On Line Processes, Outlier Rejection, and Robust Statistics

Michael Black
Anand Rangarajan

Research Report YALEU/DCS/RR-993
October 1993

YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

On Line Processes, Outlier Rejection, and Robust Statistics

Michael J. Black* and Anand Rangarajan,†

*Xerox Palo Alto Research Center, Palo Alto, CA
†Department of Computer Science, Yale University, New Haven, CT

October 14, 1993

Abstract

The modeling of spatial discontinuities for problems such as surface recovery, segmentation, image reconstruction, and optical flow has been intensely studied in computer vision. While "line-process" models of discontinuities have received a great deal of attention, there has been recent interest in the use of robust statistical techniques to account for discontinuities. This paper unifies the two approaches. To achieve this we generalize the notion of a "line process" to that of an analog "outlier process" and show how a problem formulated in terms of outlier processes can be viewed in terms of robust statistics. We also characterize a class of robust statistical problems for which an equivalent outlier-process formulation exists and give a straightforward method for converting a robust estimation problem into an outlier-process formulation. We also show how prior assumptions about the spatial structure of outliers can be expressed as constraints on the recovered analog outlier processes and how traditional continuation methods can be extended to the explicit outlier-process formulation. These results indicate that the outlier-processes approach provides a general framework which subsumes the traditional line-process approaches as well as a wide class of robust estimation problems. Examples in surface reconstruction, image segmentation, and optical flow are used to illustrate the use of outlier processes and to show how the relationship between outlier processes and robust statistics can be exploited. An appendix contains a catalog of common robust estimators and their equivalent outlier-process formulations.

Submitted as a "Regular Paper."

1 Introduction

The modeling of spatial discontinuities for problems such as surface recovery, segmentation, image reconstruction, and optical flow has been intensely studied in computer vision. In particular "line-process" models of discontinuities have been popular due, in part, to their intuitive and physical appeal, as well as their ability to model spatial properties of discontinuities. More recently, the use of robust statistics in computer vision has become popular and, at first glance, it is not at all clear that line-process approaches and robust statistics have anything in common. The goal of this paper is to show that they are closely related and that, by bringing this relationship to light, each approach can benefit from the other. Moreover, we propose a new framework based on analog or binary "outlier processes" which subsumes traditional line process approaches and a wide class of robust estimation approaches.

To achieve these goals, we first generalize the notion of a "line process" to that of an "outlier process". We then show how a problem formulated in terms of outlier processes can be viewed in terms of robust statistics. Finally, we will characterize a class of robust statistical problems for which an equivalent outlier-process formulation exists and derive a straightforward mechanism for converting the robust estimation problem to the outlier-process problem. The resulting formulation, with explicit outlier processes, is more general than the original robust estimation problem in that the new framework allows constraints on the spatial organization of outliers to be introduced. Additionally, the framework extends the continuation methods used by the robust approaches to the explicit outlier processes formulation.

Many "reconstruction" problems in computer vision are initially posed as least-squares estimation problems. For example, consider the simple problem of reconstructing a smooth one-dimensional surface $\mathbf{u}$ given a set of noisy measurements $\mathbf{d}$. The problem can be formulated using a *data term*, E_D, that enforces fidelity to the measurements and a *smoothness,* or *regularization term*, E_S, that embodies assumptions about the spatial variation of the data:

$$\begin{aligned} \min_{\mathbf{u}} E(\mathbf{u},\mathbf{d}) &= E_D(\mathbf{u},\mathbf{d}) + E_S(\mathbf{u}), \\ &= \sum_i \left[(u_i - d_i)^2 + \lambda(u_{i-1} - u_i)^2\right], \end{aligned} \tag{1}$$

where λ is a constant that controls the relative importance of the two terms.

It is well known that this least-squares formulation of regularized problems leads to smoothing across spatial discontinuities and it is this problem that motivates the use of line

processes to inhibit smoothing at discontinuities. The standard "line process" formulation allows discontinuities in the regularization term:

$$\min_{\mathbf{u},\mathbf{l}} E(\mathbf{u},\mathbf{d},\mathbf{l}) = \sum_i \left[(u_i - d_i)^2 + \lambda \left((u_{i-1} - u_i)^2 \, l_i + \Psi(l_i) \right) \right], \tag{2}$$

where l_i is typically a binary valued line process that is 1 when no discontinuity is present between i and $i-1$ and 0 when a discontinuity exists.[1] The function $\Psi(l_i)$ can be thought of as the "penalty" for introducing a discontinuity. Line-process approaches will be reviewed in detail in Section 3. In particular, we will be interested in cases involving *analog* line process where $0 \leq l_i \leq 1$.

The line process allows us to use a general model of spatial smoothness and account for violations of the model by introducing spatial discontinuities. These approaches typically assume that the noise in the data measurements is Gaussian and lead to the least-squares formulation of E_D. In many situations this may be a poor assumption and we would like to be able to ignore "bad" measurements. We do this by introducing a *measurement process*, $\mathbf{m}$, that is analogous to the line process, but applied to the data term. Hence we generalize the notion of line process to that of an *outlier process* that is applied to both data and spatial terms:

$$E(\mathbf{u},\mathbf{d},\mathbf{m},\mathbf{l}) = \sum_i \left[(u_i - d_i)^2 \, m_i + \Psi_D(m_i) + \lambda \left((u_{i-1} - u_i)^2 \, l_i + \Psi_S(l_i) \right) \right], \tag{3}$$

where m_i and l_i are outlier process and Ψ_D and Ψ_S are, possibly different, penalty functions. The term "outlier process" refers to the fact that the data and spatial errors that exceed some threshold can be thought of as "outliers", and the purpose of the outlier process is to *reject* these measurements.

The outlier processes can be eliminated from the objective function above by either minimizing over them (as done by Blake and Zisserman [9]) or by integrating them out (as is done in mean-field approaches [17]). The result is an objective function:

$$E(\mathbf{u},\mathbf{d}) \;=\; \sum_i \left[g_D(u_i - d_i) + g_S(u_{i-1} - u_i) \right], \tag{4}$$

[1] It is more common to write the smoothness term as

$$(u_{i-1} - u_i)^2 \, (1 - l_i) + \beta \, l_i$$

where β is a constant and $l_i = 1$ means a discontinuity exists; see, for example, [9]. We find it convenient to take the line process to be $z = (1 - l)$ and generalize the penalty term to be some function, Ψ, of z.

where g_D and g_S are functions of solely the data and spatial error respectively. The exact form of g_D and g_S depend on the form of the outlier processes (and penalty function) as well as the elimination technique; two common examples are the truncated quadratic [9] and the mean field function [17].

A more recent approach for coping with data and spatial outliers is to view the problem in terms of statistical estimation. As mentioned above, the simple formulation of the surface fitting problem corresponds to least-squares estimation. It is common knowledge that least squares lacks robustness when the data contains outliers. To deal with outliers, the least-squares problem can be reformulated using any number of robust statistical techniques. A common approach is to replace the quadratic estimators in the least-squares formulation with one of a number of *robust estimators*:

$$\min_{\mathbf{u}} E(\mathbf{u}, \mathbf{d}) \quad = \quad \sum_i \left[\rho_D(u_i - d_i) + \rho_S(u_{i-1} - u_i)\right], \tag{5}$$

where the estimators ρ_D and ρ_S are more "forgiving" than the quadratic when the data does not conform to our assumptions. Unlike the line-process approaches, which were originally designed to deal with spatial discontinuities, the robust techniques are naturally applied to both the data and spatial terms. We will see that this can significantly improve the "robustness" of the recovered solution; that is, its tolerance of outlying measurements.

Notice that Equation (5) is of exactly the same form as Equation (4). The removal of the outlier process gives a new function, ρ, which is simply a robust estimator performing outlier rejection and the resulting minimization problem is simply robust estimation. So, given an outlier-process problem with some penalty function Ψ we can construct a robust estimation problem with the appropriate estimator. What about the other direction?

As shown by Geman and Reynolds [21], for a certain class of estimators, it is possible to find the equivalent line-process formulation. We give a simple constructive proof of the result that leads to an explicit mechanism for converting robust estimators into outlier processes and penalty functions. In Appendix A, we provide a catalog of common estimators found in the computer vision literature and their outlier processes formulations.

There are situations where it proves useful to convert a robust estimation problem into an equivalent outlier process form. For example, one benefit of the traditional line-process approach is that it allows prior expectations about the spatial configurations of discontinuities to be expressed. Such approaches typically add constraints that express Gestalt notions of continuity while discouraging the formation of physically unlikely configurations

of discontinuities. Formulations that employ robust estimators lack the ability to express this sort of prior expectations. If we wish to enforce spatial continuity, we must recover an outlier process by converting the robust estimation problem into an explicit outlier-process formulation. We can then define "clique energies" [22] to express prior assumptions about the spatial continuity of the outliers. We extend the standard formulation of these spatial interaction terms to account for analog line processes and show how two spatial constraints (hysteresis and non-maxima suppression) can be implemented.

The conversion of robust estimators into outlier processes has another benefit; it allows us to create outlier processes with new, and useful, properties. For example there are a number of estimators used in computer vision that have a control parameter which varies the shape of the estimator. Varying this parameter controls the convexity of the objective function; making it convex allows it to be minimized readily and provides an approximate solution. Continuation methods (like the Graduated Non-Convexity (GNC) algorithm [9]) gradually change the shape of the estimator, making the objective function non-convex, while tracking the solution. When the outlier processes are recovered, the penalty functions of these estimators retain their control parameters resulting in novel outlier processes. As a result, the continuation methods used for solving the robust estimation problems can still be exploited to solve the explicit outlier-process formulations.

In the following section we review previous work on regularization with discontinuities, applications of robust statistics in computer vision, and attempts to bring the approaches together. Sections 3 and 4 provide brief introductions to line processes and robust statistics respectively. Section 5 unifies the line process and robust estimation approaches. In this section, outlier processes are introduced and the conversion from outlier process to robust estimation and back again is described in detail. Some practical benefits of the explicit relationship are described in Section 6 where we show how spatial interactions can be added to the recovered outlier process and how continuation methods can be extended to these formulations with explicit spatial organization constraints. Section 7 provides two sets of examples. The first illustrates the benefits of "robustifying" the data term with examples chosen from surface reconstruction and optical flow estimation. The second illustrates how a robust estimator with a control parameter is converted to an outlier process and how spatial interactions can be added to improve image reconstruction results.

2 Previous Work

Many problems in early vision are *ill-posed* [24] in that they are underconstrained and sensitive to noise, thus making it difficult, or impossible, to find unique solutions. It is common to *regularize* these problems by introducing additional constraints that encode prior assumptions, thereby reducing the class of admissible solutions [3, 36]. In particular, it is common in recovery problems to introduce a spatial coherence assumption which expresses prior knowledge about surfaces in the scene. The common assumption of spatial smoothness is frequently violated in problems involving image segmentation, surface recovery, optical flow, stereo, and image restoration. Performing regularization in cases where the data is spatially discontinuous has received a great deal of attention [9, 22, 36, 40, 52].

Geman and Geman [22] introduced the notion of a binary "line process" for modeling spatial discontinuities in image brightness. Working in the framework of Markov random fields (MRF's), they formulated constraints on the local spatial organization of discontinuities. These constraints, defined as "clique energies", were designed to prefer physically plausible configurations of discontinuities; for example, enforcing continuation. This simple and powerful idea has been applied to many problems that require the recovery of piecewise smooth regions; these include image restoration [14, 21, 22] texture segmentation [15, 20], surface recovery from depth data [36], and dense optical flow estimation [6, 7, 32, 41, 53]. Unfortunately, the introduction of line processes results in a non-convex optimization problem which Geman and Geman solved using an expensive stochastic minimization procedure.

If no spatial constraints are imposed on discontinuities, Blake and Zisserman [9] showed that the binary line processes can be eliminated from the optimization problem by first minimizing over them. The result is an objective function containing an energy functional which enforces spatial smoothness as long as neighboring points are "similar enough". Beyond some threshold however, spatial smoothness is no longer enforced. This is the notion of a "weak constraint" as introduced by Hinton [28].

The weak constraint approach still requires minimizing a non-convex objective function, but Blake and Zisserman showed how their energy functional could be generalized by the introduction of a "control parameter" which can be used to adjust the "shape" of the function. Using this parameter, they devised a *continuation method* called Graduated Non-Convexity in which the control parameter is adjusted to construct a convex approximation to the original objective function; this approximation is readily minimized. They construct a sequence

of increasingly good approximations to the original objective function and minimize each beginning with the previous solution.

Other continuation methods have been described by Leclerc [34] and by Geiger and Girosi [17] who derive their energy functional using a *mean-field approximation* to the stochastic MRF model. These continuation methods are in the same vein as various "scale-space" approaches [43]. In other related work, it has been shown that these weak-continuity approaches can be implemented in hardware using resistive networks [26].

Separately, the field of robust statistics [25, 30] has developed methods to address the fact that the parametric models of classical statistics are often approximations of the phenomena being modeled. In particular, the field addresses how to handle *outliers*, or gross errors, that do not conform to the statistical assumptions. While most of the work in computer vision has focused on developing optimal strategies for exact parametric models, there is a growing realization that we must be able to cope with situations for which our models were not designed.

Many robust statistical techniques have been applied to standard problems in computer vision [39, 48]. There are robust approaches for performing local image smoothing [4, 10], classification [13], surface reconstruction [50], segmentation [38], pose estimation [33], edge detection [35], structure from motion or stereo [54, 56], and optical flow estimation [7, 8, 46, 47]. Only recently have robust techniques been applied to problems of regularization with discontinuities.

Shulman and Hervé [49] point out that spatial discontinuities can be viewed as outliers and they formulate the regularization of optical flow using Huber's theory of robust estimation [30]. In doing so, they derive a convex optimization problem. Also, Black [5] has used *redescending estimators* [25] to perform regularization. These estimators result in a non-convex optimization problem but have better outlier rejection properties and are related to line-process approaches. Schunck [48] pointed out that regularization is a least-squares method and there has been recent work which reformulates the entire problem (not simply the regularization term) using robust statistics [7, 8].

Recently, a number of authors have taken steps towards unifying line-process approaches, robust estimation, and continuation methods. Geiger and Yuille [19] formulate the image segmentation problem with a binary process on the data term as suggested in the Introduction and use mean-field techniques [17] to integrate out both the data and line processes.

They point out that the resulting objective function has the same form as robust estimation techniques but they do not formalize the relationship between the approaches.

Geman and Reynolds [21] provide another piece of the puzzle. They introduce analog line processes with penalty functions and show how minimizing over the analog processes produces particular estimators; this generalizes the approach of Blake and Zisserman [9]. They also specify conditions on an estimator that must be satisfied if it is to have an equivalent line-process formulation. We provide a constructive proof of these conditions and make explicit the mechanism for recovering the analog process. Geman and Reynolds did not apply a measurement process and did not make explicit the connections to robust estimation or continuation methods.

Black and Anandan [7, 8] use robust estimation to account for both spatial discontinuities and measurement errors in the recovery of dense optical flow fields. They make explicit the relationship between outlier rejection and regularization with discontinuities. They also point out the relationships between robust estimation and continuation methods.

Finally, Rangarajan and Chellappa [44] show how an analog line process can be recovered for a general class of estimators. They derive an analog process for the GNC estimator and show how the penalty function retains the continuation parameter of the estimator. They then add spatial interactions among the analog line processes to perform hysteresis and non-maximum suppression. The control parameter can be used to extend continuation methods to the case where there is an explicit line process with spatial interactions. As with the Geman and Reynolds result, Rangarajan and Chellappa do not address the robustness of the data term or connect the approach to robust estimation. Black [5] introduces analog "outlier processes" and exploits the results of Rangarajan and Chellappa [44] to "close the loop," showing how to convert between robust estimation problems and outlier process formulations.

3 Line Processes

To introduce the idea of binary line processes we will consider a simple example of reconstructing a smooth surface $\mathbf{u}$ from noisy depth data $\mathbf{d}$. Assume that the data is an $n \times n$ image, and each pixel (or site), s, has a set of neighbors $t \in \mathcal{G}_s$. For a first-order neighborhood system, $\mathcal{G}_s$, these are just the sites to the North, South, East, and West of site s. The set of all sites in the image is

$$S = \{s_1, s_2, \ldots, s_{n^2} \mid \forall w \; 0 \leq i(s_w), j(s_w) \leq n-1\},$$

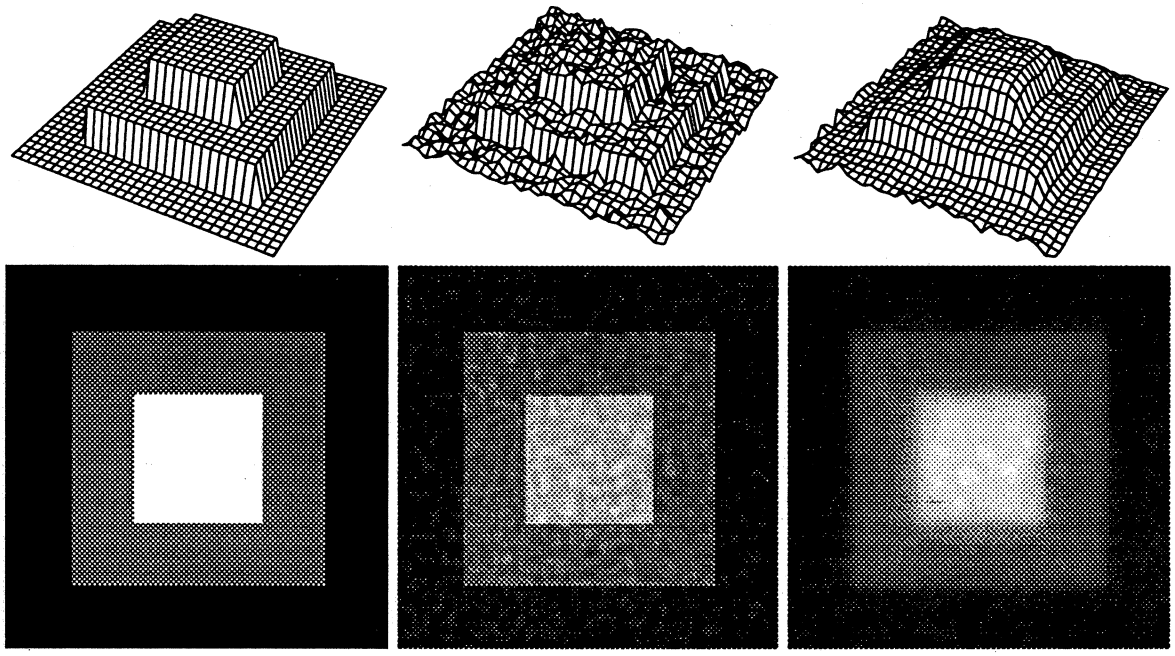


Figure 1: Surface reconstruction example (see text); *(a)* Original surfaces. *(b)* Corrupted surfaces. *(c)* Least-squares recovery. The top row shows a plot of the surfaces, while the bottom row shows the depth encoded as an intensity image (with near surfaces being brighter).

where $(i(s), j(s))$ denotes the pixel coordinates of site s.

To recover a smooth surface $\mathbf{u}$ given noisy depth data $\mathbf{d}$, we minimize the following objective function:

$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S} [(u_s - d_s)^2 + \lambda \sum_{t \in \mathcal{G}_s} (u_s - u_t)^2], \tag{6}$$

formulated as a least-squares estimation problem. The first term ensures that the recovered surface is faithful to the data, while the second term encodes our prior assumption about surface smoothness. The smoothness term tries to make the recovered surface as smooth as possible while accounting for the data. Natural scenes, however, are typically only smooth locally and contain discontinuities in depth at surface boundaries.

For illustration consider the 128×128 pixel "wedding cake" in Figure 1*a* in which the "depth" of the surfaces are 0 for the far surface, 128 for the middle layer, and 255 for the top layer. The surface is first corrupted by uniform random noise over the range $[-25.6, 25.6]$. Twenty percent of the pixels were then contaminated by uniform random noise in the range of either $[80.0, 102.4]$ or $[-102.4, -80.0]$; these points can be thought of as outliers. The corrupted wedding cake is shown in Figure 1*b*.

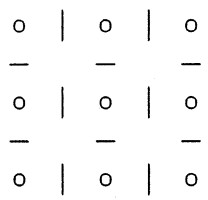


Figure 2: Arrangement of pixel sites ($\circ$) and discontinuities ($|, -$).

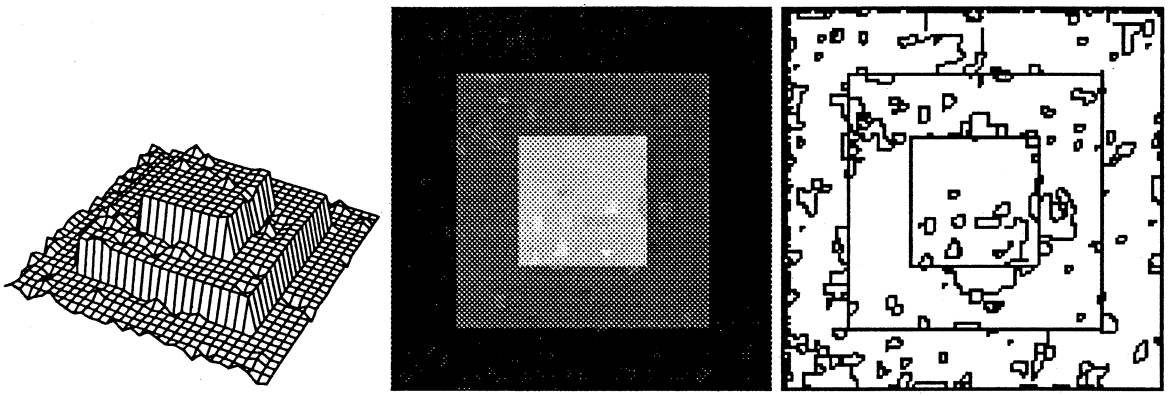
The least-squares result is shown in Figure 1*c* where λ was taken to be 6.4. This λ approximately balances the data and spatial terms if no outliers are present and results in smoothing across the depth discontinuities. The data outliers pull the solution away from the true solution resulting in a smooth but "lumpy" surface. With the least-squares approach we can increase the regularization parameter λ to reduce the effect of the data outliers but this will result in further smoothing which will obscure the depth boundaries.

Adding a spatial line process allows us to recover piecewise smooth surfaces. To do so, we define a dual $n \times n$ lattice, $S^L = (s,t)$, of all nearest neighbor pairs (s,t) in S. Figure 2 shows the pixel sites ($\circ$) in the original graph and the discontinuity processes ($|, -$) between the sites. This lattice is coupled to the original in such a way that the best interpretation of the data will be one in which the data is piecewise smooth. We have the following objective function:

$$E(\mathbf{u}, \mathbf{d}, \mathbf{l}) = \sum_{s \in S} \left((u_s - d_s)^2 + \lambda \sum_{t \in \mathcal{G}_s} [(u_s - u_t)^2 \, l_{s,t} + \beta(1 - l_{s,t})] \right), \tag{7}$$

where β is a constant penalty term and where the $l_{s,t}$ are boolean-valued line processes which indicate the presence ($l_{s,t} = 0$) or absence ($l_{s,t} = 1$) of a discontinuity between neighboring sites s and t. When no discontinuity is present, the smoothness term has the original least-squares form, but when a discontinuity is introduced, the smoothness term becomes a constant, β, regardless of the spatial error.

Minimizing this new objective function with respect to $\mathbf{u}$ and $\mathbf{l}$ gives a piecewise smooth surface with breaks where the spatial gradient is too large. Figure 3 shows that the approach preserves the discontinuities between layers of the wedding cake, but the results also illustrate how the simple line process allows the introduction of spurious spatial discontinuities. The quadratic data term draws the solution towards the data outliers and the spatial term allows discontinuities to be introduced to best fit the outlying measurements.



a *b* *c*

Figure 3: Surface Reconstruction with a spatial line-process. *(a)* Surface plot. *(b)* Height plotted as intensity. *(c)* Spatial discontinuities (outliers); black means that there is a discontinuity.

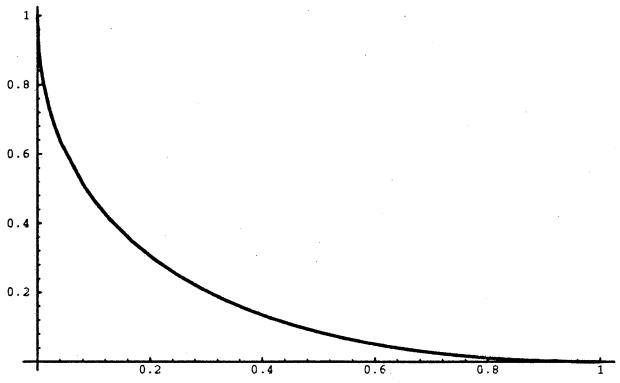


Figure 4: An example penalty function, $\Psi(l_{s,t})$, for an analog line process.

3.1 Analog Line Processes

An *analog line process* [21] is a straightforward generalization of the standard binary line process in which we take $0 \leq l_{s,t} \leq C$, for some positive constant C; that is, the line process can take on values within some range (usually $[0, 1]$). In the case of the binary line process the penalty term is typically taken to be a constant times $(1 - l_{s,t})$. For arbitrary analog line processes we generalize the penalty term to be some function $\Psi(l_{s,t})$:

$$E(\mathbf{u}, \mathbf{d}, \mathbf{l}) = \sum_{s \in S} \left((u_s - d_s)^2 + \lambda \sum_{t \in \mathcal{G}_s} [(u_s - u_t)^2 \, l_{s,t} + \Psi(l_{s,t})] \right). \tag{8}$$

An example penalty function is shown in Figure 4, where

$$\Psi(z) = (\sqrt{z} - 1)^2.$$

When a discontinuity is present ($l_{s,t} \rightarrow 0$), Ψ returns a high error; as the discontinuity lessens ($l_{s,t} \rightarrow 1$), the penalty decreases.

4 Robust Statistics

This section provides a brief introduction to robust estimation. For mathematical details the reader is referred to [25, 30, 45] and for applications in computer vision see [1, 39]. What we focus on here are the relationships between robust estimators and line processes and the heuristic application of robust estimators for outlier rejection.

As identified by Hampel *et al.* [25, page 11] the main goals of robust statistics are:

(i) To describe the structure best fitting the bulk of the data,

(ii) To identify deviating data points (outliers) or deviating substructures for further treatment, if desired.

To state the issue more concretely, robust statistics addresses the problem of finding the values for the parameters, $\mathbf{a} = [a_0, \ldots, a_n]$, that provide the best fit of a model, $\mathbf{u}(s; \mathbf{a})$, to a set of data measurements, $\mathbf{d} = \{d_0, d_1, \ldots, d_S\}$, in cases where the data differs statistically from the model assumptions.

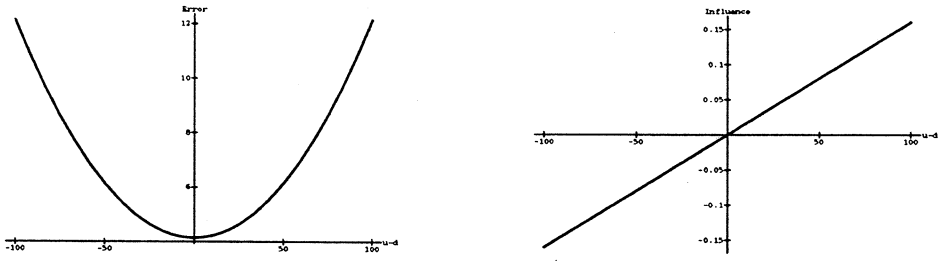
In fitting a model, the goal is to find the values for the parameters, $\mathbf{a}$, that minimize the size of the *residual* errors ($d_s - \mathbf{u}(s; \mathbf{a})$):

$$\min_{\mathbf{a}} \sum_{s \in S} \rho(d_s - \mathbf{u}(s; \mathbf{a}), \sigma_s), \tag{9}$$

where σ_s is a scale parameter, which may or may not be present, and ρ is our *estimator*. When the errors in the measurements are normally distributed, the optimal estimator is the quadratic:

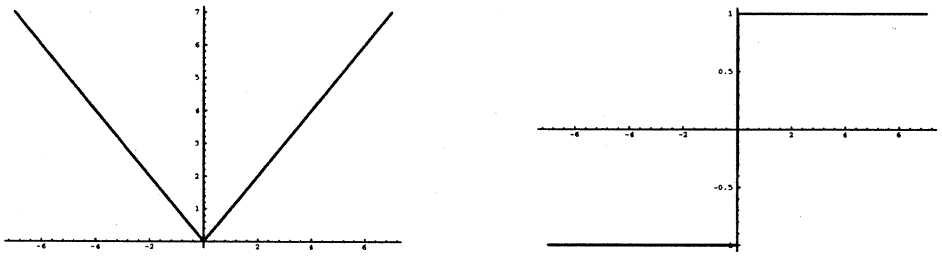
$$\rho(d_s - \mathbf{u}(s; \mathbf{a}), \sigma_s) = \frac{(d_s - \mathbf{u}(s; \mathbf{a}))^2}{2\sigma_s^2}, \tag{10}$$

which gives rise to the standard *least-squares* estimation problem. The function ρ is called an *M-estimator* since it corresponds to the *Maximum-likelihood* estimate. The *robustness* of a particular estimator refers to its insensitivity to outliers, or deviations, from the assumed statistical model.



a *b*

Figure 5: Quadratic estimator (a) and ψ-function (b).



a *b*

Figure 6: L1 norm. (a) Estimator, (b) ψ-function.

4.1 Robust Estimators

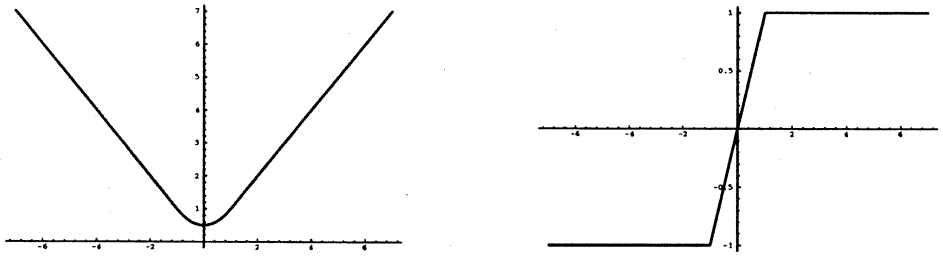
The least-squares approach is notoriously sensitive to outliers; the problem being that outliers contribute "too much" to the overall solution. Outlying points are assigned a high weight by the quadratic estimator (see Figure 5a). To analyze the behavior of an estimator, we take the approach of Hampel *et al.* [25] based on *influence functions*. The influence function characterizes the bias that a particular measurement has on the solution and is proportional to the derivative, ψ, of the estimator [25]. Consider, for example, the quadratic estimator:

$$\rho(x) = x^2, \qquad \psi(x) = 2x. \tag{11}$$

For least-squares estimation, the influence of outliers increases linearly and without bound (Figure 5b).

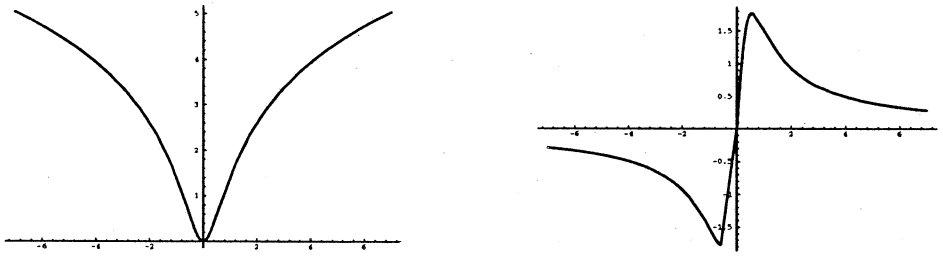
To increase robustness, an estimator must be more forgiving about outlying measurements. The most obvious first step is to replace the quadratic (or L2 norm) with the absolute value (or L1 norm):

$$\rho(x) = |x|, \qquad \psi(x) = \text{sign}(x). \tag{12}$$



a *b*

Figure 7: Huber's min-max estimator. (*a*) Estimator, (*b*) ψ-function.



a *b*

Figure 8: Lorentzian Estimator. (*a*) Estimator, (*b*) ψ-function.

In Figure 6, it is clear that outlying points are weighted less heavily by the L1 norm but the estimator is not robust in that a single large outlier can make the solution arbitrarily bad. Additionally, the L1 norm does not perform as well as the quadratic estimator when the errors are Gaussian. For this reason Huber [30] proposed the *minimax* estimator (Figure 7):

$$\rho_\epsilon(x) = \begin{cases} x^2/2\epsilon + \epsilon/2 & |x| \leq \epsilon, \\ |x| & |x| > \epsilon, \end{cases} \qquad \psi_\epsilon(x) = \begin{cases} x/\epsilon, & |x| \leq \epsilon, \\ \text{sign}(x) & |x| > \epsilon. \end{cases} \tag{13}$$

Huber's minimax estimator combines the behavior of the L2 norm when the errors are small while maintaining the L1 norm's reduced sensitivity to outliers.[2]

With Huber's estimator, outliers still have some finite influence. To increase robustness further, we will consider *redescending* estimators [25] for which the influence of outliers tends

[2]The minimax ψ-function is often written as $\psi_\epsilon(x) = \min(\epsilon, \max(x, -\epsilon))$.

to zero.[3] One such estimator is the *Lorentzian*:

$$\rho_\sigma(x) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \qquad \psi_\sigma(x) = \frac{2x}{2\sigma^2 + x^2}. \tag{14}$$

The estimator is plotted along with its ψ-function in Figure 8. Examination of the ψ-functions reveals that the estimator has a *saturating* property; that is, the influence of outliers tends to zero. The remainder of the paper will examine the relationship between these redescending estimators and outlier processes. See Appendix A for a catalog of common redescending estimators.

4.2 Robust Regularization

We now apply the robust estimation technique to our least-squares formulation of the surface recovery problem:

$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S}[(u_s - d_s)^2 + \lambda \sum_{t \in \mathcal{G}_s}(u_s - u_t)^2].$$

The first term may be violated when our model of the data is not accurate. For example, we might assume that the depth of a surface is piecewise constant. For a highly textured surface this may be a poor approximation and in fitting a constant-depth surface to the region we will encounter gross errors in $(u_s - d_s)$. Similarly, the spatial term assumes that within the neighborhood $\mathcal{G}_s$ only a single surface is present and $(u_s - u_t)^2$ will be small. Near a surface boundary, some of the spatial errors will be large; these should be rejected as outliers.

Our approach is to replace the quadratic estimator with robust estimators ρ_D and ρ_S for the data and spatial terms. This gives the following objective function:

$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S}[\rho_D(u_s - d_s) + \sum_{t \in \mathcal{G}_s} \rho_S(u_s - u_t)]. \tag{15}$$

This robust formulation of data and spatial terms has been applied to optical flow estimation [5, 7, 8], and it will be applied to image segmentation later in the paper. As with the line-process approaches the objective function may be non-convex depending on the choice of estimator.

[3] Hampel *et al.* chose a more conservative definition of redescending estimators that requires $\psi(x) = 0$, $|x| \geq r$ for some positive r.

4.3 Other Robust Approaches

There are other approaches to robust statistics that have been applied to computer vision problems (see [39] for a review). One popular approach in computer vision is the least-median-of-squares (LMedS) algorithm in which model parameters are estimated by minimizing the median of the squares of the residuals. Another common approach uses iteratively reweighted least squares (IRLS) [2, 11, 51]. While the standard least-squares approach tries to find parameters, $\mathbf{a}$, that produce small residuals, $(d_s - \mathbf{u}(s;\mathbf{a}))$, the idea of IRLS is to weight the residuals to reduce the influence of outliers. High weights (approaching unity) are assigned to "good" data and lower weights to outlying data. Other robust approaches in computer vision include the use of K-means clustering [55] and mixture models [31, 37]. All these approaches lack the clear connection to traditional line-process and weak-continuity approaches and, for this reason, we focus on the M-estimator approach.

5 Unifying Robust Estimation and Outlier Processes

This section unifies the robust estimation approaches and traditional line-process approaches. First we introduce the notion of an outlier process which is a generalization of the line process and can be applied to both data and spatial terms. We then show how these binary or analog outlier processes can be eliminated in the same way that line processes are eliminated and how this results in a robust estimation problem. Finally we derive a mechanism for converting robust estimators into outlier process, thus making the connection complete.

5.1 Outlier Processes

Recall the line-process formulation of the surface recovery problem:

$$E(\mathbf{u},\mathbf{d},\mathbf{l}) = \sum_{s\in S}\left((u_s - d_s)^2 + \lambda \sum_{t\in\mathcal{G}_s}[(u_s - u_t)^2\, l_{s,t} + \Psi_S(l_{s,t})]\right), \tag{16}$$

where $l_{s,t}$ is an analog line process, and where Ψ_S is a penalty term. This formulation accounts for violations of the spatial smoothness term, but does not account for violations of the data term. In many situations the data term, like the spatial term, is only an approximate model of the data process. For example, when computing stereo correspondence or optical flow the data term embodies an assumption of "data conservation;" that is, a portion of the scene visible in one image is visible in the other stereo image or at the next time instant.

This assumption may be violated at depth discontinuities in the scene and the resulting measurements at these locations will be erroneous. To accurately recover structure in these areas, the "bad" measurements must be ignored [7, 49].

This prompts us to generalize the notion of a "line process" to that of an "outlier process" that can be applied to both data and spatial terms. The motivation behind such a generalization is to formulate a process that performs *outlier rejection* in the same spirit as the robust estimators do. The recovery problem is then reformulated using outlier processes as follows:

$$E(\mathbf{u},\mathbf{d},\mathbf{l},\mathbf{m}) = \sum_{s\in S}\left((u_s - d_s)^2 m_s + \Psi_D(m_s) + \lambda \sum_{t\in\mathcal{G}_s}[(u_s - u_t)^2 l_{s,t} + \Psi_S(l_{s,t})]\right), \qquad (17)$$

where we have simply introduced a new process m_s and a new penalty term Ψ_D for rejecting the measurement. This process allows us to ignore erroneous information from the data term. In related work, Geiger and Pereira [18] introduce a binary measurement process that they refer to as the "sparse process". While they use the process to perform minimal visual encoding, they note its similarity to outlier rejection and robust statistics.

Returning to our surface reconstruction example, recall that the introduction of a line process prevented smoothing over the layers in the wedding cake but that spurious discontinuities were found near data outliers. We now consider the outlier-process formulation in which both data and spatial terms are allowed to be violated. The value of λ was taken to be 0.25 which means that the data and spatial terms are weighted equally. The recovered surface plot in Figure 9*a* is very close to the original surface and the number of spurious spatial discontinuities is significantly reduced (Figure 9*c*). The data points which were treated as outliers are shown as black regions in Figure 9*d*. In the case of the outlier process formulation, no parameter tuning is required to achieve an accurate solution;[4] unexpected data measurements are automatically rejected.

5.2 From Outlier Processes to Robust Estimation

The outlier-process formulation leads to an expensive joint estimation problem where one not only has to estimate $\mathbf{u}$ but also the outlier processes $\mathbf{l}$ and $\mathbf{m}$. In the case of the simple binary line-process formulation, Blake and Zisserman [9] show how the line variables can

[4]It is all too common for optimization schemes to require parameter "tuning" to achieve acceptable results. Our experience with the fully robust formulation is that it allows the parameters to be set based on the expected measurement errors, obviating the need for "tuning".

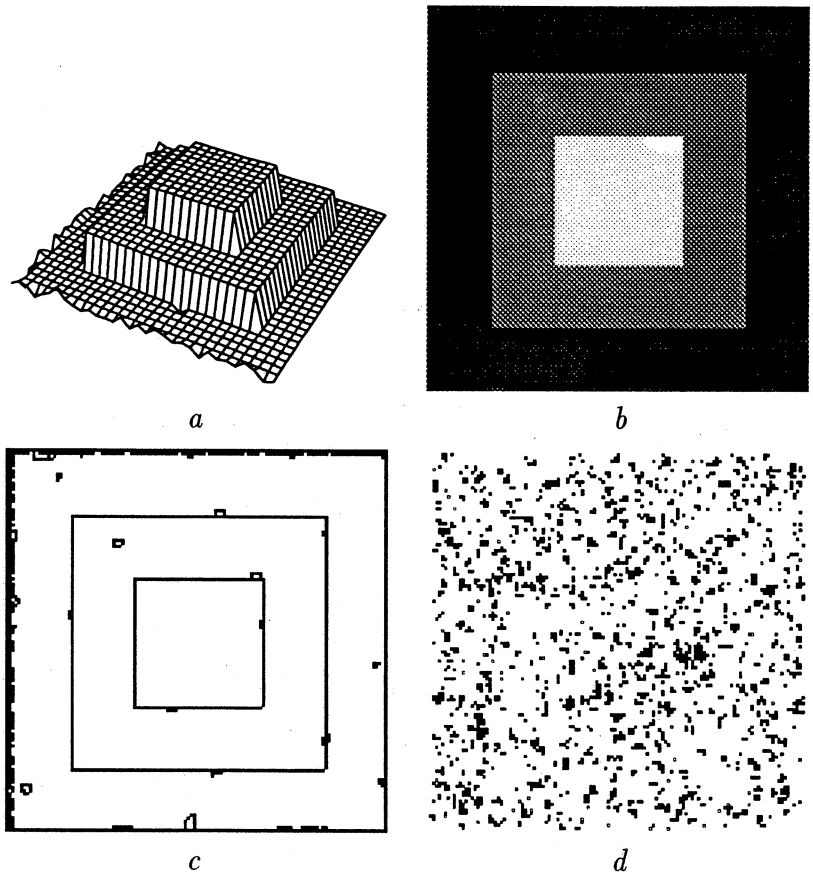


Figure 9: Surface Reconstruction with both spatial and data outlier processes. *(a)* Surface plot. *(b)* Height plotted as intensity. *(c)* Spatial discontinuities (outliers). *(d)* Data discontinuities (outliers).

be removed from the equation by first minimizing over them. They obtain a new objective function that is solely a function of $\mathbf{u}$. Exactly the same treatment can be applied to the general analog outlier-process version.

The optimization problem can be written as

$$\min_{\mathbf{u},\mathbf{m},\mathbf{l}} \left[\sum_{s \in S} [(u_s - d_s)^2 m_s + \Psi_D(m_s)] + \lambda \sum_{s \in S} \sum_{t \in \mathcal{G}_s} [(u_s - u_t)^2 l_{s,t} + \Psi_S(l_{s,t})] \right]. \tag{18}$$

Notice that the measurement term does not depend on $\mathbf{l}$ and the smoothness term does not depend on $\mathbf{m}$. Thus we can rewrite the equation as,

$$\min_{\mathbf{u}} \left[\left[\min_{\mathbf{m}} \sum_{s \in S} (u_s - d_s)^2 m_s + \Psi_D(m_s) \right] + \lambda \left[\min_{\mathbf{l}} \sum_{s \in S} \sum_{t \in \mathcal{G}_s} (u_s - u_t)^2 l_{s,t} + \Psi_S(l_{s,t}) \right] \right]. \tag{19}$$

We can now minimize with respect to each process separately; that is for each term we

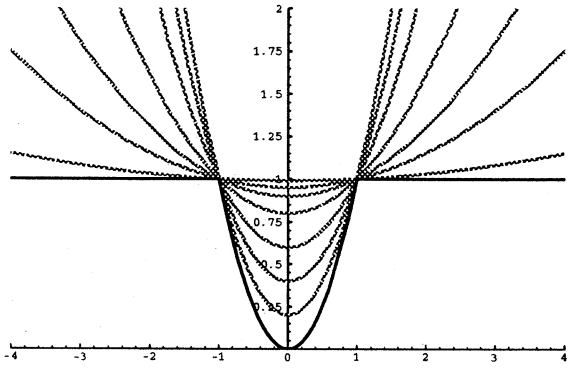


Figure 10: Family of quadratics with a common intersection. The infimum of this family is the *truncated quadratic* shown in bold.

compute:

$$\rho(x) = \inf_{0 \leq z \leq 1} (x^2 z + \Psi(z)), \tag{20}$$

where z is the outlier process. Finally, we can rewrite the minimization problem as

$$\min_{\mathbf{u}} \left[\sum_{s \in S} \rho_D(u_s - d_s)^2 + \lambda \sum_{s \in S} \sum_{t \in \mathcal{G}_s} \rho_S(u_s - u_t) \right]. \tag{21}$$

Geiger and Yuille [19] propose a similar formulation with a binary (as opposed to analog) process for the data and spatial terms based on the formulation of Geiger and Pereira [18]. They then use mean-field theory techniques to average out the binary processes. As they point out, this gives a robust estimation problem with the mean-field function as the robust estimator.

5.2.1 Example: Binary Outlier Process

Consider the simple binary outlier-process formulation $(x^2 z + \beta(1-z))$, where β is a constant. For a fixed value of z this is just a quadratic, and we can plot this function for various values of $0 \leq z \leq 1$. Figure 10 shows the function $x^2 z + \beta(1 - z)$ plotted for various values of the outlier process z, and the infimum of this family of quadratics plotted in bold.[5] This bold curve is the *truncated quadratic* derived by Blake and Zisserman [9]:

$$\rho(x, \beta) = \begin{cases} x^2 & \text{if } |x| < \sqrt{\beta}, \\ \beta & \text{otherwise,} \end{cases} \tag{22}$$

[5]While the figure shows a family of quadratics for $0 \leq z \leq 1$, the infimum is determined by the cases where $z = 0$ or $z = 1$. In this case, z is simply a binary valued line process.

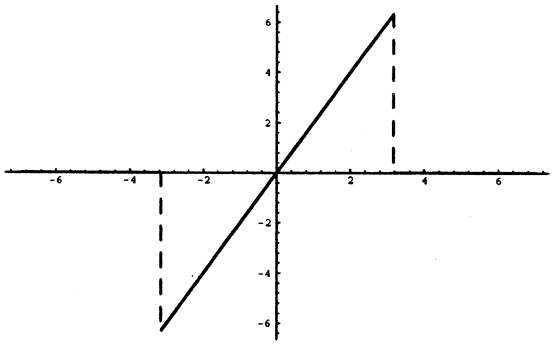


Figure 11: Truncated quadratic ψ-function.

$$\psi(x,\beta) = \begin{cases} 2x & \text{if } |x| < \sqrt{\beta}, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

We can now eliminate the outlier processes from the equation and rewrite it in terms of the truncated quadratic:

$$\min_{u} \sum_{s \in S} [\rho((u_s - d_s), \beta_D) + \sum_{t \in \mathcal{G}_s} \rho((u_s - u_t), \beta_S)]. \tag{24}$$

Notice that this is identical to the robust estimation formulation with the truncated quadratic as the estimator. This is one of the common redescending estimators used in robust statistics. Up to a fixed threshold, errors are weighted quadratically, but beyond that the estimator has a saturating property; errors receive a constant value. By examining the ψ-function (Figure 11) we see that the influence of outliers goes to zero beyond the threshold.

5.2.2 Example: Analog Outlier Process

Let us consider an analog outlier process with the penalty function from Figure 4 where $0 \leq z \leq 1$ and

$$\begin{aligned} E(x,z) &= x^2 z + \Psi(z) \\ &= x^2 z + (\sqrt{z} - 1)^2. \end{aligned}$$

Minimizing E w.r.t. z, we take the partial derivative and set it equal to zero:

$$\frac{\partial}{\partial z} E(x,z) = x^2 + \frac{\sqrt{z} - 1}{\sqrt{z}} = 0.$$

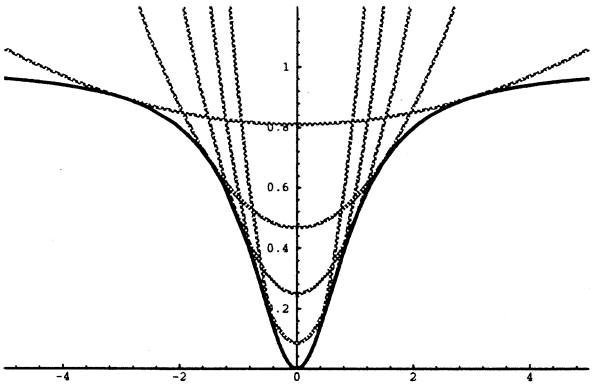


Figure 12: Geman-McClure estimator.

Solving for z, gives

$$z = \frac{1}{(x^2+1)^2}.$$

Substituting z back into $E(x,z)$ then gives

$$\rho(x) = \frac{x^2}{1+x^2},$$

which is a robust estimator proposed by Geman and McClure [23]. Figure 12 plots $E(x,z)$ for various values of z and in bold, the infimum of this family of quadratics (*i.e.* $\rho(x)$).

5.3 From Robust Estimators to Outlier Processes

Our goal is to take an objective function written in terms of robust estimators and derive a new objective function which is written in terms of outlier processes. The reason for making the outlier process explicit is two-fold. From a theoretical viewpoint, this shows the relationship between the robust approach and the outlier-process approach. From a practical view, the introduction of an outlier process allows it to be manipulated in standard ways. For example, we might begin with an objective function where the spatial smoothness term uses a robust estimator. If we wish to add prior assumptions about spatial coherence (through clique energies) then we must convert the robust formulation into one in which the outlier process is made explicit.

Consider a simple robust objective function:

$$E(x) = \rho(x) \tag{25}$$

defined in terms of some "error" x, where x, for example, might be the spatial gradient

$u_s - u_t$ or the data error $u_s - d_s$. We want to construct a new objective function:

$$E(x, z) = x^2 z + \Psi(z) \tag{26}$$

where $0 \leq z \leq 1$ is an outlier process, and Ψ is a "penalty function", such that the minimum of $E(x, z)$ with respect to x is the same as $E(x)$.

The following section provides a derivation of $\Psi(z)$ which is a simplification of the approach of Rangarajan and Chellappa [44]. The result is the same as that of Geman and Reynolds [21] but the derivation provides a constructive proof of the result. The basic result is then summarized and packaged as a simple *mechanism* for performing the conversion. The mechanism is then applied to an example in detail. A catalog of common robust estimators and their outlier processes formulations is provided in Appendix A.

5.3.1 Derivation

Our goal is to start with an objective function $E(x) = \rho(x)$ and introduce a new variable $0 \leq z \leq 1$, which will be the outlier processes, so that the solution at the minimum will be unchanged. This new objective function is

$$E(x, z) = x^2 \; z + \Psi(z).$$

To achieve this we must take a particular function $\rho(x)$ and find the appropriate function $\Psi(z)$ such that minimizing $E(x, z)$ gives the same solution as minimizing $\rho(x)$.

We must find the relationship between the error, x^2 and some outlier process $0 \leq z \leq 1$. Notice that since z is positive it should be symmetric and insensitive to the sign of the error. For these reasons we want to find some monotonic mapping, $w(z)$, which maps values of the outlier process into x^2. Given $w(z)$ written in terms of x^2 it is natural to write $\rho(x)$ in terms of x^2; that is, we define a function $\phi(x^2) = \rho(x)$ or

$$\phi(w(z)) = \rho(\sqrt{w(z)}) \tag{27}$$

since $w(z) \geq 0$.

We want the minimum of ϕ w.r.t. z to return the original function ρ. To ensure this we minimize ϕ subject to the constraint that $w(z) = x^2$ using the method of Lagrange multipliers. This gives the following function:

$$\mathcal{E}(x, z, \lambda) = \phi(w(z)) + \lambda(x^2 - w(z)). \tag{28}$$

Extremizing with respect to λ gives $w(z) = x^2$ at the minimum. Thus:

$$\mathcal{E}(x, z, \lambda) = \phi(x^2) = \rho(x) = E(x)$$

which leaves the function unchanged. Minimizing this w.r.t. z gives:

$$\frac{\partial}{\partial z}\mathcal{E}(x, z, \lambda) = \phi'(w(z))\frac{d}{dz}w(z) - \lambda\frac{d}{dz}w(z) = 0,$$

or $\lambda = \phi'(w(z))$. Substituting for λ in Equation (28) we get:

$$\mathcal{E}(x, z) = \phi(w(z)) + \phi'(w(z))(x^2 - w(z)). \tag{29}$$

We can now check to see if the minimum with respect to z also occurs at $x^2 = w(z)$. Taking the derivative w.r.t. z and setting it equal to zero gives:

$$\frac{\partial}{\partial z}\mathcal{E}(x, z) = \phi''(w(z))(x^2 - w(z))\frac{d}{dz}w(z) = 0.$$

Since $w(z)$ maps to x^2, its derivative is non-zero, so $(\partial/\partial z)\,\mathcal{E}(x, z)$ equals zero when

$$\phi''(w(z)) = 0 \quad \text{or} \quad w(z) = x^2.$$

We want the minimum at $w(z) = x^2$. Taking second derivatives with respect to z, we get

$$\begin{aligned}\frac{\partial^2\mathcal{E}(x, z)}{\partial z^2} &= \phi'''(w(z))(x^2 - w(z))(w'(z))^2 \\ &\quad - \phi''(w(z))(w'(z))^2 \\ &\quad + \phi''(w(z))(x^2 - w(z))(w'(z))^2.\end{aligned}$$

Evaluating the second derivative at the extremum point $x^2 = w(z)$, the second derivative condition for a minimum becomes

$$\frac{\partial^2\mathcal{E}(x, z)}{\partial z^2} = -\phi''(w(z))(w'(z))^2 > 0,$$

for which we require $\phi''(w(z)) < 0$. This implies that $\phi(y)$, $y \geq 0$, is concave.[6]

Since we have $\lambda = \phi'(w(z))$ we can solve for $w(z)$ giving the transformation from z to x^2 in terms of λ:

$$w(z) = (\phi')^{-1}(\lambda),$$

if the inverse exists. The simplest way to make the equality hold is to define w to be $(\phi')^{-1}$ and z to be λ.

[6]This is the same requirement imposed by Geman and Reynolds [21]

We can then rewrite the objective function in Equation (28), substituting for $w(z)$:

$$\begin{aligned}
&\phi(w(z)) + \lambda(x^2 - w(z)) \\
&= x^2\,\lambda - \lambda\, w(z) + \phi(w(z)) \\
&= x^2\, z - z\,(\phi')^{-1}(z) + \phi((\phi')^{-1}(z)) \qquad \text{(substituting } (\phi')^{-1} \text{ for } w \text{ and } z \text{ for } \lambda) \\
&= x^2\, z + \Psi(z), \\
&= E(x, z),
\end{aligned}$$

where we take the penalty function to be

$$\Psi(z) = \phi((\phi')^{-1}(z)) - z\,(\phi')^{-1}(z).$$

Notice that z is defined in terms of ϕ'. For z to be an outlier process, we want it to vary between 0 and 1, hence we require

$$\lim_{y \to 0} \phi'(y) = 1 \quad \text{and} \quad \lim_{y \to \infty} \phi'(y) = 0.$$

The results of the derivation are summarized in Figure 13 which provides a straightforward mechanism for converting robust formulations to outlier process formulations.

5.3.2 Example

To illustrate the mechanism we consider the robust surface reconstruction example:

$$E(\mathbf{u}) = \sum_{s \in S} \rho(u_s - d_s, \sigma_D) + \lambda \sum_{t \in \mathcal{G}_s} \rho(u_s - u_t, \sigma_S), \tag{32}$$

where σ_D and σ_S are constant scale parameters and where ρ is the Lorentzian estimator:

$$\rho(x, \sigma) = \log\left(1 + \left(\frac{x}{\sigma}\right)^2\right).$$

The first step is to define ϕ:

$$\phi(w) = \log(1 + w).$$

We then compute the first and second partial derivatives w.r.t. w:

$$\begin{aligned}
\phi'(w) &= \frac{1}{1 + w}, \\
\phi''(w) &= -\frac{1}{(1 + w)^2}.
\end{aligned}$$

1. Define $\phi(w) = \rho(\sigma\sqrt{x})$ where x is possibly multiplied by some scale parameter σ in ρ.

2. Compute the first and second partial derivatives ($\phi'(w)$ and $\phi''(w)$) of ϕ with respect to w.

3. If

   $\lim_{w\to 0} \phi'(w) = 1$, and
   $\lim_{w\to\infty} \phi'(w) = 0$, and
   $\phi''(w) < 0$,

   then proceed, otherwise stop: ρ does not have a simple outlier process formulation.

4. Define the outlier process $z = \phi'(w)$.

5. Solve $z = \phi'(w)$ for w giving $w = (\phi')^{-1}(z)$.

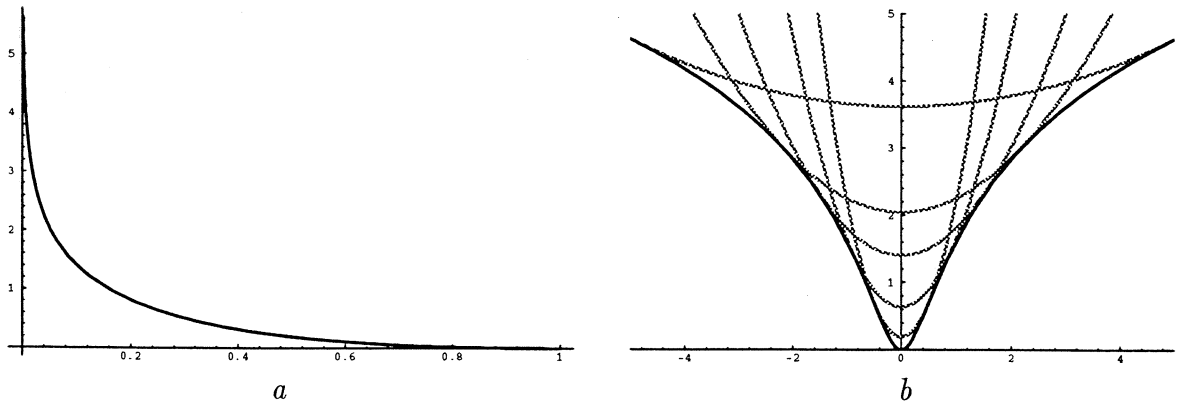
6. Define:

$$\begin{aligned} \Psi(z) &= \phi(w) - zw \\ &= \phi((\phi')^{-1}(z)) - z(\phi')^{-1}(z). \end{aligned} \tag{30}$$

7. The new objective function is:

$$E(x, z) = \sigma x^2 z + \Psi(z) \tag{31}$$

Figure 13: A simple mechanism for recovering the outlier process from a robust estimator.



a *b*

Figure 14: Lorentzian outlier process. *(a)* Penalty function. *(b)* Infimum of $E(x, z, \sigma)$.

We observe that these satisfy the conditions in Step 3 of the mechanism and we can define $z = \phi'(w)$. Solving for w gives:

$$w = (\phi')^{-1}(z) = \frac{1}{z} - 1.$$

Given z and w we can write the penalty function as

$$\Psi(z) = -1 + z + \log\left(\frac{1}{z}\right) = z - 1 - \log z.$$

The penalty function is plotted in Figure 14*a*. The outlier-process formulation for the Lorentzian is then:

$$E(x, z, \sigma) = \left(\frac{x}{\sigma}\right)^2 z + \Psi(z).$$

This function is plotted (in gray) for various values of z in Figure 14*b*; the bold curve is the Lorentzian estimator ρ which is the infimum of the family of quadratics.

We can now write the surface recovery problem using the Lorentzian outlier process:

$$\begin{aligned} E(\mathbf{u}, \mathbf{m}, \mathbf{l}, \sigma_D, \sigma_S) \;=\; & \sum_{s \in S} \left[\left(\frac{1}{\sigma_D^2} (u_s - d_s)^2 \, m_s + \Psi(m_s) \right) \right. \\ & \left. + \lambda \sum_{t \in \mathcal{G}_s} \left(\frac{1}{\sigma_S^2} (u_s - u_t)^2 \, l_{s,t} + \Psi(l_{s,t}) \right) \right], \end{aligned} \tag{33}$$

where Ψ is the penalty function defined above.

6 Exploiting the Relationship

In this section we exploit the relationship between robust estimation and outlier processes. We show that the recovery of the outlier process allows the formulation of prior assumptions on the spatial organization of outliers. We also show the extension of continuation methods to the analog outlier-process formulation with these spatial constraints.

6.1 Adding Spatial Interactions

The previous sections have shown how robust estimation and outlier processes are closely related and how we can often convert one formulation into the other. One motivation for recovering the outlier process is that it allows us to incorporate into the objective function prior assumptions on the nature of discontinuities. While numerous authors have proposed

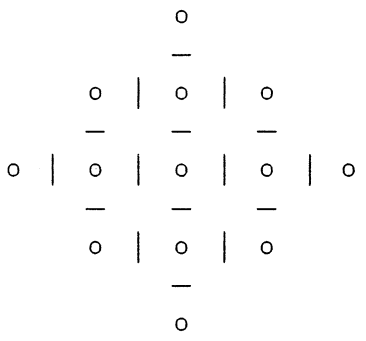


Figure 15: Neighborhood for spatial interactions.

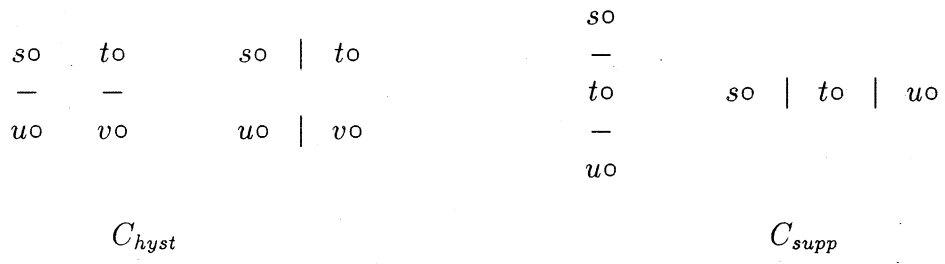


Figure 16: Cliques for spatial interaction constraints. The C_{hyst} cliques are used for hysteresis and the C_{supp} cliques are used for non-maxima suppression.

spatial coherence constraints for discrete line processes [14, 21, 22, 41], we need to generalize these results to the case of analog line processes.

We consider two kinds of interaction terms; *hysteresis* [12] and *non-maximum suppression* [42]. The hysteresis term assists in the formation of unbroken contours while the non-maximum suppression term inhibits multiple responses to a single edge present in the data. Hysteresis corresponds to lowering the penalty for creating edges at a particular orientation non-maximum suppression corresponds to increasing the penalty for creating multiple edges parallel at the same orientation.

Consider the neighborhood system in Figure 15 and cliques of the form shown in Figure 16. We define a new term which penalizes the configurations on the right in Figure 16 and rewards the configurations on the left. We define a new term E_I which encodes our prior assumptions about the organization of spatial discontinuities:

$$E_I(\mathbf{l}) = -\alpha\epsilon_1 \sum_{C_{hyst}} (l_{s,u}l_{t,v} + l_{s,t}l_{u,v}) + \alpha\epsilon_2 \sum_{C_{supp}} (l_{s,t}l_{t,u}) . \tag{34}$$

The first expression rewards continuity while the second penalizes discontinuities which are

"too close." The parameters ϵ_1 and ϵ_2 assume values in the interval $[0, 1]$ and α controls the importance of the spatial interaction term. We now minimize the following objective function which contains the data term, spatial smoothness term, and the new spatial interaction term:

$$E(\mathbf{u}, \mathbf{d}, \mathbf{m}, \mathbf{l}) = E_D(\mathbf{u}, \mathbf{d}, \mathbf{m}) + E_S(\mathbf{u}, \mathbf{l}) + E_I(\mathbf{l}). \tag{35}$$

6.2 Continuation Methods and Outlier Processes

Recall that the robust objective function (15):

$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S} [\rho_D(u_s - d_s) + \sum_{t \in \mathcal{G}_s} \rho_S(u_s - u_t)],$$

may be non-convex depending on the choice of estimator. Continuation methods provide one popular class of approaches for minimizing such non-convex functions. The idea is to choose an estimator which has a control parameter that can be used to change the shape of the estimator. This parameter is exploited to construct a convex approximation to the objective function which can be readily minimized. The minimum is then *tracked* as the control parameter is adjusted so that the objective function increasingly approximates the original non-convex estimation problem.

We can recover an analog outlier process for this type of estimator and the penalty function in the outlier-process formulation retains the control parameter of the original estimator. This allows us to apply continuation methods to the explicit outlier-process formualtions, and to do so even in the presence of spatial interations. Below we illustrate by deriving an analog outlier process for the GNC function.

6.2.1 Example: The GNC Function

Blake and Zisserman [9] construct a piecewise polynomial approximation to the truncated quadratic (Equations 22 and 23)[7]:

$$\rho(x, \lambda, c) = \begin{cases} \lambda^2 x^2 & 0 \leq \lambda^2 x^2 < \frac{c}{1+c} \\ 2\lambda |x| \sqrt{c(1+c)} - c(1 + \lambda^2 x^2) & \frac{c}{1+c} \leq \lambda^2 x^2 < \frac{1+c}{c} \\ 1 & \text{otherwise} \end{cases} \tag{36}$$

$$\psi(x, \lambda, c) = \begin{cases} 2\lambda^2 x & 0 \leq \lambda^2 x^2 < \frac{c}{1+c} \\ 2\lambda(\text{sign}(x)\sqrt{c(1+c)} - c(\lambda x)) & \frac{c}{1+c} \leq \lambda^2 x^2 < \frac{1+c}{c} \\ 1 & \text{otherwise} \end{cases} \tag{37}$$

[7]The parameterization used here is from [44].

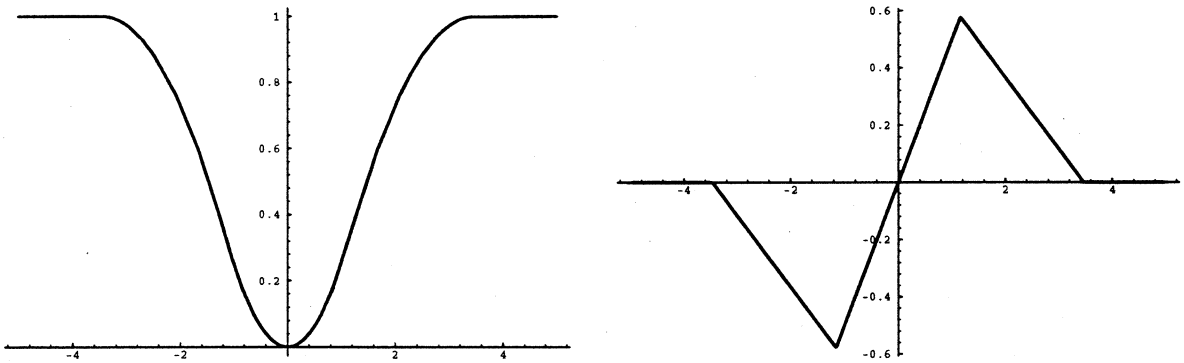


Figure 17: GNC Function ($\lambda = 0.5$, $c = 0.5$). *(a)* $\rho(x, \lambda, c)$, *(b)* $\psi-$function.

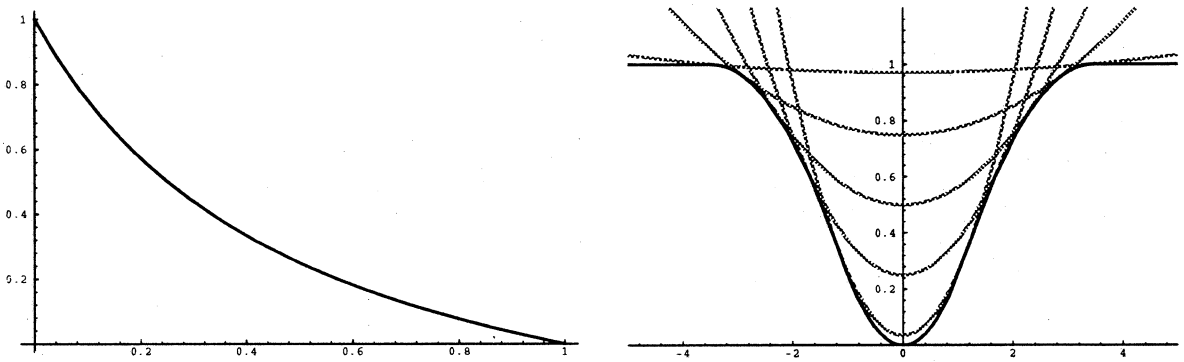


Figure 18: GNC penalty function and objective function (for $c = 0.5$).

The parameter c, controls the shape of the estimator and as $c \to \infty$ the estimator approaches the truncated quadratic. The estimator and its ψ-function are shown in Figure 17.

To recover a GNC outlier process we write down $\phi(w, c)$ and its derivatives:

$$\phi(w,c) = \begin{cases} w & 0 \le w < \frac{c}{1+c} \\ 2\sqrt{c\,w\,(1+c)} - c(1+w) & \frac{c}{1+c} \le w < \frac{1+c}{c} \\ 1 & \text{otherwise} \end{cases}$$

$$\phi'(w,c) = \begin{cases} 1 & 0 \le w < \frac{c}{1+c} \\ c\left(\sqrt{\frac{1+c}{w\,c}} - 1\right) & \frac{c}{1+c} \le w < \frac{1+c}{c} \\ 0 & \text{otherwise} \end{cases}$$

$$\phi''(w,c) = \begin{cases} 0 & 0 \le w < \frac{c}{1+c} \\ -\frac{1}{2}\sqrt{\frac{c(1+c)}{w^3}} & \frac{c}{1+c} \le w < \frac{1+c}{c} \\ 0 & \text{otherwise.} \end{cases}$$

We see that $\lim_{w\to 0} \phi'(w) = 1$, $\lim_{w\to\infty} \phi'(w) = 0$, and ϕ is concave, satisfying Step 3 of the

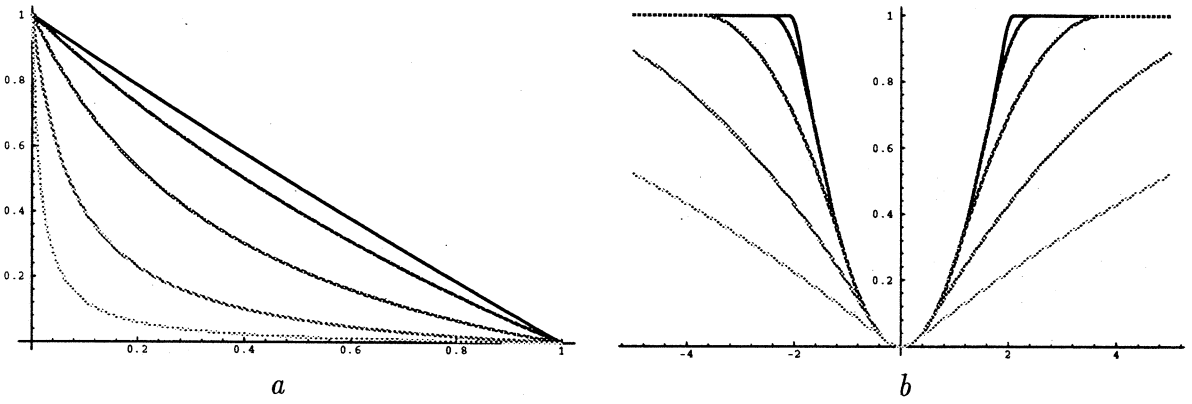


Figure 19: GNC Function. Dark indicates *high* values of the control parameter, c; light indicates *low* values of c. *(a)* The penalty function $\Psi(z, c)$ for the GNC outlier process. *(b)* The GNC-ρ function.

mechanism. Finally, deriving the penalty function we get:

$$(\phi')^{-1}(z) = \frac{c(1+c)}{(c+z)^2} \tag{38}$$

$$\Psi(z, c) = \frac{c(1-z)}{c+z} \tag{39}$$

$$E(x, \lambda, c, z) = \lambda^2 x^2 z + \Psi(z, c), \tag{40}$$

where $0 \leq z \leq 1$. The penalty function is plotted in Figure 18 as is the objective function for various values of z. The GNC function is the infimum of the family of quadratics and is displayed in bold. We have transformed the middle section alone of the GNC ϕ function. This section contains all the information. The other two sections of the GNC sequence, namely the quadratic and constant sections are degenerate in our formulation: a quadratic and a constant section correspond to outlier processes $z = 1$ and $z = 0$ respectively.

Notice that the penalty function, $\Psi(z, c)$, retains the control parameter c. Figure 19 plots the GNC function and the corresponding penalty function for various values of c. As the value of c increases, the GNC function approximates the truncated quadratic. Correspondingly as $c \to 0$ the penalty for introducing a discontinuity ($z = 0$) goes to infinity; minimizing the function then leaves us with a quadratic estimator. As c increases, the penalty function goes to $(1 - z)$ which means the objective function becomes:

$$E(x, \lambda) = \lambda^2 x^2 z + (1 - z).$$

This is just the simple binary outlier-process formulation.

6.2.2 Continuation Methods and Spatial Interactions

By deriving penalty functions with continuation parameters we can apply standard continuation methods to problems which involve spatial interaction of outlier processes. By adjusting the control parameter we can begin minimizing an objective function which gives infinitely high penalties for introducing outliers. This will mean that initially, no outliers will be introduced. Then by adjusting the control parameter, outliers begin to appear and interact.

This is one of the by-products of the relationship between robust estimation and line processes. We can begin with a problem formulated using an estimator like the GNC estimator, construct an analog outlier process, and add interactions among the spatial outliers. We can then employ the GNC continuation method in just the same way as before. The key difference is that now we can make explicit the spatial interactions if they are important. In the following section we will show experiments which illustrate this idea.

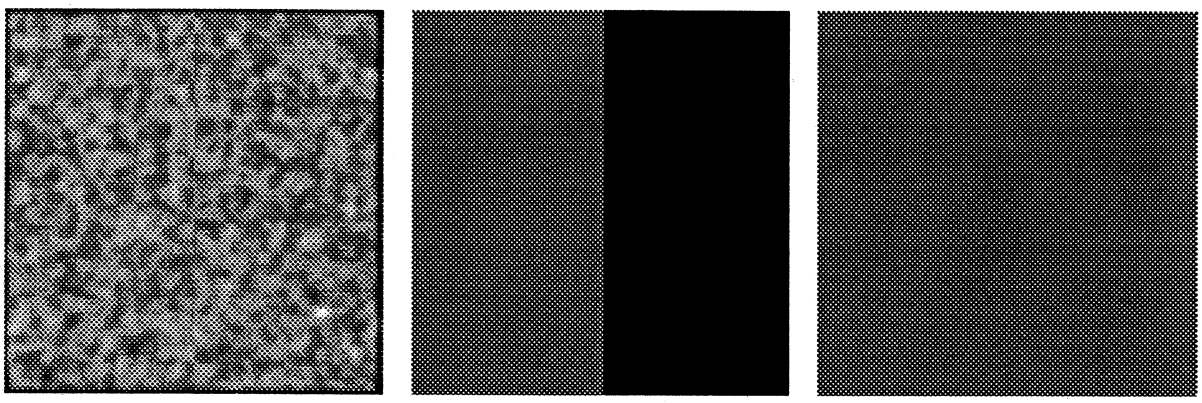
7 Outlier Processes: Experimental Results

This section illustrates the use of outlier processes and demonstrates how the outlier-process approach allows the introduction of additional constraints on spatial organization. The first two experiments show the effect of treating the data term robustly by illustrating situations in which data measurements may be viewed as outliers and rejected. The third experiment illustrates the application of continuation methods to the recovered outlier-process formulation with spatial organization constraints.

7.1 Optical Flow

Weak continuity methods have been widely applied to the problem of estimating dense optical flow, or 2D image velocity. Depth discontinuities in the scene, or the independent motion of objects, gives rise to optical flow fields that are piecewise smooth. While line proceses and weak-continuity methods have been used to preserve flow discontinuities [7, 32, 41, 49, 53], less attention has been paid to the measurement term and the assumptions it embodies. Let $I(x, y, t)$ be the image brightness at a point (x, y) at time t. The data conservation constraint can be expressed in terms of the standard *brightness constancy assumption* as follows:

$$I(x, y, t) \;=\; I(x + u\delta t, y + v\delta t, t + \delta t), \tag{41}$$



a *b* *c*

Figure 20: Random Noise Example. *a)* First random noise image in the sequence. *b)* True horizontal motion (black = -1 pixel, white = 1 pixel, gray = 0 pixels). *c)* True vertical motion.

where (u, v) is the horizontal and vertical image velocity at a point and δt is small. To recover the optical flow we want to find the vector $\mathbf{u} = (u, v)$ at each image location that minimizes the measurement error $I(x + u\delta t, y + v\delta t, t + \delta t)$ while enforcing piecewise continuity.

The brightness constancy assumption is violated in cases of transparency, shadows, reflections, and motion discontinuities. In these cases, erroneous measurements can be treated as outliers.[8] Black and Anandan [8] formulate the optical flow problem as robust estimation to account for measurement and spatial outliers:

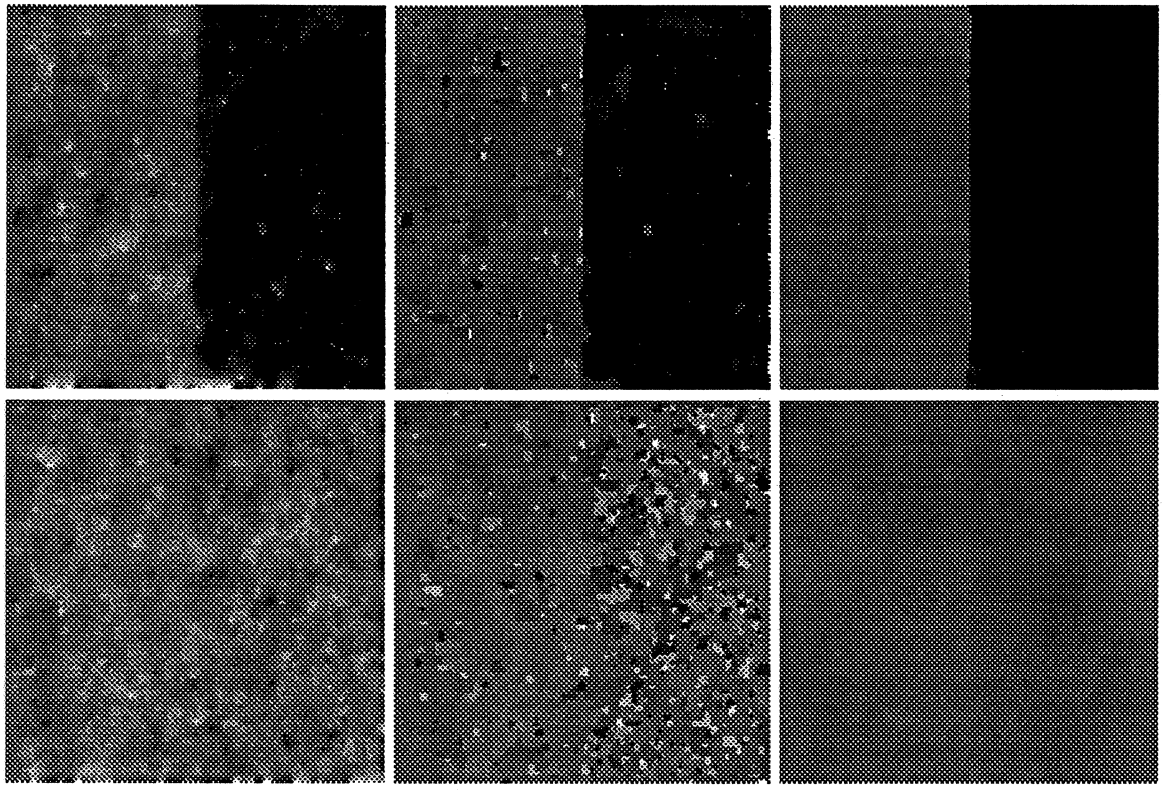
$$E(\mathbf{u}) = \sum_{s \in S} [\rho((I_x u_s + I_y v_s + I_t), \sigma_D) + \lambda \sum_{n \in \mathcal{G}_s} \rho(\|\mathbf{u}_s - \mathbf{u}_n\|, \sigma_S)], \tag{42}$$

where ρ is the Lorentzian estimator.

Here we briefly summarize some experiments with synthetic images (see [5, 8] for details). Consider the randomly textured image sequence shown in Figure 20 in which the right half of the image is translating one pixel to the left. The second image in the sequence has been corrupted with 10% uniform random noise. We compare the performance of three common approaches: a least-squares formulation (the standard Horn and Schunck formulation [29]), a version with a quadratic measurement term and robust smoothness term (see, for example, [26]), and the robust formulation.

The results are illustrated in Figure 21. The top row shows the horizontal motion and the bottom row shows the vertical motion recovered by each of the approaches (black = -1 pixel, white = 1 pixel, gray = 0 pixels). Figure 21*a* shows the noisy, but smooth, results

[8]The same types of violations occur in stereo correspondence problems.



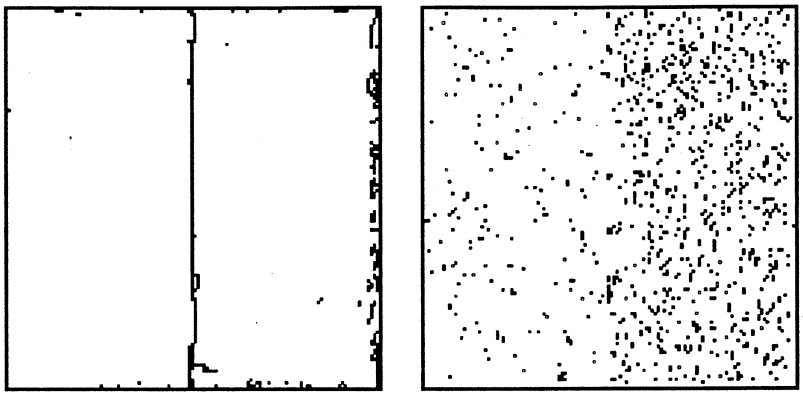
a *b* *c*

Figure 21: Effect of robust data term, (10% uniform noise). *a)* Least-squares (quadratic) solution. *b)* Quadratic data term and robust smoothness term. *c)* Fully robust formulation.

obtained by least-squares; the purely quadratic solution attempts to be faithful to both the smoothness model and the noisy data. Figures 21*b* shows the result of introducing a robust smoothness term alone (by, for example, introducing a line process). The recovered flow is piecewise smooth, but the gross errors in the data produce spurious motion discontinuities. With such a formulation, gross errors in the brightness data pull the solution away from the true flow while the line process compounds the problem by allowing discontinuities to be introduced. Finally, Figure 21*c* shows the improvement realized when outliers are rejected in both the measurement and spatial smoothness terms. Figure 22 shows where the spatial and measurement errors were treated as outliers.

7.2 Image Reconstruction

We now turn the problem of fitting a piecewise smooth brightness model to image data. Given an image brightness function I_s, $s \in S$, we want to recover a piecewise smooth surface



a *b*

Figure 22: Outliers in the smoothness and data terms, (10% uniform noise). *a)* Spatial discontinuities. *b)* Data outliers.

u that minimizes:

$$E(\mathbf{u}, I) = \sum_{s \in S} [\lambda_D \rho(u_s - I_s, \sigma_D) + \lambda_S \sum_{t \in \mathcal{G}_s} \rho(u_s - u_t, \sigma_S)], \tag{43}$$

where $\mathcal{G}_s$ is the standard first-order neighborhood system, and where we take ρ to be the Lorentzian estimator. A two level continuation method was used in which the solution was found for a coarse approximation using $\sigma_D(1)$ and $\sigma_S(1)$, then a fine level was computed with $\sigma_D(2)$ and $\sigma_S(2)$.

The assumption of piecewise smooth brightness is frequently violated in natural images. In textured regions and at intensity boundaries, where the brightness is not uniform, the use of a robust data term allows us to detect and reject the measurements that violate the uniform brightness assumption. This idea is exploited by Geiger and Pereira [18] to perform image compression. Piecewise-smooth areas contain redundant information that can be compressed, but areas that are changing rapidly (*i.e.* outliers with respect to the uniform brightness assumption) need to be detected and stored to reconstruct the image.

Figure 23 shows an image of Mission Bay in San Diego and the results of robust reconstruction. Figure 23*b* shows the recovered piecewise constant brightness model which removes a significant amount of the texture contained in the original image (Figure 23*a*). Notice that the data outliers in Figure 23*c* correspond to the textured regions in the image. Figure 23*d* illustrates the spatial outliers corresponding to brightness changes in the recovered image.

For this experiment the Lorentzian estimator was used with a two step continuation method and the parameters were as follows: $\lambda_D = 1.0$, $\lambda_S = 0.25$, $\sigma_D(1) = 75.0$ $\sigma_D(2) = 35.0$,

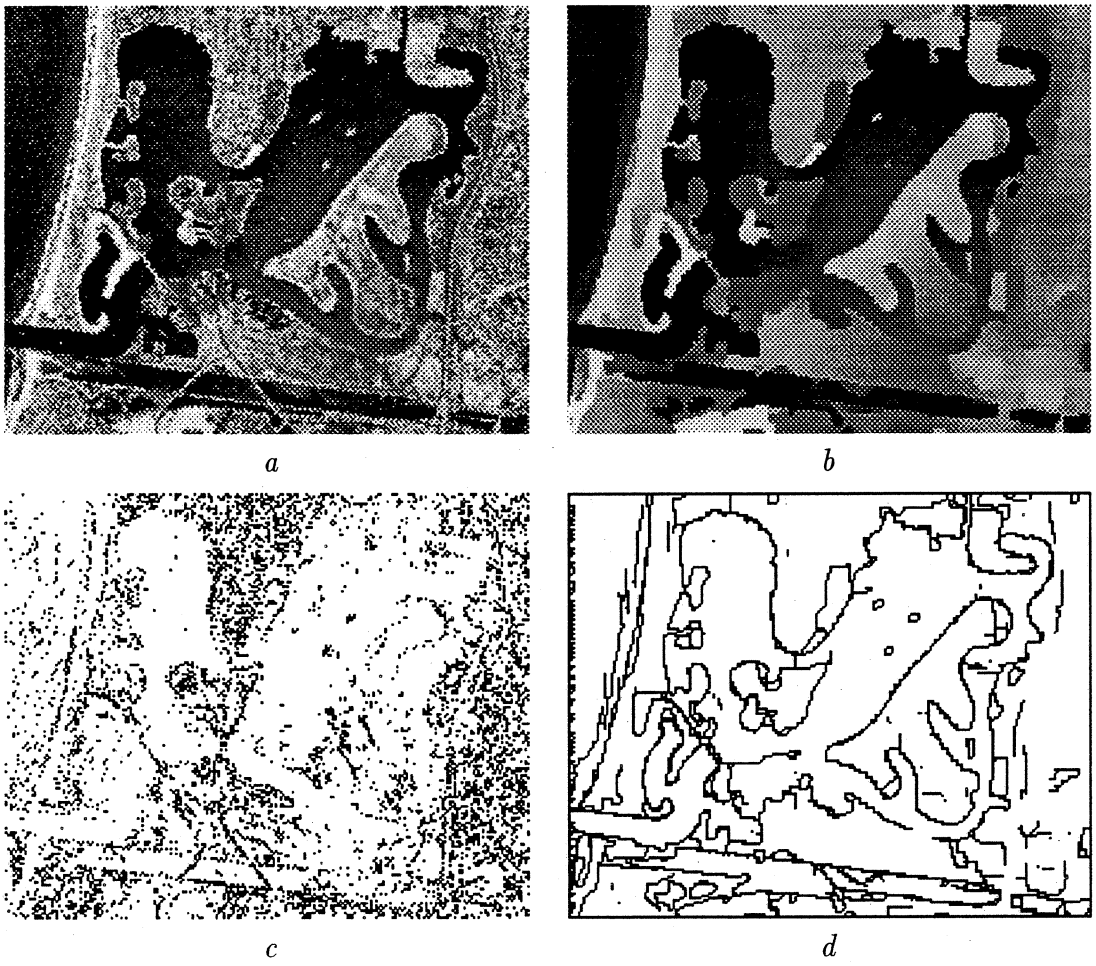


Figure 23: Aerial Image Experiment. *(a)* Original image; *(b)* Reconstructed image; *(c)* Data outliers; *(d)* Spatial outliers.

$\sigma_S(1) = 30.0$, $\sigma_S(2) = 4.5$, and 30 iterations were used at each step in the continuation method.

7.2.1 Noisy Images

We now consider an image reconstruction problem in which the image has been significantly degraded by random noise. For comparison purposes, Figure 23 shows the original (noiseless) aerial image. Figure 24 shows the same image degraded by additive white Gaussian noise (variance=175) and the results of the robust reconstruction. All parameter settings were exactly the same for the noiseless and noisy data. Visual inspection of the recovered image (in the noisy case) shows that it is very similar to the reconstruction in the noiseless case. The spatial outliers are also very similar. Notice, however, that there are significantly more

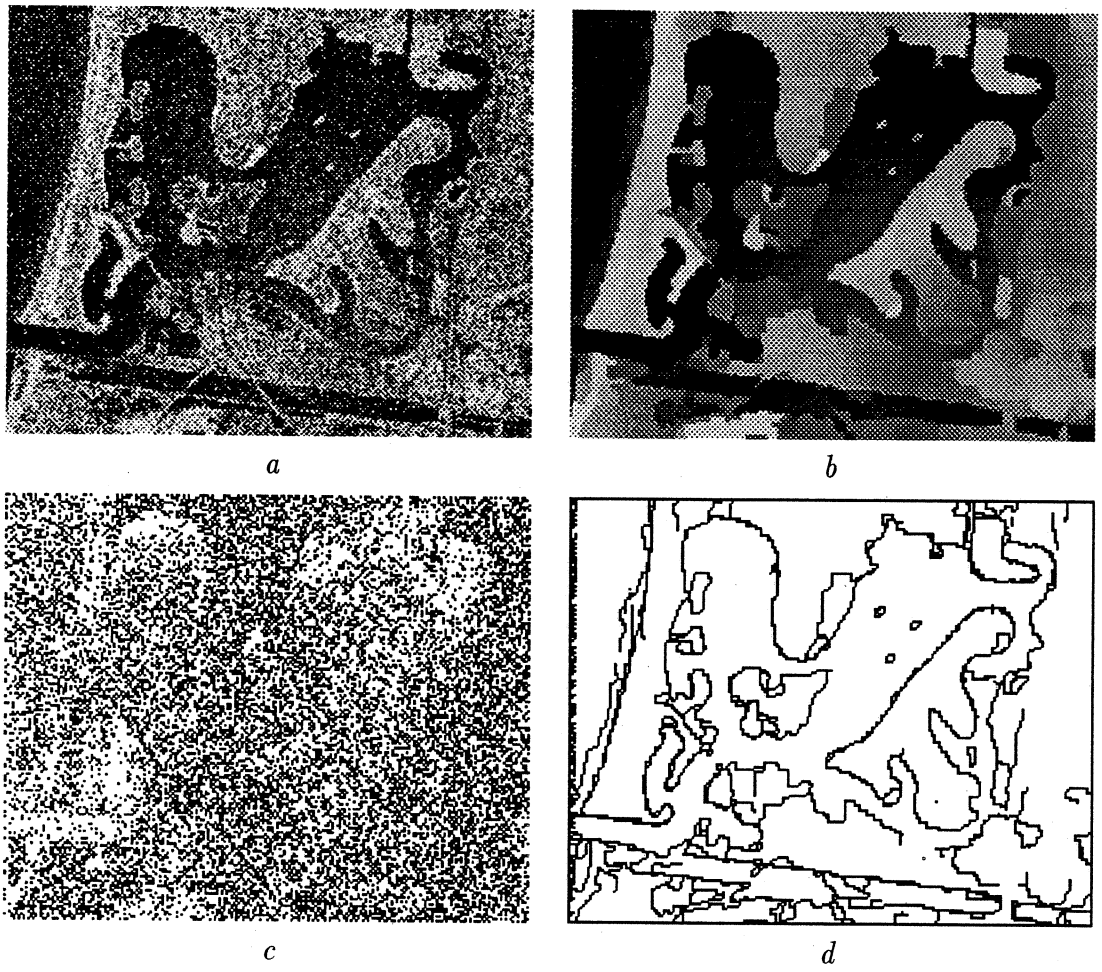


Figure 24: Degraded Image Experiment. *(a)* Corrupted image; *(b)* Reconstructed image; *(c)* Data outliers; *(d)* Spatial outliers.

data outliers which are introduced to reject the added image noise.

7.3 Spatial Interactions and Continuation Methods

In the previous experiments, no constraints on the organization of spatial outliers were used. As a result, there are gaps in the spatial outliers and many of these outliers are more than a single pixel in thickness. We now consider the addition of spatial interactions to promote hysteresis and discourage thick areas of spatial outliers. The results also illustrate the use of continuation methods in conjunction with analog outlier processes.

For these experiments we use the Mean Field function (MFF) [17] which is described in Appendix A along with its outlier process. The MFF has three control parameters: α, λ, and β. We hold α and λ fixed and change β from $\beta_0 = 0.007$ to $\beta_\infty = 7189.0$ according to the

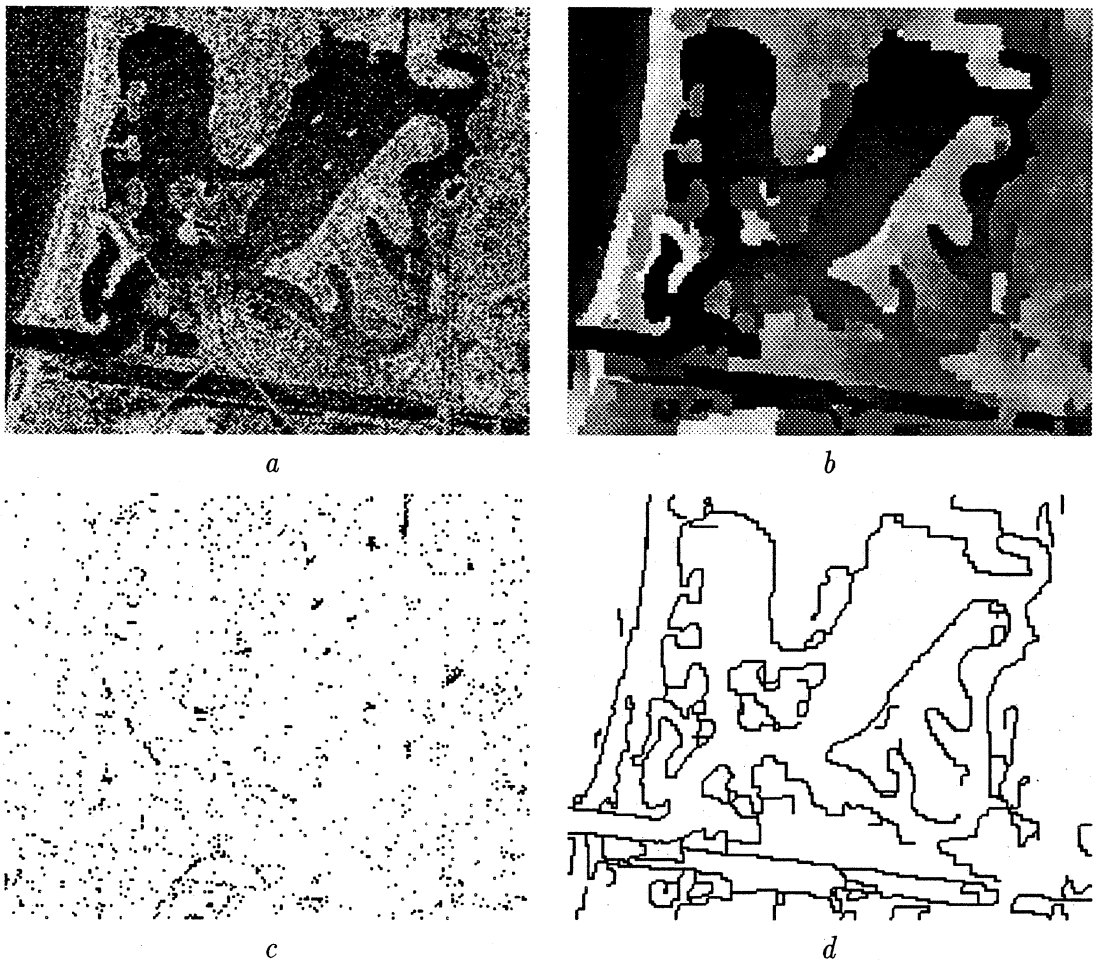


Figure 25: Mean Field Function (*without* spatial interactions). *(a)* Corrupted image; *(b)* Reconstructed image; *(c)* Data outliers; *(d)* Spatial outliers.

schedule: $\beta_i = 7\,\beta_{i-1}$. These experiments use a much more gradual continuation schedule than the simple two stage method employed in the experiments above. The MFF is applied to both the data and spatial terms with the following settings for α and λ: $\alpha_D = 1225.0$, $\lambda_D = 1.0$, $\alpha_S = 175.0$, $\lambda_S = 3.5$. For the spatial interaction term, E_I, we take $\epsilon_1 = \epsilon_2 = 0.5$ and $\alpha = 175.0$.

Figure 25 shows the results of the robust formulation before the introduction of spatial interactions. The spatial outliers recovered using the MFF are similar to those found with the Lorentzian and, in particular, there are gaps in the spatial discontinuities. Figure 26 shows the effect of introducing spatial interactions. The hysteresis term results in more closed contours; for example, the curved region of the island in the lower right portion of the image is completed in Figure 26 where it is not in Figure 25. In general, the term causes the

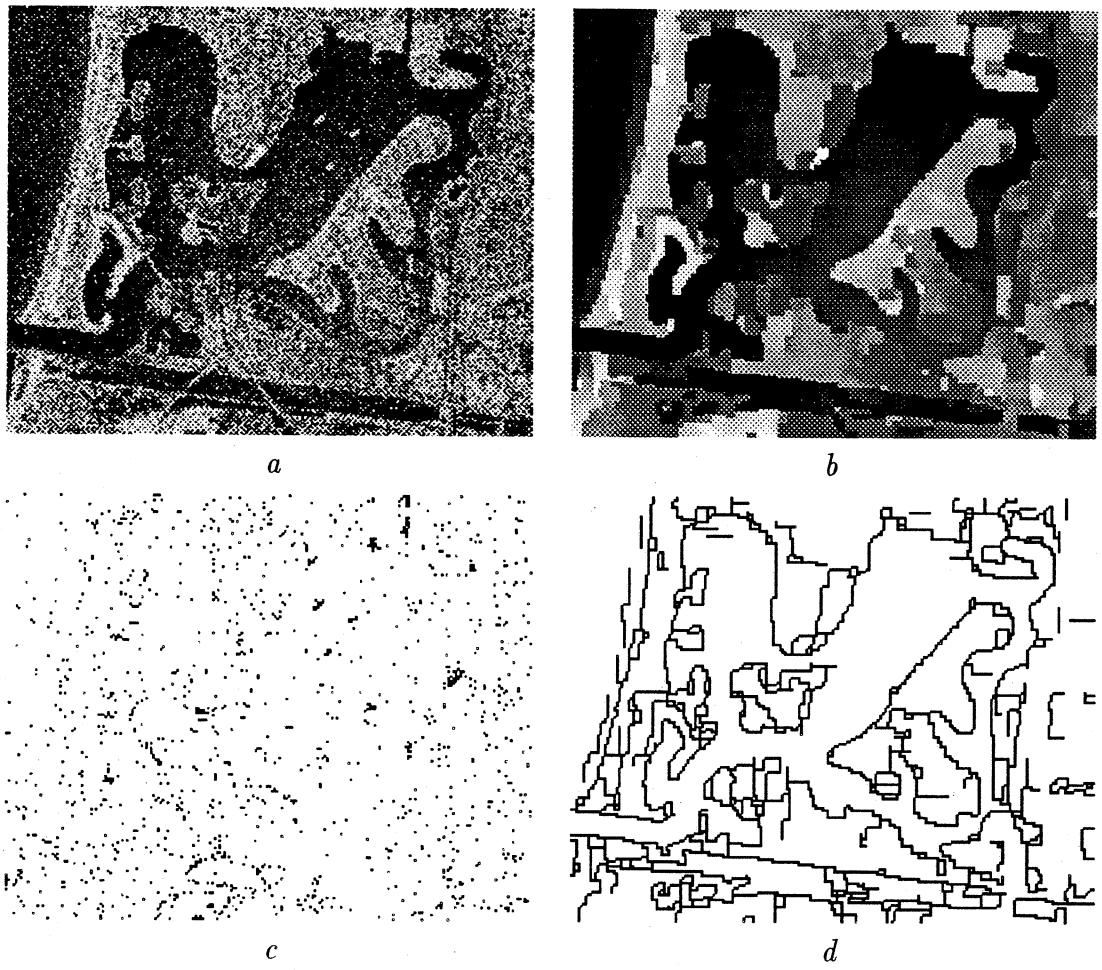


Figure 26: Mean Field Function (*with* spatial interactions). *(a)* Corrupted image; *(b)* Reconstructed image; *(c)* Data outliers; *(d)* Spatial outliers.

introduction of more spatial outliers which results in a more veridical reconstruction at the cost of perspicuity in the spatial outlier image.

8 Conclusion

The idea of weak continuity has had a great impact on many problems in early vision, including surface recovery, segmentation, image reconstruction, and optical flow estimation. Within the decade since the publication of Geman and Geman's paper [22] describing "line processes" for image restoration, the line process has become a nearly ubiquitous tool in early vision.

In the field of statistics, notions of robustness date back over one hundred years, but the 1960's and 1970's saw consolidation of the field we know as robust statistics today and as

characterized by the work of Huber [30] and Hampel *et al.* [25]. Förstner [16] was probably the first to apply robust statistical techniques to computer vision problems and, since then, there has been growing interest in robust techniques.

We have shown in this paper that the unifying concept underlying the line-process and robust-statistical approaches is the notion of outlier rejection. The generalization of line process to outlier processes makes the connection to robust statistics clear. Moreover, the elimination the outlier processes by minimization (eg. Blake and Zisserman) or by integration (eg. Mean-Field approaches) provides the connection to robust estimation approaches based on influence functions [25].

This connection is comforting since it provides a connection between two different communities, but on its own, not revolutionary. For those interested in computer vision, the power hidden in this connection lies in the ability to go in the other direction; that is, to recover an outlier process from a robust estimation problem. What this permits is the straightforward extension of results from robust statistics to problems in vision. In particular, by recovering an analog outlier process we can enforce constraints on spatial continuity which are crucial for many problems. Additionally, we can continue to use the continuation methods developed for solving non-convex optimization problems to solve the robust formulations with explicit outlier processes.

It is worth reviewing where we stand and how we arrived here. Geman and Geman [22] introduced the notion of a binary line process to account for spatial discontinuities and they formulated constraints on the spatial organization of these processes. Blake and Zisserman [9] removed the line processes by minimizing over them and consequently derived the truncated quadratic estimator. They then developed the GNC function to approximate the truncated quadratic and exploited it in a deterministic continuation method. While this is more computationally tractable than the stochastic approach used by Geman and Geman, Blake and Zisserman had to sacrifice the use of constraints on the spatial organization of the line processes. The mechanism presented here allows us to take an estimator like the GNC function and recover its analog outlier process. Notice that the outlier process for the GNC function is quite different from the binary line process used to derive the truncated quadratic. By recovering this process we can now formulate the spatial organization constraints of Geman and Geman (with some modification to account for the analog nature of the process) *without* sacrificing the continuation method of Blake and Zisserman. Moreover,

the connection to robust statistics opens up a host of new potential outlier process that have specific outlier rejection properties which may be desirable. Our experiments indicate that these connections are already paying off and that the outlier-process framework generalizes the work on line processes and extends robust estimation techniques to cope with problems in early vision.

Acknowledgements

Portions of this work were performed by M. B. at Yale University, the NASA Ames Research Center, and the University of Toronto with support from the Office of Naval Research (N00014–91–J–1577), the National Aeronautics and Space Administration (NGT–50749), and the Natural Sciences and Engineering Research Council of Canada. M. B. also thanks P. Anandan and Allan Jepson for their help, comments, and encouragement.

A. R. was supported by grants from AFOSR (AFOSR-90-0224) and DARPA-ONR (N00014--92-J-4048). A. R. thanks Rama Chellappa, Eric Mjolsness, Gene Gindi and Davi Geiger for some helpful comments.

References

[1] Proc. Int. Workshop on Robust Computer Vision, Seattle, WA, October 1990.

[2] A. E. Beaton and J. W. Tukey. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16:147–185, 1974.

[3] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, August 1988.

[4] P. J. Besl, J. B. Birch, and L. T. Watson. Robust window operators. In *Proc. Int. Conf. on Comp. Vision, ICCV-88*, pages 591–600, 1988.

[5] M. J. Black. *Robust Incremental Optical Flow.* PhD thesis, Yale Univeristy, New Haven, CT, 1992. Research Report YALEU/DCS/RR-923.

[6] M. J. Black and P. Anandan. A model for the detection of motion over time. In *Proc. Int. Conf. on Computer Vision, ICCV-90*, pages 33–37, Osaka, Japan, December 1990.

[7] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. Computer Vision and Pattern Recognition, CVPR-91*, pages 296–302, Maui, Hawaii, June 1991.

[8] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. Int. Conf. on Computer Vision, ICCV-93*, pages 231–236, Berlin, Germany, May 1993.

[9] A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press, Cambridge, Massachusetts, 1987.

[10] M. Blauer. Image smoothing with shape invariance and L_1 curvature constraints. In *Proceedings SPIE*, Boston, Mass, November 1991.

[11] N. E. Campbell. Robust procedures in multivariate analysis I: Robust covariance estimation. *Appl. Statist.*, 29(3):231–237, 1980.

[12] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[13] D. S. Chen and B. G. Schunck. Robust statistical methods for building classification procedures. In *Proc. Int. Workshop on Robust Computer Vision*, pages 72–85, Seattle, WA, October 1990.

[14] P. B. Chou and C. M. Brown. The theory and practice of Bayesian image labeling. *Int. Journal of Computer Vision*, 4(3):185–210, 1990.

[15] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1):39–55, January 1987.

[16] W. Förstner. Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision. *Computer Vision Graphics and Image Processing*, 40:273–310, 1987.

[17] D. Geiger and F. Girosi. Parallel and deterministic algorithms from MRFs: Surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5), May 1991.

[18] D. Geiger and R. A. M. Pereira. Learning how to teach or selecting minimal surface data. In J. E. Moody, S. J. Hanson, and R. P. Lippmann, editors, *Advances in Neural Information Processing Systems, 4*, pages 364–371, 1992.

[19] D. Geiger and A. Yuille. A common framework for image segmentation. *International Journal of Computer Vision*, 6(3):227–243, 1991.

[20] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):609–628, July 1990.

[21] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):376–383, March 1992.

[22] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[23] S. Geman and D. E. McClure. Statistical methods for tomographic image reconstruction. In *Proceedings of the 46th Session of the ISI, Bulletin of the ISI*, 1987.

[24] J. Hadamard. *Lectures on the Cauchy Problem in Linear Partial Differential Equations.* Yale University Press, New Haven, CT, 1923.

[25] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, New York, NY, 1986.

[26] J. G. Harris, C. Koch, E. Staats, and J. Luo. Analog hardware for detecting discontinuities in early vision. *Int. Journal of Comp. Vision*, 4(3):211–223, June 1990.

[27] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction for Poisson data using Gibbs priors. *IEEE Transactions on Medical Imaging*, vol. MI-8(2):194–202, June 1989.

[28] G. Hinton. *Relaxation and its Role in Vision.* PhD thesis, Univeristy of Edinburgh, 1978.

[29] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, August 1981.

[30] P. J. Huber. *Robust Statistics.* John Wiley and Sons, New York, NY, 1981.

[31] A. Jepson and M. J. Black. Mixture models for optical flow computation. In *Proc. Computer Vision and Pattern Recognition, CVPR-93*, pages 760–761, New York, June 1993.

[32] J. Konrad and E. Dubois. Multigrid Bayesian estimation of image motion fields using stochastic relaxation. In *Int. Conf. on Computer Vision*, pages 354–362, 1988.

[33] R. Kumar and A. R. Hanson. Analysis of different robust methods for pose refinement. In *Proc. Int. Workshop on Robust Computer Vision*, pages 167–182, Seattle, WA, October 1990.

[34] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3(1):73–102, 1989.

[35] L. Lui, B. G. Schunck, and C. C. Meyer. On robust edge detection. In *Proc. Int. Workshop on Robust Computer Vision*, pages 261–286, Seattle, WA, October 1990.

[36] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *J. of the American Statistical Assoc.*, 82(397):76–89, March 1987.

[37] G.J. McLachlan and K.E. Basford. *Mixture Models: Inference and Applications to Clustering.* Marcel Dekker Inc., N.Y., 1988.

[38] P. Meer, D. Mintz, and A. Rosenfeld. Robust recovery of piecewise polynomial image structure. In *Proc. Int. Workshop on Robust Computer Vision*, pages 109–126, Seattle, WA, October 1990.

[39] P. Meer, D. Mintz, and A. Rosenfeld. Robust regression methods for computer vision: A review. *International Journal of Computer Vision*, 6(1):59–70, 1991.

[40] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. Computer Vision and Pattern Recognition, CVPR-85*, pages 22–25, San Francisco, 1985.

[41] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(2):220–228, March 1987.

[42] R. Nevatia and K. R. Babu. Linear feature extraction and description. *Computer Graphics and Image Processing*, 13:257–269, July 1980.

[43] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[44] A. Rangarajan and R. Chellapa. A continuation method for image estimation using the adiabatic approximation. In R. Chellapa and A. Jain, editors, *Markov Random Fields: Theory and Applications*. Academic Press, Inc., 1993.

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

[46] B. G. Schunck. Image flow segmentation and estimation by constraint line clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(10):1010–1027, October 1989.

[47] B. G. Schunck. Robust estimation of image flow. In *Proceedings SPIE*, November 1989.

[48] B. G. Schunck. Robust computational vision. In *Proc. Int. Workshop on Robust Computer Vision*, pages 1–18, Seattle, WA, October 1990.

[49] D. Shulman and J. Hervé. Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, pages 81–85, Irvine, CA, March 1989. IEEE Computer Society Press.

[50] S. S. Sinha and B. G. Schunck. A two-stage algorithm for discontinuity-preserving surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(1):36–55, January 1992.

[51] G. Strang. *Linear Algebra and its Applications*. Academic Press, New York, 1976.

[52] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4):413–424, July 1986.

[53] Y. Tian and M. Shah. MRF-Based motion estimation and segmentation. Technical Report CS-TR-92-13, University of Central Florida, Orlando, FL, July 1992.

[54] A. P. Tirumalai, B. G. Schunck, and R. C. Jain. Robust dynamic stereo for incremental disparity map refinement. In *Proc. Int. Workshop on Robust Computer Vision*, pages 412–434, Seattle, WA, October 1990.

[55] J. Y. A. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proc. Computer Vision and Pattern Recognition, CVPR-93*, pages 361–366, New York, June 1993.

[56] J. Weng and P. Cohen. Robust motion and structure estimation using stereo vision. In *Proc. Int. Workshop on Robust Computer Vision*, pages 367–388, Seattle, WA, October 1990.

A Catalog of Estimators and Outlier Processes

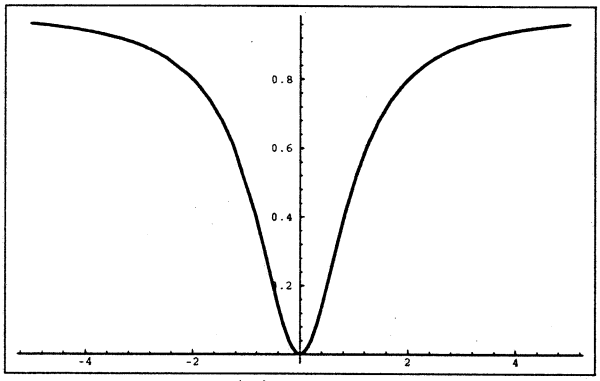
This catalog contains some of the common estimators found in the computer vision literature and their outlier-process forms. It is not meant to be an exhaustive list since new outlier processes can be easily derived using the method described in the paper While some estimators do not admit outlier processes (eg. Huber's minimax and Andrews' sine), many do, including a number of estimators used in continuation methods. The Lorentzian and the GNC outlier processes are derived in the text and are not repeated here.

One of the simplest estimators is that used by Geman and McClure [23] for image reconstruction (Figure 27). Hebert and Leahy [27] experiment with this same estimator, as well as the Lorentzian estimator, in their work on medical image reconstruction.

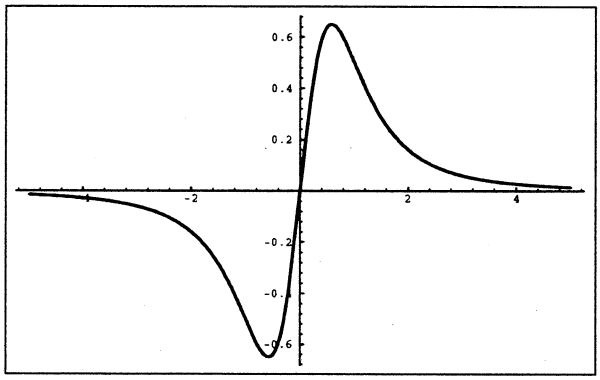
Another estimator which has been used more in the robust statistics literature than in computer vision, is Tukey's biweight (Figure 28). The recovery of the analog outlier process for Tukey's estimator illustrates the deep connections between weak continuity methods in computer vision and traditional robust estimation techniques.

Leclerc [34] derived another estimator by starting with a different formulation of the smoothness term. Consider the following objective function:

$$E(\mathbf{u},\mathbf{d}) = \sum_{s\in S}(u_s - d_s)^2 + \beta \sum_{t\in\mathcal{G}_s}(1-\delta(u_s-u_t)), \tag{44}$$



$\rho(x)$

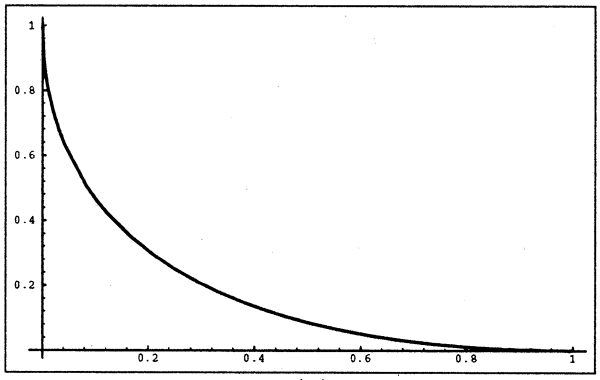


$\psi-$function

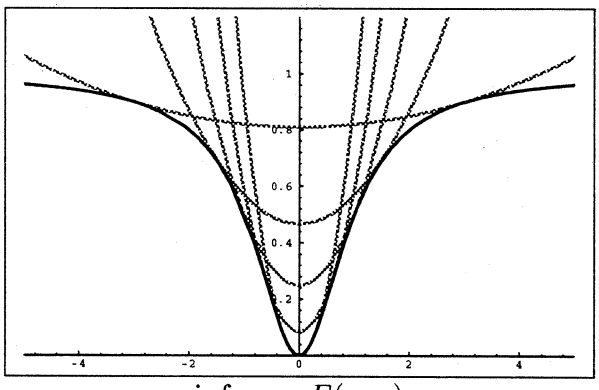
Geman and McClure:

$$
\begin{aligned}
\rho(x) &= \frac{x^2}{1+x^2} \\
\psi(x) &= \frac{2x}{(1+x^2)^2} \\
\phi(w) &= \frac{w}{1+w} \\
\phi'(w) &= \frac{1}{(1+w)^2} \\
\phi''(w) &= \frac{-2}{(1+w)^3} \\
(\phi')^{-1}(z) &= -1+\frac{1}{\sqrt{z}} \\
\Psi(z) &= (-1+\sqrt{z})^2 \\
E(x,z) &= x^2\, z + \Psi(z)
\end{aligned}
$$

$0 \le z \le 1$

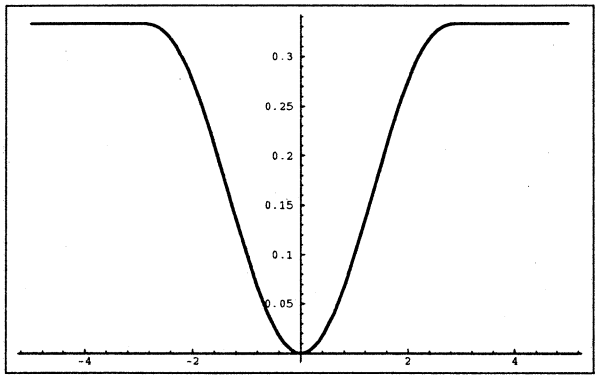


$\Psi(z)$

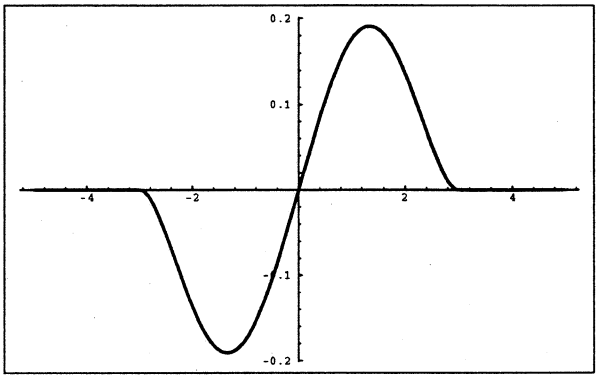


$\inf_{0\le z\le 1} E(x,z)$

Figure 27: Geman and McClure Estimator



$\rho(x)$



$\psi-$function

Tukey's Biweight:

$$\rho(x,\sigma) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2\sigma^2} + \frac{x^6}{6\sigma^4} & |x| \le \sigma \\ \frac{\sigma^2}{6} & \text{otherwise} \end{cases}$$

$$\psi(x,\sigma) = \begin{cases} x(1-(x/\sigma)^2)^2 & |x| \le \sigma, \\ 0 & \text{otherwise} \end{cases}$$

$$\phi(w) = \begin{cases} w - w^2 + \frac{1}{3}w^3 & w < 1 \\ 1 & \text{otherwise} \end{cases}$$

$$\phi'(w) = \begin{cases} 1 - 2w + w^2 & w < 1 \\ 0 & \text{otherwise} \end{cases}$$

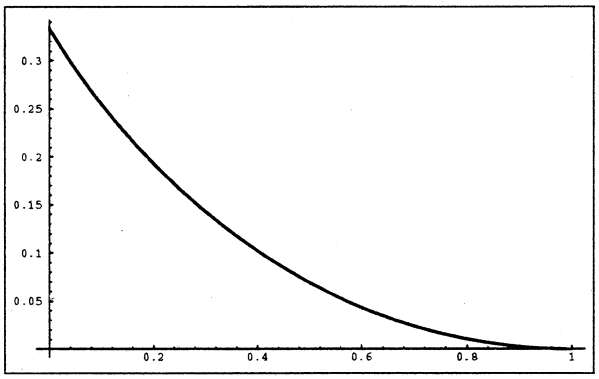
$$\phi''(w) = \begin{cases} 2w - 2 & w < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$(\phi')^{-1}(z) = 1 - \sqrt{z}$$

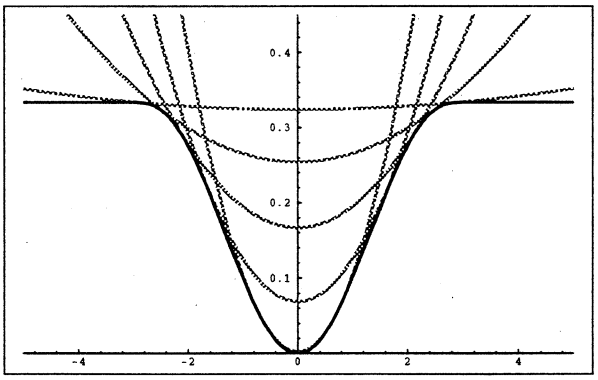
$$\Psi(z) = \frac{1}{3} - z + \frac{2}{3}z^{\frac{3}{2}}$$

$$E(x,\sigma,z) = \frac{x}{\sigma}^2 + \Psi(z)$$

$$0 \le z \le 1$$

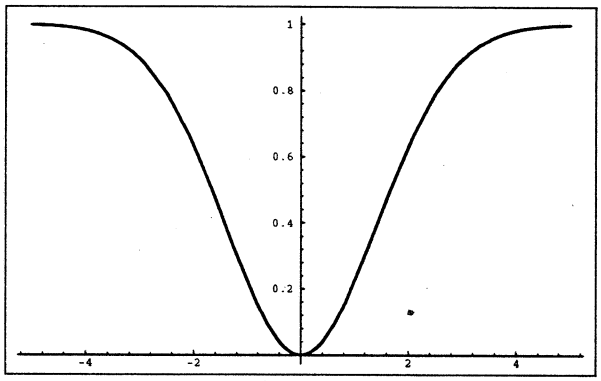


$\Psi(z)$

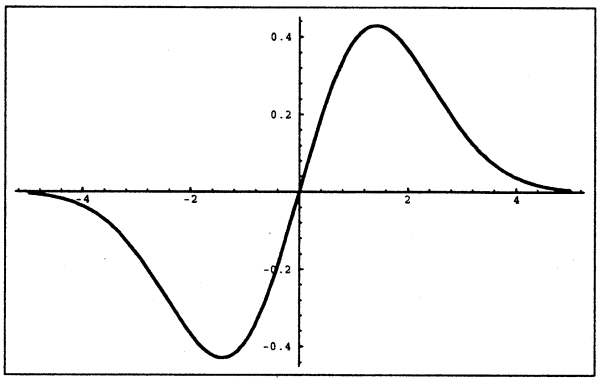


$\inf_{0 \le z \le 1} E(x,\sigma,z)$

Figure 28: Tukey's Biweight Estimator



$\rho(x)$

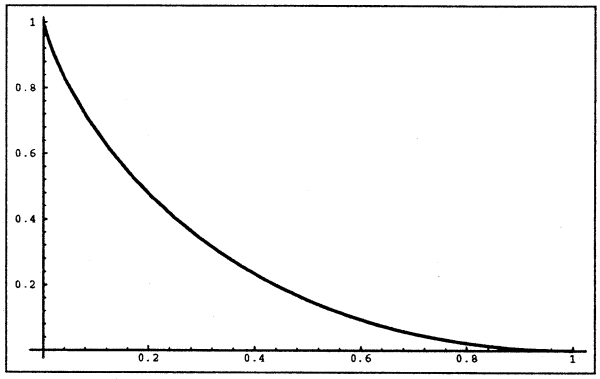


$\psi-$function

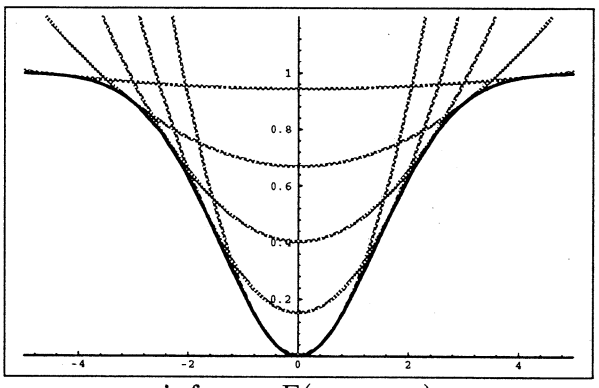
Leclerc:

$$
\begin{aligned}
\rho_{\eta,\sigma}(x) &= 1 - e^{-\frac{x^2}{(\eta\sigma)^2}} \\
\psi_{\eta,\sigma}(x) &= \frac{2x}{(\eta\sigma)^2} e^{-\frac{x^2}{(\rho\sigma)^2}} \\
\phi(w) &= 1 - e^{-w} \\
\phi'(w) &= e^{-w} \\
\phi''(w) &= -e^{-w} \\
(\phi')^{-1}(z) &= -\log z \\
\Psi(z) &= z \log z - z + 1 \\
E(x,\eta,\sigma,z) &= (\frac{x}{\eta\sigma})^2 z + \Psi(z)
\end{aligned}
$$

$$0 < z \leq 1$$



$\Psi(z)$



$\inf_{0<z\leq 1} E(x,\eta,\sigma,z)$

Figure 29: Leclerc Estimator

where β is a constant and $\delta(x)$ is the Kronecker delta function:

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

Leclerc developed a continuation strategy by approximating the delta function with a sequence of Gaussians of decreasing variance $\eta\sigma$, where η is a control parameter that is decreased from infinity to zero (see Figure 29). As η goes to zero, the estimator approaches the delta function.

Geiger and Girosi [17] derive an approximation to the truncated quadratic based on a *mean-field theory*. Instead of minimizing over the binary line processes as done by Blake and Zisserman, they integrate them out and derive the estimator in Figure 30. The parameter, β, is the same as the inverse of the temperature parameter used in simulated annealing and can be used to control the shape of the function. For small β the function behaves like the quadratic while, as β goes to infinity, the function approaches the shape of the truncated quadratic. Embedded in a continuation method, controlling β provides a deterministic annealing scheme.

Geiger and Girosi show that the mean-field estimator results from integrating out the *binary* line process. What we see in Figure 30 is that we can recover an equivalent *analog* process defined by the penalty function, Ψ. The range of the analog process however is not in $[0, 1]$, but depends on β. By recovering this analog process we can use the deterministic annealing scheme of the mean-field approach in situations where we desire spatial interactions among the outlier processes.

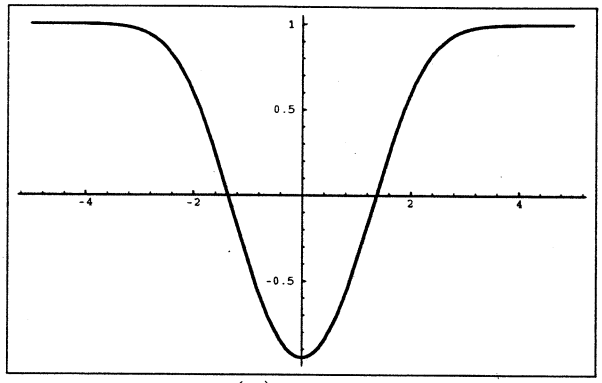
Geman and Reynolds [21] use the estimator in Figure 31 for reconstructing degraded images. In this case, $\phi'(0) = \infty$, so we can not derive an outlier process between 0 and 1; instead we take $z \geq 0$. The direct application of the mechanism for recovering the outlier process yields a rather complicated looking penalty term. Geman and Reynolds [21] propose a different parameterization of the same function:

$$E(x, z) = \alpha(z)x^2\, z + \beta(z).$$

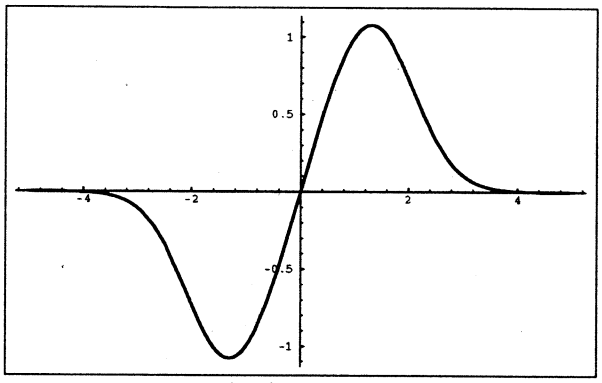
where they take:

$$\alpha(z) = \frac{z^{3/2}}{2(1 - z^{1/2})}, \quad \beta(z) = \frac{z - 3z^{1/2}}{2},$$

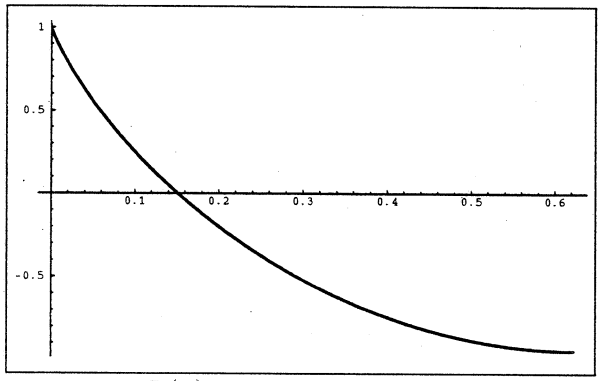
and show that their estimator is the infimum of this family of quadratics.



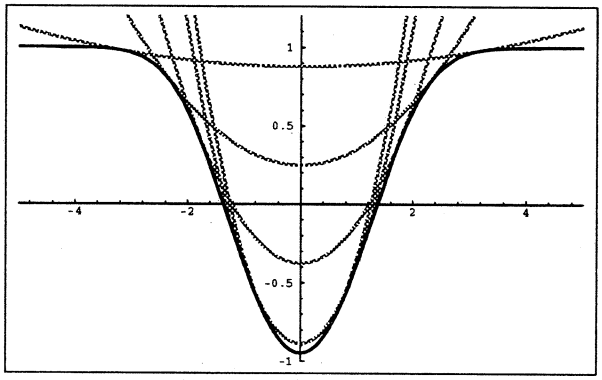
$\rho(x)$



$\psi-$function



$\Psi(z)$

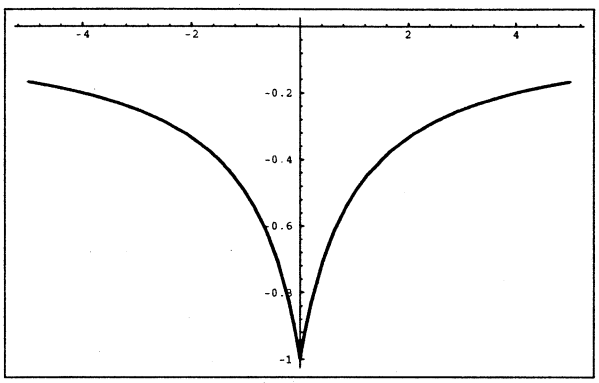


$\inf_{0\le z\le 1} E(x,\alpha,\beta,\lambda,z)$

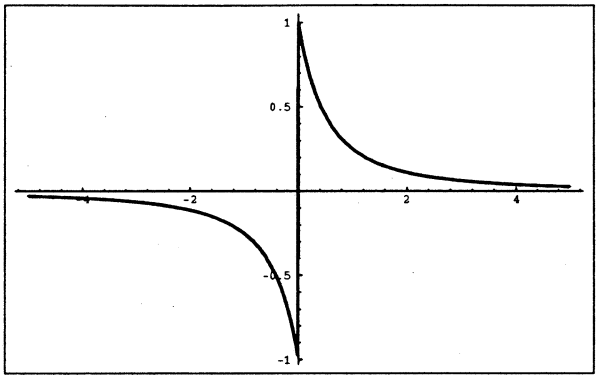
MFT:

$$
\begin{aligned}
\rho(x,\alpha,\beta,\lambda) &= -\frac{1}{\beta}\log(e^{-\beta(\lambda^2 x^2)} + e^{-\beta\alpha}) \\
\psi(x,\alpha,\beta,\lambda) &= \frac{2e^{\beta(\alpha-\lambda^2x^2)}\lambda^2 x}{1+e^{\beta(\alpha-\lambda^2x^2)}} \\
\phi(w,\alpha,\beta) &= -\frac{1}{\beta}\log(e^{-(\beta w)} + e^{-\beta\alpha}) \\
\phi'(w) &= \frac{1}{1+e^{\beta(w-\alpha)}} \\
\phi''(w) &= -\frac{\beta\, e^{\beta(w-\alpha)}}{(1+e^{\beta(w-\alpha)})^2} \\
(\phi')^{-1}(z) &= \alpha + \frac{1}{\beta}\log(\frac{1-z}{z}) \\
\Psi(z,\alpha,\beta) &= \alpha(1-z) + \frac{1}{\beta}\left((1-z)\log(1-z) + z\log(z)\right) \\
E(x,\alpha,\beta,\lambda,z) &= \lambda^2 x^2 z + \Psi(z,\alpha,\beta) \qquad 0 < z \le \frac{1}{1+e^{-\beta\alpha}}
\end{aligned}
$$

Figure 30: MFT Estimator



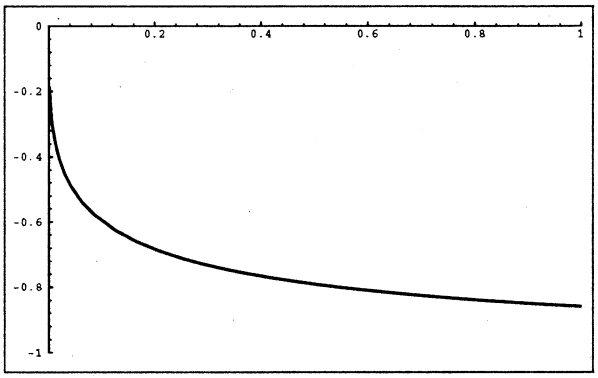
$\rho(x)$



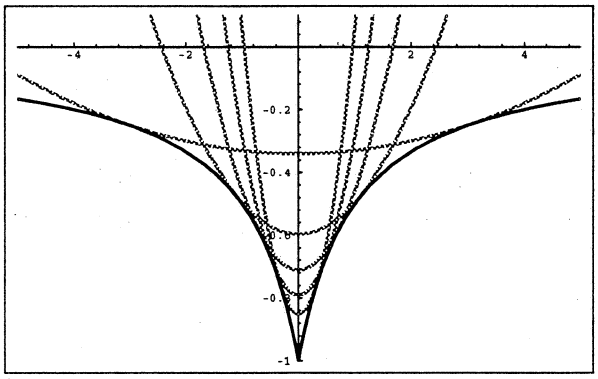
$\psi-$function

Geman and Reynolds:

$$
\begin{aligned}
\rho(x) &= \frac{-1}{1+|x|} \\
\psi(x) &= \begin{cases} -\frac{1}{(1-x)^2} & x \leq 0 \\ \frac{1}{(1+x)^2} & x > 0 \end{cases} \\
\phi(w) &= \frac{-1}{1+\sqrt{w}}, w \geq 0 \\
\phi'(w) &= \frac{1}{2\sqrt{w}(1+\sqrt{w})^2} \\
\phi''(w) &= \frac{-(1+3\sqrt{w})}{4(1+\sqrt{w})^3 w^{\frac{3}{2}}} \\
z &= \phi'(w) \\
\Psi(z) &= \phi((\phi')^{-1}(z)) - z(\phi')^{-1}(z) \\
E(x,z) &= x^2 + \Psi(z), \quad z \geq 0
\end{aligned}
$$



$\Psi(z)$



$\inf_{0 \leq z \leq 1} E(x, \sigma, z)$

where,

$$
(\phi')^{-1}(z) = \frac{2}{3} + \frac{2(-6+z)}{3z^{\frac{1}{3}}\left(27 - 144z - 8z^2 + 3^{\frac{3}{2}}\sqrt{(1+4z)^2(27+8z)}\right)^{\frac{1}{3}}} + \frac{\left(27 - 144z - 8z^2 + 3^{\frac{3}{2}}\sqrt{(1+4z)^2(27+8z)}\right)^{\frac{1}{3}}}{6z^{\frac{2}{3}}}
$$

Figure 31: Geman and Reynolds Estimator