Sectored Snakes: Evaluating Learned-Energy Segmentations

Samuel D. Fenster and John R. Kender
Dept. of Computer Science, Columbia University, New York, NY 10027, USA
[fenster | kender]@cs.columbia.edu *

Abstract

We describe how to teach deformable models to maximize image segmentation correctness based on user-specified criteria, and we present a method for evaluating which criteria work best. We present sectored snakes, a formulation that demonstrably improves upon regular snakes.

A traditional deformable model ("snake" in 2D) fails to find an object's boundary when the strongest nearby image edges are not the ones sought. But models can be trained to respond to other image features instead, by learning their probability distributions. The implementor must then decide on which of many image qualities to teach the model. To this end, we show how to evaluate the efficacy of any resulting deformable model, given a sampling of ground truth, a model of the range of shapes tried during optimization, and a measure of shape closeness.

In the domain of abdominal CT images, we demonstrate such evaluation on a simple "sectoring" of a snake, in which intensity and perpendicular gradient are observed over equal-length segments. This specific set of qualities shows a measured improvement over an objective function that is uniform around the shape, and it follows naturally from examination of the latter's failures due to images variations around the organ boundary.

1 Introduction

In [1], the authors introduced a flexible shape model that moved until it maximized some combination ("energy") of its smoothness and the cumulative strength and proximity of nearby image edges. But there was no "right" way to set the parameters of this combination [2].[1] Furthermore, it was not established how effective an objective function based on those particular quantities was in the first place. Since then, specific other features have been tried in various domains, *e.g.* matching intensity profiles at feature points [4], matching multiscale intensity and gradient at feature points [5], and centering within a region of classified pixels [6]. But few, if any, comparative studies within a domain have been done.

Let us consider what quantity should be optimized when seeking a shape in an image. To find the boundary of a specific kind of object in a cluttered scene, a shape model should know what it looks like, in a way that differentiates it from nearby object boundaries. If this is not known *a priori*, analytically or otherwise, then the model should *learn* what it looks like. That is, through training it should learn the likeliest values of selected image features associated with the boundary of that kind of object. Several researchers have trained deformable models, but none has studied the choice of training features in the general setting.

In this paper, we introduce a general method to evaluate how well training on given features will perform in a given domain. We then introduce the sectored snake, a particular improvement on the most straightforward kind of training, which is itself an improvement on standard snakes.

1.1 The need for improvement

A *deformable model* is a description of a shape which is iteratively adjusted until it best matches what is depicted in an image. "Best" is measured by an arbitrary real-valued objective function, or "energy." The process of minimizing this energy is sometimes called a "force." In the case of a contour in a 2D image, the model is called a *snake* [1].

There are domains in which traditional snakes, attracted to strongest or closest image edges, fail. The strongest edges are often not the edges of the object we seek. This is true in the domain in which we are carrying out our studies—CT images of lower abdominal organs, which are pressed up against similar organs and brighter, stronger-edged bone (Fig. 1). This is not just a problem of finding suboptimal local minima—in such cases, the wrong object satisfies the objective function better than the right one does.

1.2 Steps toward a solution

If the structure whose contours are sought does not have features (*e.g.* brightness, edge strength) that are more extreme than those of all neighboring objects, the energy being minimized will need its parameters tailored to some domain-dependent intermediate value. This value must

*This work was supported in part by DOD/ONR MURI Grant N00014-95-1-0601, by the New York State Science and Technology Foundation, and by ARPA Contract DACA-76-92-C-007.

[1] [3] used a minimax criterion to balance shape and image components of their objective function in a manner that had nothing to do with what object was sought in the domain.

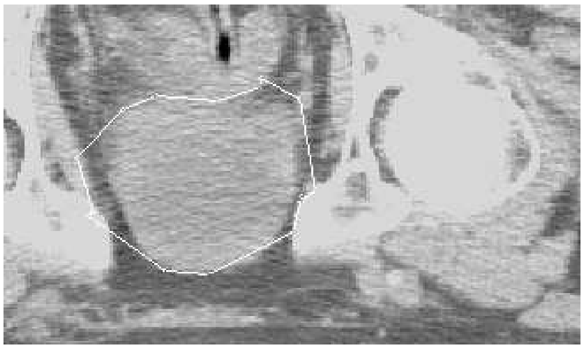

Figure 1: *An organ (bladder) next to stronger-edged bone (pelvis). A traditional snake, attracted to strongest edges, incorrectly clings to bone. The solution is to teach it to recognize the qualities of the correct boundary.*

be found through statistical measurement, which is to say, training. This requires a ground truth training set, which is a collection of images, each with the desired contour correctly specified. Such training on image features has been done in [4], whose model learned a knD Gaussian of intensities perpendicular to a shape boundary at k feature points along line segments of n pixels. Within a similar framework, [5] used a Gaussian distribution of multiscale intensity and gradient at a few points around many organs simultaneously in a human cross-section. Finally, [7] uses a 3D model incorporating the observed likelihood of nearby edges belonging to the liver or not. Here we formalize the general training framework. It easily incorporates prior shape probability, though that is not our focus.

The qualities that training teaches the model to seek may not be uniform everywhere on its boundary (Fig. 2). This calls for a spatially varying objective function. Furthermore, they may have differing degrees of variability at different places, so the function's sensitivity should vary, providing robustness where wide variation is expected. The approach of individually modeling feature points [4] would ignore most contour information. We will present sectored snakes, which address these needs within the training framework.

1.3 Measuring effectiveness

With such a choice of objective functions, we must be able to compare and evaluate their performance, which is domain-specific. Contrary to intuition, the distribution of an objective function's values for *correct* contours gives no information about its goodness for segmentation. Consider that, whatever this distribution, if incorrect contours have the *same* distribution of values, the function cannot guide a contour to the correct shape. We must also look at nearby *incorrect* shapes to assure that an objective function responds to increased correctness by getting closer to optimal. We examine an energy function's behavior for incorrect shapes, by generating such shapes from ground-truth shapes in a sample of domain images.

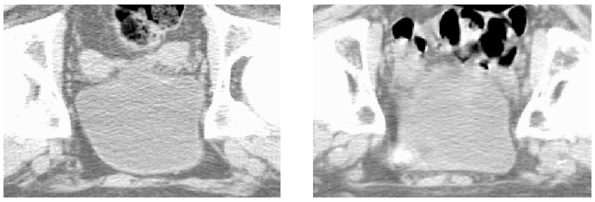

Figure 2: *Image variability. These are bladders of different patients. There are parts of their boundary that vary greatly between images, such as the top, pressed against the highly variable rectum. Other parts of the boundary may have very predictable appearance, like the sides, which are always near pelvic bone, and the seminal vesicles, on either side of the top.*

2 Domain: Radiation Treatment Planning

In our domain, object contours in the human abdomen are needed to plan radiation treatments which will destroy tumors in the prostate. In such treatments, shaped beams of radiation converge on the tumor so as to give it maximum exposure while giving surrounding healthy tissue minimal exposure. Before beam shape and direction can be optimized, the layout of all nearby organs must be known. Each of these is outlined manually, slice by slice, on CT scans, sometimes with the help of simple edge-following algorithms. This is a slow and costly process. Automated segmentation would help greatly, but, because the image is crowded and complex, it is difficult.

Thus, there is a large body of hand-segmented ground truth in this domain. This data may be used to train deformable contour models and to test the accuracy of the resulting models. For our experiments, we concentrated on median slices of the bladder. The possible proximity of a tumor gave our training data the same incidence of organ abnormalities as would be seen in the application of training results in the domain. Our images were from a study in which patients had four images taken at different stages of treatment.

To get accurate training data, we drew our own contours, with agreement from doctors that we could do better than the ones they routinely drew. (Economics and time constraints prevent them from doing better.) For greatest consistency, we trained and tested contours from the organ's middle slice (or two, if there were an even number), where each organ had four to 14 slices (averaging between nine and ten). Up to four 3D images from each of seven patients yielded 24 organ images in 3D, upon whose middle slices we drew 36 high-quality bladder contours.

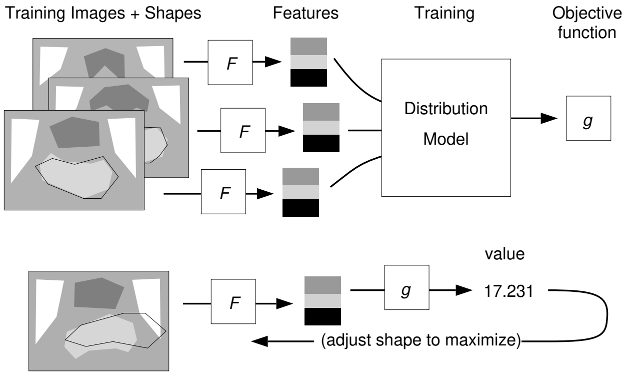

Figure 3: *To train a deformable model, we need to decide on a set of features, extracted from an image and shape by function F. We then need a continuous model of the density of feature values which might be observed. Features are extracted from correct image/shape pairs and their distribution is modeled as objective function g. Now, to find the desired shape in a new image, we find a contour in that image which produces the feature values that g says are likeliest.*

3 Learning

No segmentation method that does not use domain knowledge will be able to distinguish the sought object from neighboring ones. Since many real-world objects, including human organs, do not have image qualities that are known *a priori* or analytically, these qualities must be measured. These qualities will generally have not precisely repeated values but distributions, which must be recovered, through training, from images with known contours. Once they are known, the desired object may be found in a new image by finding the shape whose relationship to the image is likeliest, according to the known probability distributions. Although there has been plenty of work on learning the prior probability of a shape, *e.g.* [8, 3], there has been little on learning a model of shape likelihood given an image.

3.1 Formalizing how to learn

Here, we formalize the notion of learning a general deformable model objective function based on arbitrary shape and image features. Before learning the probability distribution of image qualities associated with a contour, we must select two things: which image qualities to observe, and what model of a probability distribution to fit to them.

We approach the problem in the following way (Fig. 3). We have an image $\mathbf{I}$ in which a shape $\mathbf{S}$ may or may not be a good approximation to the boundary of the object we seek. We base our guess as to the correctness of $\mathbf{S}$ on quantities which say something about it and its relation to $\mathbf{I}$. The function $F(\mathbf{I}, \mathbf{S})$ extracts this tractable set of quantities from its large input.

To know what values of F indicate a correctly placed contour, we must learn the distribution of these values for correct shapes. A probability density function can represent this. It will be a real-valued, continuous function $g(F(\mathbf{I}, \mathbf{S}))$, recovered from a finite set of training data, representing the density observed in that training set in the region of any given vector of feature values $F(\mathbf{I}, \mathbf{S})$. It will approximate the probability $\mathrm{P}(F(\mathbf{I}, \mathbf{S}))$. Before we can recover such a distribution g, we must have some representation for it.

Having picked F and a model to represent g, we can train and use a deformable model. First we must get ground truth: a set of images $\{\mathbf{I}_1, \ldots, \mathbf{I}_n\}$ of the desired object, with its correct outlines in each image, $\{\mathbf{S}_1, \ldots, \mathbf{S}_n\}$. Features from each pair, $F(\mathbf{I}_i, \mathbf{S}_i)$, are the data from which the parameters of g are recovered. The likeliest parameters are those defining a distribution that maximizes the joint probability of observing the training set's features, $\mathrm{P}(F(\mathbf{I}_1, \mathbf{S}_1) \wedge \ldots \wedge F(\mathbf{I}_n, \mathbf{S}_n))$. Since the training data were presumably drawn independently, we must thus find the function g that maximizes

$$\prod_{i=1}^{n} g(F(\mathbf{I}_i, \mathbf{S}_i))$$

Once training has given us g, a deformable model can segment an image $\mathbf{I}$ by finding the optimal shape $\mathbf{S}^*$:

$$\mathbf{S}^* = \arg\max_{\mathbf{S}} g(F(\mathbf{I}, \mathbf{S}))$$

$g(F(\mathbf{I}, \mathbf{S}))$ approximates $\mathrm{P}(F(\mathbf{I}, \mathbf{S}))$. For well-chosen F and model for g, the shape maximizing this probability will also maximize $\mathrm{P}(\mathbf{I} \wedge \mathbf{S})$, and thus also $\mathrm{P}(\mathbf{S} \mid \mathbf{I})$, since $\mathrm{P}(\mathbf{I})$ is constant as $\mathbf{S}$ varies. In a Bayesian approach, this is the maximum-likelihood shape.

A feature function will do poorly if it varies more between images than it does as a shape becomes incorrect. An adequately chosen probability model is one which is flexible enough to allow the recovered g to have high values only for feature vectors commonly produced by correct shapes.

3.2 A basic implementation of trained snakes

In our initial trained segmentation of bladder boundaries in CT scans, we made the following choices. $\mathbf{S}$ was a closed polyline $\mathbf{S}(u)$. Image quantities were observed at a scale s, using a Gaussian-blurred image $\mathbf{I}_s = \mathbf{I} * G^2_{0,s}$. The shape's relation to the image was summarized by image intensities, $\mathbf{I}_s(\mathbf{S}(u))$; and by directional image gradients normal to the shape, $\mathbf{S}^\perp(u) \cdot \nabla \mathbf{I}_s(\mathbf{S}(u))$, where $\mathbf{S}^\perp(u)$ is the unit normal to $\mathbf{S}(u)$. These features, measured along $\mathbf{S}$ at one-pixel intervals, can be considered the output of $F(\mathbf{I}, \mathbf{S})$.

We chose perhaps the simplest model of the distribution of our features—a multidimensional Gaussian. If the incidence of our chosen feature values in our domain is unimodal, a Gaussian may model it well enough. Specifically, we assumed independent, identically-distributed values of intensity and directional gradient at all points along $\mathbf{S}$. Thus, training recovered the parameters of two Gaussian distributions, $N(\mu_{\mathbf{I}}, \sigma_{\mathbf{I}})$ and $N(\mu_{\nabla}, \sigma_{\nabla})$. The joint probability of both these quantities at every point around the contour was the modeled probability of observing those features on a shape $\mathbf{S}$ in an image $\mathbf{I}$, *i.e.* $\mathrm{P}(F(\mathbf{I}, \mathbf{S}))$. This is a product of Gaussians; its negative log is

$$E_{\mathbf{I}} = \oint \frac{(\mathbf{I}_s(\mathbf{S}(u)) - \mu_{\mathbf{I}})^2}{\sigma_{\mathbf{I}}^2} du + \oint \frac{\left(\mathbf{S}^{\perp}(u) \cdot \nabla \mathbf{I}_s(\mathbf{S}(u)) - \mu_{\nabla}\right)^2}{\sigma_{\nabla}^2} du$$

$E_{\mathbf{I}}$ is the "image energy." Minimizing this "energy" is equivalent to maximizing $\mathrm{P}(F(\mathbf{I}, \mathbf{S}))$. We use straightforward gradient descent minimization over shapes $\mathbf{S}$, so we must be able to take the derivative of $E_{\mathbf{I}}$ with respect to the parameters of $\mathbf{S}$.

4 Sectored Snakes: A New Synthesis

We will show that in a cluttered domain, even the simple training described above provides a much better-behaved objective function than does the traditional snake energy. But training will only work if some aspect of this contour-image relationship is more consistent across the training set than for incorrect contours.

In using snakes trained to seek a particular boundary brightness and edge sharpness, we noticed that, unsurprisingly, a snake would behave badly where boundary qualities were locally different. Since these spatial boundary variations are often consistent across images of a particular domain object (Fig. 2), it made sense to formulate a snake that learned what to seek separately on separate portions of its length. In this, we followed [4], although they trained a small number of preselected feature points. We seek an accurate boundary everywhere, so we wish to respond to image data everywhere along the contour, rather than just at a few points.

We thus divided the contour into a fixed number of equal-length regions (sectors), each with separate training (Fig. 4). Not only would this allow the snake to be attracted to different conditions where different conditions were expected; it also would allow stronger objective function response in regions where there is less variability in conditions between images.

In a snake with M sectors, "image energy" is modified to reflect one of M Gaussian distributions, based

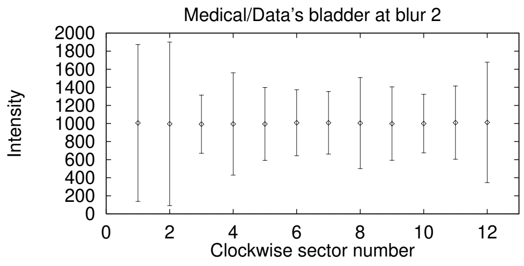

Figure 4: *Each sector of a sectored snake has separately learned a joint Gaussian of image intensity (shown here) and gradient in the normal to the snake direction. The intensity bars are tallest in regions 12, 1 and 2, corresponding to greatest intensity variability at the top of the bladder. A snake with this data would not be as strongly attracted to any particular intensity value at the top as elsewhere.*

on position along the contour. Thus, we have a function $k(u)$ which maps the shape parameter u to a sector, $1, \ldots, M$. Instead of a single pair of Gaussians, $N(\mu_{\mathbf{I}}, \sigma_{\mathbf{I}})$ and $N(\mu_{\nabla}, \sigma_{\nabla})$, we have a pair for each sector: $N(\mu_{\mathbf{I}}(k), \sigma_{\mathbf{I}}(k))$ and $N(\mu_{\nabla}(k), \sigma_{\nabla}(k))$.

The number of sectors was picked large enough to capture the number of different types of regions around the object we sought. In our tests, this was the bladder, which has pelvic bone on each side and seminal vesicles (in men) and rectum in back. If the training data were infinite, more sectors would always be better, because their training would approximate a continuously varying probability distribution of image qualities around the shape boundary, and would capture arbitrarily small variations. But since the training data is finite, more sectors also means that each sector gets less exposure to a statistically representative variety of surroundings in the image, and thus poorer training. We chose twelve sectors based on the sizes of structures that seemed consistently positioned around bladder boundaries. For our initial tests, we wanted to see if our intuition was borne out that even a simple sectoring of a snake in a consistent domain would improve our segmentation over that of spatially uniform training.

5 How to Test Goodness of Learning

The implementor of a deformable model can choose from a wide variety of objective functions. If the function is to be derived from training data, this variety equals the selection of features, and of distribution models to use for each. It is essential that we be able to judge the efficacy of any given function.

Thus far, no methodology has existed for this. The standard procedure is to run the entire deformable model segmentation procedure on a series of images, comparing each resulting shape to the ground truth for that image. But the results of this method only characterize the cumulative effect of: selection of the contour's initial position; the objective function; and the optimization method to find the best nearby contour. It cannot test the effectiveness of their separate contributions to segmentation.

We must evaluate the function over a range of *initial* and *intermediate* contour positions that might be tried during optimization. For optimization to succeed, one (obvious) condition must hold over this range of shapes: *The function must be optimal for the correct shape.* If gradient-descent optimization is used, a more stringent condition must be met: *The function must become more optimal as a shape gets closer to correct.*

To statistically characterize a function's behavior, we randomly perturb the ground-truth shape within an image to produce more- and less-incorrect contours. A perturbation model cannot densely cover the (very large) space of possible nearby shapes; but it can be used to reject objective functions in which the subset of shapes it generates *does* reveal flaws. Such a test is much more complete than running just one segmentation on each image; much more efficient than running a large number of complete segmentations; and, as a test of a single (key) component, more informative than only seeing the result of the entire optimization process for segmenting an image.

Given an image, we must compare the objective function's valuings of perturbed shapes (in the absence of ground truth) to some measure of each shape's distance to the known correct boundary (ground truth). We choose such a measure, chamfer distance, following the next section.

We must evaluate the function over a range of images, and, for each, over a range of contours that encompasses all that might be tried in the search for the correct contour. We tested six different objective functions (three different models, each at two scales) using 1000 parameterized perturbations of known bladder outlines in each of 36 different 2D images.

The relationship of objective function to known shape correctness (distance to ground truth) over the expected range of shapes tells us whether the function will perform well. For each test image, this relationship is described exactly by a scatter plot of these two quantities for many perturbed ground truth contours. We wish to test whether any deformation of a shape that decreases its energy actually brings it closer to ground truth. We get the strongest assurance of this from scatters which are most tightly clustered around some increasing function. There are various ways this relationship may be quantified. One is by calculating the correlation coefficient—the closer it is to +1, the better. Another key statistic is the number of contours having energies below that of the ground truth, thus falsely indicating that they are better solutions than the correct one; ideally, there would be none of these. Whatever statistic is chosen, one must combine its values for many images.

5.1 Range of contours: a statistical approach

We get a statistical sampling of shapes around a ground-truth shape by perturbing that ground-truth shape. The perturbations we used for objective function evaluation were a combination of translation and of scaling independently around two randomly chosen orthogonal axes centered on the contour's centroid. The translations were normally distributed with a standard deviation of 5 pixels; the logs of the scale factors were normally distributed with a standard deviation of 10% stretch/shrinkage.

5.2 A measure of distance to known ground truth

To measure a perturbed contour's distance from the correct shape, we rejected the Hausdorff distance as unstable, and the area of the symmetric difference of the enclosed regions as inefficient to calculate. We chose the chamfer distance between two boundaries [9], which is the average over one shape of distance to the closest point on the other. To calculate this, one does a distance transform on an image containing one shape, then averages the distance values intersected by the other. This measure is asymmetric, but (in our domain, at least) has a high correlation (0.977) between $d(a,b)$ and $d(b,a)$.

5.3 Putting it all together

Once we have a model of how to perturb ground-truth contours, and we have a measure of the distance between two contours, we can proceed to characterize the goodness of a deformable model's objective function. This goodness can be summarized in a plot of shape incorrectness *vs.* function value (Fig. 5).

6 Results

To the best of our knowledge, this is the first formal measurement and comparison of the effectiveness of deformable model energy functions.

Training and segmentation were done for three different snake formulations: the sectored snake described in section 4; the simple trained model described in section 3.2; and a traditional, untrained snake whose energy is gradient strength traversed by the contour, at some scale: $E_{\mathbf{I}} = -\oint ||\nabla \mathbf{I}_s(\mathbf{S}(u))||du$. Each of these was tested at two scales, $s = 2$ and $s = 4$, for a total of six different objective functions.

We evaluated each of these energy functions, using the procedure described in section 5, with 1000 perturbations of the ground truth contour in each of the 36 slices. Below we look at two statistics of the results. The first is the degree of correlation between a contour's image energy and distance from the image's ground truth contour. The second is the number perturbed contours whose image energy is lower than that of the unperturbed ground truth contour ("false positives").

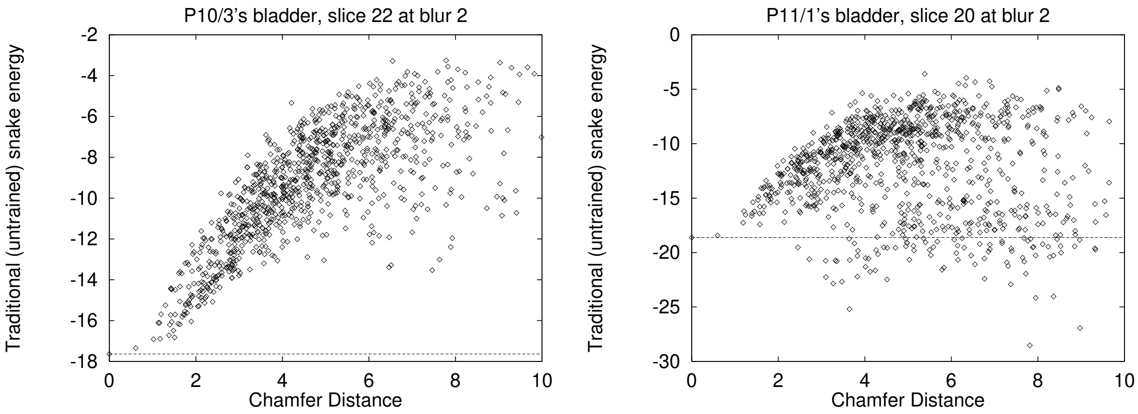

Figure 5: *Examples of plots showing how effective a deformable model's objective function will be in a domain. Given an image and the correct contour of a shape in it, a large set of perturbations of the shape are generated. Each is a point in a scatter plot of difference from correct shape* vs. *energy. A good objective function will display strong correlation with shape correctness, and will have no "false positives," contours with lower energy than the ground truth (dotted line). These plots show that the traditional snake energy works only in some cases:* Left: *a high-correlation (0.73) plot with no false positives.* Right: *A low-correlation (-0.04) plot with 8.8% false positives.*

6.1 Correlation

Our aim is to check whether contours that are further from ground truth have consistently higher (less optimal) energies. One guarantee of this is if the scatter plot of distance d from ground truth *vs.* energy E is tightly clustered around some increasing function. A high linear correlation coefficient is an indicator of this, although a low value does not preclude it. This statistic is shown, for each energy function, in Table 1.

Snakes with sectored training had slightly, but not significantly, higher correlations than unsectored trained snakes, at both image scales. Coarser (blurrier) image scale increased correlation by a small amount. But most importantly, both of the trained formulations demonstrated a much higher correlation than traditional snakes.

6.2 "False positives"

An energy function is guaranteed to produce some incorrect segmentations if it is smaller for some close but incorrect contours than it is for the ground truth. Such "false positives" occurred far less often with sectored snakes than unsectored snakes. Unsectored trained snakes, in turn, performed much better than traditional snakes. The average incidences of false positives per image for each energy function are in Table 1.

Although the coarser-scale energy functions showed correlations with distance to ground truth that were just as good as at the finer scale, they have many more false positives—a factor of ten, for the two trained formulations. This disqualifies scale $s = 4$ for use in segmentation in this domain.

Table 2 shows what portion of the 36 image slices met progressively more relaxed standards (number of false positives allowed), for each energy function. For instance, 72% of the image slices had less than 2 in 1000 bladder contour perturbations with lower sectored-snake energies than the truth. Only 28% of the images had that few glitches with the traditional snake energy.

Objective function	Scale	Avg Corr	Deviation of Corr	Avg FPs	Deviation
Sectored and trained	2	0.56	0.12	0.6%	0.8%
	4	0.61	0.13	5.8%	6.4%
Unsectored but trained	2	0.55	0.14	1.1%	1.6%
	4	0.58	0.14	10.9%	10.9%
Untrained (traditional)	2	0.15	0.33	15.4%	14.5%
	4	0.21	0.37	23.2%	16.9%

Table 1: ***Degree of correlation*** *between contour image energies and chamfer distance to correct shape: In our domain, the sectored snake's energy is slightly more correlated to shape correctness than the unsectored snake, and is much better than a traditional snake. Though correlation is higher at a blurrier scale, the following statistic shows that the bigger blur has problems.* ***Incidence of "false positives"*** *(contours with objective function values better than correct shape) for different objective functions: The sectored snake decisively outperforms the snake with unsectored training in this image domain. More blurring causes each of the three function types to make many more mistakes.*

6.3 Discussion

The evaluations show us that in our domain of cluttered images, the enhancements we made to snakes were necessary to get good segmentation results. The simple statistics we extracted from energy-distance plots showed that in this domain, a standard, untrained snake would have failed most of the time. The data also provide evidence that het-

Objective function	Scale	FPs=0	$< .2\%$	$< 2\%$	$< 5\%$
Sectored	2	53%	72%	94%	97%
and trained	4	8%	11%	56%	72%
Unsectored	2	42%	72%	89%	92%
but trained	4	8%	14%	31%	56%
Untrained	2	14%	28%	33%	42%
(traditional)	4	14%	22%	31%	33%

Table 2: *Percentage of 36 images in which fewer than the specified percentage of 1000 perturbations were false positives.*

erogeneous training (sectoring) reduces the learned objective function's false positive rate significantly over simple unsectored training.

These tables show much better performance for energies based on the fine image scale than for those on the coarse scale. Unfortunately, when using gradient-descent optimization, coarser scales are necessary to get a nonzero gradient. An unblurred energy quickly reaches zero (plus noise) as a contour departs from the correct shape, so gradient descent will not move an incorrect contour toward the solution. Since blurring seems to often prevent energy functions from being able to distinguish right and wrong boundaries, and yet blurring may be necessary for gradient descent optimization, gradual deblurring is indicated—a standard robust optimization technique recommended in [1]. In addition, there may be domains where correct contours are actually more distinguished at coarse scales.

7 Conclusion and Future Plans

We have presented a systematized framework for the training of deformable models based on any chosen shape and image features and objective function model. It allows unified analysis of the diverse deformable model formulations that have existed up to now. Within this framework, we have presented two results in the segmentation of imagery by means of deformable models: a method to evaluate the objective function; and the sectoring objective function.

Since the method is really a class of methods, perhaps the major contribution is that it permits knowledgeable experimentation over many image features, allowing researchers to compare results more accurately, determine causes of objective function failure, and devise more precise measures of model accuracy and reliability.

We intend to extend the work in several mutually supportive directions: (1) We are already starting work in another image domain. (2) The class of image features is easy to extend beyond intensity and gradient measures; in addition to higher-order measures such as texture, this method easily accommodates measures such as boundary placement relative to distant landmarks. (3) The definition of a "sector" needs to be explored, particularly with respect to whether image statistics can specify some "natural" sectoring more robustly than our 12 points of the clock face. Also to be explored are continuously spatially varying training, as opposed to discrete sectors; and the dependence between expected features at different locations. (4) We must check if some other probability models fit the distributions of some features better than a Gaussian. For instance, a multidimensional spline could be fit to training feature density data.

Acknowledgments

We would like to thank Drs. Radhe Mohan and Chen Chui of the Medical Physics Computer Services Department at Memorial Sloan-Kettering Cancer Center for their expertise and their image and contour data.

References

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proc. ICCV*, pp. 259–268, 1987.

[2] T. E. Boult, S. D. Fenster, and T. O'Donnell, "Reinterpreting physically-motivated modeling," in *Proc. ARPA Image Understanding Workshop*, (Monterey, CA, USA), Nov. 1994.

[3] K. Lai and R. Chin, "Deformable contours: Modeling and extraction," *PAMI*, vol. 17, pp. 1084–1090, Nov. 1995.

[4] T. Cootes, A. Hill, C. Taylor, and J. Haslam, "Building and using flexible models incorporating grey-level information," in *Proc. ICCV*, pp. 242–246, May 1993.

[5] B. Baldwin, *Multiscale Snakes*. PhD thesis, Courant Institute of New York University, 1997. To appear.

[6] C. A. Davatzikos and J. L. Prince, "An active contour model for mapping the cortex," *IEEE Transactions on Medical Imaging*, vol. 14, pp. 65–80, Mar. 1995.

[7] J. L. Boes, C. R. Meyer, and T. E. Weymouth, "Liver definition in CT using a population-based shape model," in *Proc. 1st Int. Conf. on Computer Vision, Virtual Reality, and Robotics in Medicine (CVRMed '95)*, pp. 506–512, Apr. 1995.

[8] C. Kervrann and F. Heitz, "A hierarchical statistical framework for the segmentation of deformable objects in image sequences," in *Proc. CVPR*, pp. 724–728, 1994.

[9] G. Borgefors, "Hierarchical chamfer matching: A parametric edge matching algorithm," *PAMI*, vol. 10, pp. 849–865, Nov. 1988.