Reinterpreting Physically-Motivated Modeling

Terrance E. Boult, Samuel D. Fenster and Thomas O'Donnell*
Center for Research in Intelligent Systems, Department of Computer Science,
Columbia University, New York, N.Y. 10027, U.S.A.

Abstract

Deformable models in the "physically-based" paradigm are almost always formulated in an ad-hoc fashion and are not related to physical reality. We reinterpret these techniques by putting them into a framework of robust statistics and using this framework to analyze the problems and ad-hoc solutions found in common physically-based formulations. These include incorrect prior shape models; bad relative weights of various energies; and the two-stage approach to minimization (adjusting global, then local shape parameters). We examine the statistical implications of common deformable object formulations. In our reformulation, the units are meaningful, training data plays a fundamental role, different kinds of information may be fused, and certainties can be reported for the segmentation results. We suggest robust statistics to combat interference from the necessarily large amount of unmodeled image information.

1 Introduction

Deformable models have proven to be a useful tool for finding shapes in an image. But the approach usually taken is to make the model simulate a physical material, rather than to derive its properties from the requirements of the segmentation problem. This paper examines the assumptions that are usually made when formulating a deformable model, and the issues that are neglected. In particular, we examine the meaning of the "energy" which is minimized. We suggest a probabilistic paradigm for formulating and analyzing energy which gives it meaning, and thus gives the resulting minimal boundary a meaning. While such a paradigm has been used profitably in many domains, we have never seen it properly explored for segmentation using deformable models. Good segmentation often requires that parameters be "learned" from training data. To date, the use of training data has been inadequately explored; we note that proper analysis of the training process is virtually synonymous with a probabilistic approach. We explore the relationship between the energy function, the model of uncertainty and the training data. This relationship is nontrivial due to lack of knowledge and issues of computational efficiency.

It is computational intractable to model a shape's probability directly from all the pixels in an image and their correlations. It is also impossibly hard to find the globally optimum shape even when one has a tractable function. Thus, one must treat some aspects of the image as having unknown, possibly "adversarial" distributions, producing image data which will lead to a false segmentation. We suggest techniques using *robust statistics* to minimize the damage caused by unmodeled data. This may involve reweighting our objective function so as to gradually trade robustness for accuracy as the optimization proceeds.

In the next section, we describe what a deformable model is. We then discuss the shortcomings of existing practice. We go on to describe the kinds of uncertainty in the process of segmenting an image, and then talk about how to formulate an "energy" which represents these uncertainties. We interpret known effects and existing practices in this framework, and suggest improved approaches.

2 What are Deformable Models?

This section provides an overview of deformable models. Those familiar with them can probably skip it.

Deformable (or *active*) models are curves or surfaces which "live" in a 2D or 3D image. The idea is that an automated process can move them so that they lie on the boundary of some structure in the image. The

*This work supported in part by ARPA Contract DACA-76-92-C-007 and by NSF PYI award #IRI-90-57951. T. Boult is now at EECS Department, Packard Laboratory, 19 Memorial Drive West, Lehigh University, Bethlehem, PA 18015.

process that moves them is a combination of "forces" that are determined by the image and by the current shape of the model. In the original formulation [Kass *et al.*-1987], an active contour ("snake") is manually placed near the desired object boundary in a 2D image. It changes shape and position, subject to internal stiffness forces and to attraction by nearby pixels of high gradient. It iteratively responds to these forces until it stabilizes at a position which is a compromise between sitting along the boundary (strongest edges) and maintaining smoothness.

A deformable model can be a discrete chain or mesh of node points connected by length (spring, tension) constraints and stiffness (curvature) constraints. This is a finite difference model (FDM). It can also be a finite element model (FEM), which is made up of continuous parametric segments or patches, connected by continuity constraints. The nodes of a snake are sometimes called "snaxels." In an FDM, forces of image attraction act only on the nodes, whereas in an FEM, they can act on any point, or all points, of an element. The model may also be represented as a continuous parameterized global shape, as a sum of global shapes (modes), or as a global shape with patches representing local deviations [Terzopoulos and Metaxas-1991].

The "forces" which animate the model are often the gradients of "potential energies" based on the deformable model's shape and on its position in the image data. If the shape is defined entirely by positions of nodes, the forces may be seen as acting on these nodes in 2- or 3-space. But, more generally, no matter how the shape is parameterized , the forces can be seen as acting on the single point in a higher-dimensional parameter space which represents the shape. An energy function's arguments are the shape parameters and the image. The energy function returns a single real number. The energy function's negative gradient, a vector in parameter space, is the "force" which moves the shape parameters in the direction which most quickly decreases the energy. In this way, a local energy minimum is sought. If the energy was properly formulated for the application domain, this energy-minimizing shape represents the "best" segmentation. We must trust that the initial placement of the model was close enough that the local minimum we found was the "right" one.

A deformable model is usually formulated as an object with simple physical properties which define its potential energy. This energy is composed of various terms; those which do not depend on the image are divided into *internal energies* and *user energies*. The rest are *external* or image energies. Typical internal energies are related to node distances, curvature and higher derivatives. These usually have their squared magnitude summed over the curve or surface. They often simulate the potential energies of mechanical entities such as springs, stiff rods or thin plates. User forces, *e.g.* *balloon forces* [Cohen-1991], push the model outward or toward some point. Typical external energies assign strengths, *e.g.* intensity or gradient magnitude, to image pixels or edgels and multiply or divide these by (often squared) distance. Energies are usually selected to make the solution of a differential equation quick or in partially closed form.

Each kind of energy can be seen as "penalizing" something, such as size, curvature, distance from image edges, or image intensities near the shape. An improvement is to penalize *deviation from the expected value* of these quantities instead.

The various energies are weighted and added up. Various methods exist to select the weights, such as cross-validation [Bates and Wahba-1982]. Differential equations are set up to simulate the model's movement. These equations sometimes include such physical quantities as mass and viscosity, to stabilize the solution [Leymarie and Levine-1993].

There are several advantages to using a deformable model to segment an image.

- It explicitly minimizes a chosen function, so it allows for arbitrary definition of the "best" solution. (Unfortunately, this flexibility is rarely exercised.) This function can easily combine many criteria.
- In continuous formulations it provides a continuous solution that best fits discrete data. Thus, data is interpolated, properties such as curvature and volume are easy to get, and rendering is easy.
- It handles missing, fuzzy, noisy or otherwise uncertain data robustly because internal forces, and global shape parameters, can keep the shape consistent (by any desired definition) in places where the data forces are weak.
- Topological consistency is assured because the manifold's topology is specified *a priori*. (This can also be a disadvantage.)
- The deforming manifold's geometric properties and responsiveness to data can be locally parameterized.

- Without prior knowledge, it can deform to fit almost any object of interest. But if the object sought is modeled *a priori*, accuracy and convergence can be improved. Objects recovered using a single model may vary considerably and may actually deform, yet accurate boundaries can be found.

3 Motivation

In this section we motivate our approach to deformable models by pointing out some of the problems that exist in current formulations. We believe the primary source of these problems is the ad-hoc approach generally taken towards recovery. It is manifested in the following ways:

3.1 Inappropriate Prior Model Shapes

Internal energy terms which penalize surface curvature may result in model "rest" or "prior" states (the expected shape in the absence of data) which in no way describe the object under recovery.

Smoothing terms which seek to minimize the total size/bending of the finite elements composing the surface are extremely common in the deformable modeling paradigm [Cohen and Cohen-1990, Leymarie and Levine-1993, Kass *et al.*-1987, Terzopoulos and Metaxas-1991]. Unfortunately, one way to minimize these values may be to shrink the element. As a result without the influence of data forces the "rest" or "prior" shapes for many of the models are unlikely to resemble the object of interest. For many snake algorithms, *e.g.* [Kass *et al.*-1987], the rest-shape is a single point. That is, given no data, the model disappears! And even in the presence of image data, there is always a force pulling inward, so the model will reach equilibrium somewhere short of the object boundary.

From one point of view a single point could be an acceptable expected shape in the case of recovering or tracking unknown objects. If however, there is some idea of what is being sought, this prior would likely be inappropriate. With good knowledge about the domain one would hope that even if there were no data the model would conform to a "likely" shape. Similarly for sparse or noisy data one would hope that "likely" shape might help account for missing regions in the dataset.

In a small subset of deformable model literature [Leymarie and Levine-1993, Lai and Chin-1994], the internal stiffness force is modified to penalize each element based on its difference from an *a priori* specified "preferred" element size. By doing so the models will not disappear in the absence of data, but can be viewed as having an expected shape of either a straight line with fixed size links (for open snakes) or a circle formed from fixed-size links (for closed).

This may be one reason that user supplied "balloon" forces, [Cohen-1991] may have become popular—they add inflationary forces which cause the rest state to become a circle, with the diameter dependent on the relative strength of the balloon forces and the stiffness forces.

Given this approach, determining an appropriate prior shape still may not be straightforward, however. Should each element have the same preferred length and curvature? How should these lengths and curvatures be determined for a particular domain? And, if these preferred lengths come as a result of examining several fits to objects of a particular class (healthy left ventricles, for example), wouldn't a *distribution* of prior shapes might be more useful in the fitting process than a single expected shape?

Vemuri's work [Vemuri and Radisavljevic-1993, Vemuri and Radisavljevic-1994] went a bit farther. He discussed uses a MAP formulation and provided rudimentary statistical interpretations of their energy formulations, including the idea of a "mean" shape. He used a simple mean and variance computation to learn good prior values for the global parameters. This prior state was, however, a rather poor approximation to his object.

3.2 Large Perturbations of Linear FEMs

Almost all of the deformable object literature in vision and graphics has used FEMs, in particular linear FEMs. These are designed for "small" perturbations from an initial state.[1] However, in the case of poor or "uninformed" priors the model may be forced to exhibit large deformations!

One can think of the linear FEM as providing a first order approximation to the stiffness properties and is generally valid only in a local neighborhood [Bathe-1982]. Inappropriately initializing a model and depending on, for example, balloon forces to push the model towards the data may result in an extreme dis-

[1] Exceptions using non-linear FEMs include [Terzopoulos and Waters-1993] and [Huang and Goldgof-1993].

tortion of even well calibrated model stiffness properties.

There has been more recent work on building global object models with FEM model on top of them. In particular, [Terzopoulos and Metaxas-1991, Park *et al.*-1994] consider fitting an superquadric to an underlying model and then using an FEM surface to account for the remaining details. In [Pentland and Sclaroff-1991], a global model using low order modes is recovered, and then a spline surface is fit on top to account for remaining details. In both cases, however, the underlying model is a generalized blob that is a very rough approximation. For most objects, the resulting deformations are still moderate in size.

3.3 Poor Fits Using Residual Data Forces

Under existing deformable model paradigms if the model is endowed with material stiffness it becomes impossible for it to deform to interpolate the data in the presence of "residual" data forces even if that data is guaranteed to be perfect!

We define residual forces as those which tend to zero as the distance between model and data tend to zero. An example of residual forces are segmented data forces (long range forces in [Terzopoulos and Metaxas-1991]) which are weighted distance between a model point an a data point. Residual data forces also occur when the image energy is the magnitude of the Gaussian-blurred gradient through which the shape passes. The image force goes to zero as an image edge is approached.

When such residual data forces combine with internal material stiffness, the model ends up sitting somewhere between the prior model and the data. In the presence of noisy data this situation can be quite acceptable. The material forces of the model combine with the data forces to determine a compromise over the surface description. Thus the price paid for preferring a smooth surface, for example, is that the data might not be interpolated.

But what if the data is virtually noiseless? Under conditions of material stiffness the model still will not interpolate the data!

On careful reading, one might notice that in many of the deformable modeling papers the reported "stiffness" parameters are set to very small values,[2] which in turns allows the surface to nearly interpolate the data. It also means, however, that in regions of little data the surface is free to wobble and curve with little internal constraint.

Another way in which researchers have circumvented this problems is by "inventing" new forces such as a "balloon forces." These forces are not predicted by the underlying physical analogy as they are neither material forces (stiffness) nor data (image) forces. Rather they are extra forces which, in an ad hoc fashion, force the equilibrium point to be "closer" to the data. (They are also used to provide an initial force to get the model near the data.)

3.4 Ill-founded Parameter Assignments

The methods in the current literature for determining the scaling values for data forces and material properties are often quite ad-hoc (when described at all) in that they are not geared towards finding the model with the "best" fit consistent with both the data and prior expectations of the object.

In the case of a stiff model being influenced by noisy data how should the the data forces and material stiffnesses be scaled such that the "best" fit is arrived at? That is, how should the "internal" and "external" forces be related? These questions are by and large ignored by papers in the deformable model community.

A related question is how should different *types* of external forces be related to one another in the physically-motivated paradigm? Forces to date have largely been based on model-data distances or on image gradients. However, forces may also be based on the similarity between gray-scale values of an edgel in a data image and the values expected by the model. Determining the proper relative strengths of these external forces is at best problematic not least because their very units differ! Even worse, the units of internal energy (stiffness) and external energy (data forces) don't even match. How can we relate intensities, for example, to smoothness?

In our prior work (surface modeling via regularization) the multi-sensor fusion problem was examined[Boult and Moerdler-1987, Moerdler and Boult-1988] using stereo and texture information.

[2]For example in [Terzopoulos and Metaxas-1991] the stiffness measures are of magnitude 10^{-6}. (Of course, since the units of the forces are not even comparable, it is possible to rescale them to any range one wishes.) What really matters are the relative sizes of the internal and external forces. Determining this from existing papers is not possible.

The scaling directly determined the success of the integration. Without a firm understanding of how the scales could be determined, fusion was extremely difficult.

3.5 "Efficiency" Hacks

Various systems for deformable object models, including [Terzopoulos and Metaxas-1991, O'Donnell *et al.*-1994a, O'Donnell *et al.*-1994b] include code for "parameter scheduling." This code is used to insure that in the initial stages of the recovery, only global parameters are adjusted. When the global parameters have settled, adjustment of the local parameters begins. There are two reasons for this parameter scheduling. The first is to reduce computation time; the second is to provide robustness—if the local parameters are allowed to adjusted too soon they can "latch" onto a strong feature that is not part of the object of interest.

While it has advantages, is this method likely to give us the "best" solution? Without some solid foundation for judging the "goodness" of a fit it is impossible to tell what the efficiency costs in terms of fitting.

3.6 Poor Justification for Results

Given the "ad-hocness" of current approaches, justification of results to users such as radiologists becomes an issue. In most physically-motivated applications the stiffness and data forces are in reality selected based what works for some small test set of data. Thus it becomes impossible to state for example that a segmentation performed yield the "best" estimate of the object contour given the data and some prior knowledge of the domain. If a deformable modeling system is to be used in clinical medicine, having a justifiable interpretation of the process is not just an academic issue, it is a prerequisite to acceptance.

4 A Robust Statistical Approach to Deformable Modeling

In this section our novel approach to deformable model recovery is presented. We claim that the weaknesses of existing methods outlined in the previous section may be overcome by recasting deformable object modeling as a robust statistics problem. In doing so it becomes possible to give meaning to both external and internal forces and thereby create a solid basis from which to interpret recovery results. In addition results may be reported with confidence levels indicating the degree of success. Finally, the training necessary for any recovery system is integrated into the paradigm at a fundamental level.

4.1 Introduction

In the context of robust statistics, more specifically M-estimators (see below), deformable modeling becomes a question of relating sensor models, data uncertainty, model uncertainty, and the uncertainty in our uncertainty distributions to arrive at a solution of maximum likelihood.

We hope this strategy will provide a solid foundation upon which to analyze existing assumptions and techniques and to propose new methods. These methods might include new force formulations or novel stiffness properties rooted not on some preconceived notions about how the prior models *should* behave but rather on actual deviations from the expected model gained from training instances. We anticipate using the results of this analysis to improve the quality of our deformable recovery algorithms and to better relate them to actual objects.

4.1.1 Giving Meaning to Forces and Stiffness

The physical analogy as often used for the "physically-based" models has weaknesses. While there are lots of impressive equations supporting the paradigm, data points are *not* spring forces acting on a spring-like contour or surface. Furthermore, the material properties of the model may be completely unrelated to those of the object undergoing recovery. One can picture a model deforming to describe a ceramic plate despite the fact that the ceramic is rigid! Because the physical analogy does not hold, we prefer to use the phrase, "physically-motivated techiques."

We note that the statistical interpretation is also very intuitive, and compatible with the physically-motivated approach. A material that is "unlikely" to deviate from its base shape is "stiff" in the physically-motivated approach because it has an energy formulation (uncertainty measure) that is more concentrated around the base shape. Data of which we are more certain will also have a more concentrated probability measure, hence larger (but shorter range) forces. If the data is known to be perfect the force would be infinitely strong resulting in the interpolation of the data.

Similarly, the stiffness of the model (*i.e.*, its predilection for maintaining its initial shape) represents the confidence in the correctness of that initial shape. If the object under recovery is unlikely to have that particular form the model stiffness should penalize a deformation towards that form. And, an object with a high expectation for having a smooth surface should be recovered by a model for which surface roughness incurs a high cost.

4.1.2 Reporting Confidence Levels

In the probabilistic framework it is possible to report confidence levels for individual results thereby enhancing the results. In any experimental science a result lacking an error bound is a useless result. Up to now errors in physically-motivated recovery have been demonstrated but not well analyzed. Such demonstrations errors give an *indication* of how well a similar case might do. By approaching recovery from a statistical point of view, however, it becomes possible to provide a confidence level appropriate for each individual experiment based on the confidences in the appropriateness of the prior model and the fidelity of the data for that experiment.

Such confidence levels provide several advantages. First, they create a means for determining the general predictability of a technique. Second, they lend a quantitative basis for disregarding certain recoveries as unlikely. Finally, they enable us to justify our results to end users. Doctors, having some knowledge of statistics (a required class in medical school), will be more likely to understand and embrace results based on statistics rather than on an imperfect analogy to the physical world.

4.1.3 Integrating Training

To achieve satisfactory results in any recovery system, training the system parameters for a particular domain is necessary. Robust statistics provides well understood facilities for this training.

Training is integrated into both the internal and external force formulations at a fundamental level since the forces themselves are probability distributions calculated via training. The relative scaling of these forces becomes a problem of relating probabilities, a basic computation in statistics. And the units of these forces are no longer at issue since they are unitless probabilities. Finally by invoking the powerful tools of robust statistics an appropriate prior model may be arrived at.

4.2 The Many Kinds of Uncertainty

Regardless of our segmentation technique, we ultimately want to recover a shape that maximizes some likelihood. The image data is (hopefully!) not uncorrelated to the shape we seek. Thus, it makes sense to either:

- Try and formulate the likelihood that a recovered shape is correct, as a function of the image and other available information. An algorithm that can maximizes this function produces optimal results. Or:
- Without being certain of this function or how to maximize it, make sure our existing algorithm plausibly approximates the maximum likelihood shape.

To accomplish either of these goals, we must be able to characterize the probability that a given shape is the one that produced the image. A brute-force approach would fit a probability distribution function to an enormous collection of ground-truth shape-image pairs. But the space of functions, and, indeed, the space of inputs, is too big for us to take this approach.

Instead, we must use our knowledge of the domain to boil the shape and image down to a small number of useful features, and model the probability as a function of these. Our choice of features will reflect our knowledge of the objects being imaged, of the imaging process, and of the cost incurred by each feature.

If a deformable model is to seek a shape with maximum probability, given the image, it must not ignore any important kinds of knowledge. Each provides a significant bias to the energy function, which affects the goodness of the recovered shape. We say that the function is defined by the *uncertainty* of many features of the problem domain.

When we say "uncertainty," we mean any aspect of the problem which has variability. Such items are correlated with the image produced and with the shape that we are trying to find and thus affect our estimate. Here we describe each such item.

- The most obvious of these is variability in what the shape parameters of the sought-after object might be. It is useful to know the *a priori* likelihood of a shape. And we wish to find correlations between the shape and as much else as possible.

- The obvious correlation is with the varying positions and shapes of surrounding objects. For instance, if a medical image of a prostate is always directly below the bladder, then a hypothesized prostate position is highly unlikely if the pixels above it are not bladder-colored.
- The model must clearly have certainty data about the appearance (coloring and texture) of the imaged objects as well as their shape parameters. If the color of the bladder in the above example were highly variable, the probability of a hypothesized prostate shape would *not* be much affected by the color of the pixels above it.
- Sensor noise plays an important role in determining the certainty of a shape. Modalities such as ultrasound and MRI can be very noisy, and feature detection algorithms can introduce noise. If the sensor's mapping of position or color from object to image is very uncertain, image data will have less credence compared to prior knowledge of shape. Noise can affect certainty differently in different parts of an image.
- The ground truth shape for a particular image may be uncertain! In the case of a medical image, we can only compare a machine segmentation to a doctor's, and different doctors may draw the outline differently.

The amount of variability in these factors controls how much their values affect a shape's probability. The correlation between them is also important—items that are highly variable, yet highly correlated, should strongly affect the estimate of a segmentation's likelihood.

There is one other kind of uncertainty: Our uncertainty models will themselves be uncertain. In most domains, it is impossible to accurately know the numerical probability that an object has a specified shape, given an image. Furthermore, we may not be able to efficiently calculate that which we *could* know. But deformable models work because we can come up with approximate uncertainty models that behave properly when their inputs are close to the correct solution. Thus, our certainty about our estimate of probability will change during our search for the maximum-probability shape.

Because of this, we must analyze which components of our probability function we know well, and in what portions of their domain (image and shape) they are accurate.

4.3 The Probability Framework

In this section we explore the probability formulation of deformable models. We shall discuss prior work in this area *after* discussing most of the issues, so that it can be examined in the light of these issues.

The probability we want to maximize with our deformable shape is $\mathrm{P}(S \mid I)$, the probability of S being the "right" shape given that I is the image. Of course, since any single configuration has infinitesimal probability, P actually represents a probability density function. We are trying to find the following:

$$\begin{aligned}\arg\max_S \mathrm{P}(S \mid I) &= \arg\max_S \frac{\mathrm{P}(S \wedge I)}{\mathrm{P}(I)} \\ &= \arg\max_S \frac{\mathrm{P}(I \mid S)\mathrm{P}(S)}{\mathrm{P}(I)}\end{aligned}$$

When looking for the S that maximizes this, we can ignore the denominator, $\mathrm{P}(I)$, because it is constant with respect to S. It is the probability of image I occurring, without knowing S. It is useful to decompose $\mathrm{P}(S \wedge I)$ into $\mathrm{P}(I \mid S)\mathrm{P}(S)$ because each of the two factors can be efficiently approximated. Let us examine each factor separately.

$\mathrm{P}(S)$ is an *a priori* shape model, a function which measures the inherent likelihood of the shape regardless of the image. It can penalize unlikely sizes, positions and curvatures. Note that this corresponds to the *internal energy*—all components of the energy which do not depend on the image.

$\mathrm{P}(I \mid S)$ penalizes a shape if the image does not correspond to it. The *image energy* is derived from this. It can be simply and efficiently modeled as a function of only those image features that are close to shape S. This is a plausible move if nearby features (*e.g.* intensity and gradient at different resolutions) are found to have roughly the same distribution of strength *vs.* distance from S regardless of what S is. However, this assumption, along with what data is being thrown away, should be examined before this simplification is used.

4.4 Mapping Probability to Energy

In the framework of *maximum likelihood estimation* (MLE), the problem of maximizing a probability is often converted to the equivalent problem of minimizing its negative log. So, for instance, we may have a set of independent Gaussian observations y_i, each of known variance σ^2, generated from known x_i by a function $y(x;\mathbf{a})$ with unknown parameters $\mathbf{a}$. Finding $\mathbf{a}$ that maximizes the joint probability of the observations is equivalent to solving a least squares minimization:

$$\begin{aligned}
&\arg\max_{\mathbf{a}} \mathrm{P}(Y_1 = y_1 \wedge Y_2 = y_2 \wedge \ldots) \\
&\quad = \arg\max_{\mathbf{a}} \prod_i \mathrm{P}(Y_i = y_i) \\
&\quad = \arg\max_{\mathbf{a}} \prod_i e^{-\frac{1}{2}(y_i - y(x_i;\mathbf{a}))^2/\sigma^2} \\
&\quad = \arg\min_{\mathbf{a}} \sum_i \tfrac{1}{2}(y_i - y(x_i;\mathbf{a}))^2/\sigma^2
\end{aligned}$$

Thus, minimizing a sum of energies can be equivalent to maximizing a product of probabilities if $E = -\ln \mathrm{P}$. It is useful to have a sum because it simplifies integrals and partial derivatives (which give us the force). This same probability-energy equivalence is well-known from the Gibbs distribution used for Markov random fields. It is useful to us as a way of formulating the energy of a deformable model, or of analyzing the probabilities implied by an energy formulation.

In all formulations we are aware of, several kinds of energy are added, and they need some kind of relative weighting. Since the units of the individual terms are unrelated, it is not at all clear what these weights should be! They must be empirically adjusted. But with a probability-energy equivalence, there is no question of weighting. Once the separate probabilities are known, adjusting them with weights makes no sense. In other words, the "calibration" of relative energy strengths takes place during the estimation of probability distributions (perhaps from training data, perhaps from models of the domain and sensor). This framework makes it easy to intelligently integrate different kinds of information, for instance images from multiple sensors.

Also, the energy minimum achieved is equivalent to a confidence level in the final answer. Thus, the absolute certainty of the locally minimum-energy shape is known, and a poor segmentation can be rejected automatically.

4.5 Stiffness Problems

Consider a shape S for which $\mathrm{P}(I \mid S)$ is at a local maximum but $\mathrm{P}(S)$ is not. If both are positive and differentiable then their product will not be at a local maximum. Thus, if S is the shape that best explains the data, but not the *a priori* most likely shape, then the segmentation process will not choose this shape. How far off the maximum will be depends on how sloped is $\mathrm{P}(S)$ and how sloped is $\mathrm{P}(I \mid S)$ in the neighborhood of the maximum.

What does this mean intuitively? That even for a high-likelihood, noiseless image, the segmentation will err. The problem is that a smooth energy maximum produces a force that goes to zero. When any other force is added, the total force at the (former) maximum now points elsewhere. There are two ways to fix this:

- Make $\mathrm{P}(I \mid S)$ have a nondifferentiable peak when S approaches a suitably strong image edge. This way, imposing a bias from $\mathrm{P}(S)$ does not change the location of the maximum; the image force does not approach zero at the edge.
- Make sure $\mathrm{P}(S)$ is nearly flat for any reasonably likely shape S. This makes the internal force close to zero for such S's.

Do we *want* $\mathrm{P}(S)$ to bias the maximum of $\mathrm{P}(I \mid S)$? We probably do not want it to bias S's position toward some *a priori* position—we want to trust the image edges. But we may want to bias the shape away from bumpiness. In this case, $\mathrm{P}(S)$ should be flat over a reasonable range of *positions* but not necessarily flat with respect to "excess" *curvature*. Obviously, if a probability is to have almost no slope over some range of a parameter, it cannot be Gaussian in that parameter. So the shape's position in the training images should not be fitted by a Gaussian model.

Some formulations define the image energy as linear in the distance to the nearest edge. Such a linear potential satisfies the first of the two conditions above—it is pointy; its gradient, the force, does not approach zero near an edge. Inverse distance also has this desirable property. But squared distance lacks it.

We may get unsatisfactory segmentations if the prior shape model, $\mathrm{P}(S)$, is not good. For example, if (as is usual practice) the function penalizes any kind or amount of curvature, the model will be biased away from the object boundary in even the most exemplary image, unless that object happens to be a circle, sphere, line or plane. In some formulations, a smaller object has less of a curvature penalty, so in the absence of strong image forces, the model

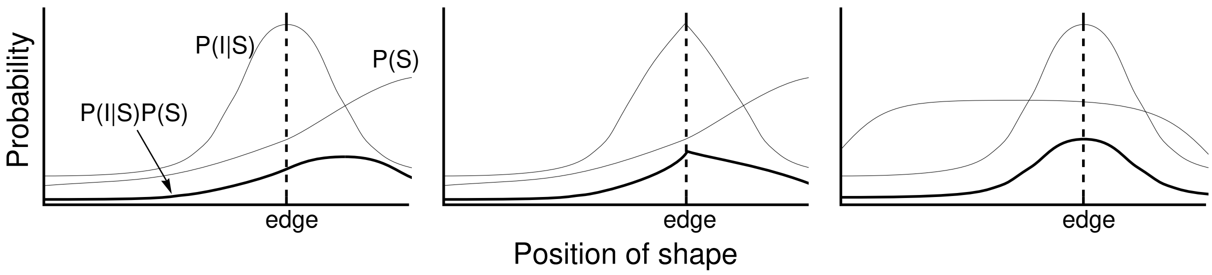

Figure 1: *When a shape's prior probability is combined with the probability that its boundary could have produced the image edges, the shape that maximizes the resulting probability may not lie on the edges. The cure is either an image probability that does not level off at its maximum, or a prior probability that is flat around the shape in question.*

shrinks to a point! This is counteracted in an ad-hoc fashion by introducing *balloon forces* which press in the opposite direction [Cohen-1991].

If there is an expected shape, the correct thing to penalize is deviation from that shape.

Terzopoulos and Metaxas [1991] used a 3D model that penalized local deviations from a superquadric. This model can only be successful if the expected shape is a superquadric. Otherwise, the outcome (the maximum-likelihood boundary will be positioned somewhere between the image edges and the superquadric) is undesirable.

4.6 Finding a Gradient to Follow

If we were to accurately model the distribution of edge distance from underlying shape in a non-noisy image, we would usually find that the edges were all within a few pixels of the shape, unless they were caused by some other object. Thus, the probability of an edge, given a shape, would reach zero (or some constant nonzero noise level) a few pixels away from the shape. It is worth noting that a Gaussian would be a very inappropriate model. The optimal shape would need to lie very close to the image edges.

But we encounter a problem when we place an initial approximate shape in the image. Under such a probability distribution, the image force (probability or energy gradient) that is to deform the shape will be zero unless the initial approximation is correct to within a few pixels. Thus, though we would like the shape that maximizes this probability, the image force will not get us there. The initial shape may get trapped in some small local maximum, or may not move at all. A reasonable solution is to blur the image energy, at least initially, so the gradient from the strong energy minimum reaches the model. Later on, though, we will want to be using the correct energy function.

Thus, even if our probability model is correct, our minimization process may need to use knowledge about the properties of the domain, rather than simply following a gradient.

4.7 Robust Statistics

But in many applications, image edges are not just caused by the object we seek or by sensor noise. There are usually other objects in the scene. Thus, an image energy which reacts to edges as if they were made by the object we seek becomes more or less accurate depending on the distribution of unrelated edges. To model image probability correctly, image energy would have to discriminate between "bad" edges and "good" edges. While we could make some headway by incorporating expected color, gradient and other properties, we would still have little hope of reaching the desired energy minimum by simply following the energy gradient.

So instead of attempting to model extrinsic edges, we may treat them as largely unknown, and try to figure out how to limit the information that goes into image energy so that such "adversarial" edges can do minimal damage. This would make the image energy a *robust statistic.*

A robust statistic is an approximation of some other statistic which is too sensitive to "adversarial" (unmodeled) data. For instance, an average can be perturbed arbitrarily much by a single outlying datum; yet, in the absence of outliers, the average, or "central tendency," may be what we are interested in. The median is a measure of central tendency that asymptotically approximates the average, but it is more robust—a fixed number of outliers can only change it by a fixed amount, no matter how bad those outliers are. We are interested in two qualities of a robust statistic:

- *Efficiency* measures how well the robust statistic approximates the desired statistic.
- The *breakdown point* measures what percentage of bad data it takes to degrade the result a certain amount. (The average has a breakdown point of zero.)

Robust techniques are generally used to deal with uncertainty in the underlying probability distribution or to handle "outliers" (data that does not fit the uncertainty model.) Recasting deformable object modeling as a robust M-estimation problem will help us to analyze the existing assumptions and techniques, and to propose new ones. Robust estimations are becoming used in vision, with M-estimates and "median" based techniques the most common, *e.g.* see [Meer *et al.*-1991, Mirza and Boyer-1993].

4.7.1 M-Estimators and Deformable Models

Let us consider the definition of an M-estimate as defined in P. Huber's classic text ([Huber-1981, page 43]) on robust statistics:

> *Any estimate T_n, defined by an minimum problem of the form:*
>
> $$\sum \rho(x_i; T_n) = \min! \qquad (1)$$
>
> *or by an implicit equation*
>
> $$\sum \psi(x_i; T_n) = \mathbf{0}, \qquad (2)$$
>
> *where ρ is an arbitrary function, $\psi(x;\theta) = (\partial/\partial\theta)\rho(x;\theta)$, is called an M-estimate or maximum likelihood type estimate. [Note that for the choice $\rho(x;\theta) = -\log f(x;\theta)$ gives the ordinary ML estimate].*

In the above definition, T_n is the set of n dependent parameters of a model, each x_i is an independent parameter (data) and "min!" is a compact way of expressing a minimum over all dependent parameters, T_n.[3] The interpretation here is that ρ is a collection of uncertainties and one seeks to find the set of parameters with minimal cumulative uncertainty. The uncertainty captured in ρ could be data uncertainty, modeling uncertainty, or even uncertainty about uncertainty models used in the computation—the form is quite general. If one replaces $\rho(x;\theta)$ with $-\log f(x;\theta)$ the formulation defines maximum likelihood estimation, hence the name maximum likelihood-type estimator or M-estimator for this more general formulation. Note that robust statistics books are quick to point out that the definition in terms of ψ may only find local minima. A classic example of an M-estimate is when ρ is the sum of squared distances between data points and associated model points. This is least squares modeling, which is the maximum-likelihood estimator for i.i.d. Gaussian noise. Robust M-estimators include weighted least-squares and trimmed least squares, which are sometimes called S-estimators).

We are not the first to note the relationship between deformable models (splines), energy formulations and uncertainties. The relation is over 25 years old—see [Kimeldorf and Wahba-1970]. The Gibbs energy formulation defines such a relationship and has been used in many areas including vision, especially in MAP/MRF studies; see [Szeliski-1989, Szeliski and Terzopoulos-1989, Geiger and Girosi-1991, Vemuri and Radisavljevic-1993]. All past work (as known to the PIs), however, relates energy directly to probabilities, and not to the robust statistical M-estimate formulation—the Gibbs energy formulations relates energy to probability by always assuming a maximum likelihood estimation (MLE), rather than robust M-estimation. The difference may be subtle, but we believe it is important. In particular, some of the ad-hoc but basically successful techniques introduced in deformable object models cannot be justified as model/data uncertainty, but they can be justified in terms of robust estimation.

The issues of robustness are related to accuracy of prior knowledge, and to outliers and their impact. For the current uses of deformable modeling, we do not think "uncertainty" models are sufficiently well understood. We believe that viewing them as a robust estimation will help. Outliers are data that cannot be accounted for with the uncertainty model—they lie outside the model. They are a real problem in nonrigid object modeling and can be caused by many things, such as edge features or texture inside the organ, other nearby organs, imager ghosting, etc. Thus the difference between formulation as robust

[3] The definition denotes the data uncertainty distribution with the symbol f, but in this paper we will denote this probability distribution as P.

M-estimation and pure Gibbs-based Bayesian estimation can be profound.

By far one of the most difficult issues in statistical estimation is "noise" that is not really noise but rather a signal from unmodeled data. For example, in ventricle modeling, edges from the papillary muscles or other heart chambers can be locally difficulty to distinguish from the inner heart wall, and have been shown to cause unwanted attractions and modeling errors. Object shadows are another example of difficult "noise." Such under-modeled problems wreak havoc with most regression techniques since the "noise" may be, from a statistical point of view, indistinguishable from a true signal. Only domain-specific knowledge, *e.g.* a partial model of the unmodeled data, can distinguish the two. By using *a priori* knowledge and training data, we believe that we can derive appropriate reweighting and trimming schemes to provide locally-tuned robust estimators.

4.7.2 Reinterpreting Data Forces

Let us now reinterpret common image forces by viewing them as robust M-estimates. The decompositions we give are not unique, but give the flavor of what can be learned by the reinterpretation.

A common long-range force, considered in many papers including [Terzopoulos and Metaxas-1991, O'Donnell *et al.*-1994a], is the scaled distance between nodes and data-points, *assuming a given correspondence*. This force is commonly used for "pre-segmented" data, where the correspondences are computed (usually the nearest point) rather than being given by a user.

To convert from a force to an energy, we integrate. So if the force function for a singe point is $\beta_i d$, it yields a "quadratic energy" of the form $\sum \beta_i d^2 + \gamma$ for some constant of integration γ. Scaling, boundary conditions and choice of measurement units would determine gamma, but for simplicity let us assume $\gamma = 0$. This then defines the ρ function for the M-estimator. Thus this "force" model yields an MLE when the distribution between the data and the original node placements is given by $\log \mathrm{P} = \rho(u;T) = \sum \beta_i d^2$ which, after proper scaling, results in the conclusion that P is Gaussian! (If $\gamma \neq 0$, we have a Gaussian-like distribution with fatter or thiner tails depending on the sign of γ.) By taking the M-estimator point of view we see that a "linear" long distance force is a component of the MLE that assumes the data displacements are Gaussianly distributed (once again

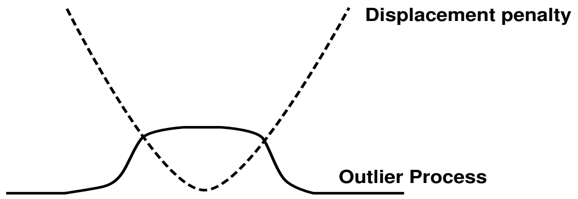

Figure 2: *The penalty (energy) associated with edge displacement grows with distance from the shape boundary. However, an outlier process suggests penalties which go to zero far from the model because at great distances it is likely that the edge is not associated with the model and is hence an outlier on the "edge-position" certainty function.*

implying that we want to keep these displacements small.) A first consequence of this analysis is that, if we believe the correspondences, the proper magnitude of the force or spring constant is given by statistical techniques. Furthermore if the localization of the points is not uniform, say because image is not sampled in square pixels, then one should not just use the distance, but rather each point should have a weighted based on the distance in each direction. This result suggests forces that are stronger in the directions where the "feature" are more accurately localized—the proper scaling of these directionally-scaled forces is directly given by the data covariance. For example, force should not penalize significantly for distance to edge-like features along the edge direction.

The above derivation assumed that the point correspondences were given. The interpretation begins to get more complex as this assumption is relaxed, and this is often where the robust M-estimate, as opposed to ML estimates, comes in to play. If one looks at the "energy" for simple image forces, the most common is the sum (integral) of the gradient of a image evaluated on boundary of the shape. Recall that this measure is the probability of the data given the shape. Note that almost all deformable papers use the "edge" location, not the real data. The current techniques do not use edge orientation and use absolute strength (rather than expected edge strength). The likelihood of the edge given the surface, however, depends in a complex way on the edge shape (because of the blurring). The form of these forces, while moderately successful, has been completely ad hoc, being driven by what is convenient to compute. If viewed as a MLE (*i.e.* using Gibbs energy), it has some odd implications: That only the boundary of the model can cause edges—there are no edges inside of an object and no other (non-modeled) objects; that

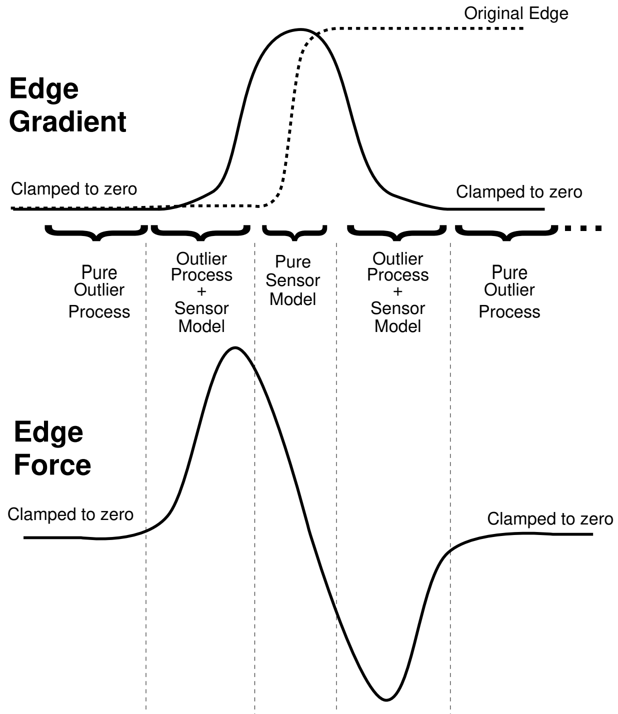

Figure 3: *Current "image" edge forces use the gradient of a smoothed version of the image. These can be reinterpreted as "short-range" edge forces with robust bi-weight functions.*

if two edges are nearby, it is more likely the model boundary generated the stronger edge; that the orientation of the edge does not affect its likelihood; and, because of blurring, that an image with a square outline is not a local maximum for a square shape. Note, however, that the forces in general resemble more traditional robust measures such as the Tukey bi-weight [Huber-1981], with a potential function growing locally but then decaying after the distance between data and model points becomes larger than some parameter; see figure 3. Their justification is not based on viewing the energy as probability, but a mixture of local probability measures and use of the current estimate to "robustly" eliminate outliers. The odd implications above occur because current models only consider the "edge" position, and because they mix data uncertainty (*e.g.* edge position) with uncertainty about the edge even being associated with the model (*i.e.* likelihood that the the edge an outlier). By properly recognizing the two components of the "force" as local estimate of certainty, and by robust reweighting, we can permit greater flexibility in determining the parameters and provide better interpretations to the results.

The first things this suggests are using expected values for edge strength, and incorporating edge orientation and sign (even local profile) information. In terms of robustness, we should recognize that in regions with more clutter the local estimator should be required to have a higher statistical breakdown-point even if it must sacrifice some statistical efficiency [Rousseeuw and Leroy-1987].

It also suggests a simple idea, currently being explored, wherein one develops partial models of the regions surrounding the object of interest so that extraneous edges can be anticipated, even if they are not "modeled." These "partially modeled" edges will be used in conditioning matches and reweighting, but will not be "fit" by the modeling process. For example, if an edge is likely to occur near the edge of interest, *e.g.* the edge of a shadow is expected, we could either use its location to help determine the object edge (if we knew enough and were willing to compute it), or (more cheaply) we could simply insure that the "reweighting" function had sufficiently small support to ignore it.

Note that we believe it is important to separate the sensor model and uncertainty modeling from the robust fitting aspects of the problem. Both are important but they are separate issues. Proper sensor and data uncertainty modeling is critical to allow meaningful statistical measures to be used—it directly affects the quality of the fitting results. The robustness is important to making the fitting process faster and more automated, and hence strongly effects the amount of effort required on the part of the user, but it only mildly effects the fit quality. We believe it is important to first understand the desired answer, and only then worry about approximations to increase the speed. For robust M-estimates of complex parameterizations, formal proofs of "optimality," or even proofs of convergence, are, unfortunately, quite difficult. Thus, while reformulation will provide a better foundation and better implementations, it will still be based on algorithms which require experimental verification.

4.7.3 New Data "Force" Models.

A particular class of new force measures suggested by the statistical point of view is what we call conditional forces. Recall that the uncertainty measure we are estimating is P$(I \mid S)$, the probability of the data given the shape. It follows that the "force" between a data point and a node or element should be conditioned on the likelihood that these actually correspond. Simple conditioning of the match, *e.g.*

requiring a particular sign of contrast, is just the beginning. We could/should use the shape and sensor models to generate "synthetic" data in a local region and compared this to the measured data. This would, however, require a very large amount of computation, so simpler approximations, such as using local contrast parameters, edge orientation, and curvature and intensity profile slope or curvature (edge sharpness), will after calibration, be used to define data forces.

We also point out that the uncertainty computation in which we are interested is global, and that simply summing local "forces" is tantamount to assuming the associated sources of uncertainty are independent (when conditioned on the current shape approximation). This assumption will be relaxed to include covariance computations of "neighboring" points; the cost/benefit of larger size neighborhoods will have to be studied. With advances in computing, larger and more accurate models will be feasible.

An area where it is clear that there is dependency in the edge information, not currently captured in the shape models, are areas of texture. In the area of medical imaging, in particular for internal organs, the image data is often highly textured. We will be investigating the use of this texture to condition matches, and hence defining new "forces" and simultaneously extending our models to include expected texture statistics. It is unlikely that we could predict the actual texture pattern, but many texture measures have been studied for image segmentation, and the parameters can be used to define an expected texture measure.

An issue that will be of particular importance in these investigations is that of calibration. To make meaningful interpretations, we will need "models" of sensor error and prior model uncertainty. In the medical applications, careful experimentation will be necessary to derive such models. This may be the most important aspect of this reformulation—it provides an interpretation where the calibration/learning of model parameters is meaningful. This training will determine the strength of our foundation; if the training data is representative, we will have justifiable methods!

4.7.4 Reinterpreting Parameter Scheduling

As mentioned in section 3.5 various systems for deformable object models, including include code for "parameter scheduling." This code is used to insure that in the initial stages of the recovery, only global parameters are adjusted. When the global parameters have settled, adjustment of the local parameters begins. We have seen three reasons to change the energy formulation during the course of shape optimization:

- We may want to give it an artificial gradient that does not really reflect likelihood, so as to create a force which guides the shape toward maximum likelihood.
- The final shape we want to recover may have many parameters. It may be very slow (and unnecessary) to adjust all of these parameters at every iteration of the deformation. So we only adjust coarse shape parameters at first.
- The probability model which is most accurate when the shape is almost correct may not be the one which is most accurate earlier on.

From a robust statistical point of view this scheduling is tantamount to reducing the dimension of the approximate distribution until the "error" is sufficiently small to warrant a better approximation. The lower dimension also provides more efficient computation. The fixed terms/directions have effective weights of zero, and are thus ignored. We note that this is a simple form of iterative reweighting, and more effective schemes have been studied in the robust statistics literature [Rousseeuw and Leroy-1987, Huber-1981, Mirza and Boyer-1993]. In particular, the reweighting should also depend on the inherent variability of the parameters. In addition, issues such as selection of the appropriate "order" (dimension) of the model have been addressed, see [Akaike-1973, Bozdogan-1987]. The iterated reweighting schemes should provide better quality and, hopefully, better speed.

4.8 The Cootes-Taylor and Baldwin Models

Cootes and Taylor have a simple Gaussian deformable model [1993] which was put into a conceptually clearer probability framework by Baldwin [1994] in as-yet unpublished work. They label a fixed set of N feature points, defining a discrete "contour," in each 2D training image. Each 2D contour can be thought of as a single point p in $2N$-space, the parameter space of contours. They find the mean contour $\bar{p}$ and the covariance matrix C for the $2N$ variables. $\bar{p}$ and C represent the parameters of a Gaussian distribution of training contours in $2N$-space. C models the correlations between contour point coordinates.

They then do a principal component decomposition on the distribution of contours, *i.e.* they find the eigenvectors of C. Each eigenvector represents a principal axis in the distribution of training contours in $2N$-space. The corresponding eigenvalue is the variance (spread) of the distribution along this axis. Any N-vertex contour is a point p in $2N$-space that has a *Mahalanobis distance* $(p - \bar{p})^{\mathrm{T}} \times C^{-1}(p - \bar{p})$, which can be directly related to the multi-dimensional Gaussian probability.

The authors look at the distributions of pixel values along one dimension near each of the N 2D points in a contour. They calculate covariance between pixels near each contour point, but not between those of different contour points. They then search for the contour that maximizes this grey-level Mahalanobis distance alone, P($I \mid p$), while merely bounding the search by imposing an arbitrary maximum on the positional Mahalanobis distance, P(p). From a Bayesian point view, ignoring P(p) is inappropriate, but from a robust statistical perspective, they have little confidence in the prior model and have replaced it with a very broad measure that is clamped outside some range.

Baldwin, however, uses P(p) as an internal energy which penalizes deviation from the correlations gathered during training. He models image energy, P($I \mid p$), using the covariances of grey values at each contour vertex, at M different image resolutions. Thus it is another Gaussian, this one in $M \times N$-space. He maximizes the product P($p \mid I$) of these two Gaussians—multiplying the probabilities is adding the Mahalanobis distances which are his internal and image energies. The assumption that the distributions are normal may cause the problems mentioned earlier, but otherwise this system is a simple and instructive example of the use of the probability formulation. Since the optimization simply relies on image gradients, it may not be robust to poor initial placement. From a "robust" point of view, the algorithm might benefit by reweighting the various scales during maximization.

4.9 Other Work

Kervrann and Heitz [1994], extend Cootes' Gaussian modal representation of a discrete contour. They track moving objects. They model the displacement from the mean contour as Gaussian, but they also model the difference between neighboring diplacements as a Gaussian Markov process. This is their prior shape energy. Their image energy is the number of pixels outside the contour which have changed by more than a threshold amount since the last frame (or since some reference frame), minus the same count for pixels inside the contour.

Like Cootes, they estimate global model pose before allowing the model to deform. Unlike Cootes, they do this by sampling the parameters randomly, using relaxation *á la* simulated annealing. They alternately do the same for perturbations to the principal deformation components, using their Markov model.

This is a pure probabilistic formulation with no special concession to robust statistics. The hierarchy of rigid transformation followed by deformations in a few principal modes, stochastically sampled, provide robustness and avoid local minima.

The work of Vemuri [1993, 1994] goes a bit farther, using a probabilistic framework for continuous shape models. The models are based on the wavelet decomposition. The papers discuss the Gibbs formulation and provide rudimentary statistical interpretations of their energy formulations, including the idea of a "mean" shape. But they do not consider robust techniques, nor do they consider the statistical interpretation of their "image" forces, nor do they discuss calibration of data forces which were presumed to be Gaussian. They do, however, use a simple mean and variance computation to "train" the model's prior of the global parameters. Their prior was, however, a very poor approximation to their object.

In [Chakraborty *et al.*-1994], a probabilistic framework is used to integrate the results of a region classification into a deformable model. Region classification is done using a Markov random field (MRF). Then contour is deformed based on a combination of two image energies—the line integral of a smoothed image gradient, and an area integral of the region inside the contour. This second term penalizes the contour for containing pixels classified as belonging outside, and vice versa.

Despite equating these energies with the logs of probabilities P($I_g \mid p$) and P($I_s \mid p$), the authors still add them weighted by arbitrary constants. They say, "...K_1 and K_2 are the weighting constants which signifies the relative importance of the two terms in the above equation. Normally, one should choose them such that the contributions from each of the terms are comparable." When one treats the energies this way, they no longer represent probabilities in any quantitatively meaningful way.

The feature of note here is that they gain robustness and tolerate poorer initial contour estimates by integrating two kinds of image information.

Although [Neuenschwander *et al.*-1994] does not use a probability formulation, they have a very clever method for overcoming the problem that image forces are only accurate very close to the desired object. They specify an initial position and orientation for their snake only at its endpoints, letting the rest hang where it may, unaffected by image force. Then they gradually turn on the image force starting at the ends and moving toward the middle. The image force does not attract the snake to the wrong object, since the force is only turned on at each snaxel once the previous one has settled, presumably close to the desired edge. This is an elegant example of a way to modify image forces gradually to make them robust against unrelated edges.

5 Conclusion

While various researchers have related model energy to probability, we believe that for many vision tasks, the use of "robust statistics," in particular M-estimators, is more appropriate that straightforward Bayesian analysis. This statistical framework provides a natural mechanism for "learning" the *a priori* model of objects to be segmented/recovered and provides standard measures of quality and significance of the recovered models.

Existing work using deformable models has shown them to be quite useful. However they suffer from various limitations, one of which is the "justification-by-analogy" of the formulation. By reinterpreting deformable object recovery from the view of robust statistics, we have found justification for various existing "ad hoc" but important aspects of deformable modeling techniques. In addition, this viewpoint suggests numerous enhancements to deformable modeling techniques.

References

[Akaike, 1973] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Proc. Second International Symposium on Information Theory*, pages 267–281, 1973.

[Baldwin, 1994] Bernard Baldwin. personal conversations on thesis work, 1994. Mr. Baldwin is at the Courant Institute of New York University. He did his research in cooperation with the Memorial Sloan-Kettering Cancer Center.

[Bates and Wahba, 1982] D. Bates and G. Wahba. Computational methods for generalized cross-validation with large data sets. In C.T.H. Baker and G.F. Miller, editors, *Treatment of Integral Equations by Numerical Methods*, pages 283–296. Academic Press, New York, 1982.

[Bathe, 1982] K.J. Bathe. *Finite Element Procedures in Engineering Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[Boult and Moerdler, 1987] T.E. Boult and M.L. Moerdler. An experimental system for the integration of information from stereo and multiple shape from texture algorithms. In *Proc. SPIE Conf. on Intelligent Robots and Computer Vision*. SPIE, Nov 1987. SPIE # 848-16.

[Bozdogan, 1987] H. Bozdogan. Model selection and akaike's information criterion (aic): General theory and its analytical extensions. *Pyschometrika*, 52(3):345–370, 1987.

[Chakraborty *et al.*, 1994] A. Chakraborty, L.H. Staib and J.S. Duncan. Deformable boundary finding influenced by region homogeneity. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 624–627, Seattle, WA, June 1994.

[Cohen and Cohen, 1990] L.D. Cohen and I Cohen. A finite element method applied to new active contour models and 3d reconstruction from cross-sections. In *Proc. of the IEEE Int. Conf. on Computer Vision*, pages 587–591, Osaka, Japan, 1990. IEEE.

[Cohen, 1991] L.D. Cohen. On active contour models and balloons. *Computer visoin graphics and image processing: Image Understanding (CVGIP:IU)*, pages 211–218, March 1991.

[Cootes *et al.*, 1993] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. In *Proceedings of the 13th International Conference on Information Processing in Medical Imaging*, Flagstaff AZ, June 1993. Springer-Verlag.

[Geiger and Girosi, 1991] D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrf's: Surface reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(5):674–693, May 1991.

[Huang and Goldgof, 1993] W.C. Huang and D. Goldgof. Nonridged motion analysis using non-linear fineite element modeling. In *Geometric Methods in Computer Vision*, volume 2031, pages 404–415. SPIE, 1993.

[Huber, 1981] P. Huber. *Robust Statistics*. Wiley, New York, 1981.

[Kass *et al.*, 1987] M. Kass, A. Witkin and D. Terzopoulos. Snakes: Active contour models. In *Proc. of the IEEE Int. Conf. on Computer Vision*, pages 259–268, London UK, 1987. IEEE.

[Kervrann and Heitz, 1994] C. Kervrann and F. Heitz. A heirarchical stastical framework for the segmentation of deformable objects in image sequences. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 724–727, Seattle, WA, June 1994.

[Kimeldorf and Wahba, 1970] G.S. Kimeldorf and G. Wahba. A correspondence between bayesian estimation on stocastic processes and smoothing by splines. *Annals of Math. Stat.*, 41(2):495–502, 1970.

[Lai and Chin, 1994] K.F. Lai and R.T. Chin. Deformable contours: Modeling and extraction. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 601–608, Seattle, WA, June 1994.

[Leymarie and Levine, 1993] F. Leymarie and M.D. Levine. Tracking deformable objects in the plane using an active contour model. *IEEE Trans. on Pattern Analysis and Machine Intelligencee*, 15(6):617–633, June 1993.

[Meer *et al.*, 1991] P. Meer, D. Mintz, A. Rosenfeld, and D.Y. Kim. Robust regression methods for computer vision: A review. *Inter. J. Computer Vision*, 6(1):59–70, 1991.

[Mirza and Boyer, 1993] M.J. Mirza and K.L. Boyer. Performance evaluation of a class of m-estimators for surface parameter estimation in noisy range data. *IEEE Trans. on Robotics and Automation*, 9:75–85, 1993.

[Moerdler and Boult, 1988] M.L. Moerdler and T.E. Boult. The integration of information from stereo and multiple shape-from texture algorithms. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 514–529. IEEE, 1988.

[Neuenschwander *et al.*, 1994] W. Neuenschwander, P. Fua, G. Székely, and O. Kübler. Initializing snakes. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 658–663, Seattle, WA, June 1994.

[O'Donnell *et al.*, 1994a] T. O'Donnell, X.S. Fang, T.E. Boult, and A. Gupta. The extruded generalized cylinder: A deformable model for object recovery. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, June 1994.

[O'Donnell *et al.*, 1994b] T. O'Donnell, A. Gupta and T.E. Boult. A periodic generalized cylinder model with local deformations for tracking closed contours exhibiting repeating motion. In *The International Conference on Pattern Recognition*, Nov. 1994. to appear.

[Park *et al.*, 1994] J. Park, D. Metaxas and A. Young. Deformable modles with parameter functions: Application to heart-wall modeling. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 437–442, Seattle, WA, June 1994.

[Pentland and Sclaroff, 1991] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(7):715–729, July 1991.

[Rousseeuw and Leroy, 1987] P.J Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. J. Wiley, New York, New York, 1987.

[Szeliski and Terzopoulos, 1989] R. Szeliski and D. Terzopoulos. Constrained fractals. *Computer Graphics*, 23(3):51–60, 1989. (SIGGRAPH).

[Szeliski, 1989] R. Szeliski. *Baysian Modeling of Uncertanity in Low-level Vision*. Kluwer Academic Publishers, Boston, MA, 1989.

[Terzopoulos and Metaxas, 1990] D. Terzopoulos and D. Metaxas. Dynamic 3d models with local and global deformations: Deformable superquadrics. In *Proc. of the IEEE Int. Conf. on Computer Vision*, pages 606–615, Osaka, Japan, 1990.

[Terzopoulos and Metaxas, 1991] D. Terzopoulos and D. Metaxas. Dynamic 3d models with local and global deformations: Deformable superquadrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(7):703–714, July 1991.

[Terzopoulos and Waters, 1993] D. Terzopoulos and K. Waters. Analysis and synthesis of facial images sequences using physical and anotomical models. *IEEE Trans. on Pattern Analysis and Machine Intelligencee*, 15:569–579, 1993.

[Vemuri and Radisavljevic, 1993] B.C. Vemuri and A. Radisavljevic. From global to local, a continuum of shape models with fractal priors. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 307–313, NYC NY, June 1993.

[Vemuri and Radisavljevic, 1994] B.C. Vemuri and A. Radisavljevic. Multiresolution stochastic hybrid spahe models with fractal priors. *ACM TOGS*, 1994. To appear, special issue on Interactive Sculpting.