Direct Motion Interpretation and Segmentation Based on the Robust Estimation of Parametric Models

X. Qian, A. Mitiche
INRS-Télécommunications, 16,Place du Commerce, Ille-des-soeurs, Verdun, P. Q., H3E 1H6, Canada

Abstract

A direct motion interpretation and segmentation method based on a motion model and robust statistics is proposed: estimation of optical velocities is consistent with the assumption of rigidity of environmental objects and accounts for motion discontinuities. The method has been tested both on several synthetic and real image sequences.

I. INTRODUCTION

One of the primary goals of motion analysis is to determine the 3-dimensional (3-D) structure of environmental objects and their movement relative to the viewing system. Usually, optical velocities are first estimated from the image spatio-temporal changes, and then interpreted in terms of the 3-D variables of motion and depth. However, one can combine these two steps and determine depth and 3-D motion "directly" from the spatio-temporal changes [1]. Once depth and 3-D motion determined, motion segmentation can be sought that is consistent with the assumption of environmental rigidity. For correct motion segmentation, motion estimation must account for motion discontinuities.

Motion estimation is an ill-posed problem [2]. A well posed problem is often obtained by regularization [3, 4]. Regularization can take a deterministic form [3] or a stochastic form [5]. In either case, if motion discontinuities are not accounted for, blurring of motion estimates occurs at these boundaries. To account for motion discontinuities, a variety of constraints have been proposed. For instance, the oriented-smoothness constraint [6] attenuates smoothing across strong intensity edges, which are identified with motion edges. Simplified, computationally more efficient versions of the oriented smoothness constraint have been developed [7], but at the cost of significant computational complexity. Motion discontinuities can also be processed in stochastic regularization by introducing variables to designate motion edges [5, 8]. This also increases significantly computational complexity which is already inherently high. Simpler, but still efficient methods use implicitly or explicitly outliers detection and rejection, outliers appearing inherently at motion boundaries [9–11].

Schunck [9] argued strongly in favor of robust regression for the outlier detection and rejection. A robust algorithm [9, 12] should be able to cope with noise, other data distortions and outlier occurrence. Proposed robust algorithms are of various complexity. Some can be of significant computational complexity (e.g., clustering methods [13]). Others, such as M-estimators, have low computational cost, but have a sensitivity to outliers that is proportional to the number of unknown [14]. Conspicuously, the least median of squares ($LMedS$) method has low sensitivity to outliers, can account for noise and other data distortions, and be implemented efficiently [12, 15].

In this paper, we present a direct motion interpretation and segmentation algorithm based on robust statistical estimation of a rigidity-based parametric motion model. Moving objects are subsequently segmented by adaptive K-means clustering.

II. MODEL AND ALGORITHM

A. Motion Estimation and Parametric Model

The Horn and Schunck gradient equation:

$$f_x u + f_y v + f_t = 0, \tag{1}$$

relates the spatial (f_x and f_y) and the temporal (f_t) derivatives of the image brightness function to optical velocity, (u, v) at each point. A reflection of the aperture problem, each such equation determines only the component of optical velocity in the direction of the gradient.

For both u and v, we substitute their expression in terms of depth and the parameters of motion in space, the environment assumed rigid:

$$\begin{cases} u = -xy\omega_x + (1+x^2)\omega_y - y\omega_z + \frac{\tau_x - x\tau_z}{Z} \\ v = -(1+y^2)\omega_x + xy\omega_y + x\omega_z + \frac{\tau_y - y\tau_z}{Z} \end{cases} \tag{2}$$

where (x, y) are the image coordinates of point (X, Y, Z) in space. τ_x, τ_y and τ_z are the translation components and ω_x, ω_y and ω_z are the rotational components of rigid motion, respectively [16]. If we substitute this expression of optical velocity into Eq.(1), we obtain an expression relating the brightness pattern spatial and temporal derivatives to the kinematic screw and depth:

$$f_x(-xy\omega_x + (1+x^2)\omega_y - y\omega_z + \frac{\tau_x - x\tau_z}{Z}) +$$
$$f_y(-(1+y^2)\omega_x + xy\omega_y + x\omega_z + \frac{\tau_y - y\tau_z}{Z}) + f_t = 0 \quad (3)$$

Let $\hat{\tau}_i = \tau_i/Z$, $i = x, y$ and z. Effecting this change of variables in Eq.(3) we obtain the following linear equation:

$$f_x(-xy\omega_x + (1+x^2)\omega_y - y\omega_z + \hat{\tau}_x - x\hat{\tau}_z) +$$
$$f_y(-(1+y^2)\omega_x + xy\omega_y + x\omega_z + \hat{\tau}_y - y\hat{\tau}_z) + f_t = 0 \quad (4)$$

To solve for these components we need to write a system of at least six equations.

We assume that environmental surfaces are piecewise planar so that $1/Z(x, y, z)$ is constant over a local image patch centered on each image pixel.

The linear system of equations written for the points in such a patch must be rank-sufficient; i.e., at least six spatial gradient must exist and have different directions. One must also ensure that the system is not ill-conditioned; ill-conditioning is likely to occur in approximately uniform image pattern regions.

When a motion boundary occurs, points in the patch are sampled across the boundary and the assumption that a single rigid body is observed is not valid. Therefore a single linear fit using these point is inappropriate as up to 50 % of these points will give rise to outliers. In addition, violations of the image/motion model assumption, such as brightness constancy and affine motion, can be viewed as "outliers" [17].

There is imprecision in image coordinate and image pattern derivatives because of optical sensing noise, discretization and algorithmic approximations for the derivatives. However, most of these noises can be modeled as Gaussian noise. We need to choose a robust estimator such that it is insensitive to both Gaussian noise and outliers.

To solve some of above problems, a statistically estimator is required at least. We shall retain the $LMedS$ estimator. Once pure parameters have been estimated, multiple moving objects can be segmented and 3-D motion parameters and structure for each object can be recovered.

B. $LMedS$ Estimator

In $LMedS$ regression, the estimates of the model parameters are given by nolinear minimizing the median of the squared residuals:

$$Min_{(\vec{a})} \ [med_i \ (r_i^2)], \quad (5)$$

where $\vec{a}$ is the unknown parametric vector to be estimated. r_i is the residual at point i with respect to the $LMedS$ fit.

The minimization of median squared residuals cannot be obtained analytically and there requires respected evaluations for different subsamples of size p drawn from the n observations. In principle one could repeat the above procedure for all possible subsample of size p, of which there are C_n^p. Unfortunately, a complete trial would rapidly become impracticable for large n and p values. In many application, it become infeasible. Some efficiency is possible, however, by adapting a Monte-Carlo approximation. Wousseeuw and Leroy [12] determined the minimum number m of subsamples required to obtain a given probability α of drawing at least one subsample containing only good observations from a sample containing a fraction ϵ of outliers. By requiring α to be sufficiently close 1, m can be determined for given values of p and ϵ:

$$\alpha = 1 - (1 - (1-\epsilon)^p)^m. \quad (6)$$

For such a subsample, index by $J = (i_1, ...i_p)$, one can determine the regression surface through the p point and denote the corresponding vectors of coefficients by $\vec{a}_J$ This step amounts to the solution of a system of p linear equations in p unknowns. For each $\vec{a}_J$ one also determines the corresponding $LMedS$ objective function with respect to the whole data set.

$$med_{(i)}[z_i - f(x_i, y_i, \vec{a}_J)]^2 \quad (7)$$

is calculated. Finally, one will retain the trial estimate for which this value is minimal.

The breakdown point of $LMedS$ is $\frac{int[\frac{n}{2}]-p+2}{n}$. The breakdown point is the smallest percentage of data that can be incorrect to an arbitrary degree and not cause the estimation algorithms to reproduce an arbitrarily wrong estimate. Therefore, the asymptotic breakdown point is 0.5. In addition, $LMedS$ has several other interesting properties: it always yields a solution; $LMedS$ estimator can be implemented more efficiently that repeated median estimator; it is also relatively easy to identify most of the outliers. All of $LMedS$ properties well suited to the present problem.

C. Implementation and Validation

1. Derivatives

We must estimate the derivatives of brightness from the discrete set of image brightness measurement available. It is important that the estimates of f_x, f_y and f_t be consistent. That is, they should all refer to the same point in the image at the same time. Most techniques deal with the information from two frames, but

information from multiple image frames or a sequence of image can be considered. For the estimation of the image brightness partial derivatives, the operators derived from that of Prewitt have been used.

2. Robust Regression

In order to improve on $LMedS$, it is followed by a least squares (LS) fit on the outlier-free datum that it produced.

$LMedS$ algorithm must be carefully designed to remove outliers and preserve discontinuities. In the $LMedS$ regression, we should identify point i as an outlier if and only if $|r_i/\hat{\sigma}|$ is large:

$$w_i = \begin{cases} 1 & if \ \ |r_i/\hat{\sigma}| \leq 2.5 \\ 0 & if \ \ |r_i/\hat{\sigma}| > 2.5 \end{cases} \quad (8)$$

where,

$$\hat{\sigma} = C\sqrt{med_{(i)} r_i^2}, \quad (9)$$

$C = 1.4826[1 + \frac{5}{(n+p)}]$ is a constant. This means simply that case i will be retained in the weighted LS if its $LMedS$ residual is small to moderate, but discarded if it is an outlier. The bound 2.5 is arbitrary, but quite reasonable because in a Gaussian situation there will be very few residuals larger than 2.5 $\hat{\sigma}$.

As a final step, we can determine and use the solution corresponding to a LS fit of the 1-weight points. We finally present the results of both stages together.

We divided the input image into overlapping patches. The same motion parameter was assumed in each patch. We assume that in each patch there is a dominant motion, the motion corresponding to a single environmental object and that includes 50 % or more of the image points in the window. Outliers are taken to be those image points which do not correspond to the dominant motion. $LMedS$ is applied to each window. We have used a 5×5 window size with a subsample size of 6. Thus the typical motion estimation is a regression problem for $p = 6$ motion parameters with $n = 25$ data points.

To use $LMedS$ regression correctly, this method is subject to a condition that the points in the window do provide a resolvable system of linear equation that is not ill-conditioned. If it is not the case the output of $LMedS$ is labeled UNKNOWN at the center of the window. A value at each point is obtained subsequently by local averaging. To reduce the number of UNKNOWN labels, we also used the multiple-constraints method [18].

3. Segmentation

Motion segmentation consists of grouping pixels that belong to independently moving objects in the scene. We use the classic adaptive K-means to cluster 3-D motion parameters in recursive fashion to detect multiple motion regions in the scene: at each iteration, a dominant motion region is detected. Once the dominant region is identified and the motion within the region is estimated, it is eliminated and the next dominant motion is estimated from the remaining portion of the image [19].

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

To investigate the performance of the approach, we use the following variety of image sequences containing different types of camera and independent multiple object motion: 1). $SOFA$, which is a package of synthetic sequences designed for testing motion analysis applications ($http://www.cee.hw.ac.uk/\ mtc/sofa$). 2). Hamburg taxi real world sequence. 3). Highway real world sequence.

A. SOFA Synthetic image sequences

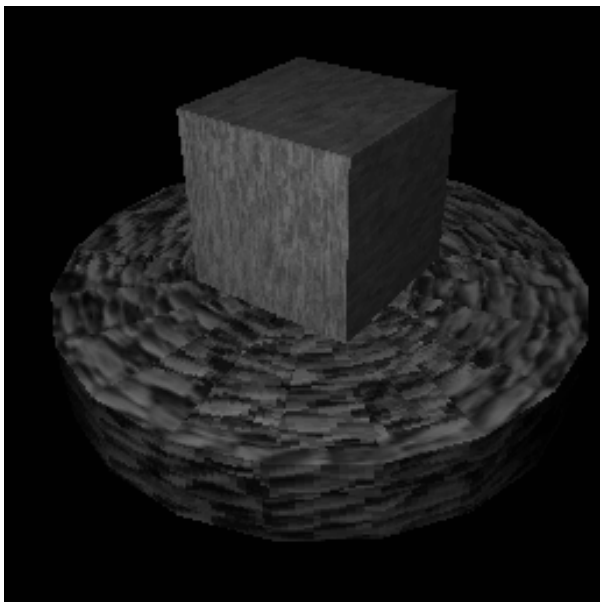

Fig.1 One frame from the SOFA synthetic image sequences which consist of a cube and a cylinder (256×256).

SOFA is a package of synthetic image sequences designed for testing motion analysis applications. All the cases, the motion is solely due to that of the camera. The scene consist of cube and cylinder, as shown in Fig. 1. The class of motion includes pure translation, pure rotation, and translation and rotation motion.

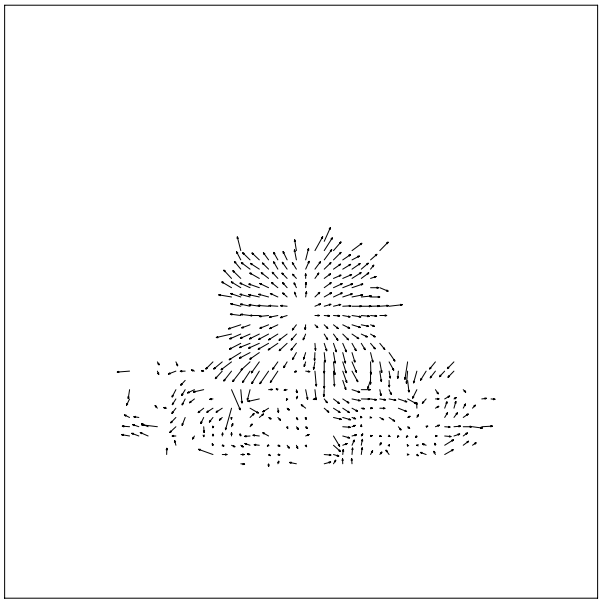

Fig.2 Optical flow using the algorithm of this work for pure translation synthetic image sequence.

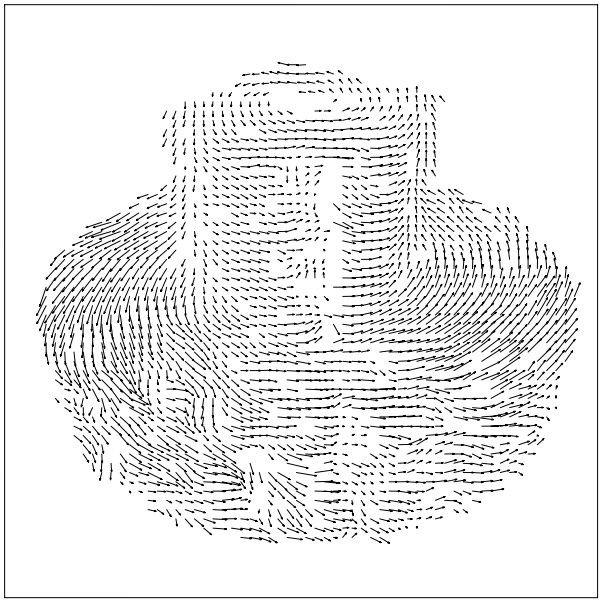

Fig.3 Optical Flow using the algorithm of this work for pure rotation synthetic image sequences.

Pure Translation: The result using our method is shown in Fig.2, where we know the true motion. In this sequence, the camera is moving forward parallel to the $z-$ axis. As we know, when the camera motion is a pure translation towards the environment, all the displacements on the image appear to emanate radially from a single point in the image. The recovered optical flow fields show this characteristic clearly, where the intersection point located at the center of the image. The recovered field quality in the cube section is better than that in the cylinder section.

Pure Rotation: This sequence was recorded using a camera which circle routed on a plane perpendicular to the $y-$ axis. The distance between camera and $y-$axis was fixed. The result using our method is shown in Fig.3. As expected, motion field for each image point follow a rotation axis is a conic. The moving discontinuities were well preserved. The result are quite close to the optimal results. It is noticed that the estimated velocity for cylinder section in pure rotation is much better than that in pure translation.

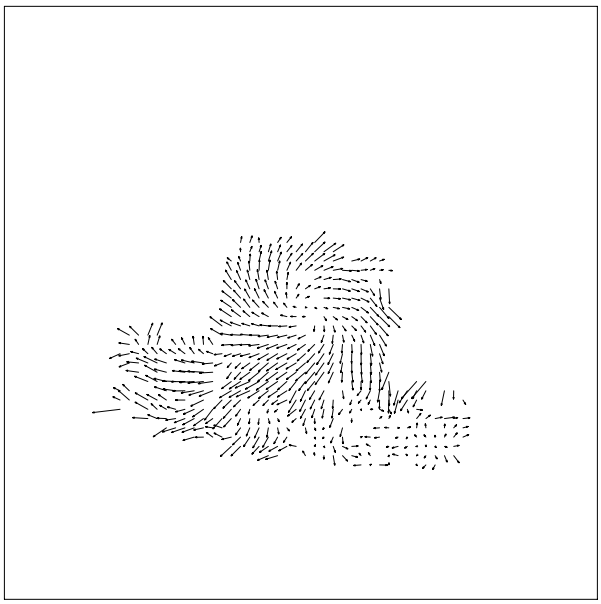

Fig.4 Optical flow using the algorithm of this work for translation and rotation synthetic image sequence.

Translation and rotation: The result on the $SOFA$ sequence for both translation and rotation motion are shown in Fig. 4. The camera translated parallel to the z-axis and rotated around the camera z-axis simultaneously. The same result with Horn and Schunk' approach is also shown in Fig. 5. As shown in figures, the differences between two results are considerable. The result of using our method is better than that obtained by Horn and Schunk's approach. Our results have good sharp in the vicinity of the motion discontinuities and can predict the optical flow almost for the whole moving object. Horn and Schunck's approach can not give good results for the region with small derivative val-

ues of brightness. The estimated optical flow with our methods are more reasonable than that of Horn and Schunck's approach.

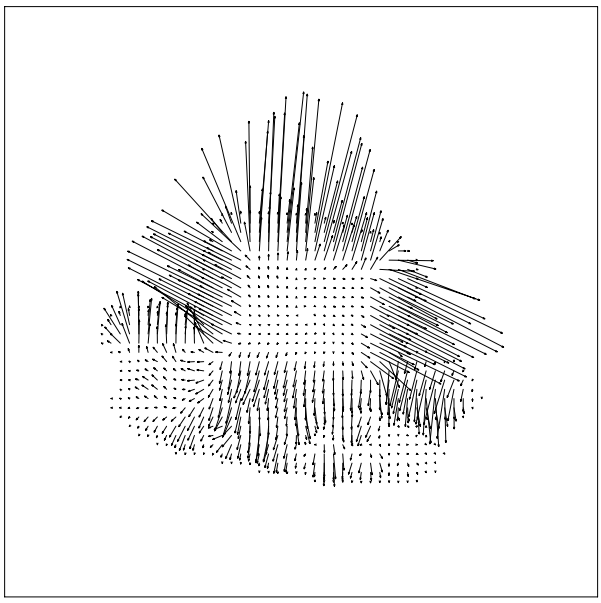

Fig.5 Optical flow using Horn and Schunk algorithm for translation and rotation synthetic image sequences.

B. Hamburg Taxi Sequence

Fig.6 One frame of Hamburg taxi image sequences.

A good real-world image sequence for multiple moving objects is Hamburg taxi. This famous scene sequence contains three important moving objects: a taxi near the center turning around the corner; a car in the lower left, driving from left to right; a van in the lower right driving from right to left. Fig. 6 show one of frame of this image sequence (256×190).

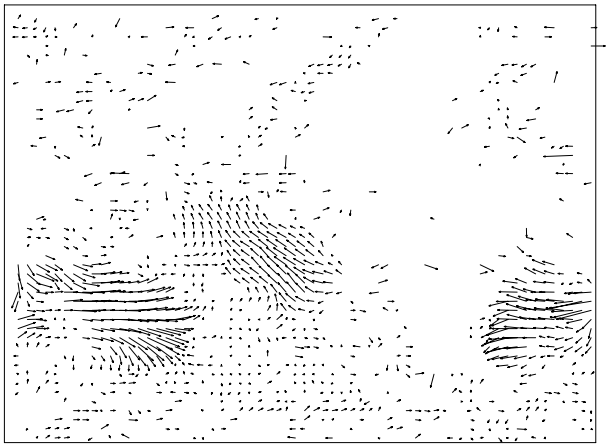

Fig.7 Optical flow using the algorithm of this work for Hamburg taxi image sequences.

The results of our algorithm drawn from 5th to 7th frames are shown in Fig. 7. One can notice that the motion estimation remains consistent in tough parts of the dynamic scenes and the independent motion of the three moving cars can be clearly distinguished. The "taxi" moving objects has better the sharp of motion discontinuities than that of "car" and "van" moving objects. Although, the shapes of motion discontinuities are not as sharp as that in synthetic image sequence. The recovered optical flow fields seem of good quality.

C. Highway Sequence

Fig.8 One frame from the original Highway sequences (256×128).

The second image sequence with multiple moving objects is a real-world highway scene showing busy traffic. One 256×128 frame of this image sequence is shown in Fig. 8. The scene consists of four important moving objects. Three of them (two cars and one truck) is going along the right side of highway. A group of cars at the left top are coming in the reverse direction. Because of different depth, it is obvious that

different moving objects have larger relative 2-D velocities in the sequence compared with Hamburg Taxi sequence. In addition, for the largest motion, the image displacement is larger than a signal pixel, which may violate the assumption of affine motion. For large and multiscale motion estimation, a standard solution is the use of multiresolution analysis with a coarse-to-fine strategy [20]. In principle, the present methods can be extended to this approach in the cost of increasing complexity. For simplicity, we assume the violation of affine motion can be viewed as "outliers".

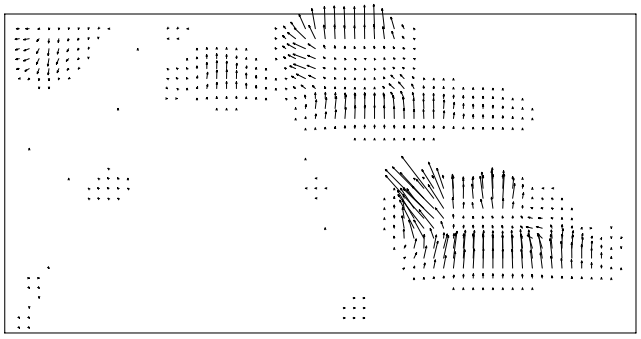

Fig.9 Optical flow using Horn and Schunk algorithm for Highway image sequences.

In Figs.9 and 10 the velocity vector fields estimated by the algorithm of this work and Horn and Schunk's algorithm. As expected, we can clearly see from these results that our method performs better than Horn and Schunk's approach. It is interesting to note that in the case of the Highway sequence, the present algorithm is able to provide good estimation in the large motion region. Robust regression has been shown to provide accurate motion estimates in a variety of situation in which affine motion assumption is violated.

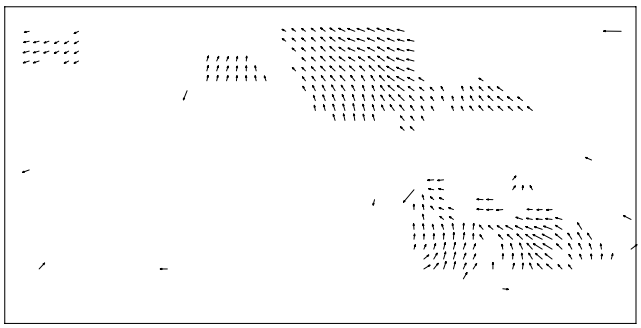

Fig.10 Optical flow using the algorithm of this work for Highway image sequences.

IV. CONCLUSIONS

In this paper, we have proposed a direct motion interpretation and segmentation method based on the robust estimation of parametric models. Our method can robustly recover 3-D motion parameters rejecting the outliers caused by moving object boundary discontinuities, noise and the violation of model assumption. In the implementation, two-stage technique, in which robust methods in the first stage to remove outliers and weighted LS methods in the second state, was able to handle outliers and Gaussian noise simultaneously. Based on the robust estimation of parametric model, the moving object segmentation was implemented by an adaptive K-means clustering algorithms. Applications of the proposed method on both synthetic and real image sequences have been demonstrated with promising results.

ACKNOWLEDGMENTS

The author would like to acknowledge the SOFA synthetic sequences courtesy of the computer Vision Group, Heriot-Watt University ($http://www.cee.hw.ac.uk/mtc/sofa$) and the $ftp.csd.uwo.ca/pub/vision$ for providing video sequences. This research was supported in part by Natural Science and Engineering Research Council of Canada under grant OGP0004234.

[1] J. Aloimonos, et al., "Direct Processing of Curvilinear Sensor Motion from a Sequence of Perspective Images", in: *Proceedings of the IEEE Workshop on Computer Vision : Representation and Analysis*, Annapolis, MD, pp.72-77 (1984).

[2] M. Bertero, et al., "Ill-posed problems in early vision", Proc.IEEE, vol 76, pp.869, (1988).

[3] B. K. P. Horn, et al., "Determining Optical Flow", Artificial Intelligence, 17, pp. 185, (1981).

[4] T. Poggio, et al., "Computational Vision and Regularization Theory", *Nature*, vol. 317, pp. 314-319, (1985).

[5] J. Konrad, et al., " Bayesian Estimation of Motion Vector Fields", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, pp.910-927, (1992).

[6] H. H. Nagel et al., "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp.565-593, (1996).

[7] P. Werkhoven et al., "The estimation of displacement vector fields by means of adaptive affine transformations", in *Proc. 8th International Conference on Pattern Recognition*, Paris, pp.798-800, (1986).

[8] F. Heitz et al., "Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields",

IEEE Trans. Patt. Anal. Mach. Intel. vol.12, pp.1217-1232, (1993).

[9] Brian G. Schunck, "Robust computational vision" , Exploratory Vision, The active eye, edited by Michael S. L., Laurence T. M. and Misha P., Springer (1992).

[10] A. Mitiche, "Computational Analysis of Visual motion" , 1994.

[11] L. Cloutrer, et al., "Segmentation and Estimation of Image Motion by a Robust Method", *Proc. IEEE International Conference on Image Processing*, Austin, Texas, pp.805-809, (1994)

[12] P. J. Rousseeuw, et al., *Robust Regression and Outliers Detection*", John Wiley, New York, (1987).

[13] C. Fennema et al., " Velocity Determination in Scenes Containing Several Moving Objects", *Computer Graphics and Image Processing*, vol. 9, pp.301-315, (1979).

[14] T. Darrell, et al., " Robust Estimation of A Multi-layered Motion Representation", In *Proceedings of IEEE workshop on Visual Motion*, Princton, pp.172-178, New York, 1991, IEEE Press.

[15] P. Meer, et al., " Robust Regression Methods for Computer Vision: A Review", *Int. J. Computer Vision*, vol. 6, pp.59-70, (1991).

[16] H. C. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene from Two Projections", *Nature*, vol. 293, pp. 133-135, (1981)

[17] F. R. Hampel, et al., " Robust Statistics: The approach Based on Influence Functions", John Wiley and Sons, New York, NY, 1986.

[18] A. D. Bimbo, et al., " Optical Flow Computation Using Extended Constraints", *IEEE Trans. on Image Processing*, vol. 5, pp.720-739, 1996.

[19] J. Y. Wang, et al., " Representing moving images with layers", IEEE Trans. Image Processing , vol. 3, Sept. 1994, pp.625-638.

[20] F. Heitz, et al., "Multiscale Minimization of Global Energy Functions in Some Visual Recovery Problems", *CVGIP : Image Understanding*, vol. 59, pp. 125-134,1994, 1992.