Modeling 2D Image Data by Robust M-estimation

V. Filova[1]*, F. Solina[2], J. Lenarčič[1]

[1]Robotics Laboratory
Jožef Stefan Institute
University of Ljubljana
Jamova 39, Ljubljana, Slovenia
Valentina.Filova@ijs.si

[2]Computer Vision Laboratory
Faculty of EE and CS
University of Ljubljana
Tržaška 25, Ljubljana, Slovenia
Franc.Solina@ninurta.fer.uni-lj.si

Abstract

The conventional least squared distance method of fitting a model to a set of data points gives unreliable results when the amount of noise in the input is significant compared with the amount of data correlated to the model itself. The theory of robust statistics formally addresses these problems and is used in this work to develop a method of separation of the data of interest from noise. It is based on iteratively reweighted least squares algorithm where Hampel redescending function is applied for weighting data. The method has been efficiently tested in modeling synthetic and real 2D image data with second order curves.

1 Introduction

The modeling process plays a key role in image understanding. The goal is to organize the data in terms of common characteristics and features and obtain concise and useful description for further processing. After selecting an appropriate model for the data the most popular approach is to use least squares analysis to estimate the model parameters. This is based under assumption that errors in the data are normally and identically distributed. However, these assumptions are frequently inappropriate. In computer vision two types of anomalities in data appear, which can not fit the Gaussian noise model: (1) a uniformly distributed error component arising from rounding and quantization, and (2) a few large deviations in the data, often thought of as outliers, indicating that all data points do not belong to the same distribution. Therefore, the least-squares (L2) estimator often gives misleading results. In order to remedy the problem of non-normal errors, new statistical techniques known as robust methods have been developed that are insensitive to such departures in the data. One the best known robust procedure is the class of maximum-likelihood-type estimators (M-estimator).

The theory of the robust M-estimator was first developed by Huber in 1964 [6]. Up to now, it has been successfully applied for general regression [10], robust curve [4] and region [2] growing, image smoothing and derivative estimation [1,7], etc. We used M-estimator build on *iteratively reweighted least squares* algorithm [1] to perform outlier rejection while modeling synthetic and real 2D image data with second order curves.

The paper is organized as follows. First, we outline the mathematical concepts of the modeling process. Next, we explain the design of the robust M-estimator. After that, we summarize our experiments. We conclude with several comments about possible applications.

*This work has been supported by the *Slovenian Ministry of Science and Technology*. The work is part of on-going research at the *Robotics Laboratory* at the Jožef Stefan Institute (IJS), Ljubljana and the *Computer Vision Laboratory* at The Faculty of Electrical Engineering and Computer Science (FER), University of Ljubljana (P2-1122-781).

2 Mathematical background

The modeling process can be defined as a *search for such parameter vector $\vec{p}$ of the model $Q(x,y) = f(\vec{p},x,y)$, which best fits the structure of the image data $Q = \{(x_1,y_1),..,(x_n,y_N)\}$* [8], where (x_i,y_i) are the 2D coordinates of the i-th edge point in the set. In this research, we have chosen the second order curve (1) to account for the structure of the data.

2.1 Model function

The general equation of second order curve $Q(x,y)$ is a linear combination of five independent basis functions $\Phi = \{x^2, xy, y^2, x, y, 1\}$:

0-7803-1772-6/94/$3.00

$$ax^2 + bxy + cy^2 + dx + ey + 1 = 0. \quad (1)$$

The residual errors can be computed as following:

$$e(x_i, y_i) = Q(x_i, y_i) - \sum_{m=1}^{5} p_m \phi_m(x_i, y_i). \quad (2)$$

The contribution of the data points to the error function (3) is not uniform. For the geometrical interpretation of the residual errors look at [9].

2.2 Robust M-estimation

A robust M-estimate for $\vec{p}$, minimizes function $\epsilon(\vec{p})$ of the deviations of the observations $Q(x_i, y_i)$ from the estimate $\hat{f}(\vec{p}, x_i, y_i)$, that is more general than the sum of squared deviations (L2 regression problem $\rho_2(x) = x^2$), or the sum of absolute deviations (L1 regression problem $\rho_1(x) = |x|$):

$$\epsilon(\vec{p}) = \sum_{i=1}^{N} \rho(\frac{e(x_i, y_i)}{s}). \quad (3)$$

Parameter s is a known or previously computed scale parameter and ρ is a robust loss function. If we let $\psi(\vec{p}, x, y) = \frac{\partial(\rho(\vec{p}, x, y)}{\partial(\vec{p})}$, then a necessary condition for a minimum is that $\vec{p}$ satisfy

$$\sum_{i=1}^{N} \psi(\frac{e(x_i, y_i)}{s}) \quad \phi_m(x_i, y_i) = 0, \qquad m = 1, 2, .., 5. \quad (4)$$

Introducing a set of weighting parameters:

$$\omega(x_i, y_i) = \begin{cases} \frac{\psi(\frac{e(x_i, y_i)}{s})}{\frac{e(x_i, y_i)}{s}} & \text{if } e(x_i, y_i) \neq 0 \\ 1 & \text{if } e(x_i, y_i) = 0, \end{cases} \quad (5)$$

the nonlinear matrix equation (4) can be rewritten as following:

$$\sum_{i=1}^{N} \phi_m(x_i, y_i) \ \omega(x_i, y_i) \ e(x_i, y_i) = 0, \qquad m = 1, 2, .., 5. \quad (6)$$

The equation (6) can be solved iteratively via several different methods, one of which is ***iteratively reweighted least squares (IRLS)***.

2.3 Iteratively reweighted least squares algorithm

Before the IRLS M-estimation scheme is applied we have to determine the weight function (look for their performance evaluation at [7]). The weighting function (see equation 5) is dependent of the scale estimate s . Although somewhat *ad hoc* the robust statistics community uses the median absolute deviation (MAD) scale estimate almost exclusively:

$$s(\vec{p}) = 1.4826 \ median(|e(x_i, y_i) - median(e(x_i, y_i))|), \quad (7)$$

where $e(x_i, y_i)$ are the residuals from the previous function fit. We compute the initial set of residuals errors by least squares (L2 regression) fit. Thus, the weighting process is govern by the scale estimate, which is updated simultaneously.

The IRLS M-estimation algorithm is as following:

- *By L2 regression compute the initial values of:*
  1. *the parameter vector $\vec{p}$ (6),*
  2. *the scale parameter s (7),*
  3. *the weighting parameters $\omega(x_i, y_i)$ (5).*
- *While $s_{new} > s_{min}$ calculate new values for:*
  1. *the parameter vector $\vec{p}$ (6),*
  2. *the scale parameter s_{new} (7),*
  3. *the weighting parameters $\omega(x_i, y_i)$ (5).*

Our experiments (see Sec. 3) and empirical studies [3] have shown that the algorithm always converges to a unique solution.

3 Experimental results

These experiments were specifically designed to meet the following goals:

- to compare the convergence speed of two different weighting functions: *Huber Minimax* and *Hampel redescending*,
- to test the *Hampel redescending* M-estimator on different quantities of noise,
- to demonstrate the power and potentional application of robust M-estimaton in outlier rejection and data modeling.

As we have already mentioned, we used second order curve to account for the structure of the data (synthetically generated and real), where elliptical shape parameters $(A, B, X_0, Y_0, \varphi)$ [9] are to be estimated.

3.1 Comparative study of two weighting functions

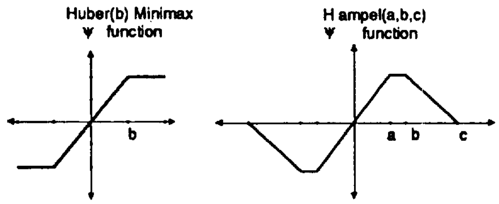

Figure 1: Two different weighting functions for robust regression.

On Fig. 1 two different weighting functions can be seen. *Huber Minimax* has been introduced by Huber [6] to yield the efficiency of L2 regression with the robustness of L1 regression. When *Huber Minimax* is used, the expression for the weights (5) becomes:

$$\omega(x_i, y_i) = \begin{cases} \frac{bs}{|e(x_i, y_i)|} & \text{if } |e(x_i, y_i)| > bs \\ 1 & \text{if } |e(x_i, y_i)| \leq bs. \end{cases} \qquad (8)$$

This form shows explicitly that all "inliers" receive unit weights $\psi(x) = 1$ for $x \leq b$ whereas questionable data points are "down weighted" proportional to the magnitude of the residual error. Although this is desirable, the weights are never exactly zero unless the residuals are infinite. In addition, the *Hampel redescending* function (see Fig. 1) has the property that $\psi(x) = 0$ for $x > c$ where c is a preselected cutoff value, also known as the finite rejection point. This allows outlier rejection (total elimination of the effects of data points with very large residual errors).

Experiment 1: On Fig. 2A we can see the results of applying *Huber Minimax* M-estimator and *Hampel redescending* M-estimator on synthetically generated ellipsoid shape data with 15 % of noise. The final solution (14 iterations) depends grossly on the initial fit obtained with L2 regression and the scale estimate s. When the data is less noisy (9%) and the initial fit better (see Fig. 2B), *Hampel redescending* M-estimator very fast converges to right solution (only 8 iterations were needed).

The experiment (see Fig. 2) demonstrates the high convergence speed of the *Hampel redescending* M-estimator due to outlier rejection. Therefore, we chose *Hampel redescending* function for weighting data in further research [5].

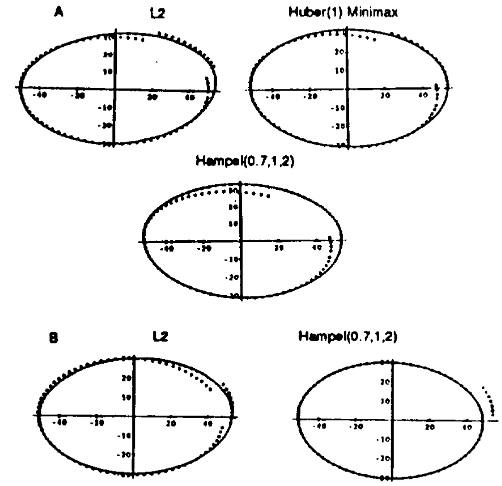

Figure 2: Experiment 1.

3.2 Behavior of Hampel redescending M-estimator

Experiment 2: This experiment was carried out to test the behavior of the robust M-estimator built on *Hampel(0.7,1,2) redescending* function while incrementing the percentage of noise in data. The tuning parameters ($a = 0.7, b = 1, c = 2$) have been chosen as in [1]. Besl [1] stated that the detailed shape of the weighting function is not as important as the cutoff value c. Two ellipsoid shape data clusters *Ell1* and *Ell2* were generated synthetically. The data cluster *Ell1* was to be modeled, while the data cluster *Ell2* was treated as noise. The noise in data was incremented in five steps from 21% to 100%. Tab. 1 lists the results obtained after 5 and 10 iterations.

The experiment demonstrates the ability of the estimator to cope with 100% bad data. Thus, it reaches the best possible value for breakdown point $\epsilon^* = 50\%$ (***breakdown point*** is the limiting fraction of arbitrarily bad data that the estimator can cope with [1]).

3.3 Hampel redescending M-estimator applied to real data

Experiment 3: In Fig. 4 the results of modeling the arm part projections with second order curves can be observed. The problem of separating one part from another is efficiently solved by considering each part as an individual object and using Hampel redescending M-estimator to perform outlier rejection.

Noise [%]	Method	# iteration	Estimated parameters					Rejected data [%]	Scale parameter s
21	L2	1	157.5	35.0	355.0	238.0	-13.7	0	0.001326
21	Hampel(0.7,1,2)	5	140.1	35.0	364.9	237.5	-15.5	27	0.0
44	L2	1	173.7	40.0	310.3	235.3	-8.25	0	0.008216
44	Hampel(0.7,1,2)	5	143.0	34.9	363.9	237.7	-15.5	51	0.000111
67	L2	1	151.6	52.3	302.2	220.7	-5.9	0	0.011422
67	Hampel(0.7,1,2)	5	192.1	32.9	350.8	239.8	-12.6	50	0.002090
90	L2	1	159.7	61.3	289.4	210.0	-4.0	0	0.017187
90	Hampel(0.7,1,2)	10	145.9	34.7	363.5	237.8	-15.4	80	0.000519
100	L2	1	184.3	57.6	265.8	212.8	-1.5	0	0.022112
100	Hampel(0.7,1,2)	10	143.5	34.8	364.5	237.6	-15.6	88	0.000387

Table 1: Experiment 2.

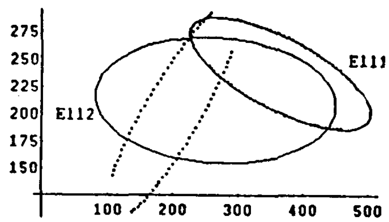

Figure 3: Experiment 2.

4 Conclusions

A robust M estimator has been described and a few aspects of its performance have been demonstrated. It was build on *iteratively reweighted least squares* algorithm, where weighting is accomplished with *Hampel redescending* function. The scale parameter, which depend on the parameter vector to be estimated is updated simultaneously yielding better convergence speed. The estimator has the best possible *break down* point $\epsilon^* = 50\%$. We demonstrated the power of the estimator in outlier rejection and data modeling. We emphasized the potentional application of the approach in data segmentation.

References

[1] P. J. Besl, J. B. Birch, L. T. Watson, "Robust window operators", *Proc. of the 2nd International Conference on Computer Vision*, Tampa, Florida, pg. 591-600, IEEE Computer Society Press, Washington.D.C, 1988

[2] P. J. Besl, R. C. Jain, "Segmentation via variable-order surface fitting", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 2, 1988

[3] J. B. Birch, "Some convergence properties of iteratively reweighted least squares in the location model, *Commun. Stat.* , pg. 359-369, 1980

[4] D. S. Chen, "A data-driven intermediate level feature extraction algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pg. 749-758, 1988

[5] V. Filova, "Software package for modeling 2D image data", Technical report IJS DP-6735, Jožef Stefan Institute, Ljubljana, Slovenia, 1992

[6] P. J. Huber, "Robust statistics", New York, Wiley, 1981

[7] M. J. Mirza, 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*, Vol. 9, No. 1, pg. 75-85, 1993

[8] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling "Modeling of Data", Ch.14 in the book "Numerical Recepies in C: The art of scientific computing", Cambridge University Press, Cambridge, 1988

[9] R. Safaee-Rad, K. C. Smith, B. Benhabib, "Accurate estimation of elliptical shape parameters from a grey-level image", *10th International Conference on Pattern Recognition*, Atlantic City, New Jersey, USA pg. 20-6, IEEE Computer Society Press, Washington.D.C, 1990

[10] X. Zhuang, T. Wang, P. Zhuang, "A highly robust estimator through partially likelihood function modeling and its application in computer vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 1, pg. 19-35, 1992

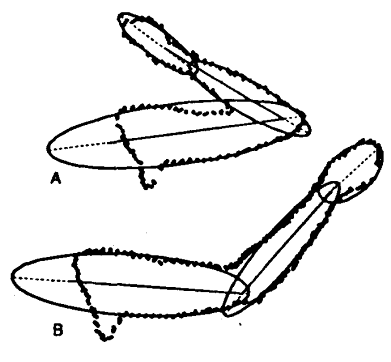

Figure 4: Experiment 3.