DICTA2002: Digital Image Computing Techniques and Applications, 21--22 January 2002, Melbourne, Australia

Robust Model Fitting in Pattern Recognition

Alireza Bab-Hadiashar,
Swinburne University of Technology,
School of Engineering and Science,
Hawthorn, Vic 3122, Australia.
a.bab-hadiashar@swin.edu.au

David Suter and Reyhaneh Hesami,
Dept. of Electrical and Computer Systems Eng.,
Monash University, Clayton Vic3168, Australia.
d.suter@eng.monash.edu.au
r_hesami@yahoo.com

Abstract

In this paper, we introduce a robust framework for model based parameter estimation. The framework is developed for a particular class of problems: one that contains many interesting examples in the computer vision area. The scheme leads to a new M-estimator, which, is shown to be optimal, in some sense, for this class of problems. This contrasts with previous works: these have employed least squares (known to be undermined by outliers), or have used an M-estimator without analysis to determine the robustness or efficiency of the estimator for the problem setting at hand. We illustrate the utility of this estimator and compare it with several previously employed M-estimators in the context of conic fitting.

1. Introduction

In many computer vision tasks, one wishes to extract information from images by a process that involves fitting a model to image measurements. For example, determining the location of lines, circles, conics and other shapes. Even tasks such as stereovision or motion estimation can be cast in such a framework. In the next section, we formulate an abstract fitting problem that is applicable to a wide variety of computer vision problems (including all those examples just mentioned) – and indeed, to a wide variety of problems throughout science and engineering generally.

Although the characterization of the problems, through our framework, is not essentially new (being very similar to that employed in [8] and [1], for example); we have generalised the setting somewhat (in terms of assumptions regarding the noise distribution). The most significant contribution, however, is that we then proceed to analyse the robustness of estimators for this class of problems. From this analysis, we derive an estimator that is efficient yet robust. The analysis, though rather standard in the statistics literature, has not been carried out before in such a setting. It provides a means for determining whether *any* given M-estimator is robust for this class of problems.

We conclude the paper by reporting experiments that confirm the robustness and efficiency of the estimator.

2. Robust Model Fitting

Let $x_1, \ldots x_n$ be independent *kx1* random vectors, with x_i having distribution with density $f(.;\mu_i,\Sigma_i)$ determined by the vector of means $\mu = E[x_i]$, the covariance matrix $\Sigma_i = E\left[(x_i-\mu_i)(x_i-\mu_i)^T\right]$ and a spherically symmetric density f_0:

$$f(x_i;\mu_i,\Sigma_i) = |V_i|\, f_0(\{(x_i-\mu_i)^T\Sigma_i^{-1}(x_i-\mu_i)\}^{\frac{1}{2}}) \quad (1)$$

In this expression $V_i^{-1} = \Sigma_i^{1/2}$ is a positive definite square root of $\Sigma_i = (V_i^T V_i)^{-1}$. In the classical case: $f_0(\|z\|) = \frac{1}{2}(2\pi)^{-k/2} e^{-\|z\|^2}$.

We also assume that the above observations are centred on a model determined by an unknown *l* by *1* parameter vector θ together with side conditions determined by a known *l* by *1* vector *u*:

$$\theta^T u(\mu_i) = 0 \text{ for } i = 1, \ldots, n. \quad (2)$$

These *n* conditions will be abbreviated $\theta^T u(\mu) = 0$, where $\mu = [\mu_1, \cdots, \mu_n]$ represents the *k* by *n* matrix of mean vectors. We allow the vectors θ and u to have some fixed constants in some (but not all) components. This permits us to include, in addition to terms bilinear in θ and u terms, which are linear in θ_i for some i and linear in u_i for some j. In the same manner, we can accommodate inhomogeneous constraints.

Under the assumption that the covariance matrices $\sum_i$ are known, we will derive the maximum likelihood estimator (MLE) of θ.

By defining:

$$\eta = [\theta\mu]^T \quad (3)$$

we denote the log-likelihood as:

$$l(\eta;x_1\ldots x_n)=\sum_{i=1}^{n}\left\{\ln\left|V_i\right|)+\ln(f_0(\left\|V_i(x_i-\mu_i)\right\|))\right\} \quad (4)$$

The MLE $\hat{\eta}=\eta(x_1,\ldots,x_n)$ will be found in steps, first by fixing θ and minimising the negative log-likelihood over the vector of means μ to obtain μ_θ and then minimising over θ the profile likelihood choice of

$$\eta_\theta=[\theta\mu_\theta]^T \quad (5)$$

More precisely, for each fixed θ let μ_θ be defined as:

$$\mu_\theta=\arg\min_{\mu}\left\{-l(\eta;x_1\ldots x_n):\theta^T u(\mu)=0\right\} \quad (6)$$

Further, for η_θ defined by (5), let

$$\hat{\eta}_\theta=\arg\min_{\eta_\theta}\left\{-l(\eta_\theta;x_1\ldots x_n):\theta^T u(\mu_\theta)=0\right\} \quad (7)$$

which is the MLE (η maximises 4). We determine the solution μ_θ of (6) by the method of Lagrange multipliers. To maximise $l(\eta;x_1\ldots x_n)$ in (4) subject to the constraints (model), it suffices to minimise, by choice of *y*, the term $\ln(f_0(\left\|V_i(x_i-y)\right\|)$ subject to the same constraints. The partial derivative with respect to *y* of the *i*th term is:

$$\frac{\partial}{\partial y}\left[-\ln f_0(\left\|V_i(x_i-y)\right\|)\right]=-w(\left\|z_i\right\|)\Sigma_i^{-1}(x_i-y) \quad (8)$$

where: $z_i=V_i(x_i-y)$ and $w(\left\|z_i\right\|)=\frac{f_0'}{f_0}(\left\|z_i\right\|)\frac{1}{\left\|z_i\right\|}$.

The minimiser, $\mu_i=\mu_{\theta,i}$, satisfying $\theta^T u(\mu_{\theta,i})=0$ also satisfies the equation:

$$w(\left\|z_{\theta,i}\right\|)\sum_i^{-1}(x_i-\mu_{\theta,i})=\lambda_i[\partial_x u(\mu_{\theta,i})]^T\theta \quad (9)$$

where λ_i is the Lagrange multiplier, $z_{\theta,i}=V_i(x_i-\mu_{\theta,i})$ and $\partial_x u(\mu_{\theta,i})$ is the *l* x *k* matrix of partial derivatives of *u*(*x*) with respect to *x* evaluated at $x=\mu_{\theta,i}$.

We now use (9) to obtain two different expressions for λ_i which can be used to determined the minimum value of $\left\|V_i(x_i-\mu_i)\right\|$ and hence the maximum value of (7).

First, we multiply both sides of (9) by $\sum_i^{1/2}$ and take the norm squared of each side to obtain:

$$\lambda_i^2=\frac{w^2(\|z_{\theta,i}\|)(x_i-\mu_{\theta,i})^T\sum_i^{-1}(x_i-\mu_{\theta,i})}{\theta^T[\partial_x u(\mu_{\theta,i})]\sum_i[\partial_x u(\mu_{\theta,i})]^T\theta} \quad (10)$$

To obtain another expression for λ_i we note that $\theta^T u(\mu_{\theta,i})=0$, which leads to the approximation

$$\theta^T u(x_i)\approx\theta^T[\partial_x u(\mu_{\theta,i})](x_i-\mu_{\theta,i}) \quad (11)$$

Next we multiply both sides of (9) by $(x_i-\mu_{\theta,i})^T$ and use the approximation (11) to obtain:

$$\lambda_i u(x_i)^T\theta\approx w(\|z_{\theta,i}\|)(x_i-\mu_{\theta,i})^T\sum_i^{-1}(x_i-\mu_{\theta,i}) \quad (12)$$

$$\|V_i(x_i-\mu_{i,\theta})\|^2=\frac{\theta^T[u(x_i)][u(x_i)]^T\theta}{\theta^T[\partial_x u(x_i)]\sum_i[\partial_x u(x_i)]^T\theta} \quad (13)$$

Where

$\|V_i(x_i-\mu_{i,\theta})\|^2=(x_i-\mu_{\theta,i})^T\sum_i^{-1}(x_i-\mu_{\theta,i})$.

To summarise: for fixed θ the μ_θ which satisfies (6) maximises the second term in (4), by minimising - $\sum_{i=1}^{n}\ln f_0(\|V_i(x_i-\mu_i)\|)$. The value of this minimum at μ_θ is

$$J(\theta;x_1,\ldots,x_n)=-\sum_{i=1}^{n}[\ln f_0(\|V_i(x_i-\mu_{i,\theta})\|)]. \quad (14)$$

This J is the same as the J_4 defined in (30) of [1], which is not surprising given that we have given a parallel argument using standard statistical notation and theory. One important difference is that we have shown the result can be derived for any symmetric $f_0(\|z\|)\propto\exp\{-\rho(\|z\|)\}$, and not just the normal $f_0(\|z\|)\propto\exp\{-(\|z\|^2)\}$. This will enable us to derive a robust estimator of θ.

Next define $l\times l$ matrices $A(x_i)=u(x_i)[u(x_i)]^T$ and $B(x_i)=[\partial_x u(x_i)]\sum_i[\partial_x u(x_i)]^T$ and the ratio:

$$r(x_i,\theta)=\left\{\frac{\theta^T A(x_i)\theta}{\theta^T B(x_i)\theta}\right\}^{1/2} \quad (15)$$

For an arbitrary ρ defining $f_0(\|z\|)\propto\exp\{-\rho(\|z\|)\}$ the function $J(\theta;x_1,\ldots,x_n)=\sum_{i=1}^{n}\rho(r(x_i,\theta))$. Hence the MLE $\hat{\theta}$ of θ for this f_0 is the solution of the estimating equation:

$$0=\sum_{i=1}^{n}\frac{\partial\rho(r(x_i,\theta))}{\partial r}\frac{\partial r(x_i,\theta)}{\partial\theta}$$

$$=\sum_{i=1}^{n}\psi(x_i,\theta)\ , \quad (16)$$

This is an M-estimating equation (see, e.g., Chapter 6 of [2]) with $\psi(x_i,\theta)=\partial\rho(r(x_i,\theta))/\partial\theta$, given by

$$\psi(x,\theta)=v(r(x,\theta))r^2(x,\theta)[\ldots]\theta, \quad (17)$$

Where $v(r)=r^{-1}\partial[\rho(r)]/\partial r$ and the quantity in brackets is

$$[...] = [\ \frac{A(x)}{\theta^T A(x)\theta} - \frac{B(x)}{\theta^T B(x)\theta}\ ] \tag{18}$$

Now the influence function of the M-estimator $\hat{\theta}$ is proportional to the ψ-function, which defines it, so we need to examine the boundedness (in x) of $\psi(x, \theta)$ in x. This can be achieved by choosing, for any c>0,

$$v(r) = v_c(r) = \begin{cases} 1 & r \le c \\ \frac{c^2}{r^2} & r > c \end{cases} \tag{19}$$

This leads, via $\partial[\rho_c(r)]/\partial r = v_c(r)r$, to

$$\rho(r) = \rho_c(r) = \begin{cases} \frac{r^2}{2} & r \le c \\ c^2[\ln(\frac{r}{c}) + \frac{1}{2}] & r > c \end{cases} \tag{20}$$

Note that ρ_c is continuous, and that the M-estimator M_{ρ_c} defined by $\psi_c(x_i, \theta) = \partial\rho_c(r(x_i, \theta))/\partial\theta$ Is fully efficient for the model $f_c(\| z \|) \propto \exp\{-\rho_c(\| z \|)\}$. Further, as $c \to \infty$ this bounded influence estimator approaches the classical MLE. For finite c we loose some efficiency in the classical case but gain robustness against outliers. It is an M-estimator that, for the class of problems posed, has the highest rate of growth of the influence function whilst still being bounded (in x) – and thus still robust to outliers. It is important to realise that there is a large family of possible M-estimators with the same growth rather ($O(\ln(r))$ - indeed, see below, two of the estimators in relatively common use have the same growth rate. These will also be robust for this class of problems and will be optimal when $f\ (\| z \|) \propto \exp\{-\rho\ (\| z \|)\}$. Likewise, there are existing M-estimators in use that have a lower growth rate and thus will also be robust for this class of problems.

3. Robust Conic Fitting

Parameter estimation in general is a crucial part of most vision algorithms and any significant improvement in this area is likely to enhance the quality of most machine vision tasks. Conic fitting is a useful prototypical problem to study. Firstly, the problem is easily understood, even by non-experts; secondly, the results are easy to display. Thus we have decided to examine the effectiveness of the above theory in computer vision applications by applying our results to the conic fitting problem when data are corrupted by both noise and outliers. Before we address the conic fitting problem, we explain the implementation of the introduced M-estimator.

3.1. M-estimator Implementation

The first step in implementing the M-estimator developed previously is to introduce a criterion for calculating the cut off value of c. Following standard statistical practice, the M-estimator is assumed to have 95% asymptotic efficiency in the case when the underlying noise in the data is unit Gaussian.

The efficiency of an estimator is commonly defined as the percentage ratio of the square of the Least Square scale to the robust estimate of scale: $efficiency = 100 \times \frac{\hat{\sigma}_{LS}^{\ 2}}{\hat{\sigma}_R^{\ 2}}\%.$

Huber [3], [4] showed that for monotone ψ function the scale can be estimated as:

$\hat{\sigma}_R^2 = \frac{E_F(\psi^2)}{[E_F(\psi')]^2}(X^T X)^{-1}$ where X is the matrix of all data for which the underlying population has a cumulative distribution function F. In order to obtain asymptotic efficiencies at unit Gaussian G, we have:

$$E_G(\psi^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \psi^2(x) e^{-x^2/2} dx \quad \text{, and,}$$

$$E_G(\psi') = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \psi'(x) e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \psi(x) x e^{-x^2/2} dx.$$

For the above M-estimator with $\psi = \begin{cases} x & |x| \le c \\ \frac{c^2}{x} & |x| > c \end{cases}$ the value of c turns out to be 1.812 for 95% efficiency.

Having found the cut off value for the nominated efficiency, we can rewrite our parameter estimation problem as: $\theta = \arg\min \sum_{i=1}^{n} \rho(r_i)$ where θ is the vector of parameters to be estimated, ρ is the robust function and r_i is the residual at every data point. To solve this minimisation problem, we take the derivative with respect to θ and we have: $\hat{\theta} : \sum_{i=1}^{n} \frac{\partial r_i}{\partial \theta} \frac{\partial \rho(r_i)}{\partial r_i} = 0$. We know that ψ is the derivative of ρ and $w_i = \psi(x_i)/x_i$, we can therefore write: $\hat{\theta}: \sum_{i=1}^{n} \frac{\partial r_i}{\partial \theta} r_i w_i = 0$. This is indeed the solution to the weighted least square problem represented by:

$$\theta = \arg\min \sum_{i=1}^{n} w_i r_i^2$$

To make the computation scale invariant, we divide the residuals r_i by the robust estimate of the scale, which is: $\sigma = (\underset{i}{med} \mid r_i - \underset{i}{med}\, r_i \mid) / 0.6745$. By defining $u_i = \frac{r_i}{c\sigma}$, the solution of our M-estimator is obtained by solving the following problem:

$$\theta = \arg\min \sum_{i=1}^{n} w_i r_i^2 \quad \text{where:} \quad w_i = \begin{cases} 1 & |u_i| \le 1 \\ \frac{1}{u_i^2} & |u_i| > 1 \end{cases}.$$

3.2. Robust Ellipse Fitting

Conic fitting is a frequently visited problem in computer vision due to the simplicity of the problem while still retaining the essential features of more complex fitting problems. In this study, we examine the effect of random noise and gross outliers in the estimation of parameters of an ellipse. In particular, we present errors on recovering the centre, the major and minor axes, and the slope of the major axis (the angle between the major axis and the horizontal axis).

We define an ellipse to be in the general form of: $Ax^2 + Bxy + Cy^2 - 1 = 0$ in which: $A = a\cos^2(\theta) + b\sin^2(\theta)$, $B = 2(a-b)\sin(\theta)\cos(\theta)$ and $C = b\cos^2(\theta) + a\sin^2(\theta)$ where a is the size of the major axis , b is size of the minor axis, and θ is the angle between the major axis and the x axis.

There are generally two approaches to the above problem in terms of defining the residuals for data points with respect to the ellipse. In the first approach, one considers the residual to be the vertical (along the y axis) distance between the data points and the ellipse. This distance is called the "algebraic distance". The algebraic distance is commonly used in the standard regression problem (i.e., response (y) as a function of explanatory variables x_i) and in that context, it the assumption that all the errors are in measurements of y. Although this assumption is often not realised in practice, it greatly simplifies the computation and therefore has been widely used.

The other approach to measuring the residuals is more realistic and assumes that errors are randomly and independently distributed in all directions and therefore seeks to minimise the orthogonal distance between a data point and the constraint. This distance is called the "geometric distance" [9]. The problem with this approach is that there is no closed form solution for obtaining the geometrical distance and therefore any scheme based on this measure will involve considerable computation. To overcome this problem, a first order approximation to this measure has been in use in computer vision for long time [5]. It is important to note here that our framework began with the assumption that the errors were distributed according to a spherically symmetric density and thus is a "geometric distance" setting. The first order approximation (11) used to derive the M-estimator is exactly the same as the above approximation. The approximation becomes exact if the constraint is linear (in the data X). Nonetheless, since the actual errors are never observable (for instance, one cannot "see" the error in the direction along the constraint surface as one can only "see" the error as a displacement off the constraint surface), we conducted experiments employing both of the algebraic and approximate geometric distance.

3.3. Implementation

The implementation of the introduced robust M-estimator for algebraic distance is straightforward and we have used the robust linear fitting routine of MATLAB which is the implementation of the algorithm described in [6] and [7]. To perform the parameter estimation for the approximated geometrical distance, we used the following algorithm:

1. Find the robust solution θ_0 using algebraic distance
2. Calculate the r_i^{k-1}, σ^{k-1} and u_i^{k-1} where k indicates the iteration step and r_i^{k-1} is calculated as:

$$r_i = \sqrt{\frac{(Ax^2 + Bxy + Cy^2 - 1)^2}{(2Ax + By)^2 + (Bx + 2Cy)^2}}$$

3. Solve the non-linear minimisation problem of: $\hat{\theta}_k = \arg\min \sum_{i=1}^{n} w_i^{k-1} (r_i^{k-1})^2$.
4. Test for convergence and stop if $|\hat{\theta}_{k-1} - \hat{\theta}_k| < Threshold$.

4. Experimental Results

In the following, we present the results of a Monte Carlo type study on estimating the parameters of an ellipse. In those experiments, we have compared the results of a few well-known M-estimators (including Huber, Fair, Cauchy, Geman-McClure and Andrews) with the one presented in this paper. The order by which the results of different M-estimators are presented is based on the descending slope of the associated influence function.

All the errors presented in the following tables are the mean absolute value of the errors calculated over 100 sets of data. The parameters of underlying models are the same in all of the experiments but we have randomly changed

the noise and also the locations of the outliers for every run.
The errors reported for direction, minor and major axes are all expressed as the percentage of deviation with respect to the associated value in the underlying model. Outliers are generated in our experiments by using uniform random distributors with 5 (near outliers) and 10 (far outliers) unit scale around the centre of the original ellipse. We employed the following Estimators:

LS: $\rho(r)=\frac{r^2}{2}$, Huber: $\rho(r)=\begin{cases}\frac{r^2}{2} & r\le c\\ c(r-c/2) & r>c\end{cases}$,

Fair: $\rho(r)=c^2[\frac{r}{c}-\ln(1+\frac{r}{c})]$,

M_{ρ_c} $\rho(r)=\begin{cases}\frac{r^2}{2} & r\le c\\ c^2[\ln(\frac{r}{c})+\frac{1}{2}] & r>c\end{cases}$, Cauchy: $\rho(r)=\frac{c^2}{2}\ln(1+(\frac{r}{c})^2)$, Geman-McClure: $\rho(r)=\frac{r^2/2}{1+r^2}$,

Andrews: $\rho(r)=\begin{cases}1-\cos(c\pi) & r\le c\pi\\ 2 & r>c\pi\end{cases}$.

Note: Of these, Huber has a higher growth rate than the theoretical optimum, Fair and Cauchy have the same growth rate, and Geman-MClure and Andrews have a lower than optimum growth rate.

Table 1. Comparing the mean of errors in estimating different parameters of an ellipse using Algebraic distance (without outliers)

Mean	LS	Huber	Fair	$M_{\rho c}$	Cauchy	Geman-McClure	Andrews
Direction	2.30	2.23	2.20	2.40	2.21	2.61	2.28
Major axis	1.57	1.59	1.54	1.62	1.73	2.03	1.58
Minor axis	3.18	2.93	2.93	2.72	2.74	2.61	2.75

Table 2. Comparing the mean of errors in estimating different parameters of an ellipse using Geometric distance (without outliers).

Mean	LS	Huber	Fair	$M_{\rho c}$	Cauchy	Geman-McClure	Andrews
Direction	2.19	2.60	2.20	3.11	2.58	2.57	2.36
Major axis	1.94	1.79	2.36	1.89	1.90	1.77	1.68
Minor axis	1.70	2.58	2.62	2.56	2.64	2.55	2.69

Table 3. Comparing the mean of errors in estimating different parameters of an ellipse using Algebraic distance (2% near outliers).

Mean	LS	Huber	Fair	$M_{\rho c}$	Cauchy	Geman-McClure	Andrews
Direction	35.4	2.7	3.4	2.3	2.2	2.5	2.3
Major axis	13.6	1.7	1.7	1.7	1.8	2.1	1.6
Minor axis	95.2	5.2	6.9	2.9	2.8	2.7	2.8

Table 4. Comparing the mean of errors in estimating different parameters of an ellipse using Geometric distance (2% near outliers).

Mean	LS	Huber	Fair	$M_{\rho c}$	Cauchy	Geman-McClure	Andrews
Direction	3.7	2.8	2.7	2.9	2.7	2.7	2.4
Major axis	2.1	1.9	1.9	1.9	1.9	1.9	1.9
Minor axis	1.9	2.8	2.8	2.8	2.8	2.7	2.7

Table 5. Comparing the mean of errors in estimating different parameters of an ellipse using Algebraic distance (5% far outliers).

Mean	LS	Huber	Fair	$M_{\rho c}$	Cauchy	Geman-McClure	Andrews
Direction	85	73	80	43	42	34.4	38
Major axis	21	19	20	11	11	9.1	11
Minor axis	510	325	379	183	176	132	213

Table 6. Comparing the mean of errors in estimating different parameters of an ellipse using Geometric distance (5% far outliers).

Mean	LS	Huber	Fair	$M_{\rho c}$	Cauchy	Geman-McClure	Andrews
Direction	44	44	44	44	43	43	40
Major axis	32	21	22	17.5	17.5	20.4	19
Minor axis	175	165	147	181	156	163	158

Conclusion

In this paper, we presented a theoretical analysis of parameter estimation for a class of problems that includes many typical computer vision tasks. We applied the theory to the estimation of the parameters of an ellipse where both noise and outliers were present. We have extensively tested and compared the presented method for cases where the number of outliers and their deviation from the underlying data are increasing. We can see that the presented method produced consistent estimates, which in most cases are more accurate than the previously established M-estimators. We have also shown that the presented estimator is not very much less efficient than the least square estimator when the data contains normal noise without any outlier.

Acknowledgment

The authors acknowledge the assistance of Elvezio Ronchetti and Robert Staudte [9] for the development of theoretical criteria for our robust model. The work of Ronchetti, Staudte and the authors, has been supported by the Australian Research Council.

References

[1] B. Chojnacki, van den Hengel, “Fitting surfaces to data with covariance information: fundamental methods applicable to computer vision,” Department of Computer Science, University of Adelaide TR99-03, 1999 1999.

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

[3] P. Huber, *Robust statistical procedures*: SIAM, Philadelphia, 1977.

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

[5] P. D. Sampson, “Fitting conic sections to "very scattered" data: An iterative refinement of the Brookstein algorithm,” *Computer Graphics and Image Processing*, vol. 18, pp. 97-108, 1982.

[6] R. E. W. P.W. Holland, “Robust regression using iteratively reweighted least-squares,” *Communications in Statistics: Theory and Methods*, vol. A6, pp. 813-827, 1977.

[7] F. L. O. B. W. H. DuMouchel, “Integrating a robust option into a multiple regression computing environment,” presented at *Computer Science and Statistics*: *Proceedings of the 21st Symposium on the Interface*, Alexandria, VA, 1989.

[8] K. Kanatani, “Statistical optimization and geometric inference in computer vision”, Phil. Trans. Roy. Soc. Lond., A—356, pp. 1303-1320, 1998.

[9] E. M. Ronchetti and R. G. Staudte “Robust Model Fitting and Selection in Computer Vision”, unpublished manuscript.