Variational Segmentation with Shape Priors

Martin Bergtholdt, Daniel Cremers, and Christoph Schnörr

ABSTRACT We discuss the design of shape priors for variational region-based segmentation. By means of two different approaches, we elucidate the critical design issues involved: representation of shape, use of perceptually plausible dissimilarity measures, Euclidean embedding of shapes, learning of shape appearance from examples, combining shape priors and variational approaches to segmentation. The overall approach enables the appearance-based segmentation of views of 3D objects, without the use of 3D models.

Key words: variational models, image segmentation, contours, statistical learning, shape clustering, shape manifolds, statistical priors, Euclidean embedding, optimization, visual perception, Bayesian inference

1 Introduction

Variational models [17, 24] are the basis of established approaches to image segmentation in computer vision. The key idea is to generate a segmentation by locally optimizing appropriate cost functionals defined on the space of contours. The respective functionals are designed to maximize certain criteria regarding the low-level information such as edge consistency or (piecewise) homogeneity of intensity, color, texture, motion, or combinations thereof.

Yet, in practice the imposed models only roughly approximate the true intensity, texture or motion of specific objects in the image. Intensity measurements may be modulated by varying and complex lighting conditions. Moreover, the observed images may be noisy and objects may be partially occluded. In such cases, algorithms which are purely based on low-level properties will invariably fail to generate the desired segmentation.

An interpretation of these variational approaches in the framework of Bayesian inference shows that the above methods all impose a prior on the space of contours which favors boundaries of minimal length. While the resulting length constraint in the respective cost functionals has a strongly regularizing effect on the generated contour evolutions, this purely geometric prior lacks any experimental evidence. In practical applications, an algorithm which favors shorter boundaries may lead to the cutting of corners and the suppression of small-scale structures.

Given one or more silhouettes of an object of interest, one can construct shape

priors which favor objects that are in some sense *familiar*. In recent years, it was suggested to enhance variational segmentation schemes by imposing such object-specific shape priors. This can be done either by adding appropriate shape terms to the contour evolution [21, 33] or in a probabilistic formulation which leads to an additional shape term in the resulting cost functional [10, 27, 22]. By extending segmentation functionals with a shape prior, knowledge about the appearance of objects can be directly combined with clues given by the image data in order to cope with typical difficulties of purely data-driven image processing caused by noise, occlusion, etc.

The design of shape priors strongly depends on ongoing work on statistical shape models [6, 12, 18]. In particular, advanced models of shape spaces, shape distances, and corresponding shape transformations have been proposed recently [36, 15, 31, 3, 19, 29]. Concerning variational segmentation, besides attempting to devise "intrinsic" mathematical representations of shape, further objectives which have to be taken into account include the gap between mathematically convenient representations and representations conforming to properties of human perception [34, 23, 1], the applicability of statistical learning of shape appearance from examples, and the overall variational approach from the viewpoint of optimization.

The objective of this paper is to discuss these issues involved in designing shape priors for region-based variational segmentation by means of two representative examples: (i) non-parametric statistics applied to the standard Euclidean embedding of curves in terms of shape vectors, and (ii) perceptually plausible matching functionals defined on the shape manifold of closed planar curves. Both approaches are powerful, yet quite different with respect to the representation of shape, and of shape appearance. Their properties will be explained in the following sections, in view of the overall goal – variational segmentation.

Section 2 discusses both the common representation of shapes by shape vectors, and the more general representation by dissimilarity structures. The latter is mathematically less convenient, but allows for using distance measures which conform to findings of psychophysics. Learning of shape appearance is described in Section 3. The first approach encodes shape manifolds globally, whereas the second approach employs structure-preserving Euclidean embedding and shape clustering, leading to a collection of locally-linear representations of shape manifolds. The incorporation of corresponding shape priors into region-based variational approaches to segmentation is discussed in Section 4.

We confine ourselves to parametric planar curves and do not consider the more involved topic of shape priors for implicitly defined and multiply connected curves – we refer the reader to [21, 33, 4, 27, 3, 9] for promising advances in this field. Nevertheless, the range of models addressed are highly relevant from both the scientific and the industrial viewpoint of computer vision.

2 Shape Representation

One generally distinguishes between *explicit* (parametric) and *implicit* contour representations. In the context of image segmentation, implicit boundary representations have gained popularity due to the introduction of the level set method, which allows to propagate implicitly represented interfaces by appropriate partial differential equations acting on the corresponding embedding surfaces. The main advantages of representing and propagating contours implicitly are that one does not need to deal with control/marker point regridding and can elegantly (without heuristics) handle topological changes of the evolving boundary.

On the other hand, explicit representations also have several advantages. In particular, they provide a compact (low-dimensional) representation of contours and concepts such as intrinsic alignment, group invariance and statistical learning are more easily defined. Moreover, as we shall see in this work, the notion of corresponding contour points (and contour parts) arises more naturally in an explicit representation. In this work, we will only consider explicit simply-connected closed contours.

2.1 Parametric Contour Representations, Geometric Distances, and Invariance

Let

$$\mathbf{c} : [0,1] \to \Omega \subset \mathbb{R}^2 \tag{1.1}$$

denote a parametric closed contour in the image domain Ω. Throughout this paper, we use the finite-dimensional representation of 2D-shapes in terms of uniform periodic cubic B-splines [13]:

$$\mathbf{c}(s) = \sum_{m=1}^{M} \mathbf{p}_m B_m(s) = \mathbf{P}\mathbf{b}(s) \;, \tag{1.2}$$

with control points $\{\mathbf{p}_i\}$ and basis functions $\{B_i(s)\}$:

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_M \end{bmatrix} \;, \qquad \mathbf{b}(s) = \begin{pmatrix} B_1(s) & B_2(s) & \dots & B_M(s) \end{pmatrix}^\top$$

Well-known advantages of this representation include the compact support of the basis functions and continuous differentiability up to second order. Yet, most of our results also hold for alternative explicit contour representations.

Using the natural uniform sampling $\{s_1, \dots, s_M\}$ of the parameter interval, we stack together the corresponding collection of curve points, to form *shape vectors* representing the contour. For simplicity, and with slight abuse of notation, we denote them again with[1]:

$$\mathbf{c} := \left(\mathbf{c}(s_1)^\top, \dots, \mathbf{c}(s_M)^\top\right)^\top \in \mathbb{R}^{2M} \tag{1.3}$$

[1] In the following, it will be clear from the context whether $\mathbf{c}$ denotes a contour (1.1) or a shape vector (1.3).

Note, that there is a one-to-one correspondence between shape vectors $\mathbf{c}$ and corresponding control points $\{\mathbf{p}_i\}_{i=1,\dots,M}$ through the symmetric and sparse positive-definite matrix: $\mathbf{B} = \begin{pmatrix}\mathbf{b}(s_1) & \dots & \mathbf{b}(s_M)\end{pmatrix}^\top$.

We consider a simple geometric distance measure between contours which is invariant under similarity transformations:

$$d^2(\mathbf{c}_1, \mathbf{c}_2) = \min_{s,\theta,\mathbf{t}} |\mathbf{c}_1 - s\mathbf{R}_\theta \mathbf{c}_2 - \mathbf{t}|^2 \tag{1.4}$$

Here, the planar rotation $\mathbf{R}_\theta$ and translation $\mathbf{t}$ are defined according to the definition (1.3) of shape vectors:

$$\mathbf{R}_\theta = \mathbf{I}_M \otimes \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix}t_1, t_2, \dots, t_1, t_2\end{pmatrix}^\top,$$

and s is the scaling parameter. The solution to (1.4) can be computed in closed-form [12, 18]. Extensions of this alignment to larger transformation groups such as affine transformations are straight-forward Furthermore, since the locations of the starting points $\mathbf{c}_1(0), \mathbf{c}_2(0)$ are unknown, we minimize (1.4) over all cyclic permutations of the contour points defining $\mathbf{c}_2$.

2.2 Matching Functionals and Psychophysical Distance Measures

It is well-known that there is a gap between distance measures with mathematically convenient properties like (1.4), for example, and distance measures which conform with findings of psychophysics [34]. In particular, this observation is relevant in connection with shapes [23].

Given two arc-length parametrized curves $\mathbf{c}_1(t), \mathbf{c}_2(s)$, along with a diffeomorphism $t = g(s)$ smoothly mapping the curves onto each other, then corresponding studies [1] argued that matching functionals for evaluating the quality of the mapping g based on low-order derivatives, should involve stretching $g'(s)$ and bending (change of curvature) of the curves (cf. Figure 1).

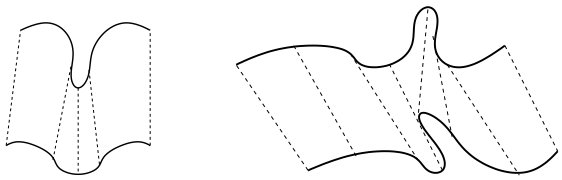

FIGURE 1. Stretching and bending of contours does not affect perceptually plausible matchings.

As a representative, we consider the matching functional [1]:

$$E(g;\, \mathbf{c}_1, \mathbf{c}_2) = \int_0^1 \frac{[\kappa_2(s) - \kappa_1(g(s))g'(s)]^2}{|\kappa_2(s)| + |\kappa_1(g(s))g'(s)|} ds + \lambda \int_0^1 \frac{|g'(s) - 1|^2}{|g'(s)| + 1} ds \tag{1.5}$$

where $\kappa_1(t), \kappa_2(s)$ denote the curvature functions of the contours $\mathbf{c}_1, \mathbf{c}_2$. The two terms in (1.5) take into account the bending and stretching of contours, respectively (see Figure 2).

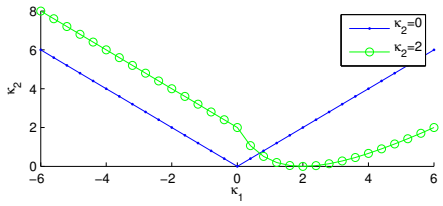

FIGURE 2. Local matching cost The local cost for bending of the matching functional (1.5) as a function of the κ_1, for two values of κ_2. Note how in the case $\kappa_2 = 2$, relatively lower costs for $\kappa_1 \approx 2$ allow for significant bending, without affecting matching too much.

Functional (1.5) favors perceptually plausible matching because it accounts that often object are structured into nearly convex-shaped parts separated by concave extrema. In particular, for non-rigid objects, parts are likely to articulate, and the matching functional produces articulation costs only at part boundaries.

From the mathematical viewpoint, functional (1.5) is invariant to rotation and translation of contours, and also to scaling provided both contours are normalized to length one. This is always assumed in what follows below. Furthermore, by taking the q-th root of the integral of local costs, where $q > 2.4$, (1.5) defines a *metric* between contours [1]:

$$d_E(\mathbf{c}_1, \mathbf{c}_2) := \min_g E(g;\, \mathbf{c}_1, \mathbf{c}_2)^{1/q} \tag{1.6}$$

Clearly, this distance measure is mathematically less convenient than (1.4). This seems to be the price for considering findings of psychophysics. However, regarding variational segmentation, we wish to work in this more general setting as well. For a discussion of further mathematical properties of matching functionals, we refer to [32].

The minimization in (1.6) is carried out by dynamic programming over all piecewise-linear and strictly monotonously increasing functions g. Figure 3 illustrates the result for two human shapes.

3 Learning Shape Statistics

Based on the shape representations described in Section 2, we consider in this section two approaches to the statistical learning of shape appearance from examples. The common basis for both approaches are Euclidean embeddings of shapes.

The first approach uses the embedding of shape vectors into Reproducing Kernel Hilbert Spaces by means of kernel functions, leading to a non-parametric global representation of shape manifolds. The second approach uses embeddings

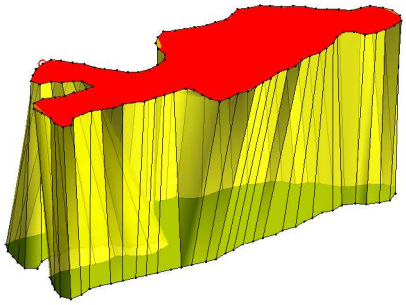

FIGURE 3. Matching by minimizing (1.5) leads to an accurate correspondence of parts of non-rigid objects, here illustrated for two human shapes.

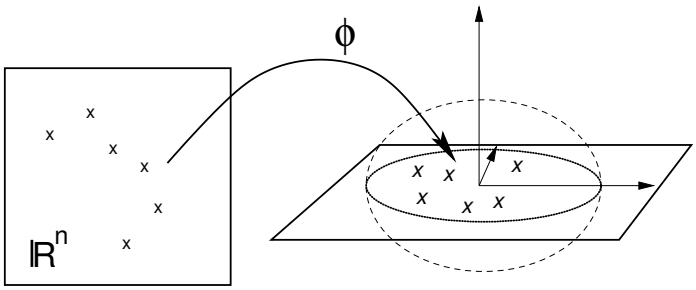

FIGURE 4. Gaussian density estimate upon nonlinear transformation to features space.

of dissimilarity structures by multidimensional scaling, along with a cluster-preserving modification of the dissimilarity matrix. Subsequent clustering results in a collection of local encodings of shape manifolds, and in corresponding aspect graphs of 3D objects in terms of prototypical object views.

3.1 Shape Distances in Kernel Feature Space

Let $\{\mathbf{c}_n\}_{n=1,\ldots,N} \in \mathbb{R}^{2M}$ denote the shape vectors associated with a set of training shapes. In order to model statistical shape dissimilarity measures, it is commonly suggested to approximate the distribution of training shapes by a Gaussian distribution, either in a subspace formed by the first few eigenvectors [6], or in the full $2M$-dimensional space [10]. Yet, for more complex classes of shapes – such as the various silhouettes corresponding to different 2D views of a 3D object – the assumption of a Gaussian distribution fails to accurately represent the distribution underlying the training shapes.

In order to model more complex (non-Gaussian and multi-modal) statistical distributions, we propose to embed the training shapes into an appropriate Reproducing Kernel Hilbert Space (RKHS) [35], and estimate Gaussian densities there – see Figure 4 for a schematic illustration.

A key assumption in this context is that only scalar products of embedded shape vectors $\phi(\mathbf{c})$ have to be evaluated in the RKHS, which is done in terms of a kernel function:

$$K(\mathbf{c}_1, \mathbf{c}_2) = \langle \phi(\mathbf{c}_1), \phi(\mathbf{c}_2) \rangle \tag{1.7}$$

Knowledge of the embedding map $\phi(\mathbf{c})$ itself is not required. Admissible kernel functions, including the Gaussian kernel, guarantee that the Gramian matrix

$$\mathbf{K} = \big\{ K(\mathbf{c}_i, \mathbf{c}_j) \big\}_{i,j=1,\ldots,N} \tag{1.8}$$

is positive definite [35]. This "non-linearization strategy" has been successfully applied in machine learning and pattern recognition during the last decade, where the RKHS is called *feature space*.

Based on this embedding of given training shapes, we use the following Mahalanobis distance:

$$J_S(\mathbf{c}) \;=\; (\phi(\mathbf{c}) - \phi_0)^\top \; \Sigma_\phi^{-1} (\phi(\mathbf{c}) - \phi_0) \;, \tag{1.9}$$

where ϕ_0 is the empirical mean, and Σ_ϕ is the corresponding covariance matrix. Note that all evaluations necessary to compute $J_S(\mathbf{c})$ in (1.9) can be traced back to evaluations of the kernel function according to (1.7). Furthermore, by exploiting the spectral decomposition of the kernel matrix $\mathbf{K}$ in (1.8), we regularize the covariance matrix Σ_ϕ with respect to its small and vanishing eigenvalues, thus defining two orthogonal subspaces as illustrated in Figure 4 on the right. For further details, we refer to [8].

3.2 Structure-Preserving Embedding and Clustering

Based on the matching functional (1.5) and the corresponding distance measure $d_E(\mathbf{c}_1, \mathbf{c}_2)$ defined in (1.6), we consider an arbitrary sample set $\{\mathbf{c}_n\}_{n=1,\ldots,N}$. To perform statistical analysis, we wish to compute an Euclidean embedding $\{\mathbf{x}_n\}_{n=1,\ldots,N}$ such that $\|\mathbf{x}_i - \mathbf{x}_j\| = d_E(\mathbf{c}_i, \mathbf{c}_j) \;,\; \forall i, j$. Such an embedding exists iff the matrix $\mathbf{K} = -\frac{1}{2}\mathbf{Q}\mathbf{D}\mathbf{Q}$, with the dissimilarity matrix $\mathbf{D} = (d_E(\mathbf{c}_i, \mathbf{c}_j)^2)$ and the centering matrix $\mathbf{Q} = \mathbf{I} - \frac{1}{M}\mathbf{e}\mathbf{e}^\top$, is positive semidefinite [7]. The vectors $\mathbf{x}_n$ representing the objects (contours) $\mathbf{c}_n$ of our data structure can then be computed by a Cholesky factorization of $\mathbf{K}$.

Figure 5 shows the eigenvalues of $\mathbf{K}$ for four different objects. The graphs illustrate that the contours are "almost embeddable" since only few and small eigenvalues are negative. This fact is caused by the powerful matching which tightly groups given curves, and is performed by evaluating the distance measure d_E. The standard way then is to take the positive eigenvalues only, and to compute a *distorted* embedding.

In view of subsequent clustering, however, a better alternative is to regularize the data structure by shifting the off-diagonal elements of the dissimilarity matrix: $\tilde{\mathbf{D}} = \mathbf{D} - 2\lambda_N(\mathbf{e}\mathbf{e}^\top - \mathbf{I})$. For the resulting embedding, it has been shown [26] that the group structure with respect to subsequent k-means clustering is preserved.

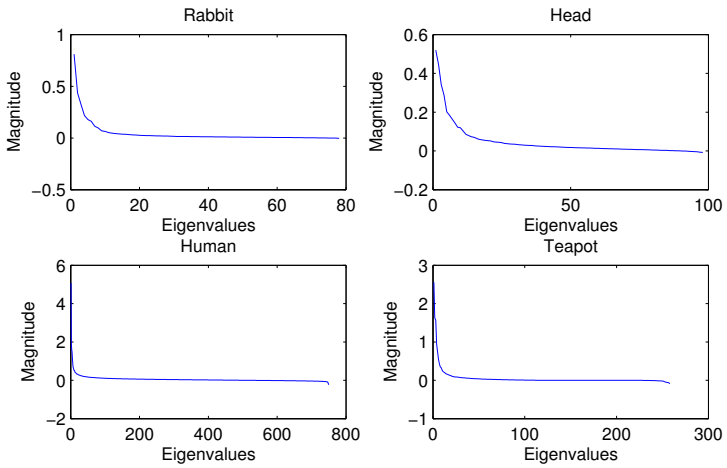

FIGURE 5. Eigenvalues of the matrices $\mathbf{K}$ corresponding to the shapes of four different objects.

Figure 6 shows a low-dimensional – and thus a heavily distorted – projection of the embedded shapes of the rabbit. For the purpose of illustration, only shapes corresponding to a single (hand-held) walk around the view-sphere are shown on the left, along with cluster centers as prototypical views of the object. In this way, we compute high-quality aspect graphs for *general* objects, without any restrictions discussed in the literature [2, 25].

On the right, Figure 6 also shows a clustering of 750 human shapes. In general, when using simple geometric distance measures, the many degrees of freedom of articulated shapes would require many templates for an accurate representation. The matching distance (1.6), however, accounts for part structure and, therefore, the principal components of the measure seem to be closer related to topological shape properties. For example, the clusters on the left are all "single-leg" prototypes, whereas on the right we find only clusters with two legs. The second principal component seems to account for the viewing direction of the human, which changes from left to right along a vertical direction through the plot.

4 Variational Segmentation and Shape Priors

4.1 Variational Approach

We consider partitions $\Omega = \overline{\Omega(F)} \cup \Omega(B)$ of the image domain into foreground and background, respectively. Our objective is to compute an optimal partition in terms of a planar closed curve $\mathbf{c}(s) = \partial\Omega(F)$ based on the corresponding restrictions of the image function $F = I|_{\Omega(F)}, B = I|_{\Omega(B)}, G = I|_{\mathbf{c}(s)}$, and by using models $\mathcal{H} = (\mathcal{H}_F, \mathcal{H}_B, \mathcal{H}_G, \mathcal{H}_S)$ for these components, including a shape

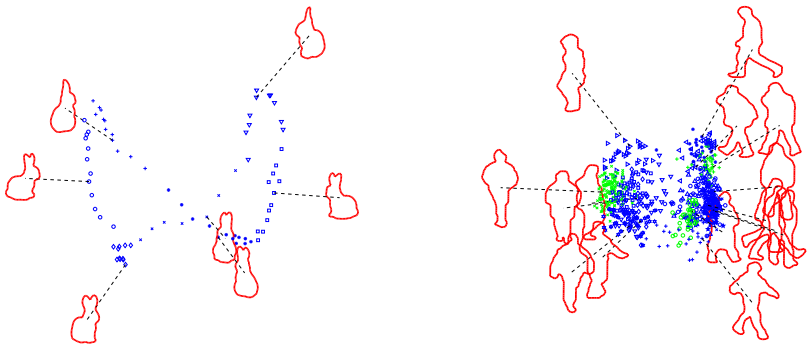

FIGURE 6. Clustering of the views of the rabbit sequence and the human shapes, projected to the first two principal components. The clusters are indicated by prototypical shapes (cluster centers) dominating a range of corresponding views.

prior $\mathcal{H}_S$ for the separating curve $\mathbf{c}(s)$.

The variational approach is to compute the *Maximum A-Posteriori (MAP)* estimate of the contour $\mathbf{c}$, given the image data I, and using the models $\mathcal{H}$:

$$\hat{\mathbf{c}}(s) = \arg\max_{\mathbf{c}(s)} P(\mathbf{c}(s)|I,\mathcal{H}) \tag{1.10}$$

We use Bayes' rule to obtain:

$$\begin{aligned} P(\mathbf{c}(s)|I,\mathcal{H}) &= \frac{P(I|\mathbf{c}(s),\mathcal{H})P(\mathbf{c}(s)|\mathcal{H})}{P(I|\mathcal{H})} \\ &\propto P(F|\mathbf{c}(s),\mathcal{H}_F)P(B|\mathbf{c}(s),\mathcal{H}_B)P(G|\mathbf{c}(s),\mathcal{H}_G)P(\mathbf{c}(s)|\mathcal{H}_S)\ , \end{aligned}$$

where we have also split up the image likelihood $P(I|\mathbf{c}(s),\mathcal{H})$ into three parts, assuming independence of these parts, given the contour $\mathbf{c}(s)$. Moreover, we assume independence of the various models. This assumption is appropriate in the single object – single object class scenario considered here.

The common form of the foreground model is:

$$P(F|\mathbf{c}(s),\mathcal{H}_F) \propto \exp(-J_F)\ , \quad J_F(\mathbf{c}) = \int_{\Omega(F)} d_F\big(F(\mathbf{x})\big)d\mathbf{x}\ ,$$

where the functional J_F depends on the contour $\mathbf{c}$ through the domain of integration $\Omega(F)$, and d_F is any measure of homogeneity of the foreground image data F, i.e. object appearance. Typically, d_F is a parametric model, a semi-parametric (mixture) model, or even a non-parametric model of the local spatial statistics of the image data, or some filter outputs. Note that d_F depends on $\mathbf{c}$ through the domain of integration, too. Similarly, we have:

$$\begin{aligned} P(B|\mathbf{c}(s),\mathcal{H}_B) &\propto \exp(-J_B)\ , \quad J_B(\mathbf{c}) = \int_{\Omega(B)} d_B\big(B(\mathbf{x})\big)d\mathbf{x}\ , \\ P(G|\mathbf{c}(s),\mathcal{H}_G) &\propto \exp(-J_G)\ , \quad J_G(\mathbf{c}) = \oint_c d_G\big(G(\mathbf{x})\big)ds \end{aligned}$$

In the following, we do not consider boundary models $P(G|\mathbf{c}(s), \mathcal{H}_G)$, but focus in the following two sections on shape models $P(\mathbf{c}(s)|\mathcal{H}_S)$, the main topic of this paper.

In order to solve (1.10), we minimize $-\log P(\mathbf{c}(s)|I, \mathcal{H})$, which entails to compute the derivatives of the above functionals with respect to $\mathbf{c}$, that is changes of the shape of the domain $\Omega(F)$. Let $\mathbf{v}(\mathbf{x})$ be a small and smooth vector field such that $(I + \mathbf{v})(\mathbf{x})$ is a diffeomorphism of the underlying domain. Then standard calculus [30, 11] yields:

$$\langle J_F'(\mathbf{c}), \mathbf{v} \rangle = \int_{\Omega(F)} d_F'\big(F(\mathbf{x})\big) d\mathbf{x} + \oint_{\mathbf{c}} d_F\big(F(\mathbf{x})\big)(\mathbf{n} \cdot \mathbf{v}) ds \;, \tag{1.11}$$

where $\mathbf{n}$ is the outer unit normal vector of $\Omega(F)$. Analogously, we compute the derivative of the background functional J_B.

If d_F depends on *parameters* which are estimated within $\Omega(F)$, then computing d_F' amounts to apply the chain rule until we have to differentiate (functions of) image data which do *not* depend on the domain (see, e.g., [16] for examples). As a result, the right hand side of (1.11) involves boundary integrals only. If, however, d_F more generally depends on *functions* which, in turn, depend on the shape of $\Omega(F)$, e.g. through some PDE, then the domain integral in (1.11) involving the unknown domain derivative d_F' can be evaluated in terms of a boundary integral by using an "adjoint state". See [28] for details and a representative application.

Finally, we set the normal vector field $v_n := \mathbf{n} \cdot \mathbf{v}$ equal to the *negative* integrand of the overall boundary integral resulting from the computation of J_F', J_B', and evolve the contour:

$$\dot{\mathbf{c}} = v_n \mathbf{n} \quad \text{on } \partial\Omega(F) \tag{1.12}$$

Inserting (1.2) yields a system of ODEs which are solved numerically.

Evolution (1.12) constitutes the data-driven part of the variational segmentation approach (1.10), conditioned on appearance models of both the foreground object and the background. In the following two sections, we describe how this approach is complemented in order to take into account statistical shape knowledge of object appearance.

4.2 Kernel-based Invariant Shape Priors

Based on the shape-energy (1.9), the shape-prior takes the form:

$$P(\mathbf{c}|\mathcal{H}_S) \propto \exp(-J_S)$$

Invariance with respect to similarity transforms is achieved by restricting the shape energy functional J_S to *aligned shapes* $\hat{\mathbf{c}} = \hat{\mathbf{c}}(\mathbf{c})$ with respect to the mean shape, which result from given shapes $\mathbf{c}$ by applying to them the translation, rotation and scaling parameters defining the invariant distance measure (1.4):

$$J_S(\mathbf{c}) = J_S[\hat{\mathbf{c}}(\mathbf{c})]$$

To incorporate the statistical shape-knowledge into the variational segmentation approach, we perturb the evolution (1.12) by adding a small vector field directed towards the negative gradient of J_S:

$$\mathbf{v} = -\varepsilon \frac{\mathrm{d}J_S}{\mathrm{d}\hat{\mathbf{c}}} \frac{\mathrm{d}\hat{\mathbf{c}}}{\mathrm{d}\mathbf{c}}$$

For further details, we refer to [8].

4.3 Shape Priors based on the Matching Distance

Related to the KPCA approach (Sections 3.1, 4.2), we use a non-parametric density estimate for the posterior of $\mathbf{c}$ given the training samples $\mathbf{c}_1, \ldots, \mathbf{c}_N$:

$$P(\mathbf{c}|\mathcal{H}_S) = p(\mathbf{c}|\mathbf{c}_1, \ldots, \mathbf{c}_N)$$

Given the Euclidean embedding $\mathbf{x}_1, \ldots, \mathbf{x}_N$ of the training samples (cf. Section 3.2), the kernel-estimate of the probability density evaluated at $\mathbf{x}$ reads:

$$p(\mathbf{x}) \approx p_N(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{V} K\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right) , \tag{1.13}$$

where $K(\cdot)$ is a normalized non-negative smoothing kernel. A kernel with compact support, favored in practice, is the Epanechnikov kernel in d-dimensions:

$$K(\mathbf{x}) = \begin{cases} \frac{1}{2} V_d^{-1} (d+2)(1 - \mathbf{x}^\top \mathbf{x}) & \text{if } \mathbf{x}^\top \mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases}$$

where V_d is the volume of the d-dimensional unit sphere. To increase the posterior probability of $\mathbf{c}$, we have to move in the gradient direction of the density estimate:

$$\nabla p(\mathbf{x}) = \frac{k}{N h^d V_d} \frac{d+2}{h^2} \left(\frac{1}{k} \sum_{\mathbf{x}_i \in \mathcal{B}_h(\mathbf{x})} \mathbf{x}_i - \mathbf{x} \right)$$

where $\mathcal{B}_h(\mathbf{x})$ is the ball with radius h centered at $\mathbf{x}$, and k is the number of samples $\mathbf{x}_k$ in $\mathcal{B}_h(\mathbf{x})$. This leads to the well-known *mean-shift* $\mathbf{x} \longrightarrow \frac{1}{k} \sum_{\mathcal{B}_h(\mathbf{x})} \mathbf{x}_i$ [14, 5].

By virtue of the embedding $\|\mathbf{x}_i - \mathbf{x}_j\| = d_E(\mathbf{c}_i, \mathbf{c}_j)$ (see Section 3.2), we may interpret this as computing the *Fréchet mean* [20]:

$$\hat{\mathbf{c}} = \arg\min_{\tilde{\mathbf{c}}} \int d_E(\tilde{\mathbf{c}}, \mathbf{c})^2 \mathrm{d}\mu(\mathbf{c})$$

of the empirical probability measure μ on the space of contours $\mathbf{c}$, which is equipped with the metric (1.6). As a result, we perturb the evolution (1.12) by adding a small vector field $\mathbf{v} = \varepsilon(\hat{\mathbf{c}} - \mathbf{c}),\ 0 < \varepsilon \in \mathbb{R}$, and thus incorporate statistical shape-knowledge into the variational segmentation approach.

4.4 Experimental Results

Both approaches to the design of shape priors allow to encode the appearance of objects. Applying the variational framework for segmentation, the models are automatically invoked by the observed data and, in turn, provide missing information due to noise, clutter, or occlusion. This bottom-up top-down behavior was verified in our segmentation experiments.

In Figure 7 we see segmentation results for two image sequences showing a rabbit and a head, computed with and without a shape prior. We can see that both shape priors can handle the varying point of view and stabilize the segmentation. Where data evidence is compromised by occlusion (**a**)-(**d**), shadows (**e**)-(**f**), or difficult illumination (**g**)-(**h**), the shape prior can provide the missing information. For the segmentation in Figure 8, we learned the shape prior model 4.3 using 750 human shapes. The shapes in the sequence are not part of the training set. The obtained results encourage the use of shape-priors for the segmentation and tracking of articulated body motion as well.

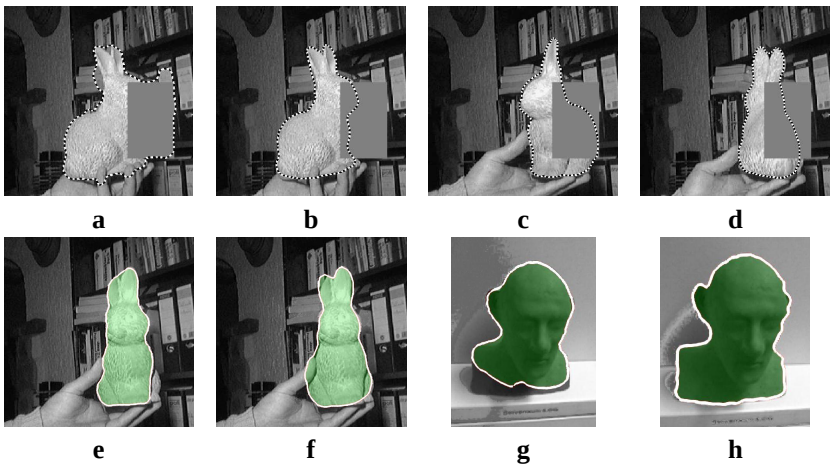

FIGURE 7. Top row: prior from Section 4.2, segmentation without the prior (**a**), with the prior (**b**), two more views with the prior (**c**), (**d**). Bottom row: prior from Section 4.3, segmentation without the prior (**e**), (**g**) and, with the prior (**f**), (**h**)

5 Conclusion and Further Work

We investigated the design of shape priors as a central topic of variational segmentation. Two different approaches based on traditional shape-vectors, and on contours as elements of a metric space defined through a matching functional, respectively, illustrated the broad range of research issues involved. The use of

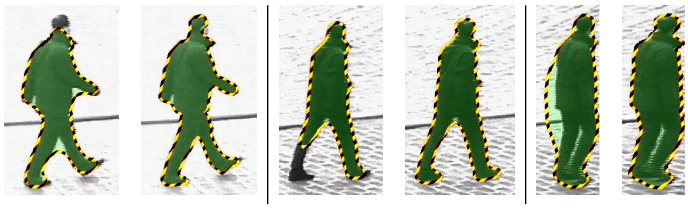

FIGURE 8. Sample screen shots of a human walking sequence. First image is without a shape prior, second image is result obtained with a shape prior, for each image pair respectively.

shape priors allows for the variational segmentation of scenes where pure data-driven approaches fail.

Future work has mainly to address the categorization of shapes according to classes of objects, and the application of this knowledge for the interpretation of scenes with multiple different objects.

Acknowledgment. We thank Dr. Dariu Gavrila, DaimlerChrysler Research, for making available the database with human shapes to the CVGPR group.

6 REFERENCES

[1] R. Basri, L. Costa, D. Geiger, and D. Jacobs. Determining the similarity of deformable shapes. *Vision Res.*, 38:2365–2385, 1998.

[2] C. Bowyer, K.W. and Dyer. Aspect graphs: An introduction and survey of recent results. *Int. J. Imaging Systems and Technology*, 2:315–328, 1990.

[3] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Journal of Foundations Of Computational Mathematics*, 2004. To appear.

[4] Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. Gopinath, R. Briggs, and E. Geiser. Using prior shapes in geometric active contours in a variational framework. *Intl. J. of Computer Vision*, 50(3):315–328, 2002.

[5] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE PAMI*, 17(8):790–799, 1995.

[6] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models – their training and application. *Comp. Vision Image Underst.*, 61(1):38–59, 1995.

[7] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, London, 2001.

[8] D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, 2003.

[9] D. Cremers, N. Sochen, and C. Schnörr. Multiphase dynamic labeling for variational recognition-driven image segmentation. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3024 of *LNCS*, pages 74–86, Prague, 2004. Springer.

[10] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion Snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Intl. J. of Computer Vision*, 50(3):295–313, 2002.

[11] M. Delfour and J. Zolesio. *Shapes and Geometries: Analysis, Differential Calculus, and Optimization*. SIAM, 2001.

[12] K. Dryden, L. and Mardia. *Statistical Shape Analysis*. J. Wiley & Sons, Chichester, 1998.

[13] G. Farin. *Curves and Surfaces for Computer–Aided Geometric Design*. Academic Press, San Diego, 1997.

[14] K. Fukunaga and L. D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Info. Theory*, IT-21:32–40, 1975.

[15] D. Gavrila. Multi-feature hierarchical template matching using distance transforms. In *Proc. of IEEE International Conference on Pattern Recognition*, pages 439–444. Brisbane, Australia, 1998.

[16] S. Jehan-Besson, M. Barlaud, and G. Aubert. Dream^{2}s: Deformable regions driven by an eularian accurate minimization method for image and video segmentation. *Int. J. Computer Vision*, 53(1):45–70, 2003.

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

[18] D. Kendall, D. Barden, T. Carne, and H. Le. *Shape and shape theory*. Wiley, Chichester, 1999.

[19] E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE PAMI*, 26(3):372–383, 2004.

[20] H. Le and A. Kume. The fréchet mean shape and the shape of the means. *Adv. Appl. Prob. (SGSA)*, 32:101–113, 2000.

[21] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. CVPR*, pages 316–323. IEEE Comp. Soc., 2000.

[22] W. Mio, A. Srivastava, and X. Liu. Learning and bayesian shape extraction for object recognition. In *European Conf. on Computer Vision*, volume 3024 of *LNCS*, pages 62–73, Prague, 2004. Springer.

[23] D. Mumford. Mathematical theories of shape: do they model perception? In *Geom. Methods in Computer Vision*, volume 1570, pages 2–10. SPIE, 1991.

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

[25] J. Rieger and K. Rohr. Semi-algebraic solids in 3-space: A survey of moldelling schemes and implications for view graphs. *Image and Vision Computing*, 12(7):395–410, 1994.

[26] V. Roth, J. Laub, M. Kawanabe, and J. M. Buhmann. Optimal cluster preserving embedding of nonmetric proximity data. *IEEE PAMI*, 25(12):1540–1551, Dezember 2003.

[27] M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Proc. of the Europ. Conf. on Comp. Vis.*, volume 2351 of *LNCS*, pages 78–92, Copenhagen, May 2002. Springer, Berlin.

[28] C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. Computer Vision*, 8(2):153–165, 1992.

[29] E. Sharon and D. Mumford. 2d-shape analysis using conformal mapping. In *Proc. CVPR*, pages 350–357, Washington, D.C., 2004. IEEE Comp. Soc.

[30] J. Simon. Differentiation with respect to the domain in boundary value problems. *Numer. Funct. Anal. Optimiz.*, 2:649–687, 1980.

[31] H. D. Tagare, D. O'shea, and D. Groisser. Non-rigid shape comparison of plane curves in images. *J. Math. Imaging Vis.*, 16(1):57–68, 2002.

[32] A. Trouveé and L. Younes. Diffeomorphic matching problems in one dimension: designing and minimizing matching functionals. In *Proc. 6th European Conf. Computer Vision*, LNCS, pages 573–587. Springer, 2000.

[33] 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.

[34] A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.

[35] G. Wahba. *Spline models for observational data*. SIAM, Philadelphia, 1990.

[36] L. Younes. Computable elastic distances between shapes. *SIAM Journal on Applied Mathematics*, 58(2):565–586, 1998.