Robust Parameter Estimation in Computer Vision

Charles V. Stewart
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, New York 12180–3590
stewart@cs.rpi.edu

January 19, 1999

1 Introduction

The goal of computer vision algorithms is to extract geometric, photometric and semantic information from image data. This may include the position and identity of an object [2, 8, 20, 38, 49, 59] the motion of a camera attached to a car or an autonomous vehicle [1, 6, 18], the geometry of object surfaces [10, 14, 71], or the transformations necessary to build a large composite image (a mosaic) from a series of overlapping images of the same scene [36, 70]. The processes used to extract this information each require some form of parameter estimation — to describe intensity edge curves, motion models, surface normals and curvatures, and Euclidean, affine and projective transformation models [24].

An intensity image typically contains 250,000 pixels or more, with each pixel storing an 8-bit grey level or three 8-bit components of a color vector. Range images, where pixels record X, Y, Z scene coordinates as measured by special purpose sensors, may contain as many measurement vectors. Each pixel measurement is subject to small scale random variations (noise) caused by the processes of sensing and digitization. This huge volume of data implies that parameter estimation techniques in vision are heavily over-constrained, even for problems where low-level feature extraction such as edge detection is used as a pre-processing step. This in turn implies that parameter estimation problems in vision should be solved by least-squares or, more generally, maximum likelihood estimation (MLE) techniques.

Unfortunately, computer vision data are rarely drawn from a single statistical population as required for effective use of MLE. Intensity and range images may contain light and depth measurements from multiple scene surfaces. A moving camera, which induces apparent image motion, may also image an independently moving object, inducing a second apparent motion. In both cases, multiple structures (populations) are represented in the image data. Additionally, some image data may be measurements that are difficult to assign to any population. These data are gross errors ("outliers") which may be caused by specular highlights, saturation of sensors, or mistakes in feature extraction techniques such as edge and corner detectors. It is important to note that these gross errors may be arbitrarily large and therefore can not be "averaged out", as is typically done with small-scale noise. As illustrated in Figure 1, when image data are drawn from multiple populations or outliers, application of MLE can produce nonsense results. Much work in computer vision, therefore, has been directed at separating (segmenting) data into distinct populations prior to or during parameter estimation. Complete segmentation, however, is not possible without parameter estimation, because the process of assigning data points to populations depends, at least partially, on the parameters describing the structure of each population. A simple illustration of this is that it is impossible to know which points belong to a line until the line parameters are known.

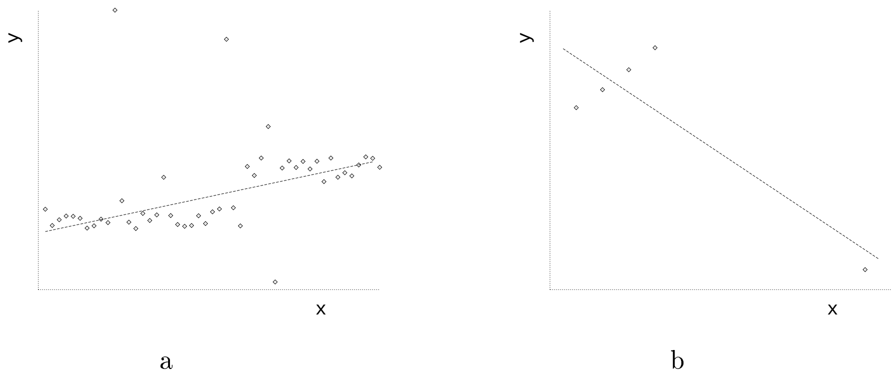


Figure 1: Simple examples demonstrating the effects of (a) multiple image structures (data populations) plus outliers and (b) a single outlier (a "leverage point") on linear least-squares fits. If the example in (a) is thought of as the plot of the cross-section of a range image (Section 3), then the x coordinate represents image position, the y coordinate represents the depth measurement, and the two populations correspond to measurements from two different surfaces, one closer to the sensor and perhaps occluding the other.

This difficulty has sparked growing interest by the computer vision community in the use of robust estimation techniques, which have been developed over the past 25 years in both the computer vision and the statistics literatures [25, 27, 34, 35, 52, 62]. These techniques are attractive because they are specifically designed to accommodate data from multiple populations when estimating the parameters describing the dominant population. Ideally, the parameters estimated should not differ substantially from those estimated via MLE for the dominant population in isolation.

This paper provides a tutorial introduction to robust parameter estimation in computer vision. The paper starts with a summary of commonly used robust estimation techniques, and then describes how they have been applied and extended. Three applications are considered in detail. The first is estimating the parameters describing low-order surface geometry from range data [7]. The second is fundamental matrix estimation — the problem of using corresponding points to establish the relationship between two different images of a scene taken with uncalibrated cameras [50, 80]. The third problem is building a "mosaic" image of a human retina by matching and combining overlapping retinal images into a single, larger image [5]. The last two problems are closely related.

In thinking about the techniques presented, it is important for the reader to note that robust estimators are not necessarily the only or even the best technique that can be used to solve the problems caused by outliers and multiple populations (structures) in all contexts. Specialized heuristics for handling occasional outliers appear throughout the computer vision literature, and in some cases, most notably when estimating surfaces from raw range or intensity images, multiple populations may be treated as an edge detection or segmentation problem (see textbook discussions in [28] for example). On the other hand, since robust estimation techniques have been designed to handle outliers and multiple populations and since these problems are pervasive, knowledge of robust estimation and its limitations is important in addressing parameter estimation problems in computer vision.

2 Robust Estimation

The first step in describing robust estimators is to state more clearly what is meant by robustness. Several measures of robustness are used in the literature. Most common is the *breakdown point* [62] — the minimum fraction of outlying data that can cause an estimate to diverge arbitrarily far from the true estimate. For example, the breakdown point of least-squares is 0 because one bad point can be used to move the least-squares fit arbitrarily far from the true fit. The theoretical maximum breakdown point is 0.5 because when more than half the data are outliers they can be arranged so that a fit through them will minimize the estimator objective function.

A second measure of robustness is the *influence function* [27, 34] which, intuitively, is the change in an estimate caused by insertion of outlying data as a function of the distance of the data from the (uncorrupted) estimate. For example, the influence function of the least squares estimator is simply proportional to the distance of the point from the estimate. To achieve robustness, the influence function should tend to 0 with increasing distance.

Finally, though not a measure of robustness, the efficiency of a robust estimator is also significant.[1] This is the ratio of the minimum possible variance in

[1]There is a potential confusion between the statistical notion of efficiency and the computational notation of efficiency, which is associated with the algorithm implementing a robust estimator. Where the meaning is not clear from the context, the phrases "statistical efficiency"

an estimate to the actual variance of a (robust) estimate [47], with the minimum possible variance being determined by a target distribution such as the normal (Gaussian) distribution. Efficiency clearly has an upper bound of 1.0. Asymptotic efficiency is the limit in efficiency as the number of data points tends to infinity. Robust estimators having a high breakdown point tend to have low efficiency, so that the estimates are highly variable and many data points are required to obtain precise estimates.

Robust estimators are usually defined and analyzed in terms of either linear regression or estimation of univariate or multivariate location and scatter [27, 34, 62]. (Most computer vision problems requiring robust estimation are similar to regression problems.) To set a general context, let $X = \{\mathbf{x}_i\}$ be a set of data points (vectors) and let $\mathbf{a}$ be a k dimensional parameter vector to be estimated. The objective functions used in robust estimation, like those used in MLE, are defined in terms of an error distance or residual function, denoted by $r_{i,\mathbf{a}} = r(\mathbf{x}_i; \mathbf{a})$. Ideally this should be a true geometric distance — e.g. the Euclidean distance between a point $\mathbf{x}_i$ and the curve determined by $\mathbf{a}$ — or better yet, a Mahalanobis distance if the covariance matrix of $\mathbf{x}_i$ is known.

2.1 M-Estimators

While many variations on robust estimation have been proposed in the statistics literature, the two main techniques used in computer vision are M-estimators and least-median of squares. M-estimators are generalizations of maximum likelihood estimators and least-squares [27, 34]. In particular, the M-estimate of $\mathbf{a}$ is

$$\hat{\mathbf{a}} = \operatorname*{argmin}_{\mathbf{a}} \sum_{\mathbf{x}_i \in X} \rho(r_{i,\mathbf{a}}/\sigma_i) \tag{1}$$

where $\rho(u)$ is a robust loss function which grows subquadratically and is monotonically non-decreasing with increasing $|u|$. Also, σ_i^2 is the variance (scale) associated with the scalar value $r_{i,\mathbf{a}}$. Constraints on $\hat{\mathbf{a}}$ are easily incorporated.

The minimization in (1) is solved by finding $\mathbf{a}$ such that

$$\sum_{\mathbf{x}_i \in X} \psi(r_{i,\mathbf{a}}/\sigma_i) \frac{dr_{i,\mathbf{a}}}{d\mathbf{a}} \frac{1}{\sigma_i} = 0.$$

where $\psi(u) = \rho'(u)$. A common, though certainly not the only, next step [29, 32] [34, pages 179-192] is to introduce a weight function w where $w(u) \cdot u = \psi(u)$ and solve

$$\sum_{\mathbf{x}_i \in X} w(r_{i,\mathbf{a}}/\sigma_i) \frac{1}{\sigma_i^2} \frac{dr_{i,\mathbf{a}}}{d\mathbf{a}} r_{i,\mathbf{a}} = 0. \tag{2}$$

This leads to a process known as "iteratively reweighted least-squares" (IRLS) which alternates steps of calculating weights $w_i = w(r_{i,\mathbf{a}}/\sigma_i)$ using the current

---

and "computation efficiency" will be used.

estimate of **a** and solving (2) to estimate a new **a** with the weights fixed. Initial estimates of **a** may be obtained in a variety of manners, including non-robust least-squares or other robust estimators discussed below.

The many M-estimators that have been proposed differ in the shape of the functions $\rho(\cdot)$ and, as a result, $\psi(\cdot)$ and $w(\cdot)$. Three common functions are listed in Table 1, with weight functions plotted in Figure 2. The ψ functions, essentially, are proportional to the influence function. Hence, ψ functions tending to 0 most quickly ("hard redescenders" in the terminology of [32]), such as the Beaton and Tukey function, allow the most aggressive rejection of outliers. This has been found to be important in computer vision problems [68] when outliers have small residual magnitudes such as in the range 4σ to 10σ. Unfortunately, when redescending ψ functions are used the objective function, $\sum_{\mathbf{x}_i \in X} \rho(r_{i,\mathbf{a}}/\sigma_i)$, is non-convex, implying that IRLS will converge at local minima.

Beaton and Tukey [4]	$\rho(u) = \begin{cases} \frac{a^2}{6}\left[1-\left(1-\left(\frac{u}{a}\right)^2\right)^3\right] & \lvert u\rvert \le a \\ \frac{a^2}{6} & \lvert u\rvert > a \end{cases}$	$\psi(u) = \begin{cases} u\left[1-\left(\frac{u}{a}\right)^2\right]^2 & \lvert u\rvert \le a \\ 0 & \lvert u\rvert > a \end{cases}$
Cauchy [32]	$\rho(u) = \frac{b^2}{2}\log\left[1+\left(\frac{u}{b}\right)^2\right]$	$\psi(u) = \frac{u}{1+(u/b)^2}$
Huber [34, Chapter 7]	$\rho(u) = \begin{cases} \frac{1}{2}u^2 & \lvert u\rvert \le c \\ \frac{1}{2}c(2\lvert u\rvert - c) & c < \lvert u\rvert \end{cases}$	$\psi(u) = \begin{cases} 1 & \lvert u\rvert \le c \\ c\frac{u}{\lvert u\rvert} & \lvert u\rvert > c \end{cases}$

Table 1: Three different robust loss functions, $\rho(u)$, and associated ψ functions. The "tuning parameters" a, b and c are often tuned to obtain 95% efficiency [32].

The scale values, σ_i, are used to normalize the error distances and to further weight the contribution of each point when scale varies with i. Scale values may be provided *a priori* from analysis of the sensor or process (such as edge detection) from which the data are obtained. In some cases, scale will not be known in advance and must be estimated from the data. In this case, σ_i is replaced by $\hat{\sigma}$ in the above equations. Equation 1 may be rewritten to jointly estimate **a** and σ [34, Chapter 7]. Alternatively, σ may be estimated from an initial fit prior to IRLS and re-estimated after each of the first few IRLS iterations. Robust scale estimation techniques such as the median absolute deviation [32] (Equation 4) are used. Scale must be fixed, however, before allowing IRLS to converge. The particularly difficult case of σ_i being both unknown and varying with i (heteroscedastic) will not be treated here.

An example of line parameter estimation using an IRLS implementation and the Beaton and Tukey [4] weight function is shown in Figure 2.1(a). Scale was re-estimated after each of the first three iterations and IRLS took ten iterations to converge. Convergence is typically faster with fewer iterations of scale estimation, but the result on this example is less accurate.

The breakdown point of standard M-estimators in regression is 0 because of the possibility of leverage points (Figure 1b) — outliers positioned far from the

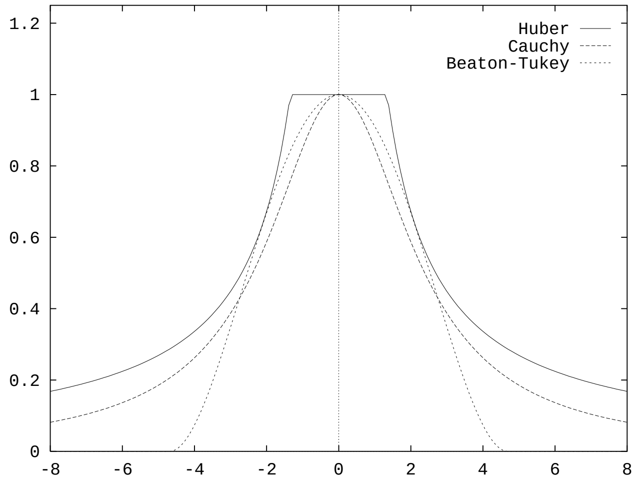


Figure 2: Plots of the weight functions for the robust loss functions given in Table 1. The horizontal axis units are scale-normalized residuals $r_{i,\mathbf{a}}/\sigma_i$.

remainder of the data in the independent variables as well as being far from the fit to the uncorrupted data. A local minimum of (1) will generally occur for a fit rotated to pass through (or near) the leverage point(s), as in Figure 1b. A generalization of M-estimates, GM-estimators (see discussion in [61]), which downgrades the influence of points based on both independent and dependent variables, can have breakdown points as high as $1/(k+1)$, where k is the length of $\mathbf{a}$.

2.2 Least-Median of Squares

Least-median of squares (LMS) is distinguished by having a breakdown point of 0.5, the highest value possible. In the notation here, the LMS estimate [60] is

$$\hat{\mathbf{a}} = \operatorname*{argmin}_{\mathbf{a}} \operatorname*{median}_{\mathbf{x}_i \in X} r^2_{i,\mathbf{a}}. \tag{3}$$

The intuition behind LMS is that up to half the data points can be arbitrarily far from the optimum estimate without changing the objective function value.

Since the median function is not differentiable, alternative search techniques are required. In the special case of fitting regression lines to points in $\Re^2$, computationally efficient algorithms are known for solving (3) exactly [23, 66]. In more general settings, a random sampling technique, developed independently in the computer vision [25] and statistics [60] literatures, is required. The idea is to randomly select a number of k-point subsets of the N data points. A parameter vector, $\mathbf{a}_s$, is fit to the points in each subset, s [58]. Each $\mathbf{a}_s$ is tested as a hypothesized fit by calculating the squared residual distance $r^2_{i,\mathbf{a_s}}$ of each

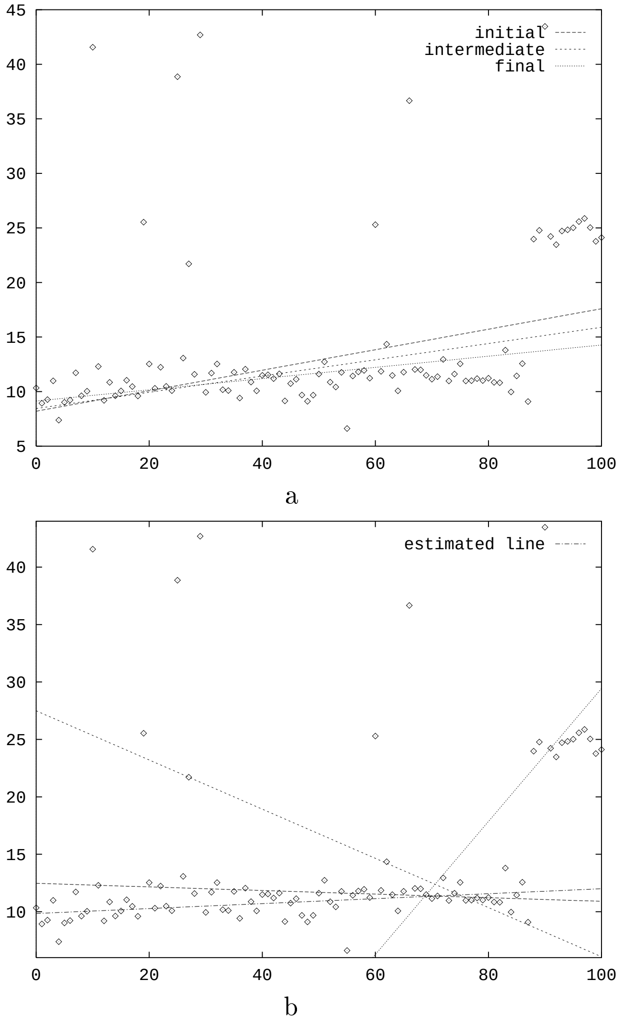


Figure 3: Example robust line estimates: (a) shows initial, intermediate and final estimated lines for an M-estimator using IRLS; (b) shows a number of lines hypothesized and tested during a random sampling implementation of LMS.

of the $N-k$ points in $X-s$ and finding the median. The $\mathbf{a}_s$ corresponding to the smallest median over S subsets is chosen as the estimate, $\hat{\mathbf{a}}$. Overall, this computation requires $O(S\,N)$ time using linear-time median finding techniques [21].

A crucial parameter in this algorithm is S, the number of subsets. S must be large enough to have a high probability of including at least one subset containing all "good" data points — points that are measurements from the structure of interest. If p is the minimum fraction of good points, then to a first approximation, the probability of a subset containing all good points is p^k. From this, it is easy to see that the probability that at least one of the S subsets contains all good points is

$$P_g = 1 - (1-p^k)^S.$$

By choosing a desired value of P_g, the minimum value of S can be found. As an example with $P_g = 0.99$, if $k=3$ and $p=0.5$ then $S=35$; if $k=6$ and $p=0.6$ then $S=97$; and if $k=6$ and $p=0.5$ then $S=293$. Clearly, the required number of subsets increases dramatically with increasing k and decreasing p. An example of line parameter estimation using a random sampling implementation of LMS is shown in Figure 2.1(b). Several hypothesized and tested lines are shown, including the final estimate.

The "Median Absolute Deviation" scale estimate may be obtained from the estimate $\hat{\mathbf{a}}$ and used to gather data points to refine the fit. The scale estimate is

$$\hat{\sigma} = 1/\Phi^{-1}(0.75)\big(1+\frac{5}{N-k}\big)\sqrt{\operatorname*{median}_{\mathbf{x}_i \in X-s^*} r^2_{i,\hat{\mathbf{a}}}}, \tag{4}$$

where s^* is the subset used to form $\hat{\mathbf{a}}$, Φ^{-1} is the inverse of the cumulative normal distribution, and $1+5/(N-k)$ is a finite sample correction factor [62]. The result is an unbiased estimate of $\hat{\sigma}$, which means that if all N points are sampled from a normal distribution with variance σ^2, then $\hat{\sigma} \to \sigma$ as $N \to \infty$. The estimated scale may be used to refine $\hat{\mathbf{a}}$: the data points such that $r^2_{i,\hat{\mathbf{a}}} < (\theta\hat{\sigma})^2$, for constant θ typically around 2.5, can be identified and used to calculate a least-squares estimate of $\mathbf{a}$. This is important because LMS has low statistical efficiency.

LMS may be generalized in a number of ways based on the order statistics of the residual distances [61]. Some of these are discussed later.

2.3 Requirements for Robust Estimation

The nature of computer vision problems alters the performance requirements of robust estimators in a number of ways.

- The optimum breakdown point of 0.5 must be surpassed in some domains. While this is not possible in general, due to the definition of breakdown, it is in particular instances due to problem specific considerations. For example, fewer than half the points are from any one line in Figure 1(a), but each line is clearly visible. A robust estimator should not fail to estimate a line that approximates one of these two.

- Having enough inliers to satisfy the breakdown limit will not guarantee satisfactory results since it only guarantees that the outliers do not cause the estimate to stray arbitrarily far. For example, the estimated line shown in Figure 1(a), which is approximately the LMS estimate when just more than 50% of the points are from the lower line [68], does not represent a breakdown of the estimator, but it is clearly wrong.

- The low breakdown point of M-estimators can be misleading. A local optimum fit passing near leverage points (Figure 1(b)) can be avoided by proper (robust) initialization of IRLS iterations. (This issue is explored in [46].) Use of an M-estimator having an influence function that tends to 0 quickly, so that the weight of the leverage points is 0, will then cause outliers to be ignored.

- The emphasis is often, though not always, on tolerating large numbers of outliers rather than on statistical efficiency. One effect of this is on the choice of tuning parameters for M-estimators — they are often set much lower [9, 68] than suggested by efficiency experiments [32]. This tends to "narrow" the weight functions shown in Figure 2, reducing the influence of points on the tail of the inlier (noise) distribution — hence the loss of efficiency — but making the estimator less sensitive to outliers near the tail of the distribution.

- In some cases, computational efficiency can be of utmost concern. This could render impractical random sampling algorithms generating large numbers of samples.

2.4 Robust Estimation Techniques Developed In Computer Vision

Several robust estimation techniques have been developed in computer vision, either independent of the statistics literature or as an extension of techniques appearing there.[2] The two most important robust techniques developed independently in computer vision are Hough transforms [35, 42] and RANSAC [25, 58]. A Hough transform is a voting technique. The domain (the "parameter space") of the parameter vector $\mathbf{a}$ to be estimated is discretized, one voting bin is allocated per discrete parameter vector, and each data point $\mathbf{x}_i$ "votes" for parameter vectors $\mathbf{a}$ — i.e. the associated voting bin is incremented — for which the fit residual $r_{\mathbf{x}_i;\mathbf{a}}$ is small enough. The parameter space is searched after voting is complete to locate maxima or to locate and analyze clusters of large numbers of votes. In the former case, the objective function underlying the Hough transform is quite similar to that of an M-estimator and shares some

[2] In addition, many published algorithms include heuristics for eliminating outliers; these are too numerous and too application specific to be reviewed here.

of its limitations [68]. An advantage of Hough Transforms in general is that a thorough sampling of the parameter space is obtained. A disadvantage is that the size of the voting space is exponential in the number of parameters, rendering it impractical for many applications.

RANSAC [25, 58] has similarities to both M-estimators and LMS. Like LMS, it is a minimal subset random sampling search technique, predating LMS by three years. The objective function to be maximized, however, is the number of data points (inliers) having absolute residuals smaller than a predefined value, θ. Equivalently, this may be viewed as minimizing the number of outliers, which may then be viewed as a binary robust loss function that is 0 for small (absolute) residuals, 1 for large absolute residuals, and has a discontinuous transition at θ. Interestingly, both RANSAC and Hough transforms, by virtue of the prespecified inlier band, can be used to find structures formed by substantially fewer than half the data. A cost of this is that small, random structures can also be found [68], implying that careful post-processing analysis of structures estimated using RANSAC or Hough transforms is required.

Several extensions to LMS have been introduced in the computer vision literature [43, 53, 54]. The simplest is to just alter the fraction of inliers used, moving up or down from the 0.5 of the median. In general, reducing the fraction will reduce the breakdown point since a smaller fraction of outliers is required to corrupt the estimate, but this is advisable when there are multiple structures in the data, when there are relatively few true outliers, and when the minimum fraction of data corresponding to a single structure is known.

A more general approach, which doesn't rely on a fixed inlier fraction, is based on realizing that the scale estimate in Equation (4) can be generalized to any inlier fraction p by replacing $\Phi^{-1}(0.75)$ with $\Phi^{-1}(0.5+p/2)$. This gives an unbiased scale estimate for the residuals to the correct fit when these residuals follow a normal distribution. Suppose, however, that the data consist of k points measured from a normal distribution and $N-k$ outliers. Here, at the correct fit, the scale estimate will be an overestimate, but will be approximately constant for $p < k/N$, and will increase substantially for $p > k/N$. Also, the scale estimates will increase, especially for $p < k/N$, as the fit is moved away from the correct one. These observations have been built into two estimators, ALKS [43] and MUSE [53, 54], which are based on finding the estimate and associated fraction p that minimize a measure tied to this scale estimate.

The final extension of LMS, an estimator referred to as MINPRAN, is based on an assumed outlier distribution, which may be known or hypothesized when properties of the sensing process are known [67]. The estimated parameter vector is the one whose chosen inlier set is *least likely* to be a set of outliers over all possible inlier sets. This inlier set is determined from an inlier bound, θ, as in RANSAC, but θ is varied over all possible values to minimize the probability of randomness. Somewhat surprisingly, the computational cost of this is only $O(SN\log N + N^2)$.

3 Surface Estimation from Range Data

The first application of robust estimation in computer vision to be considered here is the problem of estimating the parameters of surfaces from range data. This problem is closely related to regression since errors in range measurements tend to be concentrated along the optical axis of the cameras. This allows the depth measurement to be treated as the only independent variable.

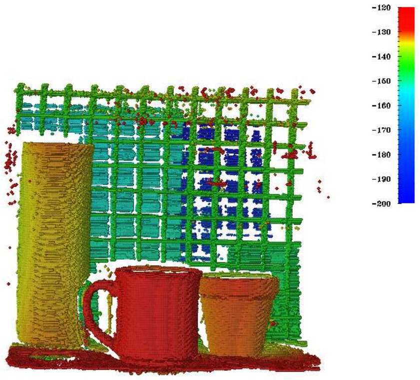
3.1 Range Data and the Estimation Problem

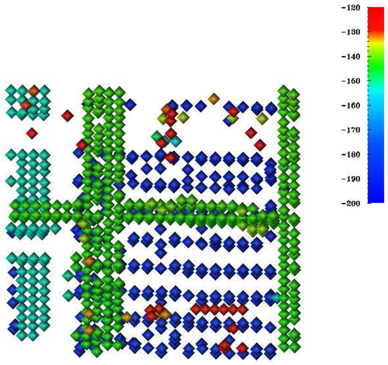
Range data are sets of measurement vectors $(X, Y, Z)^T$ from surfaces in the field of view of a range sensor [10, 57]. Each data set may contain tens or hundreds of thousands of points. There are several different types of range sensors. Many use triangulation, either from two or more cameras at known positions, or by replacing one camera with a light source that shines known planes of light into a scene, measuring $(X, Y, Z)^T$ from reflected positions in the remaining camera(s). Other techniques determine depth by measuring time of flight or phase shifts in radiation reflected from surfaces. In general, the $(X, Y, Z)^T$ measurements are usually recorded at the grid locations of an image, creating what's called a range image (Figure 4a). This range image may be either sparsely or densely populated with points. Accuracy and error of the measured points are around 1/250 to 1/1000 of the depth interval over which measurements are taken, which can vary from the order of centimeters to the order of meters. Some of the measurements may be gross errors (outliers) caused by specular highlights, by thresholding effects, or by inter-reflection of light from multiple surfaces. For more background on range sensors see [39, 57], while for background on the analysis and use of range data in computer vision see [2].

The first goal of range data analysis is to describe the surfaces from which the measurements were taken [14]. The descriptions may be polynomial models such as planar or quadratic models [10, 33, 44, 71], implicit functions such as quadrics and super-quadrics [45], or even functions of unknown form, in which case the surfaces are ultimately defined discretely, at each grid location [13, 26], or using a finite-element model [72]. Two complications make the problem particularly difficult. First, each surface will be of limited spatial extent, implying that unknown surface boundaries may need to be located. Second, there will be multiple surfaces, some abutting or even overlapping in image locations and perhaps close in depth measurements. These difficulties suggest what might be termed a "local-to-global" approach where first low-order polynomials are estimated in small and potentially overlapping image regions and then gathered into more global, complete surface descriptions. Many range image analysis techniques take this approach [9, 15, 22, 33, 44, 65].

3.2 Local Estimation of Surfaces

The goal of local estimation is to describe the surfaces represented in small image regions using planar or quadratic polynomials, identify and eliminate outliers,



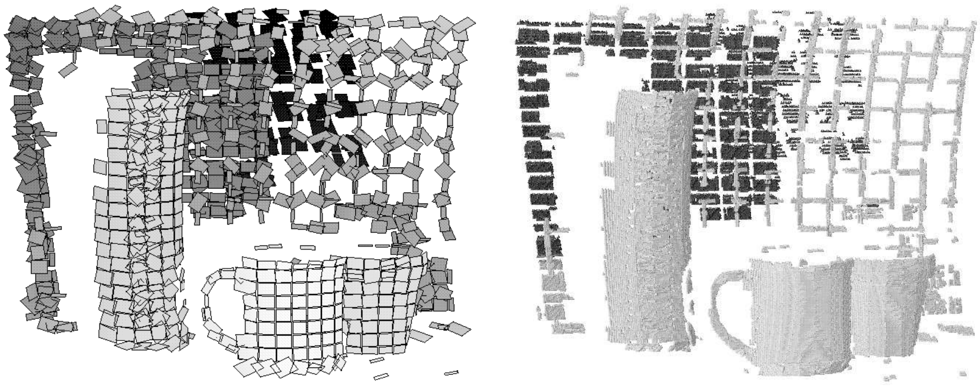


a b

Figure 4: A complicated range image data set. Image (a) shows the original range image using color to encode depth (red being closer to the viewer). The scene includes a tennis ball can, a coffee mug and a planting pot in the foreground, a wire mesh table (on its side) behind these, and two different planar surfaces in the background. Image (b) shows a small window to which local surface estimation might be applied — points are from three different surfaces.

and perhaps generate an initial estimate of surface boundaries. This is easy in regions when only one surface is present, but more difficult when there are two or more surfaces (Figure 4b). In the latter case, it is possible for fewer than 50% of the points to correspond to one surface. Ideally, all surfaces should be described.

The first work in using robust techniques for local estimation was done by Besl et al. [9] using M-estimators to robustly compute planar surface patches, and by Meer et al. [51, 52] using LMS in much the same way. Subsequently, Mirza and Boyer [55] explored a variety of M-estimators for local estimation, concluding that use of those with ψ (and weight) functions redescending to an asymptotic limit of 0 at infinity yielded the best performance on data corrupted with random Gaussian outliers. Sinha and Schunck [65] used LMS locally to eliminate outliers and estimate local variations in surface orientation. Stewart [67] introduced the MINPRAN method discussed earlier to estimate multiple, potentially-overlapping surface patches, and not requiring any patch to contribute more than 50% of the points. Better performance at small magnitude discontinuities was obtained by Miller and Stewart [53, 54] using MUSE and by Lee et al. [43] using ALKS (Section 2.4). The former incorporates techniques from MINPRAN to determine when the MUSE estimate is distinctive enough to be considered non-random; this allows effective determination of the correct number of surface estimates to describe the data set taken from small image regions. An example result using MUSE [53, 54] is shown in Figure 5.



a b

Figure 5: Results (a slightly sideways view) of local surface estimation on the range image data shown in Figure 4. Planar surface patches with bounding rectangles on their inlier sets are shown in (a) and estimated inliers are shown in (b). Outliers have clearly been eliminated and most surfaces accurately estimated. (Use of the rectangular bounding boxes in (a) introduces some artifacts.) Some surfaces from the wire mesh, unfortunately, were missed because they did not occupy a large enough fraction of the inliers to form an acceptable MUSE fit.

ALKS and MUSE produce accurate surface patches, including multiple surface patches, except when the depth change (discontinuity magnitude) between abutting surfaces is less than about 4.5σ [53], where σ is the standard deviation of the noise in the data. Most standard robust estimation techniques, including LMS and M-estimators, fail when the discontinuity magnitude is 7.5σ or higher [68]. (Figure 1a is generated from a model of a discontinuity where these estimators would fail.) The failure in all cases is an estimated surface patch that "bridges" the two real surfaces (see Figure 1a). This bridging surface is the global minimum of the objective function and is therefore not an artifact of the random sampling or IRLS search techniques [68]. In summary, the best robust estimators yield accurate local fits except at extremely small magnitude depth discontinuities, which may be caused by small-scale surface structures.

3.3 Global Estimation of Surfaces

In global estimation of surfaces, which should yield complete descriptions of surfaces and their extents, robust estimation plays a reduced role. It must at least be used in combination with other techniques and is often not used at all in current systems. There are several illuminating reasons for this. First, the appropriate model for the surface is usually unknown, implying that models must be selected automatically [17, 74]. For example, the technique of Boyer et al. [15] grows planar and quadratic surfaces from seed regions — regions of minimum variance in the data — estimated using an M-estimator. Growth is controlled using the same M-estimator, and the model is selected by a modified form of [16]. A related growth technique, which uses prediction intervals to detect finer magnitude discontinuities, is described in [53]. See [33] for a review and empirical comparison of other techniques that address this "segmentation" problem.

A second reason for the reduced role of robust estimation is that even if the surface model is known the extent of the surface must be determined. Without it, robust estimation alone has little chance of success, because a single surface will generally correspond to an extremely small percentage of the data in the entire range image and therefore the potential for "bridging fits", as discussed above, is even more substantial than in local estimation. This is the reason for growing from seed regions, as in [10, 15, 71].

A third reason is the surface may be too complicated to be described in a closed-form and can only be described at discrete locations or using a spline or finite-element model [14]. Here, while robust estimation can help eliminate outliers [65], the more common use is to convert least-squares error norms on the data and on the smoothness of the reconstructed surface into robust error norms based on M-estimator loss functions [12].

3.4 Final Comments on Surface Estimation

Two final comments about the nature of range data and surface estimation are important in order to raise issues that have been incompletely addressed in the literature but are important for practical applications. These affect the use of robust estimation and other techniques. First, most sensors produce quantized data, often with precisions as low as 8 bits. At the extreme, any "noise" in the data is buried in the quantization. This differs significantly from the assumptions under which estimators, robust or not, are typically analyzed. Analytical results should therefore be used with caution when predicting estimator performance on this quantized range data. Second, the variance in (non-quantized) range data may vary spatially and with depth, even across a single surface [53, 54]. If this "heteroscedasticity" were known in advance, it could be built into the M-estimator objective function (as written in Equation 1 above), but prior knowledge is difficult to obtain. The effect of such heteroscedasticity on high-breakdown point estimators such as LMS has not been studied in computer vision.

4 Estimation of the Fundamental Matrix

Consider two different images of the same scene, taken from two different viewpoints. A great deal of information may be gleaned from these images about the relative positions of the camera(s) when the images were taken, about the structure of the scene, and, if the images were taken at different times, about changes in position of any scene objects. This information is encoded in differences between image positions of scene points. By extracting distinctive features, such as intensity corners, in each image, and then matching them across the two images, a set of correspondences can be established. Each correspondence is a pair of image positions, one from each image, and the points in each pair are hypothesized to correspond to the same scene point. These correspondences are used to estimate image-to-image transformations, camera motion, and scene structure. What is actually estimated depends on what's known in advance about the camera, the motion and the scene.

This section and the next consider two such problems of estimating image-to-image transformations based on image-to-image correspondences. In each case the correspondence set is generally contaminated by numerous outliers. Robust techniques are used to estimate the transformation parameters and identify the outlier (incorrect) correspondences. The estimated transformation may then be used to return to the initial feature set and identify new correspondences which are much more likely to be correct. This aspect of the problem differs from most applications of robust estimation where the input to the estimation process is fixed.

4.1 Introduction to the Fundamental Matrix

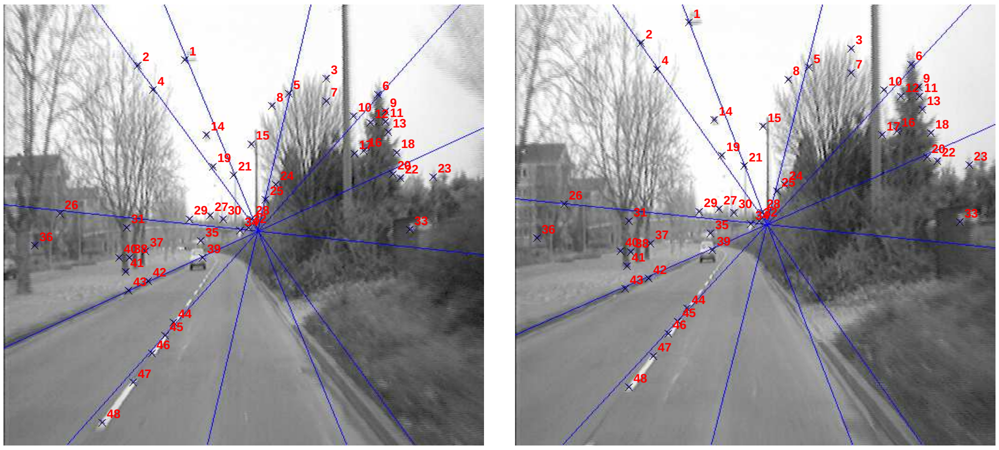
When two images are taken by a single uncalibrated camera or by two different, uncalibrated cameras, the relationship between the images can be captured in the "fundamental matrix" [50, 80]. This matrix imposes a linear constraint on the image positions of corresponding features. Once the fundamental matrix is known, the scene geometry can be reconstructed up to a projective transformation, with additional constraints (e.g. the corner of a room where three orthogonal faces meet) leading to affine or Euclidean reconstructions [24].

Let $\{\tilde{\mathbf{x}}_i\}$ be a set of feature point locations in image 1 and let $\{\tilde{\mathbf{x}}'_i\}$ be a set of corresponding feature locations in image 2, with each location vector written in homogeneous coordinates. The features are intensity corners [81] or other distinctive locations detected by an interest operator [63]. Correspondence for each $\tilde{\mathbf{x}}_i$ is found by searching an area of image 2 established by the range of scene depths and camera motions determined *a priori*. Within this area, which is sometimes as small as 30 pixels on a side [76], but could be much larger depending on what restrictions are placed on camera motion, $\tilde{\mathbf{x}}'_i$ is the location of the image feature most similar to that of $\tilde{\mathbf{x}}_i$, as decided by, for example, correlation of the surrounding image regions.

The fundamental matrix, $\mathtt{F}$, is a 3×3, rank 2 matrix such that for each i,

$$\tilde{\mathbf{x}}_i^T \mathtt{F} \tilde{\mathbf{x}}'_i = 0. \tag{5}$$

Because it is homogeneous and rank deficient, $\mathtt{F}$ has 7 degrees of freedom and therefore can be estimated from at least 7 correspondence pairs. For any point, $\tilde{\mathbf{x}}'$, in image 2, $\mathtt{F}\tilde{\mathbf{x}}'$ defines a line in image 1 through which the point corresponding to $\tilde{\mathbf{x}}'$ must pass. Similarly, for any point, $\tilde{\mathbf{x}}$, in image 1, $\mathtt{F}^T\tilde{\mathbf{x}}$ defines a line in image 2 through which the point corresponding to $\tilde{\mathbf{x}}$ must pass. All such lines, known as "epipolar lines", pass through the "epipoles", which are the null



a b

Figure 6: A pair of images and some of the epipolar lines resulting from the robust fundamental matrix technique of [81]. Points numbered are correspondences. Non-robust estimation results in substantial skewing of these lines and the position of the epipole. These images were provided by Zhengyou Zhang.

spaces of $\mathtt{F}^T$ and $\mathtt{F}$, respectively. Figure 6 shows a pair of images and epipolar lines. The epipole is the focus of expansion if the inter-camera motion is purely translational motion.

4.2 Estimating the Fundamental Matrix

Estimating the parameters of $\mathtt{F}$ is first considered when there are no outliers and then again in the presence of outliers. The first step is to rewrite (5) $\tilde{\mathbf{x}}_i^T \mathtt{F} \tilde{\mathbf{x}}_i'$ as

$$\mathbf{z}_i^T \mathbf{f}$$

where if $\tilde{\mathbf{x}}_i = (x_i, y_i, 1)^T$ and $\tilde{\mathbf{x}}_i' = (x_i', y_i', 1)^T$, then $\mathbf{z}_i^T = (x_i x_i', x_i y_i', x_i, y_i x_i', y_i y_i', y_i, x_i', y_i', 1)$ and $\mathbf{f}$ contains the appropriately ordered parameters of $\mathtt{F}$. This gives what's typically called a fitting error [28, Chapter 11] or an algebraic distance [31]. Summing the squares of these algebraic distances over all i gives

$$\sum_i (\mathbf{z}_i^T \mathbf{f})^2 = \mathbf{f}^T (\sum_i \mathbf{z}_i \mathbf{z}_i^T) \mathbf{f} = \mathbf{f}^T \mathtt{Z} \mathbf{f}.$$

With appropriate centering and normalization of the data [30], $\mathbf{f}$, and therefore $\mathtt{F}$, may be estimated directly from this as the unit eigenvector corresponding to the minimum eigenvalue of $\mathtt{Z}$. Enforcing $\det(\hat{\mathtt{F}}) = 0$ is achieved by computing the singular value decomposition of $\hat{\mathtt{F}}$, setting the smallest singular value to 0, and recalculating $\hat{\mathtt{F}}$.

Some methods minimize an objective function closer to a geometric distance [76]. Letting $r_i = \mathbf{z}_i^T \mathbf{f}$ be the algebraic residual and $r_{i,x}$, $r_{i,y}$, $r_{i,x'}$ and $r_{i,y'}$ be its partial derivatives, then an approximation of the geometric distance of (x_i, y_i, x_i', y_i') to the manifold defined by $\mathbf{f}$ is

$$\frac{r_i}{(r_{i,x}^2 + r_{i,y}^2 + r_{i,x'}^2 + r_{i,y'}^2)^{1/2}}.$$

This geometric distance value is incorporated into estimation of $\mathbf{f}$ by scaling the algebraic residuals by the weights

$$w_i = \frac{1}{(r_{i,x}^2 + r_{i,y}^2 + r_{i,x'}^2 + r_{i,y'}^2)^{1/2}}.$$

This leads to the objective function

$$\sum_i (w_i \mathbf{z}_i^T \mathbf{f})^2,$$

which is minimized iteratively in exactly the same manner as IRLS (Section 2.1).

4.3 Robust Estimation

Robust estimation of the fundamental matrix is important because the matching process is unreliable and the presence of an independently moving object in the field of view will induce a distinct fundamental matrix for the image projections of object points. Outlier percentages of 25-50% or more are not unrealistic. Application of robust estimation techniques appears straightforward, at first, but several issues arise. Some of these have been addressed in two prominent papers on the topic [76, 81]. These issues are discussed first for M-estimators and then for LMS.

M-estimators require a robust starting point and a weight function that tends to 0 quickly. Even with an unrealistically small percentage of outliers (e.g. 10-15%), the initial estimate obtained from least-squares does not allow M-estimation to converge to the correct estimate [76]. Because of the relatively small image search area for correspondences and because of the noise in estimated match positions, outlier correspondences tend to be close to correct epipolar lines. This implies that the domain of convergence for the correct estimate will be relatively small, in turn making it unlikely that the least squares fit used to initialize IRLS will be within this domain. The proximity of outlier correspondences to epipolar lines shows why the robust weight function must tend to 0 quickly. These observations are all practical realizations of the theoretical properties of M-estimators — a 0 breakdown point but also the ability to withstand outliers when robustly initialized and used with an appropriate weight function.

The difficulties associated with LMS are slightly different. The first problem is that the correct fundamental matrix is often nearly degenerate. (One example of such a degeneracy is points from a planar surface.) This has two implications.

First, this degeneracy should be avoided in the random samples themselves. Zhang et al. [81, 80] divide the image into non-overlapping regions (buckets) and prevent a random sample from incorporating more than one match from any one bucket; spreading the random samples across the image reduces the likelihood of a degenerate sample. Torr and Murray [76] develop a rank test for degeneracy and consider degenerate samples no further. Second, LMS, with its low statistical efficiency, will often yield an unreliable estimate of $\mathtt{F}$, even from a correct set of correspondences. Therefore, refinement of the LMS estimate is crucial. Zhang et al. [81] apply a least-squares estimate to the matches falling within $2.5\hat{\sigma}$ of the epipolar lines in each image, $\hat{\sigma}$ being taken from the median of the summed squared epipolar distances. Torr and Murray [76] take the LMS estimate as the starting point for an M-estimator using a MAD scale estimate (Equation 4) refined using an expectation maximization (EM) algorithm.

The second problem in using LMS concerns the random sampling procedure itself. Zhang et al. [81] generate 8 correspondence samples from which $\mathtt{F}$ can be instantiated with the appropriate constraints. Torr and Murray [76] present a method that instantiates $\mathtt{F}$ from 7 correspondence samples. Either way, the number of samples required is quite high, with 382 or 588 required for 7 point samples to obtain a 0.95 or 0.99 probability of obtaining a good sample and 766 or 1177 required for 8 point samples. This expense could be prohibitive for some applications.

The final issue in using LMS is that the number of outlier correspondences could in fact be greater than 50%. There are two interacting causes of the large number of outliers: (1) intensity corners are relatively weak features and easily mismatched; (2) large inter-image motions can lead to large matching search ranges and therefore more mismatches. (The results shown here are based on relatively small motions.) Torr and Murray [76] suggest use of RANSAC, which has the concurrent limitation of requiring establishment of a prior inlier bound, when more than 50% of the matches may be outliers. The recent techniques [43, 53, 54] that surpass the 50% breakdown limit in a probabilistic sense without prior knowledge of an inlier bound have not yet been used in fundamental matrix estimation. More generally, the effect of large motions and large numbers of features requires further exploration.

Example results of robust fundamental matrix estimation are shown in Figures 6 and 7. Figure 6, from [81], shows a pair of images and the robustly estimated epipolar lines. Figure 7, from [76], shows the motion vectors for matched corner features, and then separates them into those consistent (inliers) and inconsistent (outliers) with the fundamental matrix estimated from them. Notice that most matches from the independently moving person are eliminated as outliers and do not affect the fundamental matrix estimation. In general, however, unlike the local surface estimation problem, the conditions under which independent motions, producing distinct fundamental matrices, can be effectively handled is not yet fully understood [75, 77].

Figure 7: Fundamental matrix and motion estimation from an image sequence of a camera panning to follow a moving person: (a) shows one of these images, (b) shows the motion vectors obtained from correspondences superimposed on this image, (c) shows correspondences consistent with the estimated epipolar geometry (represented by the fundamental matrix), and (d) shows correspondences inconsistent with the estimated epipolar geometry. These images were provided by Phil Torr.

5 Construction of Retinal Image Mosaics

A new application of robust estimation is in the construction of image mosaics from overlapping images of the human retina. Starting from a series of images, the goal is to produce a composite that is much larger than any individual image and shows the retina as a whole. Example images and a resulting mosaic are shown in Figure 8. This has numerous possible applications in opthalmology.

The primary issue is calculating the transformation, T, mapping the coordinate system and therefore the pixels of a given image I_m to a reference image I_r which will form the "center" of the mosaic. Most mosaic construction techniques in computer vision formulate T as an affine transformation in two dimensions [6] or as the model of the apparent (image) motion of a planar surface [1]. These yield 6-parameter and 8-parameter transformation models, respectively. Unfortunately, these models yield substantial mapping errors for the curved surface of the retina. As a result, a 12-parameter transformation model is required, which models the motion of a quadratic surface imaged using weak perspective (scaled orthographic) projection. The derivation of this model is similar to the derivation of the 8-parameter planar surface motion model [1] and is omitted here. To see how the transformation works, let $\mathbf{x_m} = (x_m, y_m)^T$ be the coordinates of a point in I_m, and define $\mathbf{X}(\mathbf{x_m}) = (x_m^2, y_m^2, x_m y_m, x_m, y_m, 1)$. Then, the transformed point location in I_r is

$$\mathbf{x_r} = \mathtt{T}\mathbf{X}(\mathbf{x_m})$$

where $\mathtt{T}$ is a 2×6 matrix. For optimal estimates of $\mathtt{T}$, this yields transformation errors averaging about one pixel.

Similar to fundamental matrix estimation, $\mathtt{T}$ is estimated by establishing correspondence between image features located in I_m and I_r. Features are bifurcations in blood vessels, found through a tracing procedure [19]. Example images are shown in Figure 8 and extracted features are shown in Figure 9a. Establishing correspondence between features, unfortunately, is substantially more difficult than in fundamental matrix estimation, because the search range is not known in advance and the degree of overlap between I_m and I_r may vary substantially. The result is that initially nearly every possible feature pair must be considered as a correspondence and this correspondence set must be culled during estimation of T.

5.1 Hierarchical Transform Estimation

The combination of the 12-parameter transformation model and the vast majority of matches being incorrect is too difficult to handle directly. Instead, low-order approximations to T are estimated to reduce the correspondence set and focus the final robust estimation of T. This hierarchy of transformation models differs from standard hierarchical (multiresolution) techniques in motion estimation which start from low-resolution versions of images and gradually increase this resolution [6].

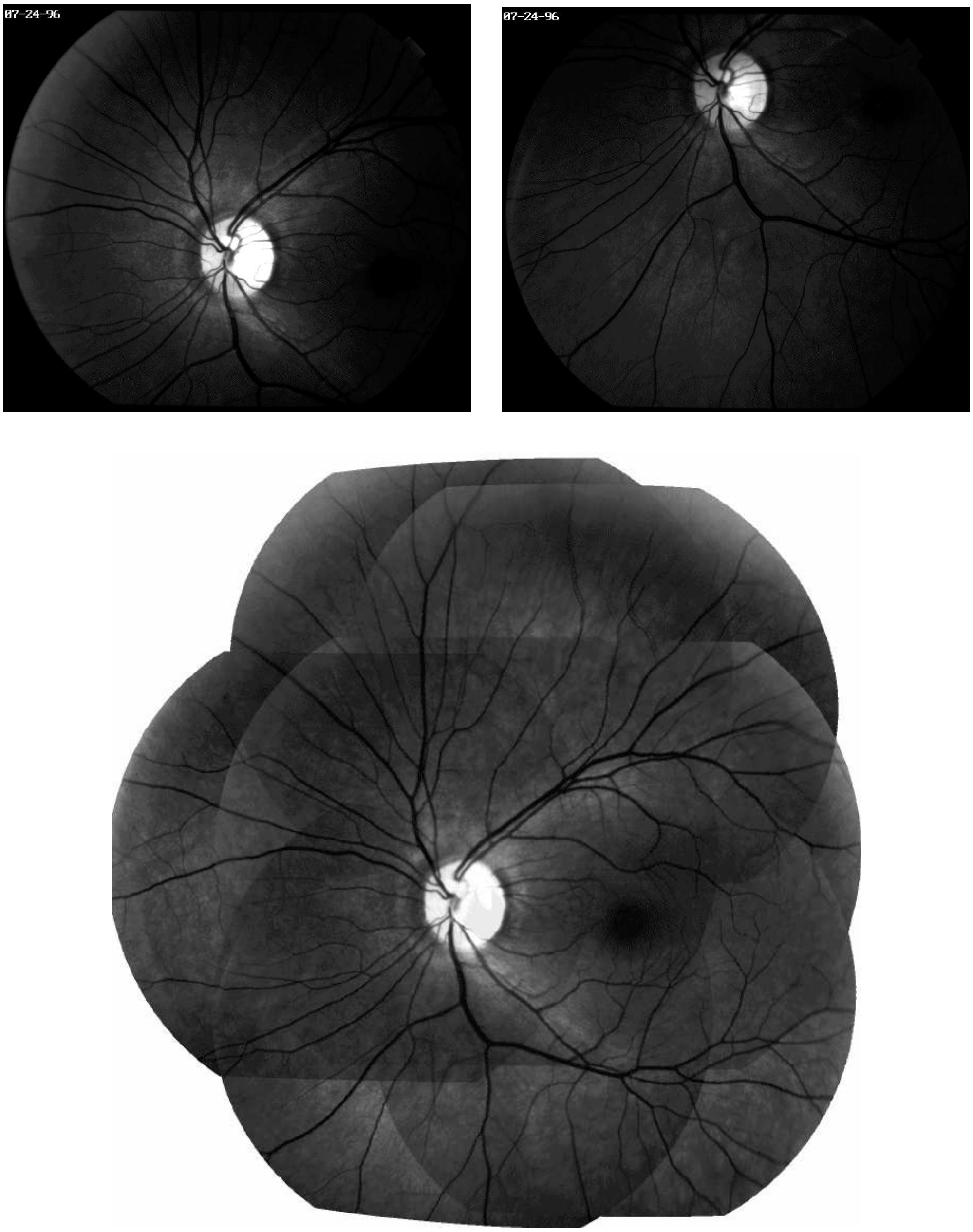


Figure 8: Two retinal images and a mosaic of many images estimated using the robust mapping technique described here.

The first approximation to T is a simple translation:

$$\mathbf{x_m} = \mathbf{x_r} + \mathbf{t}.$$

Each possible match produces a hypothesized image translation vector, $\mathbf{t} = \mathbf{x_m} - \mathbf{x_r}$, and these vectors are recorded in a two-dimensional histogram with overlapping bins. The radius of each bin equals the experimentally determined maximum position error that could occur using the translation model. For each hypothesized $\mathbf{t}$, all bins overlapping $\mathbf{t}$ are found. A weight determined by an image similarity measure (e.g. cross-correlation) between matching features is added to each overlapping bin. At the end, the translation vector corresponding to the peak bin is taken as the translation estimate and all matches falling into this bin are retained for estimating the next finer approximation to the transformation. For some features several matches could be retained, while for others, particularly those outside the overlap between images, there may be no matches. Figure 9b shows two images aligned using the translation model.

$\mathtt{T}$ is next approximated using a 6-parameter affine model [6, 37]:

$$\mathbf{x_m} = \mathtt{A}\mathbf{x_r} + \mathbf{t},$$

where $\mathtt{A}$ is a 2×2 matrix with no imposed constraints. Here, LMS is used to robustly estimate the parameters of $\mathtt{A}$ and $\mathbf{t}$. Features outside the region of image overlap determined by the translation model (plus modeling error) are not considered during LMS. In selecting the matches to form a random sample, image features are first selected randomly and then a correspondence for each selected feature is chosen at random from its correspondences that were retained following translation estimation. Other than this, application of LMS is straightforward. Figure 9c shows two images aligned using an estimated affine model.

The final stage is estimation of the full 12-parameter model using an IRLS implementation of an M-estimator. The scale and residuals used in the first iteration of IRLS are taken from the optimal LMS affine transformation. In effect, the affine model estimated by LMS is used to initialize the M-estimate of the full model. This represents a novel twist on robust initialization of an M-estimator, since it is done with the robust estimate of a lower order model. The scale value taken from the affine model is re-estimated following the first iteration of IRLS and then fixed for the remaining iterations.

The most interesting aspect of the computation in the final stage is in the weighting. Let $(\mathbf{x}_{m,i}, \mathbf{x}_{r,j})$ be a match, let $r_{i,j} = \|\mathbf{x}_{r,j} - T\mathbf{X}(\mathbf{x}_{m,i})\|$ be the fit residual (the Euclidean distance between the transformed $\mathbf{x}_{m,i}$ and $\mathbf{x}_{r,j}$), let $w_{i,j}$ be the robust weight, and let $s_{i,j}$ be the match similarity measure. The weight calculation requires two steps. First, for each match the robust weights and similarity measures are multiplied to obtain

$$w^*_{i,j} = w_{i,j} s_{i,j}.$$

Second, these weights are scaled based on how they stack up against competing matches. Competing matches are other matches (in I_r) for feature point i (which

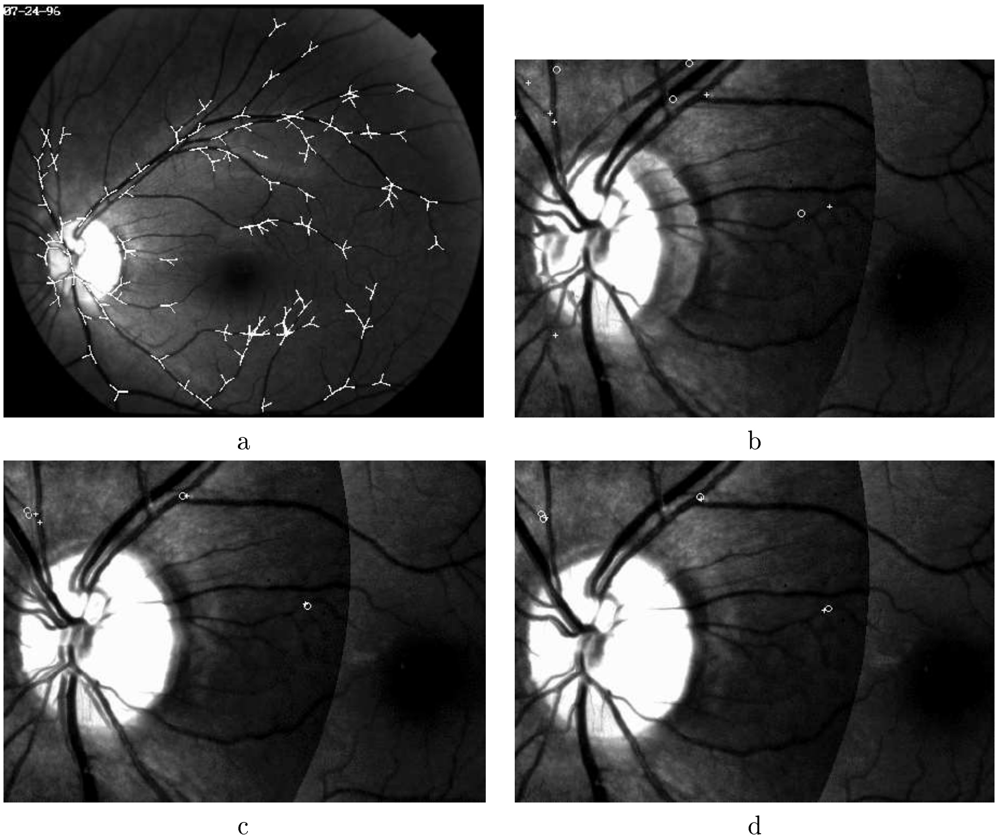


Figure 9: Hierarchical retinal image transformation estimation results. Image (a) shows features extracted from one image, (b) shows the results of applying the estimated 0th order translation only model, (c) shows the results of applying the estimated 1st order, 6-parameter, affine model, and (d) shows the results of applying the final estimated 2nd order, 12-parameter quadratic model. In each of (b), (c) and (d), a small region around the optic disk is shown to emphasize modeling errors.

is from I_m) or other matches (in I_m) for feature point j (which is from I_r). The scaling factor is

$$\eta_{i,j} = \frac{w^*_{i,j}}{w^*_{i,j} + \sum_{k \neq i} w^*_{k,j} + \sum_{m \neq j} w^*_{i,m}}$$

This scaling factor will be at or near 1.0 if no other matches for i or j have large weights, and will be much less than 1.0 if they do. The point is to downgrade the influence of ambiguous matches. The final weight used in IRLS will be $\eta_{i,j} w^*_{i,j}$. One benefit of this is that it allows matches that were rejected early to re-enter the computation — if they produce extreme residuals, their weights will be 0 and the computation will proceed as though they didn't exist.

Figure 9d shows the final alignment between two images based on the robust estimate of the 12-parameter model. Figure 8 shows a retinal mosaic of many different images. The effects of the non-linear mapping are most easily seen on the boundaries of the images.

6 Discussion and Conclusions

This paper has summarized several important robust parameter estimation techniques and outlined their use in three applications in computer vision. Many other applications have been considered in the computer vision literature as well, including optical flow and motion estimation [11, 22, 64], edge and feature detection [48, 56, 58], and pose estimation [29, 41]. Most applications of robust estimators, like the three emphasized here, are based on least-median of squares, M-estimators, and their variations.

The observations about robust parameter estimation in the foregoing discussion can be summarized in three important points.

1. The theoretical limit of the 50% breakdown point can and must be surpassed in certain applications. This can be done through use of RANSAC or Hough transforms if a prior error bound is available, through adaptive techniques based on scale estimates such as ALKS [43] and MUSE [53, 54], or by special-purpose techniques such as the hierarchy of models used in retinal mosaic construction. Care must be taken to ensure that the results are meaningful, especially when the estimated structure includes only a small percentage of the data as inliers. For example, MUSE [53, 54] incorporates a randomness test based on MINPRAN [67] to ensure the estimated structures are significant.

2. Without a robust initial estimate, as provided by a high-breakdown estimator such as LMS, M-estimation is likely to yield poor results. This reflects the low breakdown point of M-estimators. Hence, robust initialization is especially important when outliers are numerous. These outliers cause a non-robust least-squares estimate to be far from correct. Their presence requires that a weight function such as that of Beaton and Tukey

[4], which tends to 0 quickly, be used. When using such weight functions, lower tuning parameters than recommended in the statistics literature should be used, emphasizing outlier resistance over statistical efficiency. In a practical though not a theoretical sense, robust initialization is less important when few outliers are expected and, especially, when leverage points are non-existent.

3. The nature of the outliers that might arise in the data should be considered carefully. Most important, of course, is that outliers to one correctly estimated structure (population) are often inliers to (one or more) other structures. In this case, successful use of robust estimation depends on the distribution of points between structures, the physical proximity of the point sets, and the similarity between the structures themselves. For example, small magnitude depth and orientation discontinuities in range data can lead to skewed surface estimates — "bridging fits" — when using current robust estimators. Similar difficulties are likely in fundamental matrix estimation when images include proximate objects undergoing slightly different motions; the extent of the problem has not yet been fully explored.

The last of these problems is the most difficult. The current best approach to addressing it is to use mixture model formulations [73] in which multiple structures are simultaneously and robustly estimated, and data are dynamically assigned to different structures. This has been studied most heavily in motion and fundamental matrix estimation [3, 40, 77, 78, 79], but should be used increasing elsewhere. Aside from the added complexity, an important limitation of mixture models is that they are most effective when all structures in the data can be appropriately modeled. This is sometimes difficult or impossible and perhaps shouldn't be necessary when only a single estimate is required, e.g., when estimating a planar surface model for a roof, models for the leaves and branches of trees near the roof should not be required. Unmodeled structures must be treated as outliers to all models, which reduces to the original robust parameter estimation problem. Further work is clearly needed.

A new approach currently under investigation combines robust parameter estimation with boundary estimation in a single objective function [69]. This makes explicit the fact that structures in computer vision, e.g. object surfaces, have limited spatial extent. The objective function effectively treats points far outside boundaries as outliers, but allows structures to "grow" toward regions of data that are close to the extrapolated estimate. All of this results from the objective function minimization process. While still in early stages of development, this technique has already shown promise of eliminating some of the problems in local and global surface estimation discussed earlier.

In summary, robust parameter estimation is an important though incompletely solved problem in computer vision. Computer vision data are numerous, corrupted by noise and outliers, and usually are drawn from multiple statistical populations (structures). Because robust estimation techniques are designed to handle corrupted and incompletely modeled data, they provide an attractive set of tools for computer vision estimation problems. These tools can not be applied to vision problems successfully, however, without careful thought about the populations (structures) to be estimated and the frequency and nature of outliers. Successful application can lead to computer vision techniques that accommodate substantial variations in the data. This is an important development in extending the overall reliability of computer vision technology.

Acknowledgements

The author gratefully acknowledges the support of the National Science Foundation under awards IRI-9217195 and IRI-9408700. The author also thanks Dr. Kishore Bubna for feedback on an earlier draft of this paper and thanks the anonymous reviewers for their thoughtful comments.

References

[1] G. Adiv. Determining 3-d motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(4):384–401, 1985.

[2] F. Arman and J. Aggarwal. Model-based object recognition in dense range images. *Computing Surveys*, 25(1):5–43, 1993.

[3] S. Ayer and H. Sawhney. Layered representation of motion video using robust maximum likelihood estimation of mixture models and MDL encoding. In *Proceedings IEEE International Conference on Computer Vision*, pages 777–784, 1995.

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

[5] D. E. Becker, A. Can, J. N. Turner, H. L. Tanenbaum, and B. Roysam. Image processing algorithms for retinal montage synthesis, mapping and real-time location determination. *IEEE Transactions on Biomedical Engineering*, 45(1):105–118, 1998.

[6] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings Second European Conference on Computer Vision*, pages 237–252, 1992.

[7] P. Besl. *Surfaces in Range Image Understanding*. Springer-Verlag, 1988.

[8] P. Besl and R. Jain. Three-dimensional object recognition. *Computing Surveys*, 17(1):75–145, 1985.

[9] P. J. Besl, J. B. Birch, and L. T. Watson. Robust window operators. In *Proceedings IEEE International Conference on Computer Vision*, pages 591–600, 1988.

[10] P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:167–192, 1988.

[11] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proceedings IEEE International Conference on Computer Vision*, pages 231–236, 1993.

[12] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 19(1):57–91, 1996.

[13] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[14] R. Bolle and B. Vemuri. On three-dimensional surface reconstruction methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):1–13, 1991.

[15] K. L. Boyer, M. J. Mirza, and G. Ganguly. The Robust Sequential Estimator: A general approach and its application to surface organization in range data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:987–1001, 1994.

[16] H. Bozdogan. Model selection and Akaike's information criterion (AIC): The general theory and its analytic extension. *Psychometrika*, 52(3):345–370, 1987.

[17] K. Bubna and C. V. Stewart. Model selection and surface merging in reconstruction algorithms. In *Proceedings IEEE International Conference on Computer Vision*, pages 895–902, 1998.

[18] P. J. Burt et al. Object tracking with a moving camera: An application of dynamic motion analysis. In *Proceedings IEEE Workshop on Visual Motion*, pages 2–12, 1989.

[19] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam. Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. *IEEE Trans. on Biomedial Engineering (to appear)*, 1998. Accepted, to appear.

[20] R. Chin and C. Dyer. Model-based recognition in robot vision. *Computing Surveys*, 18(1):67–108, 1986.

[21] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. McGraw-Hill, 1990.

[22] T. Darrell and A. Pentland. Cooperative robust estimation using layers of support. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:474–487, 1995.

[23] H. Edelsbrunner and D. L. Souvaine. Computing least median of squares regression lines and guided topological sweep. *Journal of the American Statistical Association*, 85:115–119, 1990.

[24] O. Faugeras. Stratification of 3-dimensional vision: Projective, affine, and metric representations. *Journal of the Optical Society of America, A*, 12(3):465–484, 1995.

[25] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *CACM*, 24:381–395, 1981.

[26] W. Grimson. *From Images to Surfaces: A Computational Study of the Human Early Visual System.* MIT Press, 1981.

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

[28] R. Haralick and L. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley, 1992.

[29] R. M. Haralick, H. Joo, C.-N. Lee, X. Zhuang, V. G. Baidya, and M. B. Kim. Pose estimation from corresponding data. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1426–1446, 1989.

[30] R. Hartley. In defence of the 8-point algorithm. In *Proceedings IEEE International Conference on Computer Vision*, pages 1064–1070, 1995.

[31] R. Hartley. Minimizing algebraic error in geometric estimation problems. In *Proceedings IEEE International Conference on Computer Vision*, pages 469–476, 1998.

[32] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6:813–827, 1977.

[33] A. Hoover, G. Jean-Baptiste, X. Jiang, P. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D. Eggert, A. Fitzgibbon, and R. Fisher. An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.

[34] P. J. Huber. *Robust Statistics.* John Wiley & Sons, 1981.

[35] J. Illingworth and J. Kittler. A survey of the Hough transform. *CVGIP*, 44:87–116, 1988.

[36] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Proceedings IEEE International Conference on Computer Vision*, pages 605–611, 1995.

[37] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, 1994.

[38] D. Jacobs. Matching 3-d models to 2-d images. *International Journal of Computer Vision*, 21(1-2):123–153, January 1997.

[39] R. Jarvis. A perspective on range finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):122–139, 1983.

[40] S. Ju, M. Black, and A. Jepson. Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 307–314, 1996.

[41] R. Kumar and A. R. Hanson. Robust methods for estimating pose and a sensitivity analysis. *CVGIP: Image Understanding*, 1994.

[42] V. Leavers. Survey: Which Hough transform? *CVGIP*, 58(2):250–264, 1993.

[43] K.-M. Lee, P. Meer, and R.-H. Park. Robust adaptive segmentation of range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2):200–205, 1998.

[44] A. Leonardis, A. Gupta, and R. Bajcsy. Segmentation of range images as the search for geometric parametric models. *International Journal of Computer Vision*, 14:253–277, 1995.

[45] A. Leonardis, F. Solina, and A. Macerl. A direct recovery of superquadric models in range images using recover-and-select paradigm. In *Proceedings Third European Conference on Computer Vision*, pages A:309–318, 1994.

[46] S. Li. Robustizing robust M-estimation using deterministic annealing. *Pattern Recognition*, 29(1):159–166, 1996.

[47] B. W. Lindgren. *Statistical Theory.* Chapman and Hall, Inc., 1993.

[48] L. Liu, B. G. Schunck, and C. C. Meyer. On robust edge detection. In *Proceedings of the International Workshop on Robust Computer Vision*, pages 261–286, 1990.

[49] D. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31(3):355–395, 1987.

[50] Q. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–75, 1996.

[51] P. Meer, D. Mintz, and A. Rosenfeld. Least median of squares based robust analysis of image structure. In *Proceedings of the DARPA Image Understanding Workshop*, pages 231–254, 1990.

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

[53] J. V. Miller. *Regression-Base Surface Reconstruction: Coping with Noise, Outliers, and Discontinuities.* PhD thesis, Rensselaer Polytechnic Institute, 1997.

[54] J. V. Miller and C. V. Stewart. MUSE: Robust surface fitting using unbiased scale estimates. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 300–306, 1996.

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

[56] N. Netanyahu, V. Philomin, A. Rosenfeld, and A. Stromberg. Robust detection of straight and circular road segments in noisy aerial images. *Pattern Recognition*, 30(10):1673–1686, 1997.

[57] D. Nitzan. Three-dimensional vision structure for robot applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:291–309, 1988.

[58] G. Roth and M. D. Levine. Extracting geometric primitives. *CVGIP: Image Understanding*, 58:1–22, 1993.

[59] C. A. Rothwell. *Object Recognition Through Invariant Indexing.* Oxford Science Publications, 1995.

[60] P. J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79:871–880, 1984.

[61] P. J. Rousseeuw. Introduction to positive-breakdown methods. In G. Maddala and C. Rao, editors, *Handbook of Statistics, Vol. 15: Robust Inference*, pages 101–121. Elsevier/North-Holland, 1997.

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

[63] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Proceedings IEEE International Conference on Computer Vision*, pages 230–235, 1998.

[64] B. G. Schunck. Image flow: Fundamentals and algorithms. In W. N. Martin and J. K. Aggarwal, editors, *Motion Understanding: Robot and Human Vision*, pages 23–80. Kluwer Academic Publishers, 1988.

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

[66] D. L. Souvaine and J. M. Steele. Time- and space-efficient algorithms for least median of squares regression. *Journal of the American Statistical Association*, 82:794–801, 1987.

[67] C. V. Stewart. MINPRAN: A new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:925–938, 1995.

[68] C. V. Stewart. Bias in robust estimation caused by discontinuities and multiple structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8):818–833, 1997.

[69] C. V. Stewart, K. Bubna, and A. Perera. Estimation of model parameters and boundaries by minimizaing a joint, robust objective function. Technical report, Department of Computer Science Technical Report, Rensselaer Polytechnic Institute (in preparation), 1998.

[70] R. Szeliski. Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, 16(2):22–30, 1996.

[71] G. Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1115–1138, 1991.

[72] D. Terzopoulos. The computation of visible-surface representations. *PAMI*, 10(4):417–438, 1988.

[73] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, New York, 1985.

[74] P. Torr. An assessment of information criteria for motion model selection. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 47–52, 1997.

[75] P. Torr and D. Murray. Statistical detection of independent movement from a moving camera. *Image and Vision Computing*, 11:180–187, 1993.

[76] P. Torr and D. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[77] P. Torr and A. Zisserman. Concerning Bayesian motion segmentation, model averaging, matching and the trifocal tensor. In *Proceedings 5th European Conference on Computer Vision*, pages 511–527, 1998.

[78] J. Y. A. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 361–366, 1993.

[79] Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 520–526, 1997.

[80] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, 1998.

[81] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, 1995.