A Highly Robust Regressor and its Application in Computer Vision

Ebroul Izquierdo
Electronic Engineering Department
Queen Mary and Westfield College, University of London
London E1 4NS, UK
e.izquierdo@elec.qmw.ac.uk

Abstract

This paper introduces a highly efficient model for general regression with unknown error distribution function. The model is derived from a bi-criteria optimisation problem combining the best properties of least squares and least absolute deviation. The solution of this problem leads to a robust M-estimator that can be applied in a large range of computer vision tasks. The technique has been designed to overcome the extreme lack of robustness and low efficiency observed when conventional approaches are used to solve fundamental ill-posed computer vision problems. The performance of the method has been assessed by recovering the 3D-scene structure from stereoscopic images. In this context, several experiments have been conducted. Some selected results are reported in this article.

1 Introduction

Fitting a regression curve to observation data is a basic technique used to solve a large variety of computer vision tasks. Due to the noisy and discrete nature of digitised images, all low-level image analysis methods developed so far are subject to errors. Thus, a realistic assumption about the errors in the results obtained by any feature extractor, motion (correspondence) estimator, object recogniser, etc., must be used. The most commonly used approach assumes a Gaussian distribution of errors. Under this assumption the least-squares estimator (LS) achieves optimal results in the sense of being minimum variance unbiased and maximum likelihood. The problem with this estimator is its extreme sensitivity to deviations from the Gaussian model. Since outliers resulting from large observation errors, or noisy cannot be avoid, it is necessary to consider robust procedures modifying the LS schema. The most known robust estimators are the class of maximum-likelihood type estimators (M-estimators) and the median based estimators (repeated median and least median of squares). The last two belong to the class of robust regressors with the highest possible breakdown point (50%). The breakdown point is defined as the maximal fraction of outliers the estimator can deal with [4]. Unfortunately, median based estimators suffers of extremely low efficiency even if additional strategies are used to reduce the computational cost [12].

BMVC 2000 doi:10.5244/C.14.44

Other robust estimators such as the L-estimator and the R-estimator have been also used to fit models to data obtained from digital images. Nevertheless, the lack of robustness and efficiency is the common drawback of these techniques.

In this paper a Parametric Model Fitting estimator (PMF) for general linear regression is presented. The motivation behind the presented research has been the absence of an estimator that perform robustly, is easy to handle and has lower algorithmic complexity. The PMF-model is obtained by using a convex combination of least squares and least absolute deviation (LAD). A one-parameter quadratic programming model is used to derive a solution to the problem of fitting a curve to given observation data. The set of efficient points for such a problem is constructed by solving the linear complementary problem. In this way a set of linear regressors is generated. One of them is selected by evaluating the computed coefficients of determination.

The introduced technique is assessed by several experiments in the contex of camera calibration and structure from stereo. The goal of this task is to recover the camera location with respect to the scene and to use these parameters to estimate the scene structure. The 3D structural information can be then used to generate a 3D model of the original scene [7]. Currently, there exists a large collection of application domains for these technologies including: medical imaging, augmented reality, immersive telepresence, etc. Other applications of the PMF-model including motion and disparity estimation, image segmentation, etc., are envisaged in further developments [5].

2 The Mathematical Model

We consider the classical linear regression model given by $y_i = x_{i1}\theta_1 + \ldots + x_{im}\theta_m + e_i$ $i = 1,\ldots,n$ where the location parameter $\Theta = (\theta_1,\ldots,\theta_m)^T$ has to be estimated. Usually the error distribution function F is unknown. Nevertheless, most authors assume that the error distribution is Gaussian or Laplacian or some fixed combination of both. Unfortunately, the errors in data obtained from digital images rarely satisfy one of these assumptions. For this reason a more efficient strategy should focus of the definition of an estimator which is optimal under both LS and LAD simultaneously. Following this argument we assume that the true distribution of the errors is some combination of the standard normal η and the Laplace distribution λ. Thus,

$$F_\xi = (1-\xi)\eta + \xi\lambda ,$$

where ξ is the degree of contamination of the Gaussian model. In the sequel we assume that the design variables x_{ij} have a marginal distribution P_d on the set of Borel sets on $\Re^m$. Furthermore, we suppose that the support of P_d is not contained in a plane of dimension less than *m*, i.e., the *m* variables are needed to guarantee the fitting goodness of the regression curve. The regression distribution is defined as a probability distribution P on the Borel sets $\Omega(\Re^{m+1})$ of $\Re^{m+1}$. If P_d satisfies this hypothesis the corresponding equation belongs to a set $S(\Re^{m+1})$ where the regression estimator $\rho : S(\Re^{m+1}) \to \wp$ is defined. $\wp \subset \Re^{m+1}$ is the regression parameter space, where $\hat{\Theta} = (\theta_1,\ldots,\theta_m,\sigma^2)^T = (\Theta, \varsigma) \in \Re^m \times \Re^+$ represent the location and scale parameters to be estimated. The model is known up to the parameter $\hat{\Theta}$.

According to the arguments stated above we formulate the following bi-criteria optimisation problem

$$\min_{\Theta} \left\{ \|e\|_2, \|e\|_1 \right\}, \tag{1}$$

where $\|.\|_k$ represents the k-norm. This problem represents the search for an estimate such that there is not a better simultaneously under LS and LAD criteria. The minimiser is approximately a ML estimator if the distribution of the errors F is adequately specified. The so stated optimisation problem will not always have a unique solution. Consequently some compromise between the goodness of both distributions should provide the selection rule. For this aim, the linear regression model which solves the corresponding Parametric Quadratic Programming Problem (PQP) is used. The calculated solution of the PQP will be denoted as "Compromise Estimate" (CE). Consequently, the parametric objective function that should be considered during the search for a CE is defined as:

$$\varphi(e) = (1-\xi)\sum_{i=1}^{n} e_i^2 + \xi \sum_{i=1}^{n} |e_i| .$$

The bounding condition $|e_i| \le \delta$ is imposed to all the residuals, to guarantee that the corresponding influence function of the estimator also becomes bounded. The defined M-estimator is not invariant with respect to magnification of the error scale. Thus, a scale parameter has to be estimated simultaneously. That is, the final cost function defining the model will be given by

$$\phi(e,\varsigma) = (1-\xi)\sum_{i=1}^{n} e_i^2 / \varsigma + \xi \sum_{i=1}^{n} |e_i| / \varsigma .$$

Estimating simultaneously Θ and ς is not trivial. Following the classical procedure introduced by Huber, we use an iteration scheme to solve the underlying problem.

3 Estimation of the Regression Parameters

The complete set of regression parameters are approximated iteratively by alternating the calculus of the location and scale parameters. The initial approximation of the location is obtained by solving the LS problem

$$\min_{\Theta} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} x_{ij}\theta_j)^2$$

subject to $\left| y_i - \sum_{j=1}^{m} x_{ij}\theta_j \right| \le \delta, \quad i = 1,\ldots,n$.

Let be $\Theta^{(0)}$ the solution of this conventional optimisation problem. The initial scale estimate is calculated from the sample median of the initial residuals $e_i^{(0)} = y_i - \sum_{j=1}^{m} x_{ij}\theta_j^{(0)}$ obtained using the initial location parameters: $\varsigma^{(0)} = med\left\{ e_1^{(0)},\ldots,e_n^{(0)} \right\}$.

Next, a set of parameters $\xi_1,...,\xi_k$ with $\xi_l \in [0,1]$ for all $l = 1,...,k$ is defined. Without loss of generality we assume that $\xi_1 < \xi_2 <,...,< \xi_k$. Fixing $\xi = \xi_1$ the following parametric programming problem is solved

$$\min(1-\xi)\sum_{i=1}^{n} e_i^2 + \xi \sum_{i=1}^{n} |e_i| , \tag{2a}$$

subject to:

$$\sum_{j=1}^{m} x_{ij}\theta_j - \frac{e_i}{\varsigma^{(0)}} \leq y_i \ , \quad \sum_{j=1}^{m} x_{ij}\theta_j + \frac{e_i}{\varsigma^{(0)}} \geq y_i \text{ and } e_i \leq \varsigma^{(0)}\delta \ . \tag{2b}$$

Let $e^{(1)}$ and $\Theta^{(1)}$ the solution of (2). A new scale estimate is obtained as

$$\varsigma^{(1)} = med\left\{e_1^{(1)},...,e_n^{(1)}\right\}.$$

If the difference $\varepsilon = \left|\varsigma^{(0)} - \varsigma^{(1)}\right|$ exceeds a prefixed threshold B (e.g., $\varepsilon > B = 10^{-4}$), (2) is again solved taken $\varsigma^{(1)}$ as scale estimate. This iteration is repeated as long as $\varepsilon > B$ or a determinate number of iterations has been carried out. At the end of this iteration process an estimate $(\Theta^{(p)}, \varsigma^{(p)})$ for location and scale is available. Using this estimate, the determination coefficient $K^{(1)}$ is calculated using the following formula:

$$K^{(1)}(\xi) = \frac{(1-\xi)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \xi\sum_{i=1}^{n}|y_i - \hat{y}_i|}{(1-\xi)\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 + \xi\sum_{i=1}^{n}|y_i - \bar{y}_i|}$$

where $\hat{y}_i$ denotes the predicted value of y_i and $\bar{y}_i$ the mean of the observations.

The goodness of the estimated parameters is verified by evaluating this coefficient. If the value of $K^{(1)}$ is not sufficiently close to 1, the whole process is again carried out for $\xi = \xi_2$, but taking $(\Theta^{(p)}, \varsigma^{(p)})$ as initial estimates for location and scale. The first estimator giving a determination coefficient sufficiently close to 1 is selected as regressor. If all determination coefficients $K^{(1)},...,K^{(k)}$ are not sufficiently close to 1, then the estimator with determination coefficient closest to 1 is selected.

To solve the parametric programming problem (2) a standard approach is used. Basically we want to solve the bi-criteria optimisation problem (1). It is well known that the set of efficient points of (1) coincides with the set of optimal solutions of the corresponding parametric programming problem [8]. For $\xi \neq 0$ (2) can be written as

$$\min \sum_{i=1}^{n} e_i^2 + \tau \sum_{i=1}^{n} |e_i| , \tag{3}$$

subject to (2b) and with $\tau + \xi \big/ 1-\xi$.

Using the Karush-Kuhn-Tucker conditions we can fix the structure and properties of the corresponding linear complementary problem (LCP). Due to the evident convexity of the objective function, solving (3) is equivalent to obtain the solution of the linear complementary problem. Simplex techniques are proposed by several authors to solve such parameter-dependent LCP tasks [1], [2]. We use the parametric principal pivoting algorithm proposed by Cottle to solve the LCP. A study about efficiency of this Simplex-based technique can be found in [2].

4 Using the PMF-Model to Recover the Epipolar Geometry in StereoVision

Let us assume two perspective images of a rigid scene taken with cameras that can be accurately approximated by a pin-hole camera model. Let (X,Y,Z) be the 3D coordinates of a visible point $\hat{p}$ of the scene in the camera coordinate system, where the *Z*-axis coincides with the optical axis, the optical center *O* lies at the origin and *f* denotes the distance between image plane and optical center (see Fig. 1). Let us define the image coordinate system (u,v), with the origin at the point $(0,0,f)$ with respect to the 3D camera coordinate system and the image plane coinciding with the plane *Z=f.* Thus, the 2D projection p of the 3D point $\hat{p}$ onto the image plane has coordinates $p=(u,v)$ (see figure 1). The coordinate system used in the computer or in the respective digitized image will be denoted (*x, y*). In this coordinate system *the principal point* (x_0,y_0) gives the position of the optical center. The cameras are first calibrated independently in order to fix the intrinsic parameters. This process is carried out using a well established method due to Tsai [10]. It uses a single view of coplanar or non-coplanar points to find the focal length, two coordinates of the image center, the first order radial lens distortion coefficient and the uncertainty factor for the scale of the horizontal scan line. In our system we use the version implemented by Willson [11].

Let us now consider the stereoscopic rig shown in Fig. 1, with the two cameras having optical centers at O_l and O_r, respectively. For any point $\hat{p}$ in the 3D space, the plane (O_l, O_r, $\hat{p}$) is called epipolar plane of $\hat{p}$. It intersects the image planes in two conjugate lines called epipolar lines. The points e_l and e_r of intersection of the image planes with the base line (O_l, O_r) are called the epipoles of the stereo rig. The epipolar planes form a pencil of planes through (O_l, O_r) and the epipolar lines form two pencils of lines through e_l and e_r respectively. If a point pair p_l and p_r are corresponding points, then O_l, O_r p_l and p_r must lie in the same plane. This is the basic property involved in the epipolar geometry. This property is called the *co-planarity constraint*.

The relationship between the two corresponding points p_l and p_r can be formulated as:

$$Z_r M_r^{-1} X_r = Z_l R M_l^{-1} X_l + T \,, \tag{4}$$

with *M* the camera intrinsic matrices, *R* the rotation matrix from left to right camera coordinate systems, *T* the translation vector from the origin of the left camera coordinate system to the origin of the right camera coordinate system, and $X=(x,y,1)^{\mathsf{T}}$ the extended vector with the coordinates of the sampling position considered in the digitized image. In equation (4) the indexes *l* and *r* indicate left and right camera systems respectively. If the camera intrinsic parameter are known, (4) can be expressed in normalized coordinates. Taking $U_l = (u_l, v_l, 1)^T = M_l^{-1} X_l$ and $U_r = (u_r, v_r, 1)^T = M_r^{-1} X_r$ (the normalized image coordinates), (4) can be written as:

$$Z_r U_r = Z_l R U_l + T \ . \tag{5}$$

The cross-product of (5) with the translation vector $T = (t_1, t_2, t_3)^T$ followed by the inner product with U_r yields to

$$U_r E U_l = 0 \ , \tag{6}$$

with $E = \tau \cdot R$, $\tau = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}$ a skew symmetric matrix constructed with the translation vector. The relation (6) is called the epipolar equation. It reflects the co-planarity constraint or the fact that the three lines (O_l , O_r), (O_l , p_l) and (O_r , p_r) lie in the same plane (see Fig. 1). The singularity of τ implies *det(E)=0.* Moreover, the property *rank(E)=2* leads to the so-called *rank 2 constraint*. The epipolar equation was discovered by Longuet-Higgins [9] in the early eighties. It defines the epipolar geometry which comprises all about a stereo rig. It allows 3D reconstruction of the scene to be carried out if the intrinsic camera parameters are known and it reduces the search for corresponding points, constraining it from the entire second image to a single line.

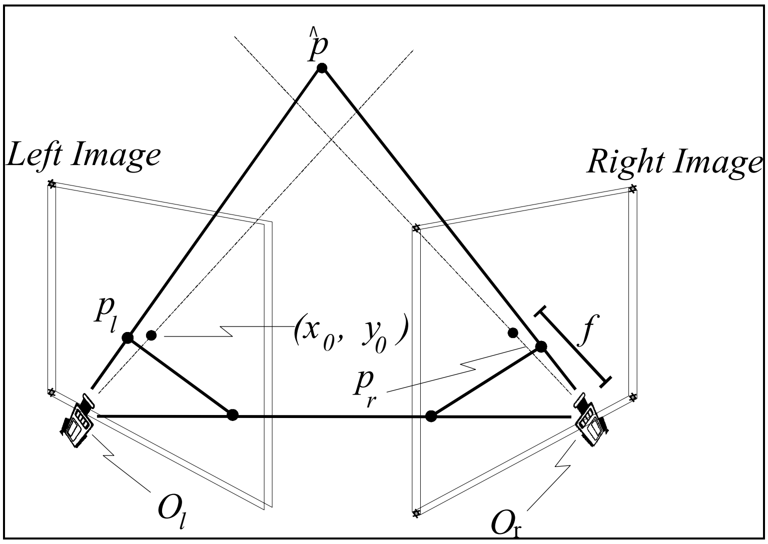

Fig 1: Stereo rig and underlying epipolar geometry

The crucial step in the process of recovering the extrinsic camera parameters and the epipolar geometry is estimation of the essential matrix from (6). The epipolar equation (6) can be written as an homogeneous linear equations in the nine unknown elements of *E:* $h^T \hat{E} = 0$, with $h = (u_l u_r, u_r v_l, u_r u_l v_r, v_l v_r, v_r u_l, v_l, 1)^T$ a vector defined by the corresponding points U_l, U_r and $\hat{E}$ the nine-dimensional vector, whose elements are the coefficients of the essential matrix. Given *n* corresponding points we obtain a system of linear equations of the form:

$$H_n \hat{E} = 0 \,. \tag{7}$$

Since $\hat{E}$ is defined up to a scale factor eight corresponding points are sufficient to calculate the essential matrix. Moreover, $rank(H_n) = k$ and *k<9*, because in other case (7) only posses the trivial solution. Estimating the epipolar geometry consists of estimating the null space $\mathrm{N}(H_n)$ of H_n. Usually, more than eight corresponding points are known but they are only estimates of the exact corresponding points. For this reason direct techniques become ineffective, because the matrix H_n build from the inexact data is not longer singular, but ill-conditioned or of deficient rank. This mean that in practice $\mathrm{N}(H_n) = \{0\}$ and *k=9*. Consequently, solving (7) becomes equivalent to solving the optimization problem:

$$\min \left\| H_n \hat{E} \right\| , \tag{8}$$

for any predefined norm $\|.\|$. Several methods for solving this problem can be found in the literature, for a comprehensive review of them we refer to [12]. Most of the conventional approaches assume than the errors in the input data is Gaussian distributed with zero mean, and use the $L_2 - norm$ to solve the problem. As mentioned before the least squares method is extremely unstable because the distribution of errors is not Gaussian and the initial data contain outliers. For this reason we expect to find a good approximation $\tilde{E}$ of E using the proposed PMF-model. The final approximation of the essential matrix is obtained by imposing the rank 2 constraint to $\tilde{E}$. This last procedure is carried out by applying the conventional truncated single value decomposition (TSVD) to $\tilde{E}$.

Fig. 2: Original stereo images from scene GWEN.

To reinforce the numerical stability of the proposed method, the input data is first normalized using the technique proposed by Hartley [3]. Summarizing, the essential matrix is estimated according to the following algorithm:

- Normalize the input data by applying Hartley's method [3]
- Use the PMF-model to solve the minimization problem (8) and use the estimated values of $\hat{E}$ to generate an approximation $\tilde{E}$ of the essential matrix
- Apply the TSVD technique to determine the matrix E_2 of rank 2 and closest to $\tilde{E}$
- Take the essential matrix as $E = E_2$.

Once the essential matrix is known the extrinsic camera parameters can be estimated using a well-established technique proposed by Longuet-Higgins [9]. Since $E^T T = 0$, T can be found up to a factor scalar by solving: $\min_T \| E^T T \|$ subject to $\| T \| = 1$. Thus, T is the unit eigenvector of EE^T corresponding to the smallest eigenvalue. The rotation matrix is estimated by solving: $\min_R \| E - \tau . R \|^2$ subject to $R^T R = I$ and *det(R)=1*. Notice that once these optimization problems have been solved sign ambiguities have to be considered. Finally, the estimation of the 3D coordinates of a point $\hat{p}$, given the location of its projections p_l and p_r in the left and right image planes and the extrinsic camera parameters is straightforward using the relation (4).

Fig. 3: Epipolar lines estimated with the proposed parametric model fitting estimator.

5 Selected Results

The proposed technique has been evaluated by estimating the epipolar geometry from several stereoscopic scenes. Currently, we are also performing several comparisons with results obtained using other previously reported methods. A more comprehensive report of this comparative evaluation is in preparation [5]. This evaluation includes results obtained using different technique form the literature as well as a direct comparison with a method reported recently by the same author [6]. Although, the method introduced in [6] was designed to satisfy physical meaningful constraints inherent to the calibration problem, the new algorithm based on the more general PMF-model supplies very similar results. In this article results obtained for the scene GWEN are reported. The original stereoscopic image GWEN is shown in Fig. 2. The epipolar lines obtained with the proposed algorithm are shown in Fig. 3. In this representation both images are shown in the background. Superimposed on the left image the matched points are highlighted, on the right image the epipolar lines corresponding to these points are drawn. In Fig. 4 the epipolar lines obtained using the parametric linear optimization problem described in [6] are shown. The differences between the results of the two methods are absolutly insignificant.

6 Summary and Further Work

A highly efficient and robust model to approximate general regression problems with unknown error distribution function has been presented. The motivation behind the presented research has been the absence of general robust and low complexity estimators for computer vision tasks. The presented model is derived from a bi-criteria optimisation problem combining the best properties of least squares and least absolute deviation. The performance of the method has been assessed by recovering the 3D-scene structure from stereoscopic images. In this context, several experiments have been conducted. Some selected results are reported in this article. Further developments include a comprehensive assessment of the introduced regressator and its suitability for other computer vision related tasks like motion and disparity estimation, segmentation, etc.

Fig. 4: Epipolar lines estimated with the linear parametric regularisation technique introduced previously in [6].

References

[1] S. Allende, S. and C. N. Bouza, "A parametric mathematical programming approach to the estimation of the coefficients of the linear regession model", *Parametric Optimization and Related Topics III,* (Guddat, et.al. Eds.) Verlag P.Lang. Frankfurt am Main, 1993, pp 9-20.

[2] T. S. Arthanari and Y. Dodge, “Mathematical Programming in Statistics”, Wiley. N.York, 1981.

[3] R. I. Hartley, “In defense of the eight-point algorithm“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997, pp. 580-593.

[4] P. J. Huber, “Robust Statistics” Wiley, N.York, 1981.

[5] E. Izquierdo, “Multiple Linear Regression for Computer Vision Applications”, in preparation.

[6] E. Izquierdo, “Video-Based Camera Registration for Augmented Reality”, Proc. Int. Conf, on Information and Visualization IV'2000, SOAS, University of London, 19-21 July 2000.

[7] E. Izquierdo and X. Feng, “Modeling Arbitrary Objects Based on Geometric Surface Conformity“, *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Synthetic/Natural Hybrid Video Coding*, Vol. 9, No. 2, Mar. 1999, pp. 336-352.

[8] K. Lommatzsch, “Anwendungen der Linearen Parametrischen Optimierung”, Akademie Verlag, Berlin, 1979, (In German).

[9] H. C. Longuet-Higgins, “A computer algorithm for reconstructing a scene from two projections“, *Nature 293,* 1981, pp. 133-135.

[10] R. Tsai, “A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses“, *IEEE Journal of Robotics and Automation,* Vol. 3, 1987, pp. 323-344.

[11] R. Willson, “Tsai camera calibration software“ - *release 3.0b3*, Online available: http://www.cs.cmu.edu/afs/cs.cmu.edu/user/rgw/www/TsaiCode.html, 1995.

[12] G. Xu and Z. Zhang, “Epipolar Geometry in Stereo, Motion and Object Recognition“, *Kluwer Academic Publishers*, London, 1996.