Robust Techniques in Least Squares-based Motion Estimation Problems *

Raúl Montoliu and Filiberto Pla

Dept. Lenguajes y Sistemas Informáticos, Jaume I Univerisity
Campus Riu Sec s/n 12071 Castellón, Spain
[montoliu,pla]@uji.es, http://www.vision.uji.es

Abstract. In the literature of computer vision and image processing, motion estimation and image registration problems are usually formulated as parametric fitting problems. Least Squares techniques have been extensively used to solve them, since they provide an elegant, fast and accurate way of finding the best parameters that fit the data. Nevertheless, it is well known that least squares estimators are vulnerable to the presence of outliers. Robust techniques have been developed in order to cope with the presence of them in the data set. In this paper some of the most popular robust techniques for motion estimation problems are reviewed and compared. Experiments with synthetic image sequences have been done in order to test the accuracy and the robustness of the methods studied.

1 Introduction

Motion estimation and image registration are important problems in computer vision, and much effort has been paid to solve them. Video compression, video processing, image mosaicing, video surveillance, robot navigation, medical imaging, traffic monitoring, ..., are only some of the many applications where motion estimation and image registration techniques can be applied. In the literature of computer vision and image processing there are different approaches to motion estimation, nevertheless, there are still challenging open problems to make solutions faster, more robust and accurate, or more general.

The motion estimation problem can be formulated in many different ways. A well known way of solving it, is to approach it as a parametric fitting problem, where the parameters to be fitted are the motion parameters. Least squares provides a well-known way for parameter estimation. In general problems, least squares methods are based on finding the values for the parameters χ that best fit a model to a set S of r data measurements, i.e. minimizing an objective function O over a set S of r observations vectors, $S = \{L_1, \dots, L_r\}$.

$$O = \sum_{L_i \in S} (F_i(\chi, L_i))^2, \qquad (1)$$

* This work has been partially supported by project CTIDIB/2002/333 from the *Conselleria de Educació Cultura i Ciència*, Generalitat Valenciana

where $\chi = (\chi^1, \ldots, \chi^p)$ is a vector of p parameters and L_i is a vector of n observations $L_i = (L_i^1, \ldots, L_i^n)$, $i = 1, \ldots, r$.

Least squares estimators assume that the noise corrupting the data is of zero mean and implicitly assume that the entire set of data can be interpreted by only one parameter vector of a given model. It is well known that least squares estimators are vulnerable to the violation of these assumptions. Robust techniques have been developed in order to cope with the presence of outliers in the data set.

One of the oldest robust method used in image analysis and computer vision is the **Hough transform**. The Hough transform is robust to outliers and it can be used to detect multiples models, but it attempts to solve a continuous problem with a discrete method and consequently it can not produce accurate results. In addition, this algorithm needs high computational effort when the number of parameters is elevate, as in the case of using an affine model in motion estimation problems.

Another popular robust technique is the **Least Median of Squares** (LMedS) method, which must yield the smallest value for the median of squares residuals computed for the entire data set. The use of the median ensures that the estimates is very robust to outliers. The main drawback is that LMedS does not have a closed form solution. There are methods that can obtain an approximate solution, but they need high computational effort. Therefore, the computational complexity of LMedS algorithms does not allow them to be used in global motion estimation problems. Nevertheless they can be used to obtain an initial estimate of the parameters of the dominant motion (see [1]).

The **Regression Diagnostics** or **outlier rejection** method [5] tries to iteratively detect possibly wrong data and reject them through analysis of the globally fitted model. This method has three steps: determine an initial fit to the whole set of data, using a ordinary least squares estimator; reject all data whose residuals exceed a threshold; determine a new fit with the remaining data set, and repeat. The success of this method clearly depens on the quality of the initial fit. Many improvements can be added to this method. For instance, estimate the initial fit using robust statitistics [3] or add an additional step that collect inliers between the outliers previously rejected [4].

Robust statistics, also called **M-Estimators**, is one of the most popular robust techniques. M-Estimators try to reduce the effect of outliers by repacling the square residuals in Equation 1 by a kernel function ρ, as follows:

$$O = \sum_{L_i \in S} \rho(\epsilon_i), \tag{2}$$

where $\rho(\epsilon_i)$ is a symmetric, positive-definite function with a unique minimum at zero and $\epsilon_i = F_i(\chi, L_i)$. If $\rho(\epsilon_i) = \epsilon_i^2$, it is the least square estimator. To analyze the behavior of an estimator, the Hampel influence function $\psi(\epsilon) = \frac{\partial \rho(\epsilon)}{\partial \epsilon}$ can be used. For least squares estimator $\psi(\epsilon) = 2\epsilon$, i.e. the influence of the outliers increases linearity and without bound. For a comprehensible study of the performance of M-Estimators see [8]. In order to solve the robust estimation problem an iterative reweighted least squares (IRLS) technique is used. The idea of the IRLS is to assign weights w_i to the residuals at each observation L_i, where the weights control the influence of the observations in the global estimation. High weights are assigned to "good" data and lower heights to outlying data. The M-Estimator problem is converted into a equivalent weighted least squares problem as follows:

$$\sum_{L_i \in S} \rho(\epsilon_i) = \sum_{L_i \in S} w_i \epsilon_i^2. \tag{3}$$

To minimize we derivate both sides and set them equal to zero, then the following expression is obtained for each w_i:

$$w_i = \frac{\psi(\epsilon_i)}{\epsilon_i}. \tag{4}$$

Gradient weighted least squares (GWLS) [8] techniques can be also used in order to achieve robustness to outliers. GWLS technique divides the original function by its gradient with respect to the observation in order to obtain a constant variance function. The solution of the GWLS problem can be also obtained using a IRLS technique replacing the weight function by:

$$w_i = \frac{1}{\sum_{j=1..n} \left(\frac{\partial F}{\partial (L_i^j)} \right)^2}. \tag{5}$$

In real motion estimation problems many of the previous robust techniques can be combined in order to deal with their problems. For instance, in [3] robust statistics, Hough transform and outlier rejection techniques are combined; in [1] LMedS and outlier rejection techniques are combined.

In this paper, four robust motion estimation algorithms are compared, three of them use a linear least squares estimator in order to estimate the motion parameters, and each of them make use of a different robust technique in order to cope with outliers: M-Estimators, Gradient Weighted and Outlier Rejection. These algorithms are explained in the Section 2.1. The last algorithm uses a non-linear least squares estimator and a gradient weighted-based technique to cope with outliers. It is explained in the Subsection 2.2. Experiments with synthetic image sequences have been done in order to show the performance of the algorithms explained. They are shown in the Section 3.

2 Robust Motion Estimation Algorithms

In motion estimation problems, the objective function O is based on the assumption that the grey level of all the pixels of a region R remains constant between two consecutive images in a sequence (Brightness Constancy Assumption). Using the BCA the objective function is expressed as follows:

$$O_{BCA} = \sum_{(x_i, y_i) \in R} (I_1(x_i', y_i') - I_2(x_i, y_i))^2, \tag{6}$$

where $I_1(x_i', y_i')$ is the grey level of the first image in the sequence at the transformed point x_i', y_i', and $I_2(x_i, y_i)$ are the grey level of the second image in the sequence at point x_i, y_i. Here, for each point i ($i = 1 \ldots r$, with r being the number of pixels) the vector of observations L_i has three elements ($n = 3$), $L_i = (x_i, y_i, I_2(x_i, y_i))$. The vector of parameters χ depends on the motion model used.

The BCA can not be directly used using an ordinary least squares (OLS) technique since it is not linear. The well-known solution to this problem derives the optic flow equation as the function to be minimized. Using the optic flow equation the objective function is expressed as follows:

$$O_{OF} = \sum_{x_i, y_i \in R} (I_t + u_x I_x + u_y I_y)^2, \tag{7}$$

where I_x,I_y and I_t are the spatial and temporal derivates of the sequence.

Nevertheless, it is possible to directly use the BCA using a non-linear least squares-based estimator. Generalized least-squares methods (GLS) [4] are an interesting approach to extend the applicability of least-squares techniques (e.g., to non-linear problems). GLS techniques can be successfully applied in motion estimation related problems [6, 7].

2.1 OLS-based Robust Motion Estimation Algorithms

Motion Estimation The solution of the motion estimation problem using an ordinary least squares method uses Taylor series expansion that produces the well know optic flow equation. The solution of O_{OF} (Equation 7) is now obtained setting to zero the derivates with respect to each of the parameters of the motion model, and solving the resulting system of equations. The solution is obtained by solving the overdetermined linear equation system $A\chi = d$ using the closed solution $\chi = (A^t A)^{-1} A^t d$, where for affine motion χ is $\chi = (a_1, b_1, c_1, a_2, b_2, c_2)$, and A $(r \times 6)$ and d $(r \times 1)$ are expressed as follows:

$$\begin{array}{ll} A = (A_1, A_2, \ldots, A_r)^T & d = (d_1, d_2, \ldots, d_r)^T \\ A_i = (xI_x, yI_x, I_x, xI_y, yI_y, I_y)_{(1\times 6)} & d_i = (xI_x + yI_y - I_t)_{(1\times 1)} \end{array} \tag{8}$$

The OLS-based motion estimation is accurate only when the frame-to-frame displacements due to the motion are a fraction of pixel. The accuracy of the estimation can be improved using an iterative alignment procedure and a multi-resolution pyramid, see [2] for details. We will refer to this algorithm as *Hierarchical and Incremental Ordinary Least Squares* (HIOLS). In order to cope with outliers a IRLS-based technique is used. For the sake of clarity, the IRLS process is described in 4 steps:

1. Create a diagonal matrix of weights W with dimensions $r \times r$. Each w_i measures the influence of the observation L_i in the global estimation. Set all $w_i = 1$.
2. Estimate the motion parameters using the equation: $\chi = (A^T W A)^{-1} A^T W d$.

3. Improve the weights using the parameters previously estimated.
4. Repeat until some termination condition is met.

This process is integrated in the HIOLS algorithm improving the weights after each parameter estimation is performed. Three robust techniques are studied: M-Estimators (R-HIOLS algorithm), Gradient Weighted-based (GW-HIOLS) and Outlier Rejection (OR-HIOLS).

R-HIOLS. The weights are calculated using the Huber M-Estimator as follows:

$$w_i = \begin{cases} 1.0 & \text{if} \quad |\epsilon_i| \leq k \\ \frac{k}{|\epsilon_i|} & \text{if} \quad |\epsilon_i| > k \end{cases} \tag{9}$$

GW-HIOLS. The weights are calculated using the Equation 5.

OR-HIOLS. In the outlier rejection technique the outliers do not have influence in the estimation of the parameters. Now, the weight are set to 1 for inliers and to 0 for outliers. The threshold is calculated using a scale measure $s(\chi)$ based on the median of the residual as follows:

$$s(\chi) = 1.4826 * \text{median}(|\epsilon_i - \text{median}(\epsilon_i)|). \tag{10}$$

The scale estimated is used to reject outliers. $w_i = 0$ if $\epsilon_i > s(\chi)$, i.e. the observation i is considered as outlier. On the other hand, it is considered as inlier and $w_i = 1$. Other similar scale measures can be used ([1], [4]).

2.2 Generalized Least Squares

The Generalized Least Squares (GLS) algorithm is based on minimizing an objective function O (see Equation 1) over a set S of r observation vectors, $S = \{L_1, \ldots, L_r\}$. In general, this equation can be non-linear, but it can be linearized using the Taylor expansion and neglecting higher order terms. This implies that an iterative solution has to be found. At each iteration, the algorithm estimates $\Delta\chi$, that improves the parameters as follows: $\chi_{t+1} = \chi_t + \Delta\chi$. The increment $\Delta\chi$ is calculated (see [4]) using the following expressions:

$$\Delta\chi = \left(A^T(BB^T)^{-1}A\right)^{-1} A^T(BB^T)^{-1}W \qquad w_i = -F_i(\chi_t, L_i)$$
$$B_i = \left(\frac{\partial F_i(\chi_t, L_i)}{\partial L_i^1}, \ldots, \frac{\partial F_i(\chi_t, L_i)}{\partial L_i^n}\right)_{(1\times n)} \quad A_i = \left(\frac{\partial F_i(\chi_t, L_i)}{\partial \chi^1}, \ldots, \frac{\partial F_i(\chi_t, L_i)}{\partial \chi^p}\right)_{(1\times p)}$$
$$B = \begin{pmatrix} B_1 & 0 & 0 & 0 \\ 0 & B_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & B_r \end{pmatrix}_{(r\times(r\times n))} \quad A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_r \end{pmatrix}_{(r\times p)} \quad W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_r \end{pmatrix}_{(r\times 1)} \tag{11}$$

In our motion estimation problems the objective function is O_{BCA} (see Equation 6). Here, for each point i ($i = 1 \ldots r$, with r being the number of pixels) the vector of observation $L_i = (x_i, y_i, I_2(x_i, y_i))$ has three elements ($n = 3$): column, row and grey level of second image at these coordinates. The affine motion model is used in this work, which is able to cope with translations, scaling, rotation and shear of images and it is defined with a vector of $\chi = (a_1, b_1, c_1, a_2, b_2, c_2)$, ($p = 6$). Therefore, B_i, A_i and w_i are expressed as follows:

$$B_i = \left(a_1 I_x^1 + a_2 I_y^1 - I_x^2, b_1 I_x^1 + b_2 I_y^1 - I_y^2, -1.0\right)_{(1\times 3)}$$
$$A_i = \left(x_i I_x^1, y_i I_x^1, I_x^1, x_i I_y^1, y_i I_y^1, I_y^1\right)_{(1\times 6)} \quad w_i = -\left(I_1(x_i', y_i') - I_2(x_i, y_i)\right) \tag{12}$$

where I_x^1, I_y^1, are the gradient of first image at the pixel (x_i', y_i') in x and y direction, and I_x^2, I_y^2, are the gradient of second image at the pixel (x_i, y_i) in x and y direction.

Similarly to OLS estimator, a multi-resolution pyramid is used in order to cope with large motion, but the iterative nature of the GLS estimator makes unnecessary the use of the alignment process of the HIOLS algorithm. We name this algorithm: *Hierarchical Generalized Least Squares* (HGLS) (see [7] for details). The robustness of the algorithm is obtained through the matrix $B^T B$ which can be viewed as a matrix of weights. Clearly the HGLS algorithm uses a gradient weighted-based technique in order to cope with outliers.

3 Experimental work

In order to test the accuracy and robustness of the proposed methods two synthetic experiments have been carried out. In the first experiment, 100 transformed images have been created using random values of the affine parameters between the limits: $a_1, b_2 \in [0.85, 1.15]$, $a_2, b_1 \in [0.0, 0.15]$ and $c_1, c_2 \in [-10.0, 10.0]$. The reference image and an example of a transformed image are showed in Figure 1(a, b). Table 1 shows the averages of the differences between the real values and the estimated values for the affine parameters, for each method.

Table 1. Error in the estimation of the motion parameters for the first experiment.

Algorithm	a_1	b_1	c_1	a_2	b_2	c_2
R-HIOLS	9.7E-07	1.3E-06	0.0001	9.8E-07	9.1E-07	0.0011
GW-HIOLS	2.8E-06	2.2E-06	0.0003	1.6E-06	2.8E-06	0.0003
OR-HIOLS	0.0002	0.0001	0.0043	4.4E-06	6.0E-06	0.0013
HGLS	2.8E-05	3.9E-05	0.0052	7.9E-05	6.4E-05	0.0103

The second experiment have been done in order to test the robustness of the algorithms. For this purpose, a patch of 150 × 150 pixels is added to the reference image and a patch of 100 × 100 pixels is added to the transformed images. The patch undergoes a different motion than of the background. The reference image and an example of a transformed image are showed in Figure 1(d, e). The averages of the differences between the real values of the background and the estimated values are shown in Table 2. All the methods accurately extract the

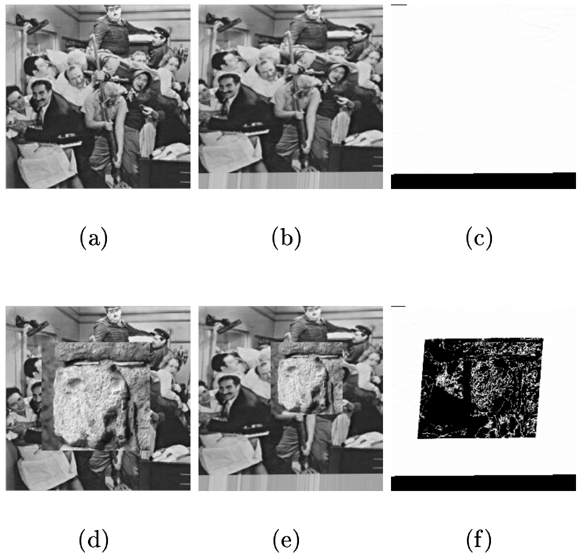

Fig. 1. Test image sequences: 1^{st} column, reference images, 2^{on} column, transformed ones. 3^{rd} column, likelihood images: dark grey values for low likelihood, i.e. outliers

motion of the background, i.e. the pixels belonging to the patch are considered as outliers, and therefore, they have not influenced the estimation of motion of the background.

Table 2. Error in the estimation of the motion parameters for the second experiment.

Algorithm	a_1	b_1	c_1	a_2	b_2	c_2
R-HIOLS	1.8E-06	2.5E-06	0.0003	2.8E-06	1.2E-06	0.0003
GW-HIOLS	7.5E-05	4.8E-05	0.0069	1.4E-05	2.9E-05	0.0057
OR-HIOLS	0.0002	0.0001	0.0167	7.8E-05	6.5E-05	0.0079
HGLS	5.6E-05	5.1E-05	0.0065	6.7E-05	7.7E-05	0.0096

The results obtained for the experiments show that all methods obtain accurate estimation of the motion parameters of the dominant motion present in the sequence, even in the case of high number of outliers as in the second experiments. No significant differences among the methods can be found. However, the results show the benefits of using a HGLS technique since it can obtain estimates as accurate as the other methods and it is more general and simpler, mainly due to the fact that it does not need the alignment process, which can introduce unexpected errors and can increase the processing time, specially in the case of large images.

In order to illustrate how the outliers have been correctly rejected, a likelihood image have been created. For each pixel, the likelihood measure $L(\chi, p_i) =$

$e^{-0.5*\frac{\epsilon_i^2}{\sigma^2}}$ (see [3]) of the pixel p_i belonging to the model estimated with parameters χ is calculated. Light grey values are used to represent high values of $L(\chi, p_i)$ in the likelihood image. On the other hand, dark grey values are used for low values of $L(\chi, p_i)$, i.e. for outliers. Figures 1(c,f) show an example of the likelihood image for the samples of the experiments. They have been created using the HGLS algorithm, but similar results would are obtained using the other algorithms. They show how the outliers have been correctly detected and rejected.

4 Conclusions

In this paper, four robust least squared-based motion estimation techniques have been explained, implemented and compared. They use M-Estimators, gradient weighted and outliers rejection techniques in order to achieve robustness in the estimation of the motion parameters.

The performance of the four algorithms have been tested using synthetic image sequences with the presence of outliers. The four methods obtain accurate estimations of the dominant motion, even in the case of an elevate number of outliers. No significant differences among the methods were found. However, the results show the benefits of using a HGLS technique since it can yield estimates as accurate as the other methods while it is more general and simpler, mainly due to the fact that it does not need the alignment process which can introduce unexpected errors and can increment the processing time, specially in the case of large images.

References

1. Alireza Bad-Hadiashar and David Suter. Robust optic flow computation. *International Journal on Computer Vision*, 29(1):59–77, 1998.
2. J.R. Bergen, P.J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *PAMI*, 14(9):886–896, September 1992.
3. M. Bober and J. V. Kittler. Estimation of complex multimodal motion: An approach based on robust statistics and hough transform. *IVC*, 12(10):661–668, December 1994.
4. G. Danuser and M. Stricker. Parametric model-fitting: From inlier characterization to outlier detection. *PAMI*, 20(3):263–280, March 1998.
5. M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motion. *IJVC*, 12(1):5–16, February 1994.
6. R. Montoliu and F. Pla. Multiple parametric motion model estimation and segmentation. In *ICIP01, 2001 International Conference on Image Processing*, volume II, pages 933–936, October 2001.
7. R. Montoliu, V.J. Traver, and F. Pla. Log-polar mapping in generalized least-squares motion estimation. In *Proccedings of 2002 IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP'2002)*, pages 656–661, September 2002.
8. Z. Zhang. Parameter-estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing*, 15(1):59–76, 1997.