G. Baratoff, H. Neumann (Eds.), Dynamische Perzeption, Ulm, Nov. 2000
Proceedings in Artificial Intelligence, Infix, Vol. 9, pp. 117–122, © Infix–Verlag

Learning of Translation Invariant Shape Knowledge for Steering Diffusion–Snakes

Daniel Cremers, Christoph Schnörr,
Joachim Weickert, and Christian Schellewald

Computer Vision, Graphics, and Pattern Recognition Group
Department of Mathematics and Computer Science
University of Mannheim, 68131 Mannheim, Germany
cremers@uni-mannheim.de, http://www.cvgpr.uni-mannheim.de

Abstract. Biological vision is characterized by an intricate interplay of external sensory input and previously acquired internal representations of the world. In this work, we present a computer vision model for image segmentation which allows to combine external visual input and internally represented shape knowledge in a unifying energy functional. We explain how prior shape information is acquired. Numerical examples show how the relative weight of prior information affects the outcome of segmentation. A method to include translation invariance clarifies the way in which shape knowledge is acquired. The behavior of our method in a real–world scenario is demonstrated.

1 Introduction

The traditional approach to low–level image processing and computer vision consists of two steps: first, the image is segmented by applying an appropriate set of filters in order to smooth out noise and enhance certain features such as edges or corners. In the second step, the segmented image is further processed in order to determine existence, location and other properties for a set of objects of interest.

In contrast to this two–step process, we shall present a method which allows to incorporate statistical knowledge about the shape of expected objects already in the segmentation process. This causes an interaction of external input and internally represented knowledge, reminiscent of the effect of the feedback architecture in biological visual pathways.

The basic idea behind these so–called *diffusion–snakes*, namely to combine shape statistics learned from example images with diffusion filtering, was introduced in [3]. Similar approaches of combining statistical shape knowledge and image segmentation have been introduced recently [7, 8]. Our approach differs from these in two ways: first, our segmentation is based on the Mumford–Shah functional. Secondly, we combine segmentation and statistical shape knowledge in one unifying energy functional rather than treating both separately.

In this paper, we shall extend our work in order to model translation invariance. Numerical examples show the effect of learned shape on the result of segmentation and demonstrate robustness against scaling. Finally, we shall present some results with real–world images, which demonstrate important properties of our approach: non–local convergence and robustness against spurious or missing image data.

2 Combining Image Segmentation and Shape Knowledge

We formulate image segmentation by a variational approach, following the work of Mumford and Shah [9] on image segmentation. For a grey–scale image $u(x)$ and a contour $C(s)$

we define the total energy as

$$E(u,C) = E_{MS}(u,C) \; + \; \alpha \, E_c(C) \, , \tag{1}$$

where the first term is the Mumford and Shah functional

$$E_{MS}(u,C) = \frac{1}{2} \int_{\Omega} (f-u)^2 \, d^2x \; + \; \lambda^2 \, \frac{1}{2} \int_{\Omega - C} |\nabla u|^2 \, d^2x \; + \; \nu |C| \, , \tag{2}$$

and the second term $E_c(C)$ is a contour energy, which is defined such that it favors contour formations that are familiar to the system from a learning procedure. How to construct the contour energy is explained in the next section.

Minimizing the Mumford–Shah part E_{MS} with respect to u and C results in a segmented image u which is similar to the input image f and smooth on the image plane Ω, except for the contour C, across which the segmented image may be discontinuous. The last term enforces that the length $|C|$ of the minimizing contour be as short as possible. The idea is that the contour C captures the boundaries of a given object during the minimization of the energy (2).

The parameter α in (1) allows to shift from a purely image–based segmentation given by the Mumford–Shah functional to a more and more knowledge–based segmentation. Increasing α will give more weight to the learned shape knowledge and thus continuously suppress any image feature which was not previously learned.

3 Shape Learning

In the learning procedure the system is trained with a set of binary images representing silhouettes of 3D–objects. The shape of an object is captured in form of a closed quadratic B–spline curve [5, 1]. Each shape is thus represented by a vector of spline control points

$$\boldsymbol{z} = (x_1, y_1, \ldots, x_N, y_N)^t \; . \tag{3}$$

The variation within the various learned shapes corresponds to a variation within the spline control point vectors. These vectors are assumed to form a multidimensional Gaussian distribution, the principal axes of which are calculated by a principal component analysis (PCA) [2, 4]. This results in a probability density function for the control point vectors of the form

$$\mathcal{P}(\boldsymbol{z}) \; \propto \; \exp\left(-\frac{1}{2} (\boldsymbol{z} - \boldsymbol{z}_0)^t \, \Sigma^{-1} \, (\boldsymbol{z} - \boldsymbol{z}_0) \right) \; , \tag{4}$$

where the mean contour $\boldsymbol{z}_0$ and the covariance matrix Σ represent the shape variation encountered in the learning set.

In general, the covariance matrix Σ will not have full rank such that an inverse is not defined. Therefore, the inverse Σ^{-1} is replaced by a modified pseudo–inverse Σ^* as explained in [3]. Basically, this regularisation consists in replacing the zero eigenvalues of Σ by some small value, for example by the smallest positive eigenvalue encountered. Thus the new matrix becomes invertible and unfamiliar shape variations produce a finite but non–zero probability density. The contour probability (4) corresponds to a contour energy

$$E_c(\boldsymbol{z}) \; = \; \frac{1}{2} (\boldsymbol{z} - \boldsymbol{z}_0)^t \, \Sigma^* \, (\boldsymbol{z} - \boldsymbol{z}_0) \, , \tag{5}$$

Adding this shape energy to the Mumford–Shah functional gives the total energy, the minimum of which gives the desired segmentation.

4 Energy Minimization

The total energy (1) must be simultaneously minimized with respect to the image u and the spline control points $\boldsymbol{z}$ which define the contour C. For these two cases we formulated Euler–Lagrange equations which we implemented as evolution equations with an artificial time parameter t. This corresponds to a gradient descent. We shall only present the resulting evolution equations for the contour control points $\boldsymbol{z}$ (given a fixed image u) and for the segmented image u (given a fixed contour). Their derivation can be found in [3].

The evolution equation obtained for the x–components of the contour control points (and a similar one for the y–components) is

$$\begin{aligned} \frac{dx_m(t)}{dt} = & \sum_{i=1}^{N} \left(\boldsymbol{B}^{-1}\right)_{mi} \left(e^+(C(s_i),t) - e^-(C(s_i),t) - \kappa(s_i,t)\right) \boldsymbol{n}_x(s_i,t) \\ & - \alpha \left[\Sigma^* \left(\boldsymbol{z} - \boldsymbol{z}_0\right)\right]_{2m-1} . \end{aligned} \tag{6}$$

The first term on the right hand side is the Mumford–Shah contribution. e^+ and e^- denote the integrand of equation (2) inside and outside the contour, κ denotes the curvature of the contour and $\boldsymbol{n}_x$ is the x–component of the normal vector. s_i is a set of discrete nodes along the contour $C(s)$ (cf. [3]) and the matrix $(\boldsymbol{B})_{mi}$ is formed by the spline basis function B_m evaluated at node s_i.

The second term on the right hand side of (6) is generated by the shape energy (5). It causes a relaxation towards the most probable shape. Moreover, the decay of various degrees of freedom is weighted by the inverse of the covariance matrix: the less familiar a shape variation from the mean shape the faster its decay.

Minimization of the total energy (1) with respect to the segmented image u leads to the following gradient descent equation:

$$\frac{du}{dt} = -\frac{dE}{du} = (f - u) + \lambda^2 \, \nabla \cdot \left(w_c(x) \, \nabla u\right) , \tag{7}$$

where the indicator function

$$w_c : \Omega \to \{0,1\}, \qquad w_c(x) = \begin{cases} 0 & \text{if } x \in C \\ 1 & \text{otherwise} \end{cases} \tag{8}$$

accounts for the contour information. Equation (7) models an inhomogeneous linear diffusion of the image u which – for fixed contour C – converges to the minimum of (1) for $t \to \infty$.

Minimization of the total energy (1) was performed by numerically iterating equations (6) and (7). The results in the following sections will be demonstrated in form of the minimizing contour C.

5 Image–Driven versus Knowledge–Driven Segmentation

Our approach of combining image information and learned shape knowledge in a single energy functional (1) allows to continuously shift from a purely image–based segmentation to one which mostly relies on the learned shape information.

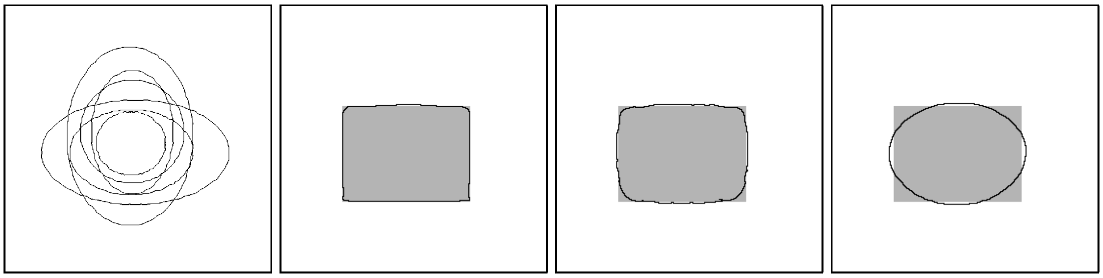

Fig. 1. Image–driven vs. knowledge–driven segmentation: the knowledge energy was calculated on a set of six ellipses (left). The input image (grey square) and the final contour (black) are shown for increasing values of the knowledge strength. From left to right: $\alpha/|\Omega| = 0.00076, 0.076, 1.526$ (cf. eqn. 1).

Figure 1 gives a simple demonstration of the effect of knowledge on the segmentation result: while for low knowledge strength the segmented image coincides with the input image (left contour) the results for larger knowledge strength are shifted towards the prior information (middle and right contour).

This effect can be used to deal with occlusion or missing information. Depending on the prior information the respective occlusion is removed from the input image and missing information is accounted for by prior knowledge (cf. Figures 2 and 3).

6 Learning Translation Invariance

By construction, the shape energy (5) does not include translation invariance: all acquired shapes are in a specific location and therefore the mean shape will also be in a specific location of the image plane. If one were to include full translation invariance in the shape energy, one way to approach this would be to normalize all object shapes in the learning procedure by centering them [4]. Then one would also center a given control point configuration before determining its contour energy.

However, in the following we present a different approach to deal with translation for two reasons:

First, one may not always be interested in full translation invariance. Maybe a given object is much more likely to appear in a specific location. One would then want to benefit from this information in a recognition process.

Secondly, the above method to include translation invariance assumes prior information about the nature of the desired invariance. The method presented here, however, fits more naturally into the framework of shape learning from examples. We discuss it separately in the following because it helps to give a better insight into the learning process.

The method of translation learning is straight forward: only *one single object* presented in the learning process is shifted along two orthogonal axes. The obtained control point configurations are included in the learning set. The result is that the contour energy becomes less sensitive to a shift of objects. Translation invariance is automatically extended to the entire scope of learned shapes, because what is *learned* by the covariance matrix are not the separate shapes but rather the deformations that take one shape into the other — in this case shifting all control points by a fixed amount in x– and y–direction.

Figure 2 demonstrates the behavior of the system before and after the proposed translation learning. Only after translation learning the minimizing contour is allowed to shift from its initial center within the image plane.

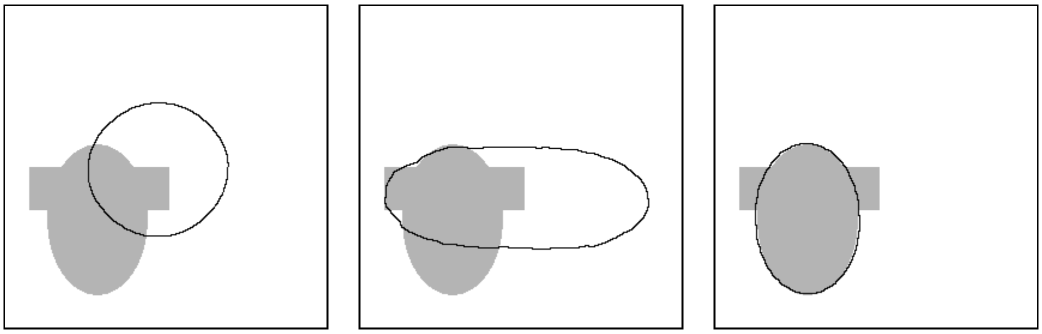

Fig. 2. Translation invariance: the input image (grey) is an ellipse occluded by a bar. The **left image** shows the initial contour. The contour energy was constructed on a set of ellipses, which will effectively restrict contour variation to elliptical shapes. The **middle image** shows the final contour *without* translation learning: note that the center of the ellipse has practically not moved. The **right image** shows the final contour *with* translation learning: the translational degrees of freedom are favored by the energy functional and the ellipse that best describes the input image is found. Due to the statistical a priori–knowledge, the bar is ignored by the diffusion–snake despite the prominent signal transitions at its boundary. Similarly, given the prior knowledge of *bars* rather than *ellipses*, one can extract the bar from the above image.

Since PCA is a *linear* approach, scaling and rotation of one of the objects will not be automatically generalized to the entire scope of shapes, as is the case for translation. Figure 3 shows, however, that some robustness against scaling is obtained: results as in Figure 2, right side, are obtained for a series of shapes on various scales.

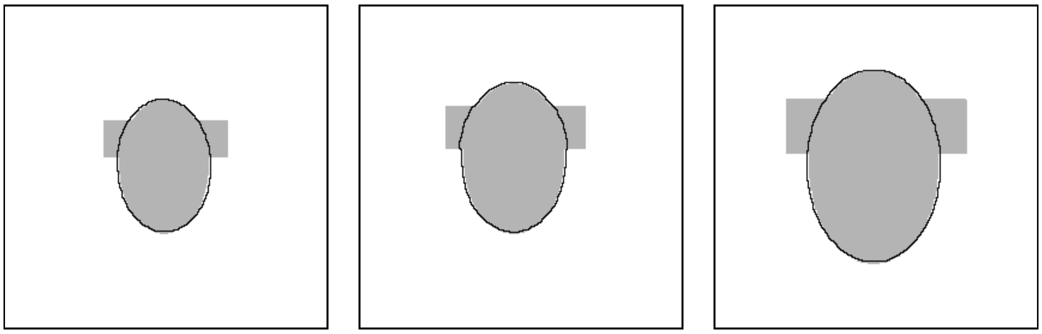

Fig. 3. Robustness against scaling: The contour energy was constructed on the set of ellipses shown in Figure 1. The input image – an ellipse occluded by a bar – is presented at three different scales. The final contour (black) shows that a certain robustness against scaling is obtained.

7 Non–Local Convergence

In order to minimize the total energy we used a gradient descent method. Therefore, the minimizing contour will always be that of the next best local minimum. However, the energy functional (1) differs from the energy functionals of most standard active contour models [6, 1] which try to fit a contour solely to the *local* image gradient.

In the Mumford–Shah functional the minimizing contour C depends on a segmented version u of the input image such that the contour evolution is entirely different from that of fitting to the local image gradient. This is exemplified in Figure 4 which shows several contours in the minimization process. The shape energy was constructed on a set of hands. Translation was enabled in the way mentioned above.

Fig. 4. Non–local Convergence: The contour energy was constructed on a set of hands. The input image and three contours (in white) show how the contour evolves from the initial contour (left image) to the final one (right image) avoiding the local minima of purely fitting to the image gradient. Note the way in which our modifications to the Mumford–Shah functional affect the process: the spline implementation restricts the contour to be closed at all times. And the prior knowledge effectively restricts the contour motion to a subspace of familiar shape variations, such that even quite prominent clutter is ignored.

8 Conclusion

We presented a model to combine image segmentation based on the Mumford–Shah functional with statistical shape knowledge in a unifying energy functional. Extending our work in [3], we gave further examples of the effect of increasing shape knowledge. We presented a method of introducing translation invariance which is naturally embedded in the learning process. It clarifies the way in which shapes are learned. We showed that some robustness against scaling of the input image is obtained. Finally, we showed encouraging results for real–world images.

In future work we will develop fast numerical algorithms to solve equation (7) and turn to statistical shape modeling beyond the standard PCA case.

References

1. A. Blake and M. Isard. *Active Contours.* Springer, London, 1998.
2. T.F. Cootes, C.J. Taylor, D.M. Cooper, and J. Graham. Active shape models – their training and application. *Comp. Vision Image Underst.*, 61(1):38–59, 1995.
3. D. Cremers, C. Schnörr, J. Weickert, and C. Schellewald. Diffusion–snakes using statistical shape knowledge. In G. Sommer and Y.Y. Zeevi, editors, *Algebraic Frames for the Perception-Action Cycle*, Lect. Notes in Comp. Sci., Kiel, Germany, Sept. 10-11, 2000. Springer-Verlag. To appear.
4. L. Dryden and K.V. Mardia. *Statistical Shape Analysis.* Wiley, Chichester, 1998.
5. G. Farin. *Curves and Surfaces for Computer–Aided Geometric Design.* Academic Press, San Diego, 1997.
6. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. of Comp. Vision*, 1(4):321–331, 1988.
7. C. Kervrann and F. Heitz. Statistical deformable model-based segmentation of image motion. *IEEE Trans. on Image Processing*, 8:583–588, 1999.
8. M.E. Leventon, W.E.L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog. (CVPR'00)*, volume 1, pages 316–323, Hilton Head Island, South Carolina, June 13–15 2000.
9. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.