Robust Regression with Projection Based M-estimators

Haifeng Chen Peter Meer

Electrical and Computer Engineering Department
Rutgers University, Piscataway, NJ 08854, USA
haifeng meer @caip.rutgers.edu

Abstract

The robust regression techniques in the RANSAC family are popular today in computer vision, but their performance depends on a user supplied threshold. We eliminate this drawback of RANSAC by reformulating another robust method, the M-estimator, as a projection pursuit optimization problem. The projection based pbM-estimator automatically derives the threshold from univariate kernel density estimates. Nevertheless, the performance of the pbM-estimator equals or exceeds that of RANSAC techniques tuned to the optimal threshold, a value which is never available in practice. Experiments were performed both with synthetic and real data in the affine motion and fundamental matrix estimation tasks.

1. An Analysis of Robust Regression

Robust regression is the generic name of techniques which estimate a parametric regression model in the presence of significant number of data points not belonging to that model, i.e., outliers. The 'secret' of robust estimation is the use of valid additional assumptions about the data. The scale of the data of interest, i.e., a measure of the noise corrupting the inliers (such as standard deviation or range), is the most frequently used additional assumption. The robust regression technique most popular in computer vision RANSAC [4] and its improved versions MSAC and MLESAC [15], [16], impose an upper bound on the scale, and the parameter estimates are found by maximizing the number of points (inliers) which can be placed within this bound. The additional assumption behind the least median of squares (LMedS) estimator [12] and similar techniques is equivalent. A lower bound is imposed on the required percentage of inliers in the data, and the parameter estimates are found by minimizing the scale of data subsets of this size.

In real applications, however, often there is not enough a priori knowledge to reliably define additional information. Embedding the robust estimator into a second optimization process over the range of possible bounds, e.g., [10], is not a general enough solution. Indeed, whenever the employed assumptions are not valid, the robust regression may yield erroneous results, which in turn can corrupt the comparison across different operating conditions. The technique described in this paper *does not require* the user to provide any scale estimate, instead it exploits an intrinsic relation between the optimization criterion and the data space.

Probabilistic sampling is the search technique of choice to minimize the optimization criterion of the robust regressions in the RANSAC family and LMedS. Elemental subsets containing the smallest number of data points which uniquely define a model parameter candidate are drawn without replacement from the data. The quality of the candidate is then assessed using all the data points and the final estimate is found by comparison over a number of such candidates [12, p.198]. This number, however, becomes unfeasible large when the outliers dominate in the data or the data is high dimensional. A possible solution is guided sampling, in which additional information about the probability of a data point being an inlier is integrated into the sampling process, e.g., [14].

It is very important to recognize that while probabilistic sampling *is a computational tool* with no relation to the optimization criterion, guided sampling *is a robust procedure* since it also exploits additional information. But reliable information to guide sampling cannot be guaranteed in many computer vision applications, and the robust regression method proposed in this paper shifts the emphasis from sampling in the input space to an efficient search in the space of the parameters.

In Section 2 by reformulating the M-estimator optimization criterion, we introduce a new generic technique, the *pbM-estimator*, which does not require the user to provide the scale estimate. A multidimensional direct search technique using simplex, described in Section 3, keeps the computational burden at a satisfactory level. In Section 4 the performance of pbM-estimator is compared to techniques from the RANSAC family for synthetic data and two vision applications: affine motion and fundamental matrix estimation.

2. M-estimate Computation Using Projection Pursuit

The principle behind the method discussed in this section was first proposed in [1], as part of a different approach limited to low-dimensional data. See also Section 5. Here we provide the implementation for arbitrary dimensional data, and introduce a new robust regression technique, the pbM-

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

estimator.

The class of *linear errors-in-variables* (EIV) models is adequate for many computer vision problems. Let $\mathbf{y}_{io}$ be the true values of the p-dimensional measurements $\mathbf{y}_i$. The following EIV model is of interest to us

$$\begin{aligned} &\mathbf{y}_{io}^\top \boldsymbol{\theta} - \alpha = 0 \qquad && i = 1, \ldots, n \\ &\mathbf{y}_i = \mathbf{y}_{io} + \delta\mathbf{y}_i \qquad && \delta\mathbf{y}_i \sim GI(0, \sigma^2 \mathbf{I}_p) \end{aligned} \tag{1}$$

where the additive zero-mean measurement noise $\delta\mathbf{y}_i$ is independent and identically distributed (i.i.d.), with an arbitrary symmetric distribution. Note that in general the scale of the noise, the standard deviation σ, is unknown.

The ambiguity up to a multiplicative constant of (1) can be eliminated by taking $\|\boldsymbol{\theta}\| = 1$. This is the Hessian normal representation of the linear constraint and it is easy to prove that in this case the algebraic and geometric distance of $\mathbf{y}_i$ from the hyperplane have the same value

$$\|\mathbf{y}_i - \hat{\mathbf{y}}_i\| = |\mathbf{y}_i^\top \boldsymbol{\theta} - \alpha| \tag{2}$$

where $\hat{\mathbf{y}}_i$ is the orthogonal projection of $\mathbf{y}_i$ on the plane.

We have shown in [2], that all the major robust regression techniques used in computer vision can be expressed as M-estimators. Taking into account (2) the M-estimator of the EIV model parameters is

$$\begin{aligned} &[\hat{\alpha}, \hat{\boldsymbol{\theta}}] = \underset{\alpha, \boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{1}{s}|\mathbf{y}_i^\top \boldsymbol{\theta} - \alpha|\right) \\ &\text{subject to} \quad \hat{\mathbf{y}}_i^\top \hat{\boldsymbol{\theta}} - \hat{\alpha} = 0 \qquad i = 1, \ldots, n \end{aligned} \tag{3}$$

where s is the scale parameter, and $\rho(u)$ is a nonnegative, even symmetric loss function, nondecreasing with $|u|$ and having the unique minimum $\rho(0) = 0$. The *redescending* M-estimator with *biweight* loss function

$$\rho(u) = \begin{cases} 1 - (1 - u^2)^3 & \text{if} \quad |u| \le 1 \\ 1 & \text{if} \quad |u| > 1 \end{cases} \tag{4}$$

is a satisfactory choice for most applications. The optimization criterion (3) of the M-estimator can be rewritten as

$$\begin{aligned} [\hat{\alpha}, \hat{\boldsymbol{\theta}}] &= \arg\max_{\alpha, \boldsymbol{\theta}} \left[1 - \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{\mathbf{y}_i^\top \boldsymbol{\theta} - \alpha}{s}\right)\right] \\ &= \arg\max_{\alpha, \boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^{n} \kappa\left(\frac{\mathbf{y}_i^\top \boldsymbol{\theta} - \alpha}{s}\right) \end{aligned} \tag{5}$$

where $\kappa(u) = c_\rho[1 - \rho(u)]$ will be called the *M-kernel function*. Note that $\kappa(u) = 0$ for $|u| \ge 1$, and that the even symmetry of the loss function (4) allows dropping the absolute values. The positive normalization constant c_ρ assures that $\kappa(u)$ integrates to one and has no bearing on the optimization. Expressing the M-estimator as in (5) yields a new approach for its computation.

Consider a direction in $\mathcal{R}^p$ specified by the unit vector $\boldsymbol{\theta}$. The orthogonal projection of a data point $\mathbf{y}_i$ on the line through the origin oriented along $\boldsymbol{\theta}$ defines the coordinate $x_i = \mathbf{y}_i^\top \boldsymbol{\theta}$ in this one-dimensional subspace. The points x_i, $i = 1, \ldots, n$, obey the an unknown probability density $f_{\boldsymbol{\theta}}(x)$, where the dependence on projection direction is made explicit. The most popular approach for density estimation is the kernel density estimator, also known as the Parzen window method in pattern recognition. See [3, Sec.4.3] for an introduction, and [17] for a more complete discussion. The density estimate at location x, computed with the kernel $\kappa(u)$ scaled to bandwidth $h_{\boldsymbol{\theta}}$ is

$$\hat{f}_{\boldsymbol{\theta}}(x) = \frac{1}{n h_{\boldsymbol{\theta}}} \sum_{i=1}^{n} \kappa\left(\frac{\mathbf{y}_i^\top \boldsymbol{\theta} - x}{h_{\boldsymbol{\theta}}}\right) . \tag{6}$$

The cardinal observation of our approach is the close similarity between (5) and (6). This allows us to replace the scale parameter s in (5) with the bandwidth $h_{\boldsymbol{\theta}}$, and the computation of the M-estimator becomes the process of finding a projection direction $\hat{\boldsymbol{\theta}}$ such that

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[h_{\boldsymbol{\theta}} \max_x \hat{f}_{\boldsymbol{\theta}}(x)\right] . \tag{7}$$

The inner maximization process searches for the *mode* of the density estimate and also provides

$$\hat{\alpha} = \underset{x}{\operatorname{argmax}} \hat{f}_{\hat{\boldsymbol{\theta}}}(x) . \tag{8}$$

The optimization procedure has a simple geometric interpretation. A hyperplane with normal $\boldsymbol{\theta}$ is placed in the densest region in $\mathcal{R}^p$ such that the points located in the proximity of the plane, i.e., the inliers, are most clustered when projected on the direction of the normal. Since orthogonal projection is used, the center of this "best" cluster provides the location of the hyperplane in $\mathcal{R}^p$, that is $\hat{\alpha}$.

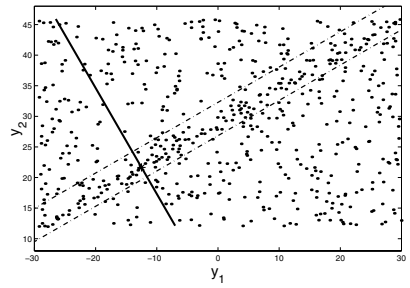
The relation (7) is the projection pursuit definition of an M-estimator. Projection pursuit solves estimation problems by seeking "interesting" low-dimensional projections of multidimensional data. The informative value of a projection is measured with a *projection index*, such as the quantity inside the brackets in (7). The papers [5] [6] are surveys of all related topics.

2.1. Adaptive Mode Search

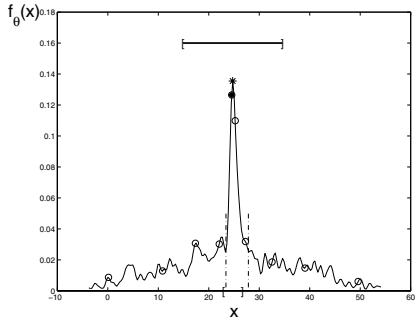
An adaptive sampling strategy is employed for a computationally efficient search for the mode of $\hat{f}_{\boldsymbol{\theta}}(x)$ in (8). This is illustrated through an example.

The 2D data in Figure 1a contains 100 inliers and 500 uniformly distributed outliers. The employed projection direction was chosen close to the normal of the true inlier structure (a line), to emphasize the idea behind the proposed method. The inliers are corrupted with noise $N(0, 2^2\mathbf{I}_2)$.

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE



(a)



(b)

Figure 1: An example of robust analysis through projection. (a) 2D data. Solid line: the projection direction. Dashed lines: bounds of the detected inlier region. (b) The kernel density estimate of the projected points. Open circles: coarse sampling. Filled circle: the largest estimate. Star: the mode detected by fine sampling in the interval marked with brackets on the abscissa. Vertical dashed lines: the basin of attraction of the mode. The bar at the top represent the size of a window with half-width $1.96\,\hat{\sigma}_{MAD}$.

Let $x_{i:n,\boldsymbol{\theta}}$, $i = 1, \ldots, n$, be the *ordered* sequence of the projections, i.e., $x_{i:n,\boldsymbol{\theta}} \leq x_{(i+1):n,\boldsymbol{\theta}}$. In the *coarse* sampling step, the density is estimated at ten locations defined by equidistance indices $j = k * \left[\frac{n}{11}\right]$, $k = 1, \ldots, 10$, where $[\cdot]$ rounds to the nearest integer. This sampling strategy assures that denser regions are better represented in the analysis (Figure 1b).

Kernel density estimation, a standard technique is employed. See for example, [17] for a complete survey. The bandwidth, i.e., the size of the employed kernel is the parameter of this technique. It can be shown that the optimal bandwidth value requires the knowledge of the density to be estimated. Several practical methods were proposed in the statistical literature to avoid this contradiction. Taking into account that we use the M-kernel corresponding to the biweight loss function (4), the following bandwidth selection formula can be derived from the results presented in [17, Sec.3.2.2]

$$h_{\boldsymbol{\theta}} = n^{-1/5} \operatorname*{med}_j \mid x_{j,\boldsymbol{\theta}} - \operatorname*{med}_i x_{i,\boldsymbol{\theta}} \mid = \frac{\hat{\sigma}_{MAD}}{1.483\, n^{1/5}} \quad (9)$$

where the dependence of the x_i on the projection direction $\boldsymbol{\theta}$ was made explicit. The bandwidth (9) is proportional to the robust median absolute deviations (MAD) scale estimate, $\hat{\sigma}_{MAD}$, frequently used in computer vision.

For the *fine* sampling step a narrow range is delineated around the location j_0 of the largest density estimate found at the previous step, $[x_{j_0:n,\boldsymbol{\theta}} - h_{\boldsymbol{\theta}},\ x_{j_0:n,\boldsymbol{\theta}} + h_{\boldsymbol{\theta}}]$. Again ten sample points are defined, but this time at equidistance intervals. To increase the accuracy of the mode localization the bandwidth of the kernel density estimator is reduced to $0.5h_{\boldsymbol{\theta}}$. The mode corresponds to the largest density among the 10 estimates (Figure 1b).

The *basin of attraction* of the mode is delimited by the nearest left and right significant local minimum in the density estimate. They are evaluated using a simple heuristic about their magnitudes. The density estimation process starts from the location of the mode and uses $0.5h_{\boldsymbol{\theta}}$ as bandwidth. If the first minimum corresponds to at least a 70 percent drop from the mode, it is accepted independent of the value of the subsequent local maximum, otherwise that value is also taken into account.

The boundaries of the basin of attraction are used as thresholds for separating inliers from outliers. Indeed, for a projection direction close to the true normal of the hyperplane, the two minima represent two low density regions in R^p where the inlier structure relinquishes its dominant presence. Note that in general the two thresholds are not symmetric relative to the location of the mode. The basin of attraction is delineated in Figure 1b. The corresponding band in the input space (Figure 1a) contains mostly points from the inlier structure and can be further processed with traditional robust regression methods.

The detection of the right bound in Figure 1b is fragile. A slight change in the projected location of a few data points could have changed this bound to the next, much more significant local minimum. However, this sensitivity is tolerated by the method. First, by the nature of the projection pursuit many different projections are investigated and thus with high probability a satisfactory band will be delineated at least for a few of the projection directions. Second, from any reasonable inlier/outlier dichotomy of the data postprocessing can recover the correct estimates.

The technique introduced above, the projection based M-estimator will be called *pbM-estimator*.

2.2. The Advantage of pbM-estimator

Interpretation of robust estimators as a search for the location of the densest (or narrowest) band is not new. Indeed, RANSAC determines the parameter estimate by trying to place a band of *user defined width* over the data such that the band contains the most number of points. In MSEC and MLESAC the data is also weighted with a function monotonically decreasing with the distance from the center of the

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

band. The LMedS estimator can be described as searching for the narrowest band containing at least half the data points [12, p.126]. In this case the user set scale value is substituted with a *preset minimal percentage* of inliers.

In our approach, the scale s of the M-estimator is replaced by the bandwidth of the kernel density estimator $h_{\boldsymbol{\theta}}$. *This is more than a simple substitution!* Under the original formulation, in redescending M-estimators the scale parameter determines a band centered on the current fit, and only the data points within this band are taken into consideration (3). The same observation is valid for the RANSAC family. If the value of the scale is chosen too large the decision about in which position does the band contain the most points will always be based at least in part on outliers. This can yield to an erroneous final estimate.

In the pbM-estimator the bandwidth $h_{\boldsymbol{\theta}}$ has a much weaker influence on the final result since it is *only* used in estimating the density of the projected points. The inlier/outlier dichotomy is determined from the shape of this density. The bandwidth is *not* the threshold for acceptance or rejection of data points in the processing. As long as the first significant local minima around the mode are related to the trough in $\mathcal{R}^p$ between the inliers and outliers, satisfactory estimates can be obtained for both $\boldsymbol{\theta}$ and α.

The MAD scale estimate of the residuals from an initial fit, $\hat{\sigma}_{MAD}$ is a frequently used choice for tuning RANSAC [15]. However, for heavily corrupted data even when based on the true fit, $\hat{\sigma}_{MAD}$ is an overestimate of the inlier noise. Compare the size of the band having as halfwidth $1.96\hat{\sigma}_{MAD}$ –the value recommended in [15]– shown at the top of Figure 1b, with the band detected by the pbM-estimator, marked by vertical bars on the density. In practice the true fit is not available and the residuals of a nonrobust total least squares (TLS) estimate are used to compute $\hat{\sigma}_{MAD}$. This yields to even more severe overestimation of the scale and to a decrease in performance is shown in Section 4.1.

The pbM-estimator requires a search in R^p for the optimal direction of projection $\hat{\boldsymbol{\theta}}$. To be computationally feasible in higher dimensions, the pbM-estimator thus must incorporate an efficient way to implement the maximization (7). This is described next.

3. Multidimensional Direct Search

Probabilistic sampling, discussed in Section 1, will be employed. For each elemental subset, i.e., a p-tuple of points in general position drawn without replacement from the data, the parameters of the model (1) are computed analytically. From $\boldsymbol{\theta}$ the value of the projection index is obtained. To refine the estimate a local search around the neighborhood of $\boldsymbol{\theta} \in R^p$ is then performed. In [13], a similar idea of refinement, the use of line search, was advocated to stabilize the performance of LMedS.

The optimization criterion (7) is nonlinear and nondifferentiable and therefore only search methods which do not rely on derivatives can be considered. Several such multidimensional unconstrained optimization techniques exist. In our context, where a search is to be performed for many elemental subsets, the ease of computations is of more concern than a high accuracy of the result. We have chosen the simplex based direct search technique proposed in 1965 by Nelder and Mead [11, Sec.10.4]. Recently significant progress was reported in the literature for this class of search methods, e.g., [9], but based on our experience there is no need to use more sophisticated (and more computationally intensive) techniques.

The vector $\boldsymbol{\theta}$ is a unit vector. The search therefore has to be restricted to the p-dimensional unit sphere in R^p. This condition is satisfied by expressing the elements of $\boldsymbol{\theta}$ in the *polar angles* $\boldsymbol{\beta} = [\beta_1\, \beta_2\, \cdots\, \beta_{p-1}]^\top$, where $0 \leq \beta_j < \pi$ for $j = 1, \ldots, p-2$, and $0 \leq \beta_{p-1} < 2\pi$,

$$\begin{aligned}
\theta_1(\boldsymbol{\beta}) &= sin\beta_1 \cdots sin\beta_{p-2}\, sin\beta_{p-1} \\
\theta_2(\boldsymbol{\beta}) &= sin\beta_1 \cdots sin\beta_{p-2}\, cos\beta_{p-1} \\
\theta_3(\boldsymbol{\beta}) &= sin\beta_1 \cdots sin\beta_{p-3}\, cos\beta_{p-2} \\
&\vdots \\
\theta_{p-1}(\boldsymbol{\beta}) &= sin\beta_1\, cos\beta_2 \\
\theta_p(\boldsymbol{\beta}) &= cos\beta_1
\end{aligned} \tag{10}$$

The search is then performed in the space of $\boldsymbol{\beta} \in R^{p-1}$.

A simplex in R^{p-1} is the volume delineated by p vertices in a nondegenerate position. For example, in R^2 the simplex is a triangle, in R^3 it is a tetrahedron. In our case, the vertices of the simplex are defined by the polar angle vectors $\boldsymbol{\beta}_k \in R^{p-1}$, $k = 1, \ldots, p$, representing p projection directions $\boldsymbol{\theta}_k \in R^p$. Each vertex is associated with the corresponding value of the projection index

$$g_k = g(\boldsymbol{\beta}_k) = h_{\boldsymbol{\theta}_k} \max_x \hat{f}_{\boldsymbol{\theta}_k}(x) \tag{11}$$

i.e., the value of the function to be maximized. We can always consider that initially the vertices are labeled such that $g_1 \leq g_2 \leq \cdots \leq g_p$.

The Nelder-Mead method is a heuristic to improve the least favorable value of the optimization criterion, g_1 in our case. This is achieved by finding a new location $\boldsymbol{\beta}_1'$ for the vertex $\boldsymbol{\beta}_1$, such that $g(\boldsymbol{\beta}_1) < g(\boldsymbol{\beta}_1')$. First, $\bar{\boldsymbol{\beta}}$ the centroid of the *other* vertices, $\boldsymbol{\beta}_k$, $k = 2, \ldots, p$, is computed. The new location is derived by one of the following operations along the direction $\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_1$: reflection, expansion, contraction. The (uninvolved) technical details are not be presented here due to lack of space, please consult one of the mentioned references.

The update procedure is now repeated for the new simplex having the vertices $\boldsymbol{\beta}_1', \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p$. The search evolves

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

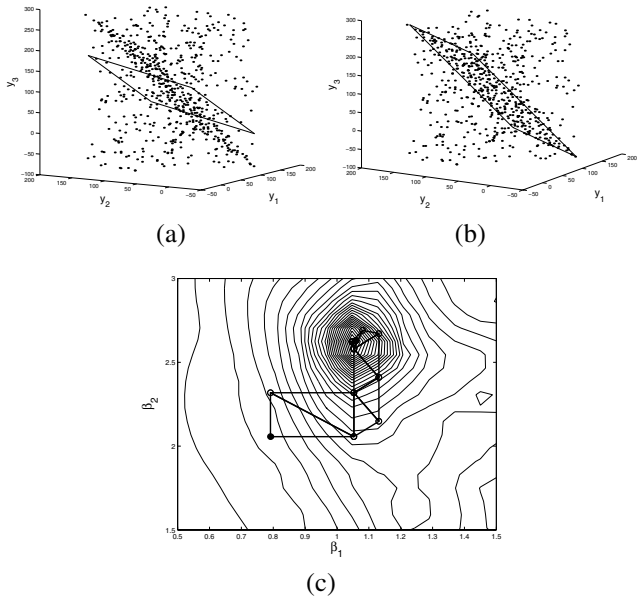


Figure 2: The role of direct search. (a) 3D data and a plane estimated from an elemental subset. (b) The 3D data with the final plane estimate. (c) The evolution of the simplex based search in the (β_1, β_2) space of polar angles. The filled circles mark the input and the output of the search.

iteratively till the variance of the function values associated with the vertices in the current simplex and/or the volume of the simplex falls below a threshold. In our context the stopping criterion is of lesser influence. Since most of the computational effort is spent at the final steps of the search, in all our experiments we limited the number of iterations to 25. At convergence, the vertex associated with the largest projection index value is taken as the output of the search.

The initial simplex is derived from the polar angle $\boldsymbol{\beta}$ representing the projection direction $\boldsymbol{\theta}$ defined by the elemental subset. The remaining $p-1$ vertices are then

$$\boldsymbol{\beta}_k = \boldsymbol{\beta} + \mathbf{e}_k * \frac{\pi}{12} \qquad k = 1, \ldots, (p-1) \tag{12}$$

where $\mathbf{e}_k \in R^{p-1}$ is a vector of 0-s except a 1 in the k-th element. Since $\boldsymbol{\theta}$ is a projection direction during the search the polar angles are allowed to wander outside the limits assuring a unique representation in (10).

To illustrate the effectiveness of the simplex based direct search a 3D example was used. In Figure 2a the data and the plane yielded by an arbitrary elemental subset (three-tuple of data points) are shown. Two polar angles are needed to define a projection direction in 3D, and in Figure 2c the evolution of the search is shown. The initial simplex is the triangle at the bottom left, the starting vertex $\boldsymbol{\beta}$ being marked with the filled circle. The search converges in ten iterations to the correct estimate shown in Figure 2b. However, it should not be concluded that in general the search is always so successful.

4. Experimental Results

Three different tasks were used to assess the performance of the pbM-estimator. In all the experiments, when the pbM-estimator is compared with estimators from the RANSAC family both use the *same* number of computational units. For RANSAC this is measured as the number of employed elemental subsets. For the pbM-estimator, to account for the iterations of the direct search procedure, it is 25 times the number of employed elemental subsets. Since often the search ends before 25 steps the number used for the pbM-estimator is actually an upper bound.

4.1. Comparison with RANSAC

In [15] the performance of different estimators in the RANSAC family is studied. It is shown that MLESAC and MSAC have very similar performance and are superior to RANSAC.

The performance of the pbM-estimator was compared with RANSAC and MSAC for eight-dimensional synthetic data obeying the EIV model (1). The 100 inlier points were normally distributed with covariance matrix $5^2\mathbf{I}_8$ around a hyperplane. The *true sample standard deviation* of the inliers σ_t was computed in each trial. A variable percentage of outliers was uniformly distributed within the bounding box of the region occupied in R^8 by the inliers. Restricting the outliers to the bounding box of the inliers is the worst case situation for the pbM-estimator.

For each percentage of outliers 100 trials were run. A trial for the pbM-estimator was based on 200 elemental subsets, thus *at most* 5000 projection index computations. A trial for RANSAC or MSAC used 5000 elemental subsets.

The output of each technique is a band containing the points *declared* inliers. The residuals of these points from the *true* fit have the standard deviation $\hat{\sigma}_{in}$. The performance of an estimator is then measured through the ratio $\hat{\sigma}_{in}/\sigma_t$. For a satisfactory result, the ratio should be very close to one.

While the pbM-estimator does not require any user set threshold, two different scale estimates were supplied to MSAC. As the best possible threshold we used $s_{opt} = 1.96\sigma_t$. To emulate the situation in practice we also used $s_{mad} = 1.96\hat{\sigma}_{MAD}$ computed in each trial based on a TLS estimate of the parameters. The graphs in Figure 3 show that for any percentage of outliers the pbM-estimator performs at least as well as MSAC tuned to the optimal scale.

4.2. Affine Motion Estimation

An M-estimator is a robust version of the total least squares (TLS) estimator, the optimal parameter estimation tech-

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

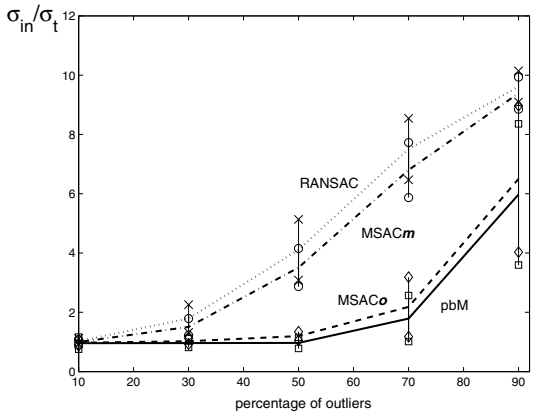


Figure 3: RANSAC vs. pbM-estimator. The relative standard deviation of the residuals function of the percentage of outliers. Eight dimensional synthetic data. The employed scale threshold: RANSAC–s_{mad}; MSAC*m*–s_{mad}; MSAC*o*–s_{opt}. The pbM-estimator has no tuning parameter. The vertical bars mark one standard deviation from the mean.

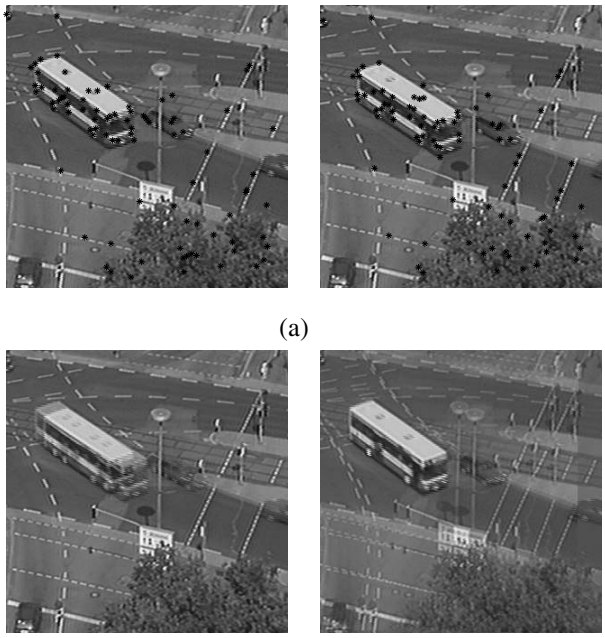
nique for the linear errors-in-variables model (1). The optimality of the M-estimator thus requires that the measurement noise $\delta\mathbf{y}_i$ be i.i.d. with covariance matrix $\sigma^2\mathbf{I}_p$. In computer vision applications this situation appears only when the elements of the vector $\mathbf{y}_i$ are *measured* quantities.

In the second experiment the problem of estimating the 2D affine transformation associated with a moving object was considered. Let $\mathbf{x}_{ko} = [x_{k1o} \;\; x_{k2o}]^\top$, $k = 1, 2$, be the (unknown) true coordinates of a pair of salient points in correspondence. The six parameter affine transformation between them

$$\begin{bmatrix} x_{21o} \\ x_{22o} \end{bmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{bmatrix} x_{11o} \\ x_{12o} \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \qquad (13)$$

can be decoupled into two three-dimensional problems, in a_{11}, a_{12}, t_1 and a_{21}, a_{22}, t_2 respectively [7], each obeying the EIV model (1). Thus, the noisy measurements of corresponding points are distributed around two planes in two different 3D spaces. The parameters of the affine transformations are estimated separately, while the points obeying the transformation must be inliers in *both* estimation processes.

The two images used in the experiment are shown in Figure 4a. Note the movement of the bus and car in the center. To obtain the data points, first a large number of point correspondences were established. As customary, a Harris corner detector was combined with correlation score maximization. Next, the point matches on the static background were identified by having zero displacement and removed. The estimation process used the remaining 87 point correspondences of which about 60% were not related to the movement of the vehicles. Since the data is low dimensional, only 10 elemental subsets, i.e., less than 250 projection index computations were needed for each pbM-estimator.



(a)

(b) (c)

Figure 4: Estimation of an affine transformation. (a) Two frames from an image sequence with marked salient points. (b) Background registration, i.e., superposition of the two images. (c) Registration using the estimated affine motion parameters.

The affine motion parameters were computed using the 22 points declared inliers by both pbM-estimators. The quality of this estimate is shown by registering the images (Figure 4c). Note the alignment of the bus and compare with the case when the background is aligned (Figure 4b).

4.3. Fundamental Matrix Estimation

The constraint capturing geometric relations in many computer vision tasks is a nonlinear function of the measurements, and to obtain an unbiased estimate of its parameters nonlinear minimization techniques must be employed. Often, the minimization starts from an initial solution computed with a TLS estimator applied to the *linearized* constraint.

Linearization is the process of embedding the nonlinear constraint into a higher dimensional space. As an example the estimation of the fundamental matrix $\mathbf{F}$ is considered. The epipolar constraint $[\mathbf{x}_{1o}^\top \;\; 1]\,\mathbf{F}\,[\mathbf{x}_{2o}^\top \;\; 1]^\top = 0$ is a bilinear expression which can be rewritten under a linear form

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

by defining the vector in $\mathbf{y}_o \in R^8$

$$\mathbf{y}_o^\top = \left[\mathbf{x}_{1o}^\top \;\; \mathbf{x}_{2o}^\top \;\; \left(\mathrm{vec}[\mathbf{x}_{1o}\mathbf{x}_{2o}^\top] \right)^\top \right] \qquad (14)$$

where the operator vec$[\cdot]$ replaces the 2×2 matrix with a column vector. The model parameters α and $\boldsymbol{\theta}$ are elements of the fundamental matrix. The linearized model however has an important difference relative to (1). It can be shown that the noise process associated with the "measurements" $\mathbf{y}$ is point dependent, i.e., heteroscedastic, and therefore the TLS estimator is no longer the optimal technique [8]. The role of the pbM-estimator is thus limited to discriminate inliers from outliers. Furthermore, the rank two constraint of $\mathbf{F}$ is also not taken into account.

First, two far apart frames from the popular *corridor* sequence (Figure 5a) were used. From the established correspondences 265 point pairs were retained. Ground truth is available for this sequence, and the histogram of the residuals (computed in 8D) is shown in Figure 5b. From the histogram 105 points were defined as inliers (Figure 5c), having the standard deviation $\sigma_t = 0.88$.

The pbM-estimator, was based on 600 elemental subsets, less than 15000 projection index computations. Using the 15000 elemental subsets, MSAC was tuned to the optimal scale and to the scale derived from the MAD estimate. The number true inliers among the points selected by an estimator and the ratio between the standard deviation of the selected points and that of the true inlier noise, are used as performance measures. The pbM-estimator clearly outperforms the MSAC in spite of being completely autonomous.

Performance Comparison - Corridor Image Pair

	selected points/true inliers	$\hat{\sigma}_{in}/\sigma_t$
MSAC (s_{mad})	219/105	42.32
MSAC (s_{opt})	98/87	1.69
pbM	95/88	1.36

The points detected by the pbM-estimator were used to obtain an unbiased fundamental matrix estimate with a nonlinear method in the package [18] discussed in [19]. The average error computed by the program relative to the ground truth was 2.4 pixels.

No ground truth is available for the *castle* sequence (Figure 6a). To have a reference fundamental matrix, first 30 reliable point matches were defined by inspection and the matrix was computed with the nonlinear estimation method. The histograms of the residuals computed in 8D from this synthesized ground truth are very similar to the ones in Figures 5b and 5c. A total of 297 point pairs were chosen from the possible matches, among which 102 were defined as inliers and 195 outliers. The inliers had $\sigma_t = 0.56$. Using the same procedure as for the *corridor* images the following results were obtained.

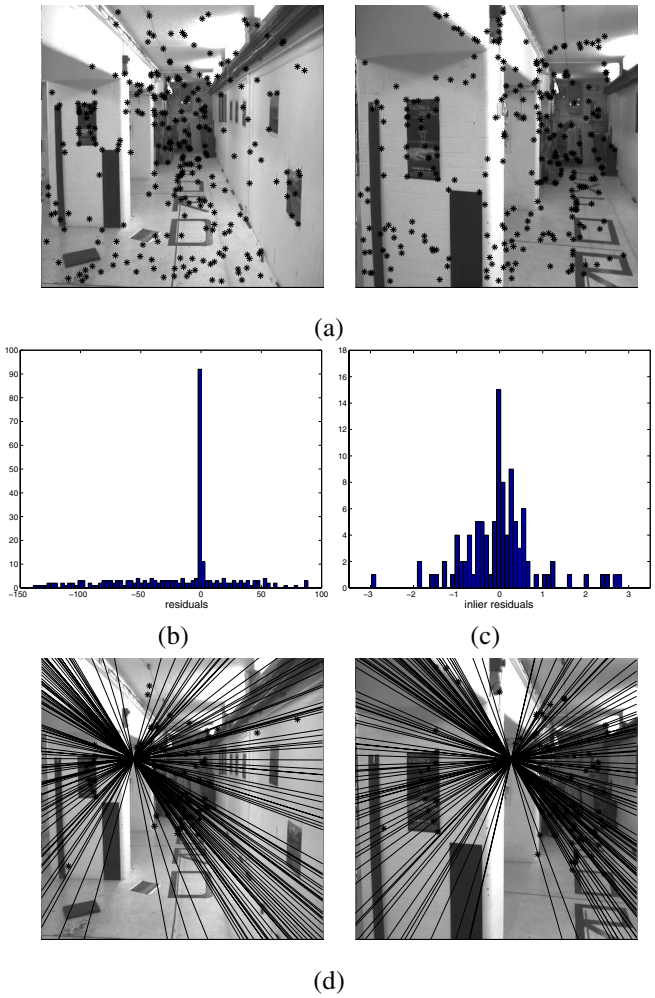


Figure 5: Estimating the epipolar geometry for two frames of the *corridor* sequence. (a) The input images with the point correspondences marked. (b) Histogram of the residuals from the ground truth. (c) Histogram of the inliers. (d) The retained points, the estimated epipoles and epipolar lines.

Performance Comparison - Castle Image Pair

	selected points/true inliers	$\hat{\sigma}_{in}/\sigma_t$
MSAC (s_{mad})	243/102	25.38
MSAC (s_{opt})	102/97	1.07
pbM	100/99	0.99

The performance of pbM-estimator exceeds that of MSAC tuned to the optimal scale which is *not available* in practice. The error associated with the nonlinear estimate computed starting from the pbM-estimate is only 0.25 pixels, probably due to the synthesized ground truth.

The influence of guided sampling was investigated by using only 200 subsets for the pbM-estimator and 5000 for RANSAC tuned with s_{opt}. When guided, the sampling was restricted to point pairs with a correlation score of at least

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

(a)

(b)

Figure 6: Estimating the epipolar geometry for two frames of the *castle* sequence. (a) The input images with the point correspondences marked. (b) The retained points, and a few epipolar lines.

0.7. Guided sampling improved the relative residual error for RANSAC from 11.73 to 2.61, and for the pbM-estimator from 3.15 to 1.32. We used a less sophisticated strategy than in [14], but for the *castle* image pair even this proved to be effective.

5. Discussion

The pbM-estimator can be used as a computational module in the robust analysis of data containing multiple structures. Prior to the analysis, however, a subset of data points must be selected in which only one structure dominates. In [1] we used for selection the *density assumption*, i.e., the local density in any inlier structure exceeds that of the outliers. The final parameter estimates were then obtained by robust clustering in the parameter space.

There are two serious limitations with that approach. First, the density assumption is clearly violated by linearized models since a mapping like (14) induces a noneuclidean metric in the linearized space where the projections are to be computed. Second, employing clustering requires that significant support was gathered for every inlier structure to be detected. This puts unnecessary burden on the data analysis. Currently we are investigating an approach which will avoid these two limitations.

Acknowledgments

The support of the National Science Foundation under the grant IRI 99-87695 is gratefully acknowledged.

References

[1] H. Chen and P. Meer. Robust computer vision through kernel density estimation. In *7th European Conference on Computer Vision*, volume I, pages 236–250, Copenhagen, Denmark, May 2002.

[2] H. Chen, P. Meer, and D. E. Tyler. Robust regression for data with multiple structures. In *2001 IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 1069–1075, Kauai, HI, December 2001.

[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, second edition, 2000.

[4] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. Assoc. Comp. Mach*, 24(6):381–395, 1981.

[5] P. J. Huber. Projection pursuit (with discussion). *Annals of Statistics*, 13:435–525, 1985.

[6] M. C. Jones and R. Sibson. What is projection pursuit? (with discussion). *J. Royal Stat. Soc. A*, 150:1–37, 1987.

[7] E.-Y. Kang, I. Cohen, and G. Medioni. Robust affine motion estimation in joint image space using tensor voting. In *16th Intern. Conf. on Computer Vision and Pattern Recog.*, volume IV, pages 256–259, Quebec City, Canada, August 2002.

[8] Y. Leedan and P. Meer. Heteroscedastic regression in computer vision: Problems with bilinear constraint. *International J. of Computer Vision*, 37:127–150, 2000.

[9] R. M. Lewis, V. Torczon, and M. W. Trosset. Direct search methods: Then and now. *J. Computational and Applied Math.*, 124:191–207, 2000.

[10] S. Z. Li. Robustizing robust M-estimation using deterministic annealing. *Pattern Recog.*, 29:159–166, 1996.

[11] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.

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

[13] D. Ruppert and D. G. Simpson. Comment on "Unmasking Multivariate Outliers and Leverage Points", by P. J. Rousseeuw and B. C. van Zomeren. *J. of Amer. Stat. Assoc.*, 85:644–646, 1990.

[14] B. Tordoff and D.W. Murray. Guided sampling and consensus for motion estimation. In *7th European Conference on Computer Vision*, volume I, pages 82–96, Copenhagen, Denmark, May 2002.

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

[16] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.

[17] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman & Hall, 1995.

[18] Z. Zhang. Software package for epipolar geometry computation. Available at `www-sop.inria.fr/robotvis/personnel/zzhang/software-FMatrix.html`.

[19] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International J. of Computer Vision*, 27:161–195, 1998.

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE