the nonlinear shape prior: It can encode very different shapes and generalizes to novel views without blurring or mixing different views. Furthermore, it improves segmentation by reducing the dimension of the search space, by stabilizing with respect to clutter and noise and by reconstructing the contour in areas of occlusion.

2. Diffusion snakes: statistical shape prior in Mumford–Shah based segmentation

In Ref. [6], we presented a variational integration of statistical shape knowledge in a Mumford–Shah based segmentation. A segmentation u of a given input image f was obtained by minimizing a joint energy functional

$$E(C,u) = E_{image}(C,u) + \alpha E_{shape}(C), \tag{1}$$

which takes into account both the low-level grey value information of the input image and a higher-level knowledge about the expected shape of the segmenting contour C. We suggested modifications of the Mumford–Shah functional E_{image} and its cartoon limit [17] which facilitate the implementation of the segmenting contour as a parameterized spline curve:

$$C_z : [0,1] \to \Omega \subset \mathbb{R}^2, \qquad C_z(s) = \sum_{n=1}^{N} \begin{pmatrix} x_n \\ y_n \end{pmatrix} B_n(s), \tag{2}$$

where B_n are quadratic, uniform and periodic B-spline basis functions [18], and $z = (x_1, y_1, \ldots, x_N, y_N)^t$ denotes the vector of control points. Shape statistics can then be obtained by estimating the distribution of the control point vectors corresponding to a set of contours which were extracted from binary training images.

In the present paper, we focus on significantly improving the shape statistics. For the low-level image information, we will therefore restrict ourselves to the somewhat simpler cartoon limit of the Mumford–Shah functional. The segmentation of a given grey value input image $f : \Omega \to [0,255]$ is obtained by minimizing the energy functional

$$E_{image}(C,\{u_i\}) = \frac{1}{2}\sum_i \int_{\Omega_i} (f-u_i)^2 \,\mathrm{d}x + \nu \mathrm{L}(C) \tag{3}$$

with respect to the constants u_i and the segmenting contour C. This enforces a segmentation of the image plane into a set of regions Ω_i, such that the variation of the grey value is minimal within each region. [1]

In Ref. [6], we proposed to measure the length of the contour by the squared L_2-norm

$$\mathrm{L}(C) = \int_0^1 \left(\frac{\mathrm{d}C}{\mathrm{d}s}\right)^2 \mathrm{d}s, \tag{4}$$

which is more adapted to the implementation of the contour as a closed spline curve than the usual L_1-norm, because it enforces an equidistant spacing of control points. This length constraint induces a rubber-band like behavior of the contour and thereby prevents the formation of cusps during the contour evolution. Since it is the same length constraint which is used for the classical snakes [23], we obtain a hybrid model which combines the external energy of the Mumford–Shah functional with the internal energy of the snakes. For this reason, we refer to the functional (3) with length constraint (4) as *diffusion snake*.

Beyond just minimizing the length of the contour, one can minimize a shape energy $E_{shape}(C)$, which measures the dissimilarity of the given contour with respect to a set of training contours. Minimizing the total energy (1) will enforce a segmentation which is based on both the input image and the similarity to a set of training shapes.

In order to study the interaction between statistical shape knowledge and image grey value information we restricted the shape statistics in Ref. [6] to a common model by assuming the training shapes to form a multivariate Gaussian

[1] The underlying piecewise-constant image model (3) can easily be generalized to incorporate higher-order grey value statistics [19], edge information [20] or motion information [21,22]. In this paper, however, we focus on modeling shape statistics and therefore do not consider these possibilities.

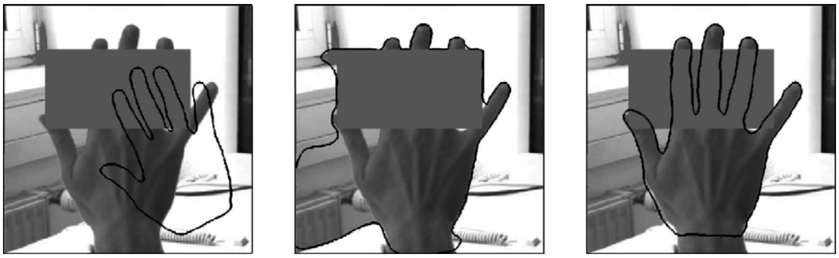

Fig. 1. Segmentation with *linear* shape prior on an image of a partially occluded hand: Initial contour (left), segmentation without shape prior (center), and segmentation with shape prior (right). The statistical shape prior compensates for misleading information due to noise, clutter and occlusion. Integration into the variational framework effectively reduces the dimension of the search space and enlarges the region of convergence.

distribution in shape space. This corresponds to a quadratic energy on the spline control point vector z:

$$E_c(C_z) = \tfrac{1}{2}(z - z_0)^{\mathrm{t}} \Sigma^{-1}(z - z_0), \tag{5}$$

where z_0 denotes the mean control point vector and Σ the covariance matrix after appropriate regularization [6]. The effect of this shape energy[2] in dealing with clutter and occlusion is exemplified in Fig. 1. For the input image f of a partially occluded hand, we performed a gradient descent to minimize the total energy (1) without ($\alpha = 0$) and with ($\alpha > 0$) shape prior. Incorporating the shape prior draws the evolving contour to a submanifold of familiar shapes. Thus the resulting segmentation process becomes insensitive to misleading information due to clutter and occlusion.

3. Limitations of the linear shape model

Unfortunately, the linear shape statistics (5) are limited in their applicability to more complicated shape deformations. As soon as the training shapes form distinct clusters in shape space—such as those corresponding to the stable views of a 3D object—or if the shapes of a given cluster are no longer distributed according to a hyperellipsoid, the Gaussian shape prior tends to mix classes and blur details of the shape information in such a way that the resulting shape prior is no longer able to effectively restrict the contour evolution to the space of familiar shapes.

A standard way to numerically verify the validity of the Gaussian hypothesis is to perform statistical tests such as the χ^2-test. In the following, we will demonstrate the "non-Gaussianity" of a set of sample shapes in a different way, because it gives a better intuitive understanding of the limitations of the Gaussian hypothesis in the context of shape statistics.

Fig. 2, left side, shows the training shapes corresponding to nine views of a right hand and nine views of a left hand, projected onto the first two principal components and the level lines of constant energy for the Gaussian model (5). Note that if the training set were Gaussian distributed, then all projections should be Gaussian distributed as well. Yet in the projection in Fig. 2, left side, one can clearly distinguish two separate clusters containing the right hands ($+$) and the left hands ($\bullet$).

As suggested by the level lines of constant energy, the first principal component—i.e. the mayor axis of the ellipsoid—corresponds to the deformation between right and left hands. This *morphing* from a left hand to a right hand is visualized in more detail in the right images of Fig. 2: Sampling along the first principal component around the mean shape shows a mixing of shapes belonging to different classes. Obviously the Gaussian model does not accurately represent the distribution of training shapes. In fact, according to the Gaussian model, the most probable shape is the mean shape given by the central shape in Fig. 2. In this way, sampling along the different eigenmodes around the mean shape can give an intuitive feeling for the quality of the Gaussian assumption.

4. Density estimation in feature space

In the following, we present an extension of the above method which incorporates a strong nonlinearity at almost no additional effort. Essentially we propose to perform a density estimation not in the original space but in the feature space of nonlinearly transformed data. The nonlinearity enters in terms of Mercer kernels [13], which have been extensively used in pattern recognition and machine learning (cf. [24,25]). In the present section, we will introduce the method of density estimation, discuss its relation to kernel principal component analysis (kernel PCA) [14], and propose estimates of the involved parameters. Finally, we will illustrate the density estimate in applications to artificial 2D data and to 200-dimensional data corresponding to silhouettes of real-world training shapes. In order not to break the

[2] A similarity invariant shape energy E_{shape} is obtained by applying the statistical energy E_c in Eq. (5) to the shape vector z after aligning it with respect to the training set. This will be detailed in Section 6.2.

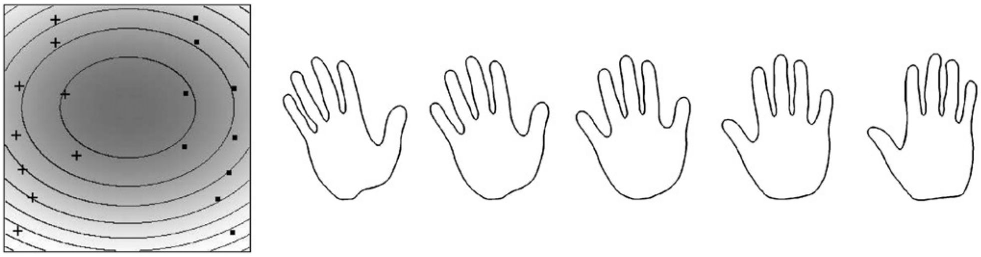

Fig. 2. *Left*: Projection of the training shapes and the estimated energy onto the first two principal components for a set containing right (+) and left (•) hands. *Right images*: Sampling along the first principal component shows the mixing of different classes in the Gaussian model. Note that according to the Gaussian model the mean shape (central shape) is the most probable shape.

flow of the argument, further remarks on the relation of distances in feature space to classical methods of density estimation are postponed to Appendix A.

4.1. Gaussian density in kernel space

Let $z_1, \ldots, z_m \in \mathbb{R}^n$ be a given set of training data. Let ϕ be a nonlinear mapping from the input space to a potentially higher-dimensional space Y. The mean and the sample covariance matrix of the mapped training data are given by

$$\phi_0 = \frac{1}{m}\sum_{i=1}^{m} \phi(z_i),$$

$$\tilde{\Sigma}_\phi = \frac{1}{m}\sum_{i=1}^{m} (\phi(z_i) - \phi_0)(\phi(z_i) - \phi_0)^{\mathrm{t}}. \tag{6}$$

Denote the corresponding scalar product in Y by the Mercer kernel [13]

$$k(x, y) := (\phi(x), \phi(y)) \quad \text{for } x, y \in \mathbb{R}^n. \tag{7}$$

Denote a mapped point after centering with respect to the mapped training points by

$$\tilde{\phi}(z) := \phi(z) - \phi_0 \tag{8}$$

and the centered kernel function by

$$\begin{aligned} \tilde{k}(x, y) &:= (\tilde{\phi}(x), \tilde{\phi}(y)) \\ &= k(x, y) - \frac{1}{m}\sum_{k=1}^{m} (k(x, z_k) + k(y, z_k)) \\ &\quad + \frac{1}{m^2}\sum_{k,l=1}^{m} k(z_k, z_l). \end{aligned} \tag{9}$$

We estimate the distribution of the *mapped* training data by a Gaussian probability density in the space Y—see Fig. 3. The corresponding energy, given by the negative logarithm of the probability, is a Mahalanobis type distance in the space Y:

$$E_\phi(z) = \tilde{\phi}(z)^{\mathrm{t}} \Sigma_\phi^{-1} \tilde{\phi}(z). \tag{10}$$

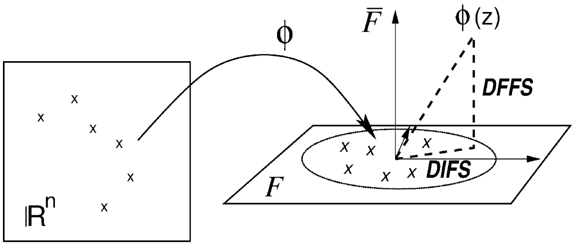

Fig. 3. Nonlinear mapping into $Y = F \oplus \bar{F}$ and the distances DIFS and DFFS.

It can be considered a nonlinear measure of the dissimilarity between a point z and the training data. The regularized covariance matrix Σ_ϕ is obtained by replacing all zero eigenvalues of the sample covariance matrix $\tilde{\Sigma}_\phi$ by a constant $\lambda_\perp$:

$$\Sigma_\phi = V \Lambda V^{\mathrm{t}} + \lambda_\perp (I - VV^{\mathrm{t}}), \tag{11}$$

where Λ denotes the diagonal matrix of nonzero eigenvalues $\lambda_1 \leqslant \cdots \leqslant \lambda_r$ of $\tilde{\Sigma}$ and V is the matrix of the corresponding eigenvectors $V_1, \ldots, V_r$. By definition of $\tilde{\Sigma}_\phi$, these eigenvectors lie in the span of the mapped training data:

$$V_k = \sum_{i=1}^{m} \alpha_i^k \tilde{\phi}(z_i), \qquad 1 \leqslant k \leqslant r. \tag{12}$$

Schölkopf et al. [14] showed that the eigenvalues λ_k of the covariance matrix and the expansion coefficients $\{\alpha_i^k\}_{i=1,\ldots,m}$ in Eq. (12) can be obtained in terms of the eigenvalues and eigenvectors of the centered kernel matrix as follows. Let K be the $m \times m$ kernel matrix with entries $K_{ij} = k(z_i, z_j)$. Moreover, let $\tilde{K}$ be the centered kernel matrix with entries $\tilde{K}_{ij} = \tilde{k}(z_i, z_j)$. With Eq. (9), one can express the centered kernel matrix as a function of the uncentered one:

$$\tilde{K} = K - KE - EK + EKE,$$

$$\text{where } E_{ij} = \frac{1}{m} \quad \forall i, j = 1, \ldots, m. \tag{13}$$

With these definitions, the eigenvalues $\lambda_1, \ldots, \lambda_r$ of the sample covariance matrix are given by $\lambda_k = (1/m)\tilde{\lambda}_k$, where $\tilde{\lambda}_k$ are the eigenvalues of $\tilde{K}$. And the expansion coefficients

$\{\alpha_i^k\}_{i=1,\ldots,m}$ in Eq. (12) form the components of the eigenvector of $\tilde{K}$ associated with the eigenvalue $\tilde{\lambda}_k$.

Inserting (11) splits energy (10) into two terms:

$$E_\phi(z) = \sum_{k=1}^{r} \lambda_k^{-1} (V_k, \tilde{\phi}(z))^2 + \lambda_\perp^{-1} \left(|\tilde{\phi}(z)|^2 - \sum_{k=1}^{r} (V_k, \tilde{\phi}(z))^2 \right). \qquad (14)$$

With expansion (12), we obtain the final expression of our energy:

$$E_\phi(z) = \sum_{k=1}^{r} \left(\sum_{i=1}^{m} \alpha_i^k \tilde{k}(z_i, z) \right)^2 \cdot (\lambda_k^{-1} - \lambda_\perp^{-1}) + \lambda_\perp^{-1} \cdot \tilde{k}(z,z). \qquad (15)$$

As in the case of kernel PCA, the nonlinearity ϕ only appears in terms of the kernel function. This allows to specify an entire family of possible nonlinearities by the choice of the associated kernel. For all our experiments we used the Gaussian kernel:

$$k(x,y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x-y\|^2}{2\sigma^2} \right). \qquad (16)$$

For a justification of this choice, we refer to Appendix A, where we show the relation of the proposed energy with the classical Parzen estimator.

4.2. Relation to kernel PCA

Just as in the linear case (cf. [26]), the regularization (11) of the covariance matrix causes a splitting of the energy into two terms (14), which can be considered as a *distance in feature space* (DIFS) and a *distance from feature space* (DFFS)—see Fig. 3. For the purpose of pattern reconstruction in the framework of kernel PCA, it was suggested to minimize a reconstruction error [27], which is identical with the DFFS. This procedure is based on the assumption that the entire plane spanned by the mapped training data corresponds to acceptable patterns. However, this is not a valid assumption: Already in the linear case, moving too far along an eigenmode will produce patterns which have almost no similarity to the training data, although they are still accepted by the hypothesis. Moreover, the distance DFFS is not based on a probabilistic model. In contrast, energy (15) is derived from a Gaussian probability distribution. It minimizes both the DFFS and the DIFS.

The kernel PCA approach has been studied in the framework of statistical shape models [28,29]. Our approach differs from these two in three ways: Firstly, our model is based on a probabilistic formulation of kernel PCA (as discussed above). Secondly, we derive a *similarity invariant* nonlinear shape model, as will be detailed in Section 6.2. Thirdly, we introduce the nonlinear shape dissimilarity measure as a shape prior in a variational framework for segmentation.

4.3. On the regularization of the covariance matrix

A regularization of the covariance matrix in the case of kernel PCA—as done in Eq. (11)—was first proposed in Ref. [15] and has also been suggested more recently in [30]. The choice of the parameter $\lambda_\perp$ is not a trivial issue. For the linear case, such regularizations of the covariance matrix have also been proposed [15,26,31,32]. There [26,32], the constant $\lambda_\perp$ is estimated as the mean of the replaced eigenvalues by minimizing the Kullback–Leibler distance of the corresponding densities. However, we believe that in our context this is not an appropriate regularization of the covariance matrix. The Kullback–Leibler distance is supposed to measure the error with respect to the correct density, which means that the covariance matrix calculated from the training data is assumed to be the correct one. But this is not the case because the number of training points is limited. For essentially the same reason this approach does not extend to the nonlinear case considered here: Depending on the type of nonlinearity ϕ, the covariance matrix is potentially infinite-dimensional such that the mean over all replaced eigenvalues will be zero. As in the linear case [6], we therefore propose to choose

$$0 < \lambda_\perp < \lambda_r, \qquad (17)$$

which means that unfamiliar variations from the mean are less probable than the smallest variation observed on the training set. In practice, we fix $\lambda_\perp = \lambda_r/2$.

4.4. On the choice of the hyperparameter σ

The last parameter to be fixed in the proposed density estimate is the hyperparameter σ in Eq. (16). Let μ be the average distance between two neighboring data points:

$$\mu^2 := \frac{1}{m} \sum_{i=1}^{m} \min_{j \neq i} |z_i - z_j|^2. \qquad (18)$$

In order to get a smooth energy landscape, we propose to choose σ in the order of μ. In practice, we used

$$\sigma = 1.5\mu \qquad (19)$$

for most of our experiments. We chose this somewhat heuristic measure μ for the following favorable properties:

- μ is insensitive to the distance of clusters, as long as each cluster has more than one data point,
- μ scales linearly with the data points,
- μ is robust with respect to the individual data points.

5. Density estimate for silhouettes of 2D and 3D objects

Although energy (10) is quadratic in the space Y of mapped points, it is generally not convex in the original space, showing several minima and level lines of essentially arbitrary shape. Fig. 4 shows artificial 2D data and the

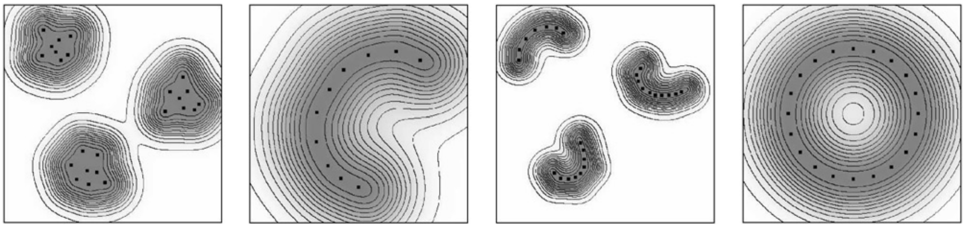

Fig. 4. *Density estimate* (10) *for artificial 2D data*. Distributions of variable shape are well estimated by the Gaussian hypothesis in *feature space*. We used the kernel (16) with $\sigma = 1.5\mu$.

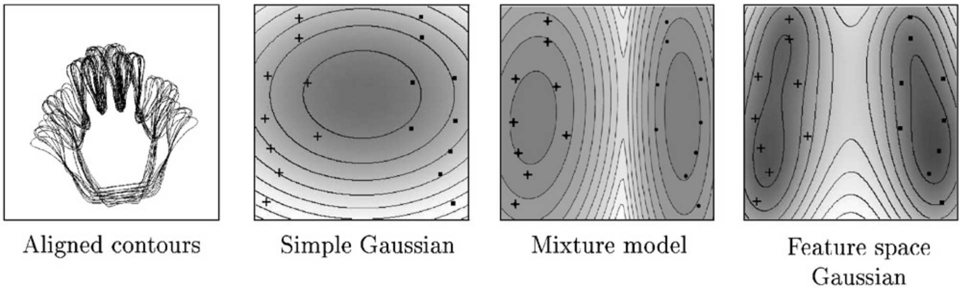

Fig. 5. *Model comparison*. Density estimates for a set of left (•) and right (+) hands, projected onto the first two principal components. *From left to right*: Aligned contours, simple Gaussian, mixture of Gaussians, Gaussian in feature space (10). Both the mixture model and the Gaussian in feature space capture the two-class structure of the data. However, the estimate in feature space is unsupervised and produces level lines which are not necessarily ellipses.

corresponding lines of constant energy $E_\phi(z)$ in the original space: The modes of the associated density are located around the clusters of the input data.

For a set of binarized views of objects we automatically fit a closed quadratic spline curve around each object. All spline curves have $N = 100$ control points, set equidistantly. The polygons of control points $z = (x_1, y_1, x_2, y_2, \ldots, x_N, y_N)$ are aligned with respect to translation, rotation, scaling and cyclic permutation [6]. This data was used to determine the density estimate $E_\phi(z)$ in Eq. (15).

For the visualization of the density estimate and the training shapes, all data was projected onto two of the principal components of a linear PCA. Note that due to the projection, this visualization only gives a very rough sketch of the true distribution in the 200-dimensional shape space.

Fig. 5 shows density estimates for a set of right hands and left hands. The estimates correspond to the hypotheses of a simple Gaussian in the original space, a mixture of Gaussians and a Gaussian in feature space. Although both the mixture model and our estimate in feature space capture the two distinct clusters, there are several differences: Firstly the mixture model is supervised—the number of classes and the class membership must be known—and secondly it only allows level lines of elliptical shape, corresponding to the hypothesis that each cluster by itself is a Gaussian distribution. The model of a Gaussian density in feature space does not assume any prior knowledge and produces level lines which capture the true distribution of the data even if individual classes do not correspond to hyperellipsoids.

This is demonstrated on a set of training shapes which correspond to different views of two 3D objects. Fig. 6 shows the two objects, their contours after alignment and the level lines corresponding to the estimated energy density (10) in appropriate 2D projections.

6. Nonlinear shape statistics in Mumford–Shah based segmentation

6.1. Minimization by gradient descent

Energy (10) measures the similarity of a shape C_z parameterized by a control point vector z with respect to a set of training shapes. For the purpose of segmentation, we combine this energy as a shape energy E_{shape} with the Mumford–Shah energy (3) in the variational approach (1).

The total energy (1) must be simultaneously minimized with respect to the control points defining the contour and with respect to the segmenting grey values $\{u_i\}$. Minimizing the modified Mumford–Shah functional (3) with respect to the contour C_z (for fixed $\{u_i\}$) results in the evolution

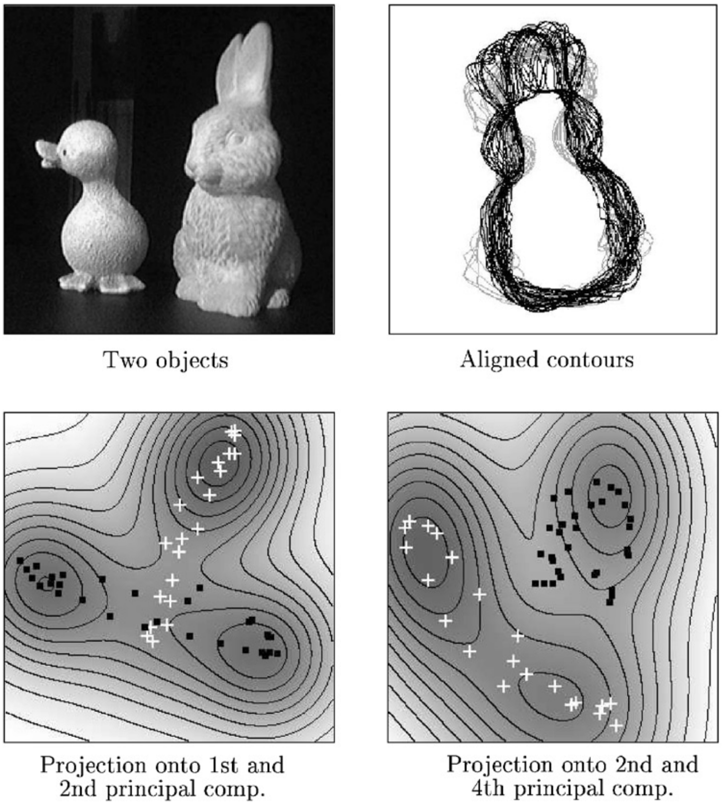

Fig. 6. *Density estimate for views of two 3D objects.* The training shapes of the duck (white +) and the rabbit (black •) form distinct clusters in shape space which are well captured by the energy level lines shown in appropriate 2D projections.

equation

$$\frac{\partial C_z(s,t)}{\partial t} = -\frac{\mathrm{d}E_{image}}{\mathrm{d}C_z} = (e_s^+ - e_s^-) \cdot n_s + v\,\frac{\mathrm{d}^2 C_z}{\mathrm{d}s^2}, \qquad (20)$$

where the terms e_s^+ and e_s^- denote the energy density $e = (f - u_i)^2$, inside and outside the contour $C_z(s)$, respectively, and n_s denotes the normal vector on the contour. The constants $\{u_i\}$ are updated in alternation with the contour evolution to be the mean grey value of the adjoining regions $\{\Omega_i\}$. The contour evolution equation (20) is transformed into an evolution equation for the control points z by introducing definition (2) of the contour as a spline curve. By discretizing on a set of nodes s_j along the contour we obtain a set of coupled linear differential equations. Solving for the coordinates of the ith control point and including the term induced by the shape energy we obtain:

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N} (\mathbf{B}^{-1})_{ij} [(e_j^+ - e_j^-) n_x(s_j,t) + v(x_{j-1} - 2x_j + x_{j+1})] - \alpha \left[\frac{\mathrm{d}E_{shape}(z)}{\mathrm{d}z}\right]_{2i-1},$$

$$\frac{\mathrm{d}y_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N} (\mathbf{B}^{-1})_{ij} [(e_j^+ - e_j^-) n_y(s_j,t) + v(y_{j-1} - 2y_j + y_{j+1})] - \alpha \left[\frac{\mathrm{d}E_{shape}(z)}{\mathrm{d}z}\right]_{2i}. \qquad (21)$$

The cyclic tridiagonal matrix **B** contains the spline basis functions evaluated at these nodes.

The three terms in the evolution equation (21) can be interpreted as follows:

- The first term forces the contour towards the object boundaries, by maximizing a homogeneity criterion in the adjoining regions, which compete in terms of their energy densities e^+ and e^-.
- The second term enforces an equidistant spacing of control points, thus minimizing the contour length. This prevents the formation of cusps during the contour evolution.
- The last term pulls the control point vector towards the domains of familiar shapes, thereby maximizing the similarity of the evolving contour with respect to the training shapes. It will be detailed in the next section.

6.2. Invariance in the variational framework

By construction, the density estimate (10) is not invariant with respect to translation, scaling and rotation of the shape C_z. We therefore propose to eliminate these degrees of freedom in the following way: Since the training shapes were aligned to their mean shape z_0 with respect to translation, rotation and scaling and then normalized to unit size, we shall do the same to the argument z of the shape energy before applying our density estimate E_ϕ.

We therefore define the shape energy by

$$E_{shape}(z) = E_\phi(\tilde{z}) \quad \text{with } \tilde{z} = \frac{R_\theta z_c}{|R_\theta z_c|}, \tag{22}$$

where z_c denotes the control point vector after centering:

$$z_c = \left(I_n - \frac{1}{n} A \right) z$$

$$\text{with } A = \begin{pmatrix} 1 & 0 & 1 & 0 & \cdots \\ 0 & 1 & 0 & 1 & \cdots \\ 1 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{23}$$

and R_θ denotes the optimal rotation of the control point polygon z_c with respect to the mean shape z_0. We will not go into details about the derivation of R_θ. Similar derivations can be found in Refs. [33,34]. The final result is given by the formula:

$$\tilde{z} = \frac{M z_c}{|M z_c|}$$

$$\text{with } M = I_n \otimes \begin{pmatrix} z_0^t z_c & -z_0 \times z_c \\ z_0 \times z_c & z_0^t z_c \end{pmatrix}, \tag{24}$$

where $\otimes$ denotes the Kronecker product and $z_0 \times z_c := z_0^t R_{\pi/2} z_c$.

The last term in the contour evolution equation (21) is now calculated by applying the chain rule:

$$\frac{dE_{shape}(z)}{dz} = \frac{dE_\phi(\tilde{z})}{d\tilde{z}} \cdot \frac{d\tilde{z}}{dz}$$

$$= \frac{dE_\phi(\tilde{z})}{d\tilde{z}} \cdot \frac{d\tilde{z}}{dz_c} \cdot \frac{dz_c}{dz}. \tag{25}$$

Since this derivative can be calculated analytically, no additional parameters enter the above evolution equation to account for scale, rotation and translation.

Other authors (cf. [35]) propose to explicitly model a translation, an angle and a scale and minimize with respect to these quantities (e.g. by gradient descent). In our opinion this has several drawbacks: Firstly, it introduces four additional parameters, which makes numerical minimization more complicated—parameters to balance the gradient descent must be chosen. Secondly this approach mixes the degrees of freedom corresponding to scale, rotation and shape deformation. And thirdly potential local minima may be introduced by the additional parameters. On several segmentation tasks we were able to confirm these effects by comparing the two approaches.

Since there exists a similar closed form solution for the optimal alignment of two polygons with respect to the more general affine group [33], the above approach could be extended to define a shape prior which is invariant with respect to affine transformations. However, we do not elaborate this for the time being.

7. Numerical results

In the following, we will present a number of numerical results obtained by introducing the similarity invariant nonlinear shape prior from Eqs. (22) and (15) into the Mumford–Shah based segmentation process as discussed above. The results are ordered so as to demonstrate different properties of the proposed shape prior.

7.1. Linear versus nonlinear shape prior

Compared to the linear case (5), the nonlinear shape energy is no longer convex. Depending on the input data, it permits the formation of several minima corresponding to different clusters of familiar contours. Minimization by gradient descent will end up in the nearest local minimum. In order to obtain a certain independence of the shape prior from the initial contour, we propose to first minimize the image energy E_{image} by itself until stationarity and to then include the shape prior E_{shape}. This approach guarantees that we will extract as much information as possible from the image before "deciding" which of the different clusters of accepted shapes the obtained contour resembles most.

Fig. 7 shows a simple example of three artificial objects. The shape prior (22) was constructed on the three aligned silhouettes shown on the top left. The mean of the three shapes (second image) indicates that the linear Gaussian is not a reliable model for this training set. The next images show the initial contour for the segmentation of a partially occluded image of the first object, the final segmentation without prior knowledge, the final segmentation after introducing the linear prior and the final segmentation upon introduction of the nonlinear prior. Rather than drawing the contour towards the mean shape (as does the linear prior), the nonlinear one draws the evolving contour towards one of the encoded shapes. Moreover, the *same* nonlinear prior permits a segmentation of an occluded version of the other encoded objects.

The bottom right image in Fig. 7 shows the training shapes and the density estimate in a projection on the first two axes of a (linear) PCA. The white curves correspond to the path of the segmenting contour from its initialization to its converged state for the two segmentation processes respectively. Note that upon introducing the shape prior the correspond-

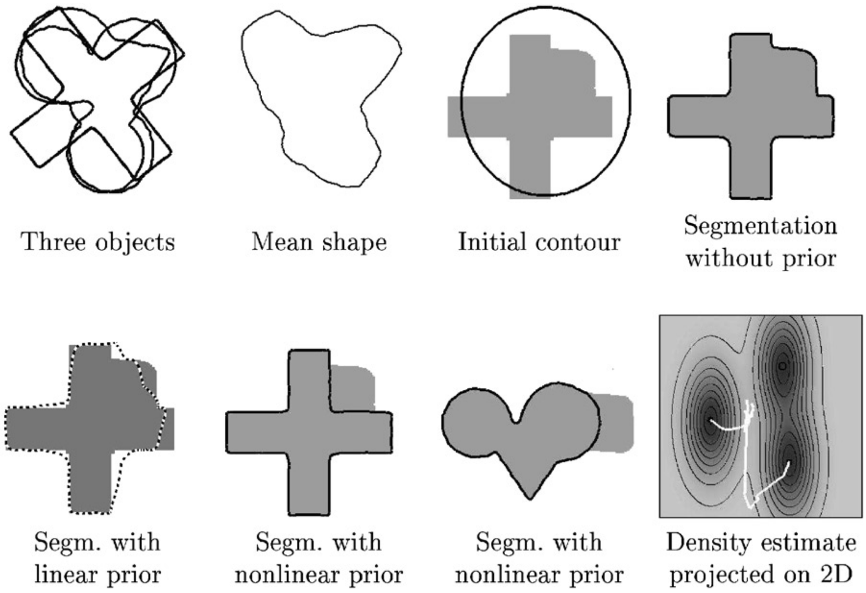
Three objects
Mean shape
Initial contour
Segmentation without prior
Segm. with linear prior
Segm. with nonlinear prior
Segm. with nonlinear prior
Density estimate projected on 2D

Fig. 7. *Segmenting partially occluded images of several objects*. While the linear prior draws the segmenting contour towards the mean shape, the nonlinear one permits the segmentation process to distinguish between the three training shapes. Introduction of the shape prior upon stationarity of the contour (top right) causes the contour to evolve normal to the level lines of constant energy into the nearest local minimum, as indicated by the white curves in the projected density estimate (bottom right).

ing contour descends the energy landscape in direction of the negative gradient to end up in one of the minima. The example shows that, in contrast to the linear shape prior, the nonlinear one can well separate different objects without mixing them. Since each cluster in this example contains only one view for the purpose of illustration, the estimate (19) for the kernel width σ does not apply; instead we chose a smaller granularity of $\sigma = \mu/4$.

7.2. Simultaneous encoding of several training objects

The following example is an application of our method which shows how the nonlinear shape prior can encode a number of different alphabetical letters and thus improve the segmentation of these letters in a given image.

We want to point out that there exists a vast number of different methods for optical character recognition. We do not claim that the present method is optimally suited for this task, and we do not claim that it outperforms existing methods. The following results only show that our rather general segmentation approach with the nonlinear shape prior can be applied to a large variety of tasks and that it permits to simultaneously encode the shape of *several* objects.

A set of 7 letters and digits were segmented (several times) without any shape prior in an input image as the one shown in Fig. 8(a). The obtained contours were used as a training set to construct the shape prior. Fig. 9 shows the set of aligned contours and their projection into the plane spanned by the first and third principal component (of a linear PCA). The clusters are labeled with the corresponding letters and digits. Again, the mean shape, shown in Fig. 8(c), indicates that the linear model is not an adequate model for the distribution of the training shapes.

In order to generate realistic input data, we subsampled the input image to a resolution of 16×16 pixels, as shown in Fig. 8(b). Such low resolution input data are typical in this context. As a first step, we upsampled this input data using bilinear interpolation, as shown in Fig. 8(c).

Given such an input image, we initialized the contour, iterated the segmentation process without prior until stationarity and then introduced either the linear or the nonlinear shape prior. Fig. 10 shows segmentation results without prior, with the linear prior and with the nonlinear prior. Again, the convergence of the segmenting contour towards one of the learnt letters is visualized by appropriate projections onto the first two linear principal components of the training contours.[3]

Fig. 11 shows results of the segmentation approach with the *same* nonlinear shape prior, applied to two more shapes. Again, the nonlinear shape prior improves the segmentation results. This demonstrates that one can encode information on a set of fairly different shapes into a single shape prior.

[3] For better visibility, the projection planes were shifted along the third principal component, so as to intersect with the cluster of interest.

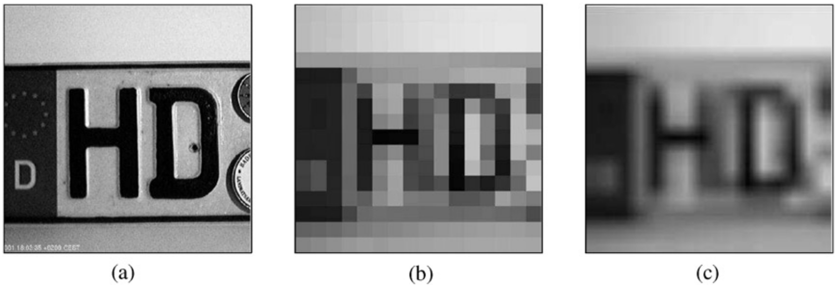

Fig. 8. (a) Original image region of 200 × 200 pixels. (b) Subsampled to 16 × 16 pixels (used as input data). (c) Upsampled low-resolution image using bilinear interpolation.

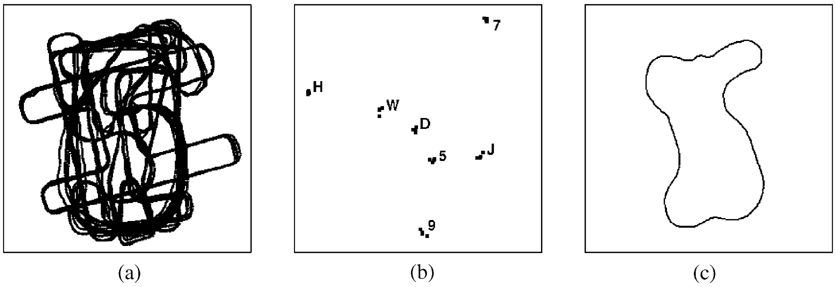

Fig. 9. (a) Aligned training shapes. (b) Projection onto the first and third (linear) principal component. (c) Mean shape.

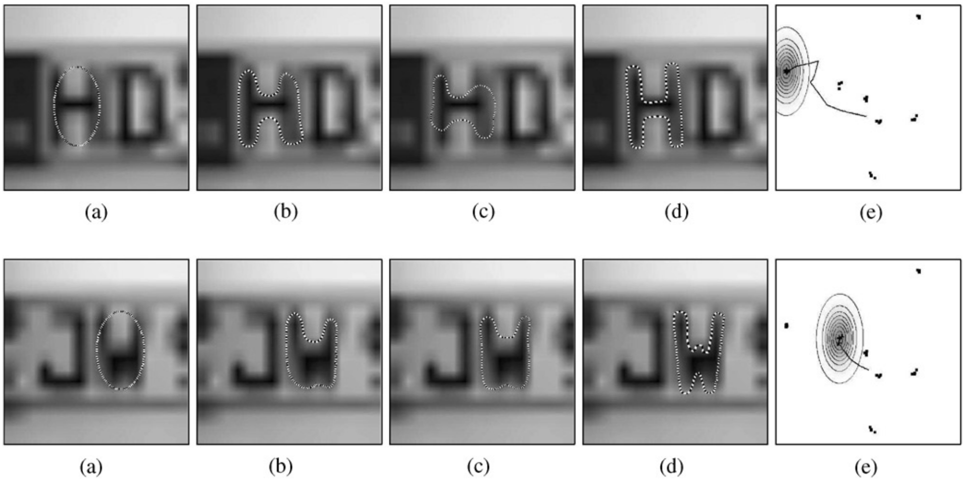

Fig. 10. Initial contour (a), final segmentation without prior (b), segmentation upon introduction of the linear prior (c), and final segmentation with the nonlinear prior (d). Appropriate projections of the contour evolution with nonlinear prior into the space of contours show the convergence of the contour towards one of the learnt letters (e).

7.3. Generalization to novel views

In all of the above examples, the nonlinear shape prior merely permitted a reconstruction of the training shapes (up to similarity transformations). The power of the proposed shape prior lies in the fact that not only it can encode several very different shapes, but also that the prior is a *statistical* prior: It has the capacity to generalize and abstract from the fixed set of training shapes. As a consequence, the respective segmentation process with the nonlinear prior is able to

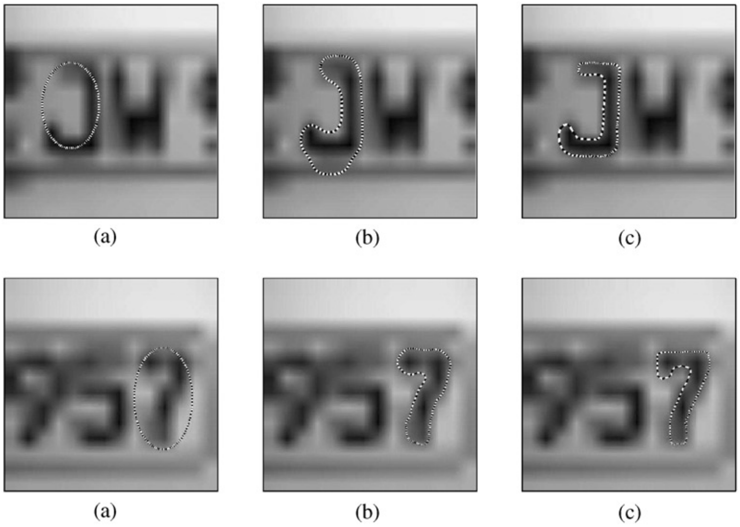

Fig. 11. Initial contour (a), final segmentation without prior (b), and final segmentation upon introduction of the nonlinear prior (c). With a single nonlinear prior, a number of fairly different shapes can be reconstructed from the subsampled and smoothed input image.

segment novel views of an object which were not present in the training set. This aspect of the nonlinear statistical shape prior will be demonstrated in the following examples.

The training set consists of nine right and nine left hands, shown together with the estimated energy density in a projection onto the first two principal components in Fig. 12, right side. Rather than mixing the two classes of right and left hands, the shape prior clearly separates several clusters in shape space. The final segmentations without (left) and with (center) prior shape knowledge show that the shape prior compensates for occlusion by filling up information where it is missing. Moreover, the statistical nature of the prior is demonstrated by the fact that the hand in the image is not part of the training set. This can be seen in the projection (Fig. 12, right side), where the final segmentation (white box) does not correspond to any of the training contours (black crosses).

7.4. Tracking 3D objects with changing viewpoint

In the following, we present results of applying the nonlinear shape statistics for an example of tracking an object in 3D with a prior constructed from a large set of 2D views. For this purpose we binarized 100 views of a rabbit—two of them and the respective binarizations are shown in Fig. 13. For each of the 100 views we automatically extracted the contours and aligned them with respect to translation, rotation, scaling and cyclic permutation of the control points. We calculated the density estimate (10) and the corresponding shape energy (22).

In a film sequence we moved and rotated the rabbit in front of a cluttered background. Moreover, we artificially introduced an occlusion afterwards. We segmented the first image by the modified Mumford–Shah model until convergence before the shape prior was introduced. The initial contour and the segmentations without and with prior are shown in Fig. 14. Afterwards we iterated 15 steps in the gradient descent on the full energy for each frame in the sequence.[4] Some sample screen shots of the sequence are shown in Fig. 15. Note that the viewpoint changes continuously.

The training silhouettes are shown in 2D projections with the estimated shape energy in Fig. 16. The path of the changing contour during the entire sequence corresponds to the white curve. The curve follows the distribution of training data well, interpolating in areas where there are no training silhouettes. Note that the intersection of the curve and of the training data in the center (Fig. 16, left side) are only due to the projection on 2D. The results show that—given sufficient training data—the shape prior is able to capture fine details such as the ear positions of the rabbit in the various views. Moreover, it generalizes well to novel views not included in the training set and permits a reconstruction of the occluded section throughout the entire sequence.

[4] The gradient of the shape prior in Eq. (15) has a complexity of $O(rmn)$, where n is the number of control points, m is the number of training silhouettes and r is the eigenvalue cutoff. For input images of 83 kpixels and $m = 100$, we measured an average runtime per iteration step of 96 ms for the prior, and 11 ms for the cartoon motion on a 1.2 GHz AMD Athlon. This permitted to do 6 iterations per second. Note, however, that the relative importance of the cartoon motion increases with the size of the image: For an image of 307 kpixels the cartoon motion took 100 ms per step. Note, however, that we did not put much effort into runtime optimization.

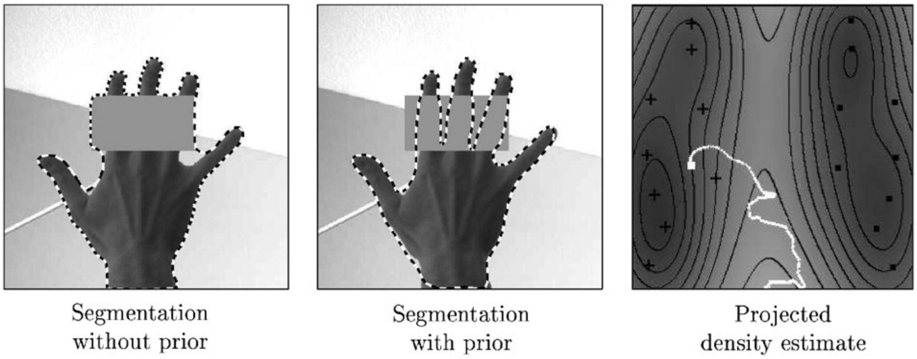

Fig. 12. *Segmentation with a nonlinear shape prior containing right* (+) *and left* (•) *hands*—shown in the projected energy plot on the right. The input image is a right hand with an occlusion. After the Mumford–Shah segmentation becomes stationary (left image), the nonlinear shape prior is introduced, and the contour converges towards the final segmentation (center image). The contour evolution in its projection is visualized by the white curve in the energy density plot (right). Note that the final segmentation (white box) does not correspond to any of the training silhouettes, nor to the minimum (i.e. the most probable shape) of the respective cluster.

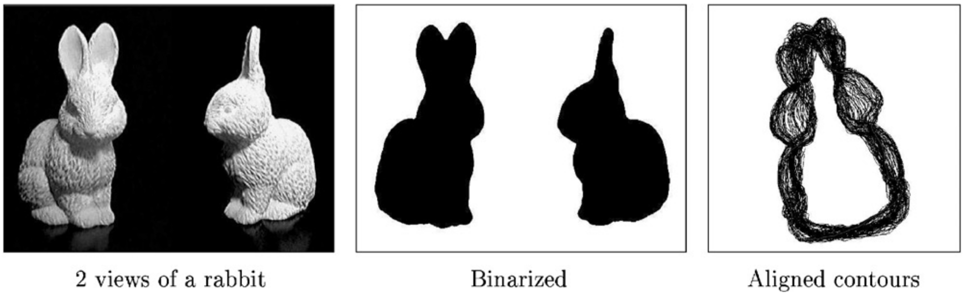

Fig. 13. Example views and binarization used for estimating the shape density.

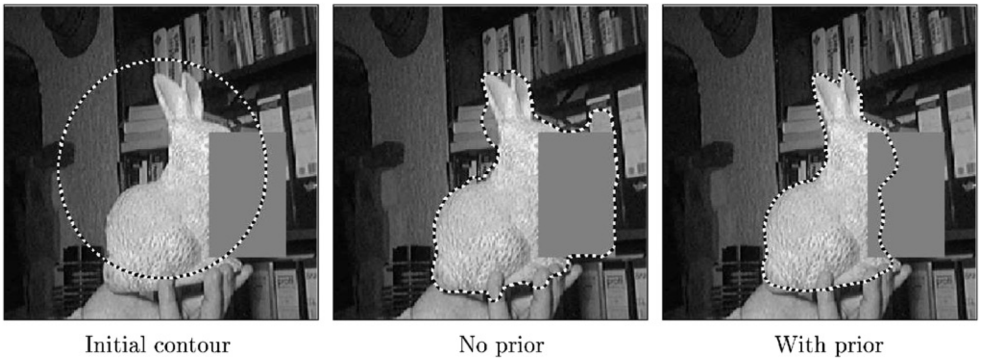

Fig. 14. *Begin of the tracking sequence*. Initial contour, segmentation without prior, segmentation upon introducing the nonlinear prior on the contour.

8. Conclusion

We presented a variational integration of nonlinear shape statistics into a Mumford–Shah based segmentation process. The statistics are derived from a novel method of density estimation which can be considered as an extension of the kernel PCA approach to a probabilistic framework. The original training data is nonlinearly transformed to a feature space. In this higher dimensional space the distribution of the mapped data is estimated by a Gaussian density. Due to the strong

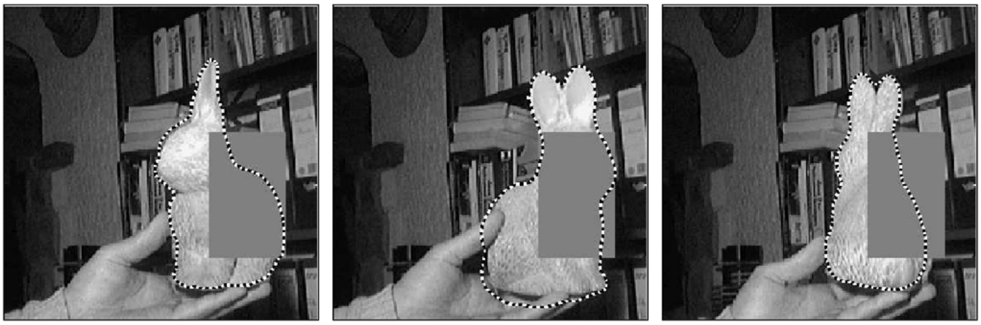

Fig. 15. Sample screen shots from the tracking sequence.

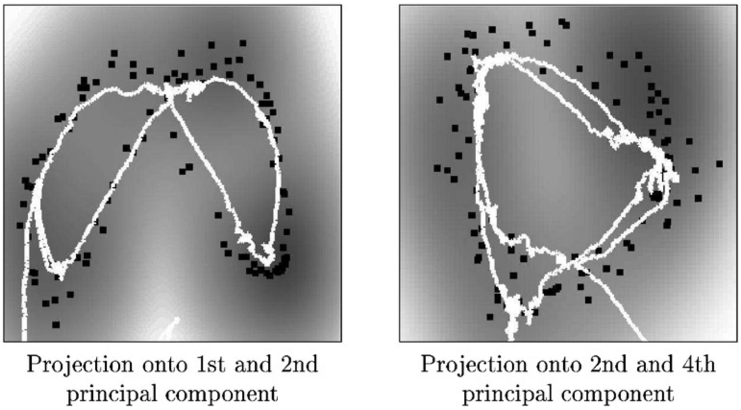

Fig. 16. *Tracking sequence visualized.* Training data (●), estimated energy density and the contour evolution (white curve) in appropriate 2D projections. The contour evolution is restricted to the valleys of low energy induced by the training data.

nonlinearity, the corresponding density estimate in the original space is highly non-Gaussian, allowing several shape clusters and banana-or ring-shaped data distributions.

We integrated the nonlinear statistics as a shape prior in a variational approach to segmentation. We gave details on appropriate estimations of the involved parameters. Based on the explicit representation of the contour, we proposed a closed-form, parameter-free solution for the integration of invariance with respect to similarity transformations in the variational framework.

Applications to the segmentation of static images and image sequences show several favorable properties of the nonlinear prior:

- Due to the possible multimodality in the original space, the nonlinear prior can encode a number of fairly different training objects.
- It can capture even small details of shape variation without mixing different views.
- It copes for misleading information due to noise and clutter, and enables the reconstruction of occluded parts of the object silhouette.
- Due to the statistical nature of the prior, a generalization to novel views not included in the training set is possible.

Finally we showed examples where the 3D structure of an object is encoded through a training set of 2D projections.

By projecting onto the first principal components of the data, we managed to visualize the training data and the estimated shape density. The evolution of the contour during the segmentation of static images and image sequences can be visualized by a projection into this density plot. In this way we verified that the shape prior effectively restricts the contour evolution to the submanifold of familiar shapes.

Acknowledgements

We thank P. Bouthemy, E. Mémin, T. Corpetti, C. Kervrann, A. Trubuil and the members of these groups for hospitality and many fruitful discussions, and J. Keuchel for comments on the manuscript.

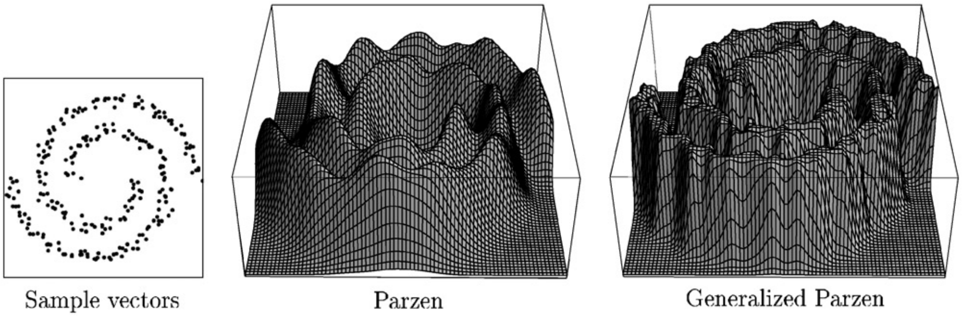

Fig. 17. Sample vectors randomly distributed on two spirals (*left*), corresponding estimates of Parzen (*middle*) and generalized Parzen (*right*) for appropriate values of the kernel width σ.

Appendix A. From feature space distance to the Parzen estimator

In this section, we will link the feature space distances which induce our shape dissimilarity measure to classical methods of density estimation. The derivation of the energy (15) was based on the assumption that the training data after a nonlinear mapping corresponding to the kernel (16) are distributed according to a Gaussian density in the space Y. The final expression (15) resembles the well-known Parzen estimator [36,37], which estimates the density of a distribution of training data by summing up the data points after convolution with a Gaussian (or some other kernel function).

In fact, the energy associated with an *isotropic* (spherical) Gaussian distribution in feature space is (up to normalization) equivalent to a Parzen estimator in the original space. With the definitions (8) and (9), this energy is given by the Euclidean feature space distance

$$E_{sphere}(z) = |\tilde{\phi}(z)|^2 = \tilde{k}(z,z)$$

$$= -\frac{2}{m}\sum_{i=1}^{m} k(z,z_i) + \text{const.} \tag{A.1}$$

Up to scaling and a constant, this is the Parzen estimator.

The proposed energy (10) can therefore be interpreted as a *generalization of the Parzen estimator* obtained by moving from a spherical distribution in feature space to an ellipsoidal one. Due to the regularization of the covariance matrix in (11), energy (10) contains a (dominant) isotropic component given by the last term in (15). We believe that this connection to the Parzen estimator justifies the assumption of a Gaussian in feature space and the choice of localized (stationary) kernels such as (16).

Numerical simulations show that the remaining anisotropic component in (15) has an important influence. Fig. 17 shows the example of a set of 2D points which were randomly sampled along two spirals (left). Middle and right image show the Parzen and the generalized Parzen for appropriate values of the kernel width σ. Note that the spiral structures are more pronounced by the generalized Parzen. However, a more detailed theoretical study of the difference between the Euclidean distance in feature space (A.1) and the Mahalanobis distance in feature space (10) is beyond the scope of this paper.

References

[1] M.E. Leventon, W.E.L. Grimson, O. Faugeras, Statistical shape influence in geodesic active contours, in: Proceedings of Conference on Computer Vision and Pattern Recognition, Vol. 1, Hilton Head Island, SC, June 13–15, 2000, pp. 316–323.

[2] D. Cremers, C. Schnörr, J. Weickert, Diffusion snakes: combining statistical shape knowledge and image information in a variational framework, in: IEEE First Workshop on Variational and Level Set Methods, Vancouver, 2001, pp. 137–144.

[3] L.H. Staib, J.S. Duncan, Boundary finding with parametrically deformable models, IEEE Trans. Pattern Anal. Mach. Intell. 14 (11) (1992) 1061–1075.

[4] C. Kervrann, F. Heitz, A hierarchical markov modeling approach for the segmentation and tracking of deformable shapes, Graphical Models Image Process. 60 (5) (1998) 173–195.

[5] T.F. Cootes, C. Breston, G. Edwards, C.J. Taylor, A unified framework for atlas matching using active appearance models, in: A. Kuba, M. Samal, A. Todd-Pokropek (Eds.), Proceedings of International Conference on Information Processing in Medical Imaging, Lecture Notes in Computer Science, Vol. 1613, Springer, Berlin, 1999, pp. 322–333.

[6] D. Cremers, F. Tischhäuser, J. Weickert, C. Schnörr, Diffusion snakes: introducing statistical shape knowledge into the Mumford–Shah functional, Int. J. Comput. Vision 50 (3) (2002) 295–313.

[7] B. Chalmond, S.C. Girard, Nonlinear modeling of scattered multivariate data and its application to shape change, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 422–432.

[8] T. Heap, D. Hogg, Automated pivot location for the Cartesian-polar hybrid point distribution model, in: British Machine Vision Conference, Edinburgh, UK, September 1996, pp. 97–106.

[9] T. Heap, D. Hogg, Wormholes in shape space: tracking through discontinuous changes in shape, in: International

Conference on Computer Vision, Edinburgh, UK, September 1998, pp. 97–106.

[10] T.F. Cootes, C.J. Taylor, A mixture model for representing shape variation, Image Vision Comput. 17 (8) (1999) 567–574.

[11] D. Hastie, W. Stuetzle, Principal curves, J. Am. Stat. Assoc. 84 (1989) 502–516.

[12] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, Philos. Trans. Roy. S. London, A 209 (1909) 415–446.

[13] R. Courant, D. Hilbert, Methods of Mathematical Physics, Vol. 1. Interscience Publishers, Inc., New York, 1953.

[14] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1998) 1299–1319.

[15] D. Cremers, T. Kohlberger, C. Schnörr, Nonlinear shape statistics via kernel spaces, in: B. Radig, S. Florczyk (Eds.), Pattern Recognition, Lecture Notes in Computer Science, Vol. 2191, Munich, Germany, September 2001, Springer, Berlin, pp. 269–276.

[16] D. Cremers, T. Kohlberger, C. Schnörr, Nonlinear shape statistics in Mumford–Shah based segmentation, in: A. Heyden et al. (Eds.), Proceedings of the European Conference on Computer Vision, Lecture Notes in Computer Science, Vol. 2351, Copenhagen, May 2002, Springer, Berlin, pp. 93–108.

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

[18] G. Farin, Curves and Surfaces for Computer-Aided Geometric Design, Academic Press, San Diego, 1997.

[19] S.C. Zhu, A. Yuille, Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 18 (9) (1996) 884–900.

[20] N. Paragios, R. Deriche, Coupled geodesic active regions for image segmentation: a level set approach, in: D. Vernon (Ed.), Proceedings of the European Conference on Computer Vision, Lecture Notes in Computer Science, Vol. 1843, Springer, Berlin, 2000, pp. 224–240.

[21] D. Cremers, C. Schnörr, Statistical shape knowledge in variational motion segmentation, Image Vision Comput. 21 (1) (2003) 77–86.

[22] D. Cremers, C. Schnörr, Motion competition: variational integration of motion segmentation and shape regularization, in: L. van Gool (Ed.), Pattern Recognition, Lecture Notes in Computer Science, Vol. 2449, Zürich, September 2002. Springer, Berlin, pp. 472–480.

[23] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vision 1 (4) (1988) 321–331.

[24] M.A. Aizerman, E.M. Braverman, L.I. Rozonoer, Theoretical foundations of the potential function method in pattern recognition learning, Autom. Remote Control 25 (1964) 821–837.

[25] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: D. Haussler (Ed.), Proceedings of the Fifth Annual ACM Workshop on Computer Learning Theory, ACM Press, Pittsburgh, PA, 1992, pp. 144–152.

[26] B. Moghaddam, A. Pentland, Probabilistic visual learning for object detection, in: Proceedings of IEEE International Conference on Computer Vision, Boston, MA, 1995, pp. 786–793.

[27] B. Schölkopf, S. Mika, A. Smola, G. Rätsch, Müller K.-R. Kernel, PCA pattern reconstruction via approximate pre-images, in: L. Niklasson, M. Boden, T. Ziemke (Eds.), International Conference on Artificial Neural Networks, Springer, Berlin, Germany, 1998, pp. 147–152.

[28] S. Romdhani, S. Gong, A. Psarrou, A multi-view non-linear active shape model using kernel pca, in: T. Pridmore, D. Elliman (Eds.), Proceedings of the British Machine Vision Conference, Vol. 2, BMVA Press, Nottingham, UK, September 1999. pp. 483–492.

[29] C.J. Twining, C.J. Taylor, Kernel principal component analysis and the construction of non-linear active shape models, in: T. Cootes, C. Taylor (Eds.), Proceedings of the British Machine Vision Conference, 2001, pp. 23–32.

[30] M.E. Tipping, Sparse kernel principal component analysis, in: Advances in Neural Information Processing Systems 13, Vancouver, December 2001.

[31] S. Roweis, EM algorithms for PCA and SPCA, in: M. Jordan, M. Kearns, S. Solla (Eds.), Advances in Neural Information Processing Systems, Vol. 10, MIT Press, Cambridge, MA, 1998, pp. 626–632.

[32] M.E. Tipping, C.M. Bishop, Probabilistic principal component analysis, Technical Report Woe-19, Neural Computing Research Group, Aston University, UK, 1997.

[33] M. Werman, D. Weinshall, Similarity and affine invariant distances between 2d point sets, IEEE Trans. Pattern Anal. Mach. Intell. 17 (8) (1995) 810–814.

[34] I.L. Dryden, K.V. Mardia, Statistical Shape Analysis, Wiley, Chichester, 1998.

[35] Y. Chen, S. Thiruvenkadam, H. Tagare, F. Huang, D. Wilson, E. Geiser, On the incorporation of shape priors into geometric active contours, in: IEEE Workshop on Variational and Level Set Methods, Vancouver, CA, 2001, pp. 145–152.

[36] F. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Stat. 27 (1956) 832–837.

[37] E. Parzen, On the estimation of a probability density function and the mode, Ann. Math. Stat. 33 (1962) 1065–1076.