

Two-Stage Robust Optical Flow Estimation

Ming Ye and Robert M. Haralick
Department of Electrical Engineering
University of Washington
Seattle, WA, 98195
{ming,haralick}@george.ee.washington.edu

Abstract

We formulate optical flow estimation as a two-stage regression problem. Based on characteristics of these two regression models and conclusions on modern regression methods, we choose a Least Trimmed Squares followed by weighted Least Squares estimator to solve the optical flow constraint (OFC); and at places where this one-stage robust method fails due to poor derivative quality, we use a Least Trimmed Squares estimator to robustify the facet model fitting.

This two-stage robust scheme produces significantly higher accuracy than non-robust algorithms and those only using robust methods at the OFC stage. On the synthetic data, the one-stage robust method has an average error of 7.7% against 24% of Black's and 19% of the pure LS method; and the two-stage robust method further reduces the error by half near motion boundaries. Advantages are also demonstrated on real data.

1. Introduction

Gradient-based optical flow estimation techniques essentially consist of two stages, estimating derivatives, and organizing and solving optical flow constraints (OFC). Derivatives are usually estimated by convolving the neighborhood data with masks [2]. OFCs are based on the optical flow constraint equation $I_x u + I_y v + I_t = 0$ (OFCE) [12][2], where I_x, I_y, I_t are spatial and temporal image intensity gradients, and (u, v) is the optical flow vector. Additional constraints are acquired from either differentiating the OFC [10] or requiring the flow field to be smooth within a certain neighborhood [12][14]. The former scheme is often integrated to a regularization method for higher robustness [27][17]. Interested readers are referred to [2][3] for quite comprehensive optical flow technique review and comparison.

Both derivative estimation and OFC stages involve estimation by pooling information in a certain neighborhood and are regression procedures in nature. Classical approaches solve both regression problems in a Least-Squares (LS) sense. As it is well-known, when the neighborhood model is violated, the results from LS regression can be arbitrarily bad. Multiple motion in a neighborhood, caused by either independent moving objects or different depths, non-translational motion, brightness variation, shadow, reflection, large image noise, . . . , all create great difficulties to those LS approaches.

To enhance resistance against erroneous derivatives and motion modeling, a host of robust methods have been introduced to the OFC stage. For instance, [4][13] explicitly model the motion in a region as having more than one layers; [21][18] search for the most prominent motion through modified Hough Transform. Initial pointer into this literature can be gleaned from [3][1]. The type of methods we are more interested in recover the motion representing the majority of the neighborhood data by solving the OFCs using robust regression tools. M-Estimators [7][16] and Least Median of Squares (LMedS) estimators [19][1] were previously employed for this purpose. Although significant improvement upon LS methods has been reported, choice of these tools usually lacks solid statistical justification and what technique is most appropriate remains an open problem.

Meanwhile, as a very similar information pooling step and a vital step for accurate motion recovery [2], derivative calculation has seldom received proper attention. Crude derivative estimators [2][7] are widely used. This fact is probably due to the thought that poor derivative estimates near occlusions are inevitable but they can be treated as gross errors by a robust OFC method [6][1]. This thought is not always true. Firstly, at some places, e.g. near immediate motion boundaries, bad derivatives can totally disrupt the OFC model so that no regression method works any more. This will be verified by experimental results in Section 5.1.2. Secondly, robust derivative estimation is fea-

sible. [5][22] used robust regression methods to preserve discontinuities in surface reconstruction. These kinds of methods can be extended to derivative computation.

Based on conclusions on modern regression methods [25] [20] [23] and characteristics of the optical flow constraints, we find that the *Least Trimmed Squares followed by weighted Least-Squares (LTS.LS)* regression technique might be more appropriate for the OFC stage. We calculate derivatives from the 3D cubic facet model to achieve higher precision [27]. This approach reveals that the derivative estimation step is also a regression problem in nature, which can be robustified when the LS technique fails. We choose a *Least Trimmed Squares (LTS)* estimator for robust facet model fitting. Preliminary experimental results show that this scheme permits correct flow recovery even at immediate motion boundaries.

The rest of this paper is organized as follows. Section 2 formulates derivative estimation and OFC solving as regression problems. Section 3,4 describes how we choose appropriate robust tools for these two problems according to their individual characteristics. Experimental results and comparison are presented in Section 5. Finally Section 6 summarizes this contribution and points out future work directions.

2. Two-Stage Regression Framework

The linear regression model assumes the relationship between the observation vector Y and the unknown parameter Θ is,

$$Y^{N \times 1} = X^{N \times M} \Theta^{M \times 1} + \xi^{N \times 1},$$

where X is the design matrix and ξ is the error vector. Θ is usually estimated as

$$\hat{\Theta} = \operatorname{argmin}_{\Theta} F(r),$$

where r is the residual error $Y - \hat{Y} = Y - X\hat{\Theta}$. The criterion function $F(r)$ differs among estimators depending on what error model is assumed. The LS estimator assumes ξ to be iid with zero mean and small variance σ^2 and uses $F(r) = \|r\|^2 = \sum_{i=1}^N r_i^2$.

In this section we show that both derivative estimation and the OFC stage of optical flow estimation can be formulated as such regression problems.

2.1. Optical Flow Constraint

Following [10] we constrain the optical flow vector $V = (u, v)'$ at (x, y, t) by

$$AV + \xi = b \quad (1)$$

where

$$A = \begin{pmatrix} I_x & I_y \\ I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \\ I_{tx} & I_{ty} \end{pmatrix} \quad b = - \begin{pmatrix} I_t \\ I_{xt} \\ I_{yt} \\ I_{tt} \end{pmatrix}.$$

We further assume that flow vectors in each small neighborhood of N pixels is constant, and hence each vector V conforms to N sets of constraints simultaneously. This constitutes our optical flow constraint [27], a linear regression model

$$A_s V + \xi = b_s \quad (2)$$

where $A_s = (A'_1, A'_2, \dots, A'_N)'$, $b_s = (b'_1, b'_2, \dots, b'_N)$, and each pair of A_i, b_i are the A, b defined by Eq.(1) at pixel i , $i = 0, \dots, N-1$. In our experiment, we choose the constant flow neighborhood size to be 5×5 , so $N = 25$.

This optical flow constraint has the advantage that a large number of equations are provided for each vector estimate on a small data support (100 equations in a $9 \times 9 \times 5$ neighborhood). The compactness is important because smaller neighborhood size means less chance of multiple motion, and a larger sample size brings higher statistical efficiency (Section 3). For example, [1] chooses much larger smoothness window size as only first order constraints are used. Second order derivatives are analogous to signals in higher frequency bands. Using constraints of them can increase the regression stability but they are usually avoided due to estimation quality concerns [2]. Nevertheless throughout experiments we find second order derivatives from the 3D cubic facet model have sufficient accuracy.

2.2. Derivatives From the Facet Model

The facet model basically characterizes each small image data neighborhood by a signal model and a noise model [11][27]. Low-order polynomials are the most commonly used signal form. We use a 3D cubic polynomial for derivative estimation. Here "3D" means that the polynomial is about the spatiotemporal variable (x, y, t) ; "cubic" means that the highest order of a term is 3. The facet model finds the polynomial coefficient vector a from the linear regression model

$$Da + \xi = J \quad (3)$$

where J is the observed image data vector (formed by traversing the neighborhood data lexicographically), and D is the design matrix composed of 20 canonical polynomial bases $(1, x, y, t, x^2, \dots, xyt)$. We use the facet model neighborhood size $5 \times 5 \times 5$ and so D has dimension 125×20 . Once a is found, the spatiotemporal derivatives are merely scaled versions of its elements. More details about derivatives from the facet model can be found in [27].

Most popular derivative estimators in optical flow estimation are neighborhood masks. They essentially come from facet models [11] of different dimension (1D, 2D or 3D), order (1st, 2nd or 3rd) or neighborhood size (2, 3 or 5). For example, the four-point central difference mask $(-1, 8, 0, -8, 1)/12$ Barron uses [2] is actually a 1D cubic facet model on a neighborhood of 5 pixels [27]. Our facet model outperforms it on most image sequences.

3. Robust OFC-Stage

An LS estimator is optimal *only* when X contains no error and noise in Y is Gaussian. When either condition has a significant violation, it can be completely disrupted. There are majorly two types of significant model violations (*gross errors*), those caused by bad Y values are *y-outliers* and those caused by error in X are *leverage points*. The performance of a regression estimator is usually characterized by *statistical efficiency* and *breakdown point*. Simply put, statistical efficiency indicates the accuracy when no gross error is present, and breakdown point is the smallest fraction of contamination that can cause the estimator to take on values arbitrarily far from the truth. These two factors work against each other. A good regression tool should have both values high [20] [25].

To choose an appropriate robust estimator for the OFC stage, let us first study its characteristics. Both the observation vector b_s and the design matrix A_s are composed of derivative estimates. Because derivatives are contaminated by error, both y-outliers and leverage points might happen. If derivative estimates are roughly equally possible to be bad, because the size of A_s is double of that of b_s , chances of leverage points are even higher than y-outliers. Hence the right estimator should be resistant to both gross errors, especially leverage points. Secondly, a significant portion of these constraints might be gross errors, when, for example, multiple motion models happen in a neighborhood, so the estimator's breakdown point should be high. Finally, the number of constraints are relatively small, therefore this estimator should have good statistical efficiency on a small sample size.

M-estimators [7][16] are resistant to y-outliers and have relatively high statistical efficiency, but they meet with computational difficulties such as initial guess dependency and non-convexity (for redescending estimators [7]), have a low breakpoint (about $1/(p+1)$ where p is the dimension of Θ), and they are vulnerable to leverage points [20][25]. The LMedS [20][19][1] estimator is resistant to both types of gross errors, has a breakdown point as high as 50%, does not need a initial guess and is guaranteed to converge. However it has extremely low statistical efficiency, which means that it tends to have very large estimation variances when no gross error is present. Using LMedS seems to trade the accuracy of majority of the flow field for that of the minority.

The Least Trimmed Squares (LTS) estimator [20] was introduced to repair the low efficiency of LMedS,

$$\hat{\Theta} = \operatorname{argmin}_{\Theta} \sum_{i=1}^h (r^2)_{i:n} \quad (4)$$

where $h < n$ and $(r^2)_{1:n} \leq \dots \leq (r^2)_{n:n}$ are the ordered

squared residuals. Recalling that a Least-Squares estimator is $\hat{\Theta} = \operatorname{argmin}_{\Theta} \sum_{i=1}^N r_i^2$, LTS allows the fit to stay away from the gross errors by excluding the largest squared residuals from the summation. Owing almost all merits of LMedS and better statistical efficiency, LTS is usually considered to be preferable to LMedS [25].

Efficiency of LTS can be further improved by a *Weighted Least-Squares (WLS)*. [20] defines an error scale as $\hat{\sigma} = C \sqrt{\frac{1}{n} \sum_{i=1}^h (r^2)_{i:n}}$, where C is a factor used to achieve consistency at Gaussian error distributions; and then makes use of the weights

$$w_i = \begin{cases} 1 & \text{if } |r_i/\hat{\sigma}| \leq 2.5 \\ 0 & \text{if } |r_i/\hat{\sigma}| > 2.5 \end{cases} \quad (5)$$

to do the WLS

$$\hat{\Theta} = \operatorname{argmin}_{\Theta} \sum_{i=1}^n w_i r_i^2. \quad (6)$$

We call the above procedure Least Trimmed Squares followed by (weighted) Least-Squares (LTS.LS), and use it to solve the optical flow constraint.

4. Robust Derivative Estimation

Now we examine what a robust tool can be used to improve the derivative quality. Since the dimension of the parameter a is high, 20 for our 3D cubic facet model, simple M-estimators cannot be used because of the low breakdown point of $1/21$. We are again led to the Least Trimmed Squares estimator. Unlike the OFC stage, where we estimate two parameters out of 100 constraint equations, in this stage, there are 20 parameters but only 125 constraint equations. With a rather small sample size, WLS can hardly improve results of LTS. So we decide to use LTS for robust facet fitting.

Also due to the problem of small sample size, LTS tends to have low efficiency than LS when there is no significant model violation. So LTS facet model should be used when failure of the LS facet cannot be tolerated by the robust OFC. Detecting failures of the LS facet model through a systematic error analysis [27] is part of our future work. In this paper we emphasize on how high boundary accuracy is enabled by robust derivative estimation.

5. Experiments and Analysis

We demonstrate on both synthetic and real data how optical flow accuracy significantly improves from our LS-based method (LS-LS) to proposed one-stage robust method (LS-LTS.LS), and from LS-LTS.LS to the two-stage robust one (LTS-LTS.LS). We also compare our results with those from

Black and Anandan’s dense regularization approach [7]. We use the default parameters as suggested in Black’s software. For more accurate results, we modified his C code to output floating-point instead of 8-bit flow vectors. We implemented LS-LTS.LS and LTS-LTS.LS using standard functions in S-Plus [23] with all parameters as default.

5.1. Experiments on Synthetic Data

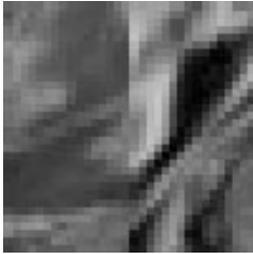


Figure 1. Central frame of the synthetic sequence

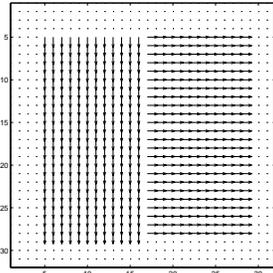


Figure 2. Correct flow field

We synthesized an image sequence to highlight motion boundary accuracy. Fig. 1 gives the central frame and the correct flow field. The image size is 32×32 and the velocity magnitude is 1 pixel/frame. To facilitate comparison, we shade all estimates with error vector magnitude larger than $\sqrt{0.004}$, a somehow arbitrary number, and call these estimates bad estimates. We calculate the average error vector magnitude for interested flow field regions. Since in this experiment the true velocity magnitude is 1, this measure multiplied by 100 is the average error percentage (AEP). We use AEP as the quantitative accuracy measures. Each optical flow constraint equation is a line $au + bv + c = 0$ in the u, v coordinate, with its distance to the true velocity an indicator of the degree of modeling imperfection. OFC cluster plots can be used to visualize derivative quality and results of different estimators.

5.1.1. Optical Flow Field Comparison

The result from LS-LS (Fig. 3(a)) has an AEP of 19.06% and almost all estimates within the boundary part, whose supports contain data from the other motion model, turned out to be bad. LS-LTS.LS reduces the AEP to 7.73% (Fig. 3(b)), but many estimates near the boundary are still compromises of the two different velocities, with the AEP of shaded pixels still as high as 47.79%.

Fig. 3(c) is an improved version of Fig. 3(b) by recomputing the vectors at the shaded pixels using LTS-LTS.LS. The AEP of the updated estimates reduces by about half to

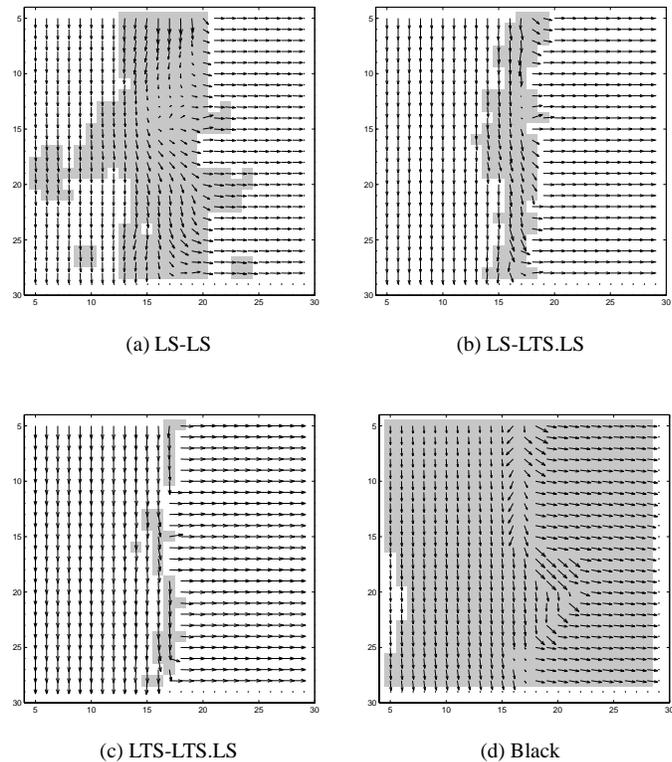


Figure 3. Optical flow estimates on the synthetic data (bad estimates are shaded)

25.08%. It has few compromises of two motion models and almost all bad estimates on the boundary take the motion of the other side. This is largely due to the randomness in the LTS implementation [23]. Even though the theoretical performance is not achieved, the boundary location error has been reduced to about 1 pixel. Such a high boundary accuracy might benefit applications such as motion segmentation.

Black’s result (Fig. 3(d)) has an high AEP of 23.89%. We are looking into why this method performs so badly here. Possible reasons are poor derivative quality, and M estimators’ inherent problems of low breakdown point and leverage point sensitivity.

5.1.2. OFC clusters—A Closer Look

We now illustrate the reasons of such accuracy contrast by examining three typical points, (5, 5), (5, 20) and (14, 17) in Fig. 3. Their true velocities are marked by black dots in Fig. 4.

(5, 5) is a point where most derivatives are of good quality, as can be told from the nice OFC cluster at the true ve-

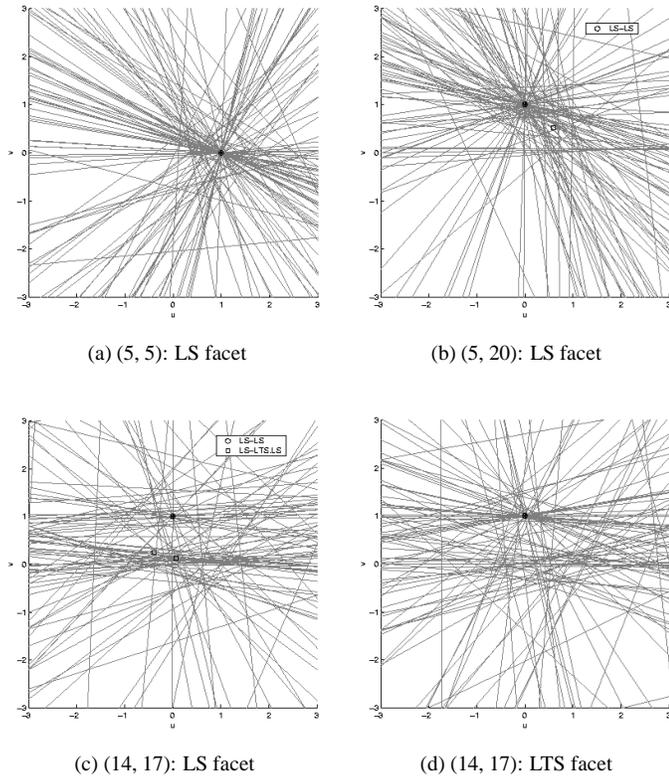


Figure 4. OFC cluster plots at three typical pixels

locity (Fig. 4(a)). However, even in this favorable case, LS-LS only yields $(0.9734, 0.0015)$ while LS-LTS.LS yields (numerically) exactly $(1, 0)$. The $9 \times 9 \times 5$ support of point $(5, 20)$ has $1/9$ conveying the left motion mode. Accordingly we observe a clear cluster at the true velocity and a small vague one at the left velocity (Fig. 4(b)). LS is totally lost in this not-so-bad case, yielding a compromise of $(0.5933, 0.518)$, as oppose to LS-LTS.LS, which gives $(-0.0051, 0.9913)$. These two cases suggest that LS-LTS.LS *significantly* outperforms LS-LS at the OFC stage.

In the above the facet model fitting error can be compensated by robust OFC. But this is not the case with $(14, 17)$, a boundary point on the right side. Fig. 4(c) shows constraint lines scattering around with two very vague clusters at $(0, 1)$ and $(1, 0)$. Estimates from LS-LS $(-0.3937, 0.2482)$ and LS-LTS.LS $(0.0708, 0.1267)$ are totally wrong. Here applying robust regression at the OFC stage alone does not help any more. The reason is that derivative estimation at most points fails and a large portion of the constraints become gross errors, so that the major optical flow constraint model does not exist. Without this model, where can a regression technique “regress” to? In contrast, Fig. 4(d)



Figure 5. Central frame of the SRI sequence

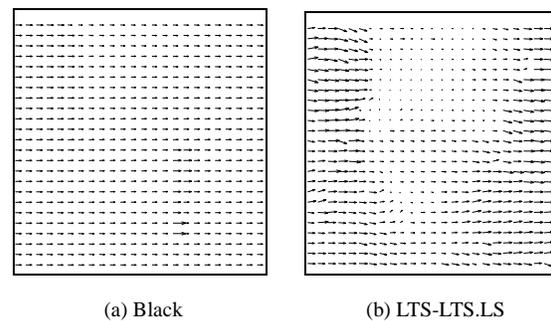


Figure 6. OF field of SRI in the box

shows the OFC plot from the robust facet model fitting. The major motion model becomes clear so that LTS.LS yields a good estimate at $(0.0109, 1.0000)$.

5.2. SRI Tree Sequence

The SRI tree sequence [2] (Fig. 5) is widely used for comparing optical flow estimation accuracy. [7] demonstrates the motion boundary preserving capability on it. However when applying Black's method to different image sequences, we find that although it preserves boundaries better in some cases, sometimes it smears boundaries more than non-robust techniques. Such inconsistency also shows up on this sequence.

To closely compare motion boundary accuracy, we show the unsampled optical flow estimates (Fig. 6) within the framed box. The middle part of the block has smaller velocity magnitude because it is farther from the camera. Black's result does not show proper foreground and background speed contrast, while the proposed LTS-LTS.LS method preserves the discontinuities well.

6. Discussion

The primary novel contribution of this work is that it incorporates robust derivative estimation into the optical flow method. It carefully analyzes the characteristics of the two regression stages and chooses robust tools correspondingly. Preliminary experimental results on both synthetic and real image sequences verified the effectiveness.

More experiments and extensive qualitative and quantitative comparison with other approaches are being carried out. So far robust facet fitting are conducted at manually chosen places. This could be automated by integrating motion boundary detection into the algorithm. A possible approach is to systematically analyze the errors in facet fitting and optical flow estimation [27], and consequently apply one- and two-stage robust methods adaptively. At the same time we are also seeking for more appropriate robust regression techniques to estimate derivative and solve optical flow constraints. We expect to apply this technique to motion segmentation and triangulation problems in the future.

References

- [1] A. Bab-Hadiashar and D. Suter. Robust optical flow estimation. *IJCV*, 29(1):59–77, 1998.
- [2] J. L. Barron, S. S. Beauchemin, and D. J. Fleet. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.
- [3] S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM Computing Surveys*, 27(3):433–67, 1995.
- [4] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. Three-frame algorithm for estimating two-component image motion. *PAMI*, 14(9):886–896, 1992.
- [5] P. J. Besl, J. B. Birch, and L. T. Watson. Robust window operators. *2nd ICCV*, pages 591–600, 1998.
- [6] M. J. Black. *Robust Incremental Optical Flow*. Yale University, Dept. of Computer Science, Research Report YALEU-DCS-RR-923, 1992.
- [7] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVIU*, 63(1):75–104, 1996.
- [8] R. M. Haralick, editor. *Proc. 1st International Workshop on Robust Computer Vision*, Seattle, WA, Oct. 1990.
- [9] R. M. Haralick, editor. *Workshop Proc. Performance vs. Methodology in Computer Vision*, Seattle, WA, June 1994.
- [10] R. M. Haralick and J. S. Lee. The facet approach to optic flow. *Proceedings of the Image Understanding Workshop*, pages 84–93, 1993.
- [11] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision, Vol. II*. Reading, MA: Addison-Wesley, 1992.
- [12] B. K. P. Horn and B. G. Schunck. Determining optic flow. *Artificial Intelligence*, 17:185–203, 1981.
- [13] A. Jepson and M. J. Black. Mixture models for optical flow computation. *Partitioning Data Sets, DIMACS Workshop*, pages 271–286, 1993.
- [14] B. D. Lucas and T. Kanade. An iterative image-registration technique with an application to stereo vision. *Image Understanding Workshop Proc.*, pages 121–130, 1981.
- [15] P. Meer, D. M. D. Y. Kim, and A. Rosenfeld. Robust regression methods for computer vision: a review. *IJCV*, 6(1):59–70, 1991.
- [16] E. Memin and P. Perez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *PAMI*, 7(5):703–19, 1998.
- [17] H. H. Nagel. Towards the estimation of displacement vector fields by “oriented smoothness” constraints. *7th ICPR*, 1:6–8, 1984.
- [18] P. Nesi, A. D. Bimbo, and D. Ben-Tzvi. A robust algorithm for optical flow estimation. *CVIU*, 62(1):59–68, 1995.
- [19] E. P. Ong and M. Spann. Robust optical flow computation based on least-median-of-squares. *IJCV*, 31(1):51–82, 1999.
- [20] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley and Sons, 1987.
- [21] B. G. Schunck. Image flow segmentation and estimation by constraint line clustering. *PAMI*, 11(10):1010–1027, 1989.
- [22] S. S. Sinha and B. G. Schunck. A two-stage algorithm for discontinuity-preserving surface reconstruction. *PAMI*, 14(1):36–55, 1992.
- [23] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S-Plus, 2nd Edition*. Springer, 1997.
- [24] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. *IJCV*, 14(1):67–81, 1995.
- [25] R. R. Wilcox. *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press, 1997.
- [26] M. Ye and R. M. Haralick. Image flow estimation using facet model and covariance propagation. *Technical Report ISL-TR-00-01*, 1999.
- [27] M. Ye and R. M. Haralick. Image flow estimation using facet model and covariance propagation. In M. Cheriet and Y. H. Yang, editors, *Vision Interface: Real World Applications of Computer Vision*, volume 35 of *MPAI*, pages 209–241. World Scientific Pub Co., 2000.