

# Variational Shape Detection in Microscope Images Based on Joint Shape and Image Feature Statistics

Matthias Fuchs

Infmath Imaging, University of Innsbruck  
6020 Innsbruck, Austria

<http://infmath.uibk.ac.at/~matthiasf>

Samuel Gerber

SCI Institute, University of Utah  
Salt Lake City, UT 84112

<http://www.cs.utah.edu/~sgerber>

## Abstract

This paper presents a novel variational formulation incorporating statistical knowledge to detect shapes in images. We propose to train an energy based on joint shape and feature statistics inferred from training data. Variational approaches to shape detection traditionally involve energies consisting of a feature term and a regularization term. The feature term forces the detected object to be optimal with respect to image properties such as contrast, pattern or edges whereas the regularization term stabilizes the shape of the object. Our trained energy does not rely on these two separate terms, hence avoids the non-trivial task of balancing them properly. This enables us to incorporate more complex image features while still relying on a moderate number of training samples. Cell detection in microscope images illustrates the capability of the proposed method to automatically adapt itself to different image features. We also introduce a nonlinear energy and exemplarily compare it to the linear approach.

## 1. Introduction

Variational approaches to detect shapes in images are based on functionals which map shape geometries to an energy that reflects how well the given shape corresponds to the image features. Mumford and Shah [18] proposed to use the mean intensity of the region defined by the shape compared to the intensity of the background as such a feature. This idea can be extended to regions of homogeneous patterns as in Chan and Vese [1]. A second important feature are the edges in images. Kass *et al.* [13] proposed the Snakes approach to fit curves to the edges of an image. Both formulations require an additional regularization term in the energy functional to ensure that the corresponding variational problem is well-posed. This term measures the regularity of the boundary of the detected region. A general



Figure 1. Mumford-Shah segmentation of a cross. *Left*: Relative regularization parameter  $\alpha = 1$ . *Right*: Relative regularization parameter  $\alpha = 5$ . The arrows indicate areas where the regularization contracts the curve too much.

form of these energies is

$$E_\alpha = I + \alpha R, \quad (1)$$

where  $I$  denotes the *fit-to-data term*,  $R$  the *regularization term* and  $\alpha > 0$  the *regularization parameter*. The fit-to-data term assigns small energies to shapes which fit to the image features whereas the regularization term favors “regular” shapes. In the approaches cited above this regularity is ensured by penalizing the length or area of the boundary. This forces the shape boundary to be bounded.

The regularization term is necessary to ensure the well-posedness of the variational problem associated with the energy functionals. In Figure 1 we illustrate the influence of the regularization parameter  $\alpha$  in the Mumford-Shah functional. This example indicates that the correct choice of the regularization parameter is important to obtain satisfying segmentation results.

These traditional methods have difficulties to correctly detect shapes that are partially occluded, on cluttered image background, or on images corrupted by too much noise. A common solution to this problem is the use of shape priors. The idea of using statistics of shapes as a basis for shape detection was introduced by Cootes *et al.* [5]. More recent approaches are Chen *et al.* [2], Cremers *et al.* [7], Fang and

Chan [8], Gastaud *et al.* [9], Leventon *et al.* [16], Rousson and Paragios [20, 21] and Tsai *et al.* [23]. Shape prior methods use training data to compute shape statistics. These statistics define a likelihood functional that maps a shape to its probability w.r.t. the shape statistics and replaces the regularization term in the traditional variational formulation. This regularization ensures that only shapes which seem to be “reasonable” with respect to the training statistics are detected.

Again, the above approaches define energies of the form (1) where  $R$  includes the statistical prior information. The regularization parameter determines the influence of the shape statistics. A high weight stabilizes the shape detection but might render it impossible to detect shapes which are very different from the training shapes (but still correct), whereas a too low weight introduces the danger of getting wrong results in case of noisy, cluttered or occluded image data. The correct choice of the parameter is not trivial and application dependent. Multiple (possibly time consuming) tests are often necessary to validate the weighting parameter. This situation is illustrated in Figure 2. There we manually annotated the cells in the central cluster in the image and estimated the mean and the covariance of the shape parameters of the training shapes. Then we minimized the Snakes functional with a regularization term defined by the statistics of the training data and compared the results for regularization parameters  $\alpha$  at three orders of magnitude. As in the previous example we observe that this approach gives satisfying results for a correctly chosen regularization parameter but fails in case of too small or too large values of  $\alpha$ .

The cited approaches further limit themselves to the use of only one kind of image feature, *e.g.* image contrast *or* edges. A combination of multiple features would again require each of them to be weighted with respect to the other and thus introduce even more parameters.

In this paper we propose an approach which does not require the explicit choice of a regularization parameter. Similar to the above works on shape priors, we train statistics on annotated data. In contrast to limiting the statistics to shapes only, we incorporate the corresponding image features from the training data. This allows us to learn the full-fledged segmentation energy and not only a regularization term. Furthermore, we avoid the choice of regularization parameters. Our method is capable of incorporating an arbitrary number of different kinds of image features. The relative importance of the various features is automatically learned from the training data. *I.e.* the trained energy gives high weight to combinations of features it learned to be representative for a class of objects and does not consider features which vary a lot across the training data. The computational effort to evaluate the resulting energy for a given shape is comparable to the methods mentioned above and

the number of required training samples is very moderate.

Learning a combination of shape and image features was proposed by Cootes *et al.* [3] in their work about Active Appearance Models. There, the complete intensity distribution inside the shapes is learnt whereas in our work we consider features obtained by integration over the shape boundary. The idea of learning an energy from multi-dimensional training data and leave the work of selecting important features to the learning process is also similar to machine learning approaches acting on raw pixel values of image data (cf. LeCun *et al.* [15, and references therein]). In comparison to these methods our approach requires significantly less training because we use shape knowledge and intelligently computed image features.

Another approach related to ours was proposed by Cremers *et al.* [6]. There, the authors learn a kernel density based on shape and image features. In contrast to our work, they focus on level set representations of shapes and the intensity distribution within shapes. They also consider the distributions of the shapes and the image features separately, whereas we treat them jointly. An approach to solve the problem of choosing the optimal regularization parameter for a given image was presented by McIntosh and Hamarneth [17]. They minimize a quadratic functional for regularization parameters which yield convex detection energies. In our setting, the problem of the optimal regularization is equivalent to the selection of the image features the shape detection is based on. This is related to Law *et al.* [14].

The outline of this paper is as follows. In the next section we introduce the *shape-to-feature map* which, for a given image, maps a shape to a vector of image features determined by this shape. The shape-to-feature map is used to learn an energy based on shape and image feature statistics. For the results in this paper we concentrated on features which can be expressed as boundary integrals along the shape outline. Section 3 is dedicated to the training of an energy for a given image, a given set of training samples and a given shape-to-feature map.

The subsequent section is devoted to experimental results. We applied our method to the detection of objects in biologic microscope images. These results were obtained by gradient and intensity based image features together with normal density estimates. In Section 5 we outline the approach we used to minimize the learnt energy. In the last section of the paper we exemplarily demonstrate the use of nonlinear density estimates to learn a shape statistic with two modes. We use this energy to detect shapes in artificial image data and compare them to the results obtained by the normal density estimate.

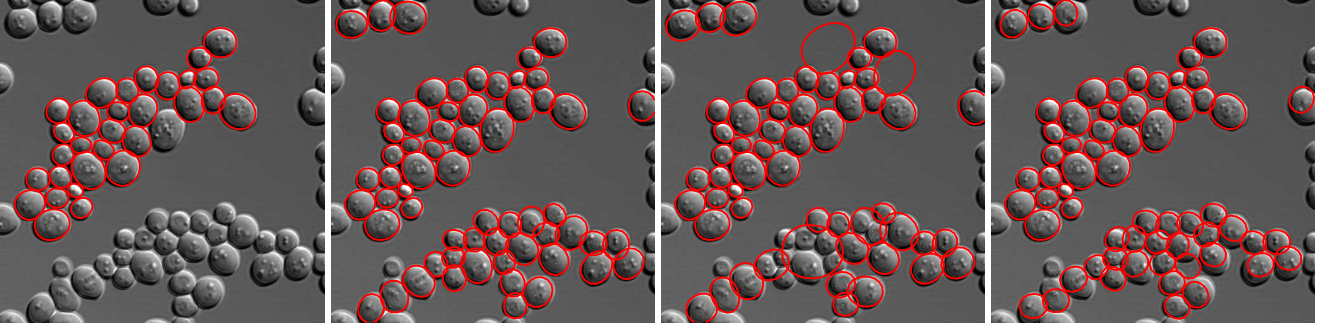


Figure 2. Edge-based segmentation with shape regularization. *Top, left*: Manually annotated training shapes. *Top, right*: Detected shapes with relative regularization parameter  $\alpha = 1$ . *bottom, left*: Detected shapes with relative regularization parameter  $\alpha = 0.1$ . *bottom, right*: Detected shapes with relative regularization parameter  $\alpha = 10$ .

### 1.1. Notation and preliminaries

In the following we always assume  $u : \Omega \rightarrow \mathbb{R}^d$  to be a (possibly vector valued) image defined on a 2-dimensional domain  $\Omega \subseteq \mathbb{R}^2$ . If  $d = 1$  then  $u$  can be interpreted as a gray-scale image. Let  $\gamma$  be a closed planar curve in  $\Omega$  without self-intersections, *i.e.*  $\gamma : S^1 \rightarrow \Omega$ , injective. We assume  $\gamma$  to be piecewise differentiable. We refer to  $\gamma$  as a *shape*. The shape  $\gamma$  is completely determined by a *shape parameter*  $p \in \mathbb{R}^m$ . This is denoted by  $\gamma(p)$ . In our case  $p$  parametrizes the medial axis of the shapes, but it might as well be a list of the coefficients of a B-spline curve or any other kind of shape parametrization.

Finally note that our work is presented for the planar case only, but generalizes to higher dimension very easily.

### 2. Shape-to-feature map

We call a map  $F$ ,

$$\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^n, \quad p \mapsto \mathbf{F}(\gamma(p)), \quad (2)$$

which maps a shape parameter to a vector of features of the image  $u$  a *shape-to-feature map*. *I.e.*  $\mathbf{F}$  depends only on the shape  $\gamma(p)$  but not on  $p$  itself. It is important to note, though, that  $\mathbf{F}$  does depend on the image  $u$ . To simplify notation and because we chose  $u$  to be fixed throughout the paper we do not denote this dependence.

In this paper we concentrate on a subclass of shape-to-feature maps, which is characterized by a special form of  $\mathbf{F}$ . In particular, we consider  $\mathbf{F}$  to be the composition  $\mathbf{F} = \mathbf{G} \circ \mathbf{H}$ , where  $\mathbf{G} : \mathbb{R}^k \rightarrow \mathbb{R}^n$  and  $\mathbf{H} : \mathbb{R}^m \rightarrow \mathbb{R}^k$ . We assume that only  $\mathbf{H}$  depends on the image  $u$  whereas  $\mathbf{G}$  is independent of the image and the shape. Each of the components  $H_i$ ,  $1 \leq i \leq k$ , of  $\mathbf{H}$  should have one of the following forms:

$$H_i(p) = \int_{\gamma(p)} a_i(u) ds \quad \text{or} \quad H_i(p) = \int_{\gamma(p)} b_i(u) \cdot dn, \quad (3)$$

where  $a_i(u) : \Omega \rightarrow \mathbb{R}$ ,  $b_i(u) : \Omega \rightarrow \mathbb{R}^2$  and  $n$  denotes the outer unit normal of  $\gamma(p)$ . *I.e.* we assume each component of  $\mathbf{H}$  to be either the integral of a scalar function along the shape boundary or the integral of a vector field along the same boundary. This construction enables us to evaluate complex image features and still give good estimates on the complexity of an evaluation of  $\mathbf{F}$  for a given shape  $\gamma(p)$ . Because  $a_i(u)$  and  $b_i(u)$  depend only on the image  $u$  but not on the shape  $\gamma(p)$  we can precompute a discretized version of them. The computation of  $H_i(p)$  then involves

- the computation of  $\gamma(p)$ , and
- the evaluation of a 1-dimensional boundary integral.

For accordingly chosen functions  $a_i(u)$  it is possible to evaluate a wide range of features such as intensity, histogram data and gradient information along the shape boundary. The integral over  $b_i(u)$  enables us to compute the same values over the region  $\Gamma(p) \subseteq \Omega$  inside a given shape  $\gamma(p)$ . Assume a scalar function  $c_i(u)$  which we want to integrate over  $\Gamma(p)$ . We first compute  $b_i(u)$  such that  $\nabla \cdot b_i(u) = c_i(u)$ . This equation constitutes an under-determined system of partial differential equations for  $b_i$  which is trivial to solve for a given image  $u$ . Then, by the divergence theorem,

$$\int_{\Gamma(p)} c_i(u) dx = \int_{\gamma(p)} b_i(u) \cdot dn. \quad (4)$$

For  $a_i(u) = 1$  or  $c_i(u) = 1$  the integrals (3) evaluate to the length of the boundary of  $\gamma(p)$  and its volume, respectively.

The map  $\mathbf{G}$  is used to combine the values of the integrals  $H_i$  to get more meaningful features. In our examples the function  $\mathbf{G}$  normalizes the integrals along the boundary  $\gamma(p)$  and the region  $\Gamma(p)$  w.r.t. the length of  $\gamma(p)$  and area of  $\Gamma(p)$ , respectively. It is further possible to write the simplified Mumford-Shah functional [18, 1] and the Snakes [13] functional in the form  $\mathbf{F} = \mathbf{G} \circ \mathbf{H}$  with  $\mathbf{H}$  being an expression of integrals as in (3).

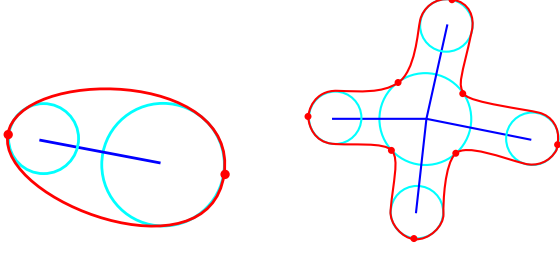


Figure 3. *Left:* The skeleton and the maximal circles of this cell shaped object are parametrized by  $p$ . The outline  $\gamma(p)$  is computed by interpolating the points on the circles. *Right:* Medial axis parametrization of a cross.

## 2.1. Shape representation

The shape model we use is based on the idea of parametrizing a shape by constructing a discrete approximation of its medial axis as proposed by Joshi *et al.* [12]. In our case, a shape parameter  $p$  parametrizes a tree-like skeleton consisting of nodes, edges connecting the nodes, and circles at the nodes. These circles are supposed to be maximal circles within the shape, *i.e.* they touch the shape in at least two points. In more detail, the components of  $p$  determine

- the position and rotation of the skeleton,
- the lengths and the angles of the edges of the skeleton, and
- the radii of maximal circles centered at the nodes of the skeleton.

We chose this model because it is a more natural parametrization of shapes than B-spline curves but still allows for complex shapes as illustrated in Figure 3.

Because we will explicitly refer to the position and rotation of shapes later on, we decide that the first three components of  $p$  determine these properties, *i.e.*

$$p = \left( \underbrace{p^1, p^2, p^3}_{\text{position, rotation}}, \underbrace{p^4, \dots, p^m}_{\text{skeleton, radii}} \right)^T. \quad (5)$$

## 3. Energy training

The main contribution of this paper is the computation of an energy  $E$  for given training shape parameters  $p_1, \dots, p_N$  and a shape-to-feature map  $\mathbf{F}$ . In this section we will introduce the variational form of the shape detection problem based on this energy and explain two different ways to train energies based on normal density estimation and kernel density estimation.

The energy  $E : \mathbb{R}^m \rightarrow [0, \infty)$  maps an unseen shape parameter  $p$  to a non-negative value which determines how well  $\gamma(p)$  fits on the image considering shape and image properties learned from the training shapes. Small values

$E(p)$  correspond to a good match. Hence, the shape detection problem of single shape can be written as

$$p = \operatorname{argmin}_{p \in D} E(p), \quad (6)$$

where  $D \subseteq \mathbb{R}^m$ . The domain  $D$  constrains the above variational problem. In all applications we chose  $D$  such that only shapes on the image domain  $\Omega$  are considered. We further can adapt  $D$  such that shapes close to training shapes or already detected shapes on the same image are not considered in the minimization problem.

In the following we explain two different approaches to compute  $E$ . As mentioned before we assume  $u$  to be a fixed image. Furthermore,  $p_1, \dots, p_N$  are the parameters of manually determined training shapes on this image. This means that we expect the shapes  $\gamma(p_1), \dots, \gamma(p_N)$  to match objects on  $u$ . Finally let  $\mathbf{F}$  be a feature map for  $u$ . Then, for a given shape parameter  $p$ , we define its *shape-feature vector*  $q(p)$  by setting

$$q(p) = (p^4, \dots, p^m, F_1(p), \dots, F_n(p))^T \in \mathbb{R}^M, \quad (7)$$

where  $M := m + n - 3$ . In other words,  $q(p)$  consists of the features for the shape determined by  $p$  and the shape parameter  $p$  excluding position and rotation. We further denote the shape-feature vectors of the training data  $p_i$  as  $q_i := q(p_i)$ ,  $1 \leq i \leq N$ .

In this paper we consider energies of the form

$$E(p) = -\log f(q(p)), \quad (8)$$

where  $f$  is a probability density on  $\mathbb{R}^M$  and depends on the training data  $p_1, \dots, p_N$ . This formulation translates the energy learning into a density estimation problem.

For the shape detection in the microscope image we assume  $q$  to be normally distributed with density function

$$f(q) = (2\pi)^{-M/2} \det(\Sigma)^{-1/2} e^{-\frac{1}{2}(q-\mu)^T \Sigma^{-1} (q-\mu)}. \quad (9)$$

with  $\mu \in \mathbb{R}^M$  and  $\Sigma$  a symmetric and positive definite  $M \times M$ -matrix. Assuming the shape-feature vectors of the training data to be independently and identically distributed w.r.t.  $f$  we compute maximum-likelihood estimators of the parameters  $\mu$  and  $\Sigma$ :

$$\mu = \frac{1}{N} \sum_{i=1}^N q_i, \quad (10)$$

$$\Sigma = \frac{1}{N-1} \sum_{i=1}^N (q_i - \mu)(q_i - \mu)^T. \quad (11)$$

By (8) the energy  $E$  of a given shape parameter  $p$  is then

$$E(p) \propto (q(p) - \mu)^T \Sigma^{-1} (q(p) - \mu). \quad (12)$$

There exist several interpretations of the above expression. For one, it equals the Mahalanobis distance between  $q$  and

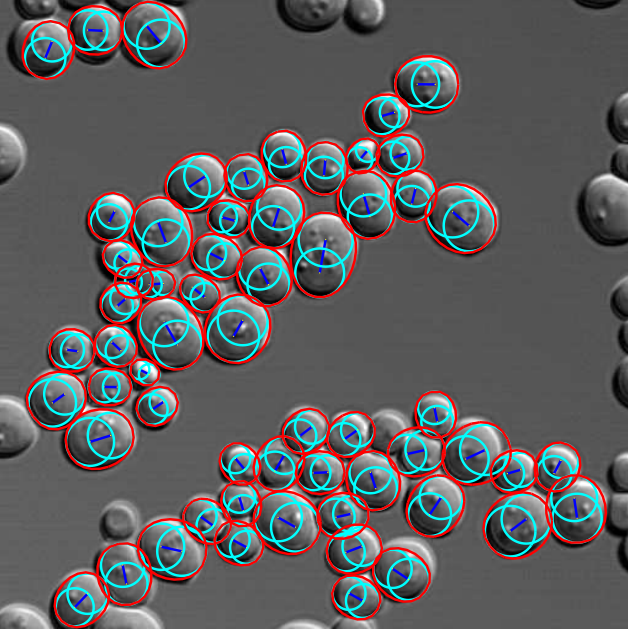


Figure 4. Microscope image ( $512 \times 512$  pixels) of yeast cells. The image shows training shapes as in Figure 2 and the detected cells.

$\mu$  w.r.t. the covariance  $\Sigma$ . Also, it can be interpreted as the squared norm of the coefficients of  $q$  w.r.t. the principal component analysis of the training shape-feature vectors. It is further important to note that (12) with  $\mu$  and  $\Sigma$  as in (10) and (11) is invariant under linear transformation of  $q$ . In particular, rescaling single components of the shape-feature vector does not change  $E$ .

## 4. Results

We used the proposed method to detect shapes in microscope image data. We manually annotated some of the objects on a given image and automatically detected the remaining ones by minimizing the energy learned from the annotated data.

The data in Figure 2 involves two major challenges. The cells on the image form a huge cluster and it is difficult to separate them with traditional methods. In addition, the shadow-like features on each cell cause extra edge information in the cells which can not be removed by smoothing. The objects in Figure 5 are more clearly set apart from each other, but the quality of their appearance varies more than in the first examples.

We computed the same feature map for both images. First, we smoothed the images with a 2-pixel-wide Gaussian kernel, denoting the result as  $u_\sigma$ . Then we defined  $\mathbf{F}$

by

$$\mathbf{F}(p) = \begin{pmatrix} \oint_{\gamma(p)} |\nabla u_\sigma| ds \\ \oint_{\gamma(p)} u ds \\ \oint_{\gamma(p)} u^2 ds \\ \oint_{\Gamma(p)} u dx \\ \oint_{\Gamma(p)} u^2 dx \end{pmatrix}. \quad (13)$$

Here,  $\oint_{\gamma(p)} ds$  and  $\oint_{\Gamma(p)} dx$  denote the integrals over  $\gamma(p)$  and  $\Gamma(p)$  normalized by the length of  $\gamma(p)$  and the area of  $\Gamma(p)$  respectively. In a nutshell, we compute the normalized values of the absolute values of the image gradients along the boundary and the normalized values of the intensities and their first moment along the boundary and inside the shape. The first moments in (13) enable us to capture variations of intensities as in the cells in Figure 4. Note that as in (4) the latter two integrals can be transformed into a boundary integral of a vector field. This also holds for the computation of the area of  $\Gamma(p)$ .

We computed the training shape-feature vectors  $q_1, \dots, q_N$  from the manually annotated objects,  $N = 31$  and  $N = 75$  in Figure 4 and Figure 5, respectively. From these training sets we estimated  $\mu$  and  $\Sigma$  as in (10) and (11) to define the energy (12).

## 5. Minimization

The solution of (6) was done iteratively. We chose the problem domain  $D$  such that only shapes on the image  $\Omega$  are considered and further removed all parameters from  $D$  which corresponded to shapes which overlap with the training shapes. By "overlapping shapes" we mean shapes whose common area is above some threshold (50 common pixels in Figure 4 and 150 pixels in Figure 5).

To find multiple shapes in the image we used Algorithm 5. In simple words, the algorithm generates random shapes and tries to improve the current detection result  $p_1, \dots, p_M$  by successively replacing previously detected shapes with new ones.

In the algorithm, by the random choice of a shape parameter  $p$  we mean selecting  $p \in D$  as follows:

- The position  $(p^1, p^2)$  and the rotation  $p^3$  are uniformly sampled on the image domain  $\Omega$  and in the interval  $[-\pi, \pi]$ , respectively.
- We compute the mean  $\mu_{4:M}$  and the covariance matrix  $\Sigma_{4:M}$  of the components  $(p_i^4, \dots, p_i^M)$ ,  $1 \leq i \leq N$ , of the training data. Then we sample  $(p^4, \dots, p^M)$  from the multivariate normal distribution with mean  $\mu_{4:M}$  and variance  $\Sigma_{4:M}$ .

The result of the above selection is accepted if  $p \in D$ . Otherwise a new candidate  $p$  is sampled.

We chose  $M$  larger than the expected number of shapes in the image. After stopping the algorithm we manually

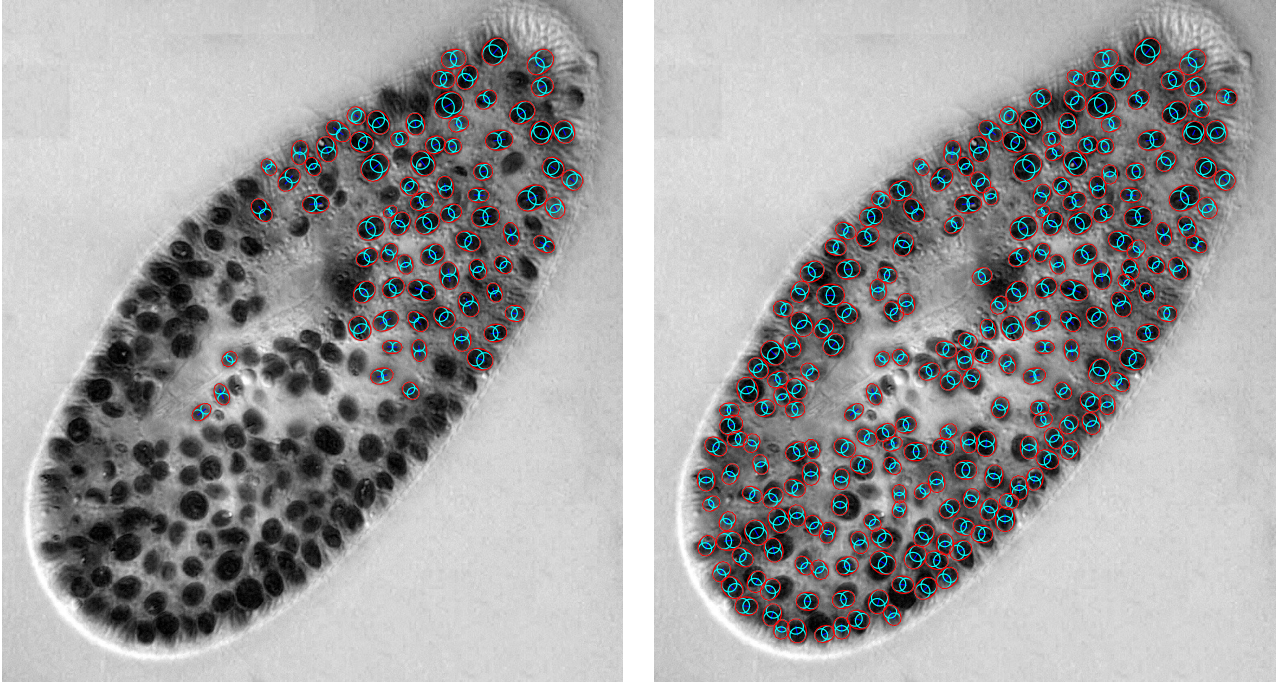


Figure 5. Microscope image ( $1536 \times 1686$  pixels) of the ciliate *Paramecium bursaria* containing many symbiotic green algae. *Left*: Manually annotated symbionts. *Right*: The training shapes and the detected symbionts.

estimated a threshold  $c_0 > 0$  such that the shapes  $\gamma(p_i)$  with  $E(p_i) \leq c_0$ ,  $1 \leq i \leq M$ , represented usable results. In many applications, the manual selection of  $c_0$  does not really pose a problem, because it is done *after* the algorithm is run. *I.e.* changes of  $c_0$  can be visualized in real-time. Furthermore, techniques to estimate  $c_0$  from the distribution of the final energies  $c_i$ ,  $1 \leq i \leq M$ , could be employed.

Compared to greedy techniques this approach is very inefficient but completely avoids local minima. Thus, it effectively demonstrates the capability of the  $E$  to detect shapes from learned shape and image features. Genetic algorithms or the combination of genetic and gradient based approaches might significantly speed up the minimization process. Model based shape detection using genetic algorithms was investigated by Hill and Taylor [10].

For this work we did not do any further investigations on alternative stopping criteria for the algorithm but ran it until the result stopped to improve. An analysis of how the number of random samples, *i.e.* the number of iterations, compares to the quality of the result requires a meaningful way to measure the usefulness of a detection result and is beyond the scope of this work.

## 6. Nonlinear density estimation

The normal density estimation approach presented in Section 3 puts limitations on the range of probability densities we are able to estimate properly. The energy (8) is not limited to normal distributions, though. We illustrate

---

### Algorithm 1 Detection of multiple shapes

---

```

choose  $M$  random shape parameters  $(p_1, \dots, p_M)$ 
 $c_i := E(p_i)$ ,  $1 \leq i \leq M$ 
repeat
  choose a random shape parameter  $p'$ 
  if  $E(p') < c_i$  for some  $1 \leq i \leq M$  then
    if  $\gamma(p')$  does not overlap with any of the shapes
       $\gamma(p_1), \dots, \gamma(p_M)$  then
         $p_i := p'$  and  $c_i := E(p')$ 
    else if  $\gamma(p')$  overlaps with  $\gamma(p_{i_1}), \dots, \gamma(p_{i_k})$  and
       $E(p') < \min(c_{i_1}, \dots, c_{i_k})$  then
         $p_{i_1} := p'$  and  $c_{i_1} := E(p')$ 
        choose  $p_{i_2}, \dots, p_{i_k}$  randomly
         $c_{i_j} := E(p_{i_j})$ ,  $2 \leq j \leq k$ 
    end if
  end if
until  $(p_1, \dots, p_M)$  stop improving significantly

```

---

the capability of the proposed method to model more complicated shape-feature distributions by the use of a kernel density estimator. For shapes, nonlinear statistics were investigated by Cootes and Taylor [4]. The kernel density with a Gaussian kernel is given by the function

$$f(q) = (2\pi)^{-M/2} \det(\Sigma)^{-1/2} \frac{1}{N} \sum_{i=0}^N e^{-\frac{1}{2}(q-q_i)^T \Sigma^{-1}(q-q_i)}. \quad (14)$$

This leaves the problem of selecting an appropriate kernel width, a task which is also known as bandwidth or window width selection [22]. Various approaches have been proposed and there is no single optimal solution in practice. For our experiments we set the variance to the diagonal matrix  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_M^2) \in \mathbb{R}^{M \times M}$ . We chose the diagonal entries of  $\Sigma$  to be the average of the squared distances from each training vector  $q_i$  to its  $K$  nearest neighbors  $Q_i$ ,  $1 \leq i \leq N$ , scaled by a parameter  $\beta > 0$ :

$$\sigma_k^2 = \frac{\beta}{NK} \sum_{i=1}^N \sum_{q_j \in Q_i} (q_i^k - q_j^k)^2, \quad 1 \leq k \leq M. \quad (15)$$

The resulting energy  $E$  of a shape parameter  $p$  is then

$$E(p) \propto -\log \sum_{i=0}^N e^{-(q(p)-q_i)^T \Sigma^{-1} (q(p)-q_i)}. \quad (16)$$

Due to the specific choice of the covariance matrix  $\Sigma$  in (15), the energy (16) is invariant to scaling of the individual components of the shape-feature vectors as long as their  $K$  nearest neighbors stay the same. The energy as formulated in (16) allows to model complex energies at the cost of increased computational effort (for large amounts of training data) as well as the problem of selecting an appropriate kernel width. Note also that in principle the proposed energy is not restricted to energies based on density estimation. One could *e.g.* use neural networks to learn an energy function as proposed in [15].

We illustrate the performance of the kernel density energy on an artificial data set. Figure 6 shows the training data. Note that the corners point downward for one half of the training data and upward for the other half. This creates a shape distribution with two major modes. We applied Algorithm 5 with  $M = 3$  to an image containing two shapes similar to the training shapes and a third straight line shape. For the computation of the shape-feature-vectors we used again the shape-to-feature map (13) and pre-smoothed the images with a 3-pixel-wide Gaussian kernel. In (15) we chose  $\beta = 10$  and  $K = 15$ .

In Figure 6 we compare the results of the minimization of the kernel density to the normal density. The kernel density energy detects the two shapes corresponding to the training data correctly and assigns a significantly higher energy to the wrong result in the middle. Note that the kernel density in energy (16) is not normalized and thus not necessarily positive. The normal density energy accurately detects all shapes but identifies the straight line shape as the best fit (assigning a significantly lower energy to this shape). Thus, in this example the normal density estimate prefers shapes it was not trained for whereas the kernel density energy correctly reflects the geometry of the training shapes.

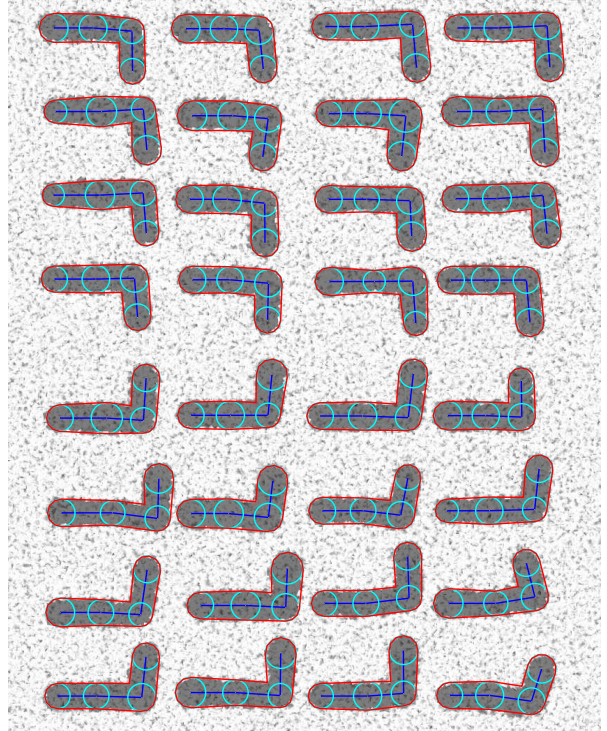


Figure 6. Training data. The two different oriented corners represent the two major modes in the shape distribution

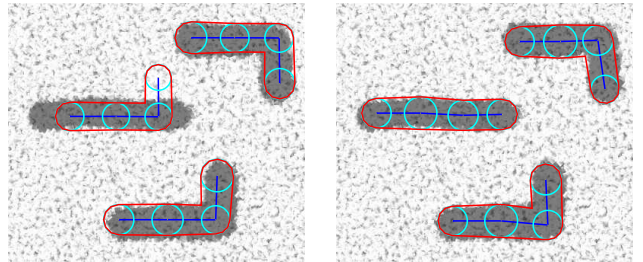


Figure 7. Detected shapes with energy values from top to bottom. *Left*: Kernel density estimation: -1.99, -0.43, -1.78 *Right*: Normal density estimation: 239.05, 87.29, 184.05

## 7. Conclusion and future directions

We suggest a novel variational formulation to shape detection based on training a task-specific segmentation energy. The underlying mathematical model of our method is very general and can be easily used for a wide range of applications. The proposed energy learns the significant shape and image features from training shapes and is able to distinguish them from non-relevant features.

The key advantage over existing approaches is the absence of an explicit regularization term. This avoids the often difficult task of choosing the optimal regularization for a given application. On the other hand, because we incorporate shape priors and rely on a meaningful selection of image features, our approach requires far less training

samples than methods solely relying on learning pixel values [19] and the training is computationally cheap.

Section 6 demonstrates that the proposed method can be easily extended to nonlinear energies to detect objects in cases where the normal density energy might deliver wrong results.

In the future we would like to investigate different energies. *E.g.* learning a kernel based on positive and negative training samples is considered. A Bayesian approach to the parametric density estimation could help to mitigate difficulties due to small training sets (over fitting). Also nonparametric techniques such as adaptive kernel density estimation or projection pursuit density estimation could further improve the performance of the method [11]. Finally, the development of efficient algorithms to minimize the learned energy is subject of ongoing research.

### Acknowledgements

This work was initiated during the special semester *Mathematics of Knowledge and Search Engines* at IPAM, UCLA, 2007, where both authors participated. M. F. has been supported by the Austrian Science Foundation (FWF, project S9202-N12). S. G has been supported by National Institutes of Health (NIH, grant RO1 EB005832) and the National Science Foundation (NSF, grant CCF-0732227). The image in Figure 5 was kindly provided by Bettina Sonntag and Monika Summerer, Institute of Ecology, Leopold-Franzens-Universität, Innsbruck, Austria. We thank Heimo Wolinski, IMB-Graz, Karl-Franzens-Universität, Graz, Austria, and Bettina Heise, FLLL, Linz-Hagenberg, Austria for the image data in Figures 2 and 4.

### References

- [1] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10(2):266–277, 2001.
- [2] Y. Chen, H. D. Tagare, S. Thiruvankadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. A. Geiser. Using prior shapes in geometric active contours in a variational framework. *Int. J. Comput. Vision*, 50(3):315–328, 2002.
- [3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. ECCV'98, Volume II*, volume 1407 of LNCS, pages 484–498. Springer, 1998.
- [4] T. F. Cootes and C. J. Taylor. A mixture model for representing shape variation. *Image Vision Comput.*, 17(8):567–573, 1999.
- [5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Comput. Vision Image Understanding*, 61(1):38–59, 1995.
- [6] D. Cremers and M. Rousson. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In *Deformable Models*, pages 447–460, 2007.
- [7] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snakes: Introducing statistical shape knowledge into the mumford-shah functional. *Int. J. Comput. Vision*, 50(3):295–313, 2002.
- [8] W. Fang and K. L. Chan. Incorporating shape prior into geodesic active contours for detecting partially occluded object. *Pattern Recognit.*, 40(7):2163–2172, 2007.
- [9] M. Gastaud, M. Barlaud, and G. Aubert. Combining shape prior and statistical features for active contour segmentation. *IEEE Trans. Circuits Syst. Video Technol.*, 14(5):726–734, 2004.
- [10] A. Hill and C. J. Taylor. Model-based image interpretation using genetic algorithms. *Image Vision Comput.*, 10(5):295–300, 1992.
- [11] J. Hwang, S. Lay, and A. Lippman. Nonparametric multivariate density estimation: a comparative study. *IEEE Trans. Signal Process.*, 42(10):2795–2810, 1994.
- [12] S. Joshi, S. Pizer, P. T. Fletcher, P. Yushkevich, A. Thall, and J. S. Marron. Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *IEEE Trans. Med. Imaging*, 21(5):538–550, 2002.
- [13] M. Kass, A. Witkin, and D. Terzopoulos. Snakes active contour models. *Int. J. Comput. Vision*, 1(4):321–331, 1988.
- [14] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain. Simultaneous feature selection and clustering using mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(9):1154–1166, 2004.
- [15] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F.-J. Huang. A tutorial on energy-based learning. In *Predicting Structured Data*. MIT Press, 2006.
- [16] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 316–323, June 2001.
- [17] C. McIntosh and G. Hamarneh. Is a single energy functional sufficient? adaptive energy functionals and automatic initialization. In *Proc. MICCAI, Part II*, volume 4792 of LNCS, pages 503–510. Springer, 2007.
- [18] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.*, 42(4):577–684, 1989.
- [19] M. Ranzato, P. Taylor, J. House, R. Flagan, Y. LeCun, and P. Perona. Automatic recognition of biological particles in microscope images. *Pattern Recognit. Lett.*, 28:31–39, 2007.
- [20] M. Rousson and N. Paragios. Shape priors for level set representations. In *ECCV 2002 Proceedings, Part II*, volume 2351 of LNCS, pages 78–92. Springer, 2002.
- [21] M. Rousson and N. Paragios. Prior knowledge, level set representations & visual grouping. *International Journal of Computer Vision*, 76:231–243, 2007.
- [22] B. W. Silvermann. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, 1986.
- [23] A. Tsai, A. Yezzi, C. Tempany, D. Tucker, A. Fan, W. E. L. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Med. Imaging*, 22(2):137–154, 2003.