Robust Regression and Outlier Detection with SVR: Application to Optic Flow Estimation

Johan COLLIEZ, Franck DUFRENOIS and Denis HAMAD

Université du Littoral Côte d'Opale
Laboratoire d'Analyse des Systèmes du Littoral
50 rue Ferdinand Buisson, BP 699
62228 CALAIS CEDEX
Franck.Dufrenois@lasl.univ-littoral.fr

Abstract

The robust regression is an important tool for the analysis of data contamined by outliers. In computer vision, the optic flow computation is considered as belonging to this kind of problem. In this paper, we discuss a robust optic flow computation based on a modified support vector regression (SVR) technique. We experimentally show that the proposed method significantly improves the robustness against outliers compared to traditional SVR. Next, we illustrate the performances of the method for the optic flow computation problem from noised image sequences.

1 Introduction

One major task of computer vision is to compute the optic flow from an image sequence. Accurate estimation of the optic flow is an important foundation for tasks, such as motion segmentation, extracting structure from motion, video coding, etc.

Usually optic flow estimation techniques are matching-based or gradient-based, and use either local or global optimization schemes. We propose a differential method based on a regression criterion. Traditional methods based on a Least-Squares criterion are non robust, which means they will fail to correctly compute optical flow when the two main assumptions are violated : brightness conservation and spatial coherence. Clearly, these assumptions may be violated in presence of factors such as motion boundaries, shadows, occlusions, and/or transparent motions. In order to provide resistant (stable) results in the presence of outliers, it is important to develop robust approaches for the optic flow computation. The breakdown point is an important property in statistics to measure the "robustness" of an estimator. It can be defined as the smallest percentage of outlier contamination that can cause the estimator to produce arbitrarily large values. For instance, the Ordinary Least-Squares estimator (OLS) has a breakdown point of 0%. Thus, during the last ten years, statiticians have developped robust estimators such as: M-estimators [4], Least Median Squares (LMedS) estimator [7], Least Trimmed Squares (LTS) estimators [14], etc. These estimators are more performant than LS and have a demonstrated breakdown point less than 50%. More recently, some new promising estimators have

been proposed by the vision community to improve the robustness when data contain more than 50% outliers. For example in [12], the authors propose a new robust estimator : the Variable Bandwidth Quick Maximum Density Power Estimator (VBQMDPE) which can tolerate more than 50% outliers. A crucial part of this algorithm is the use of the mean shift procedure with an adaptive bandwidth.

In this paper, we propose to solve the optic flow computation problem with a modified SVM regression technique (SVR). Based on the Vapnik's theory on Support Vector Machines, SVR is now a well-established method for function regression. However, the traditional SVR is not effective in dealing with outliers in training data commonly encounter in practical applications. Thus few outliers result in a poor regression. The basic idea of the proposed method consists in gradually partitioning data into outliers and inliers, and thus refining the estimation with the inliers. In this study, data are regarded as outliers when their absolute residuals is higher than a given threshold. We experimentally show that the proposed method significantly improves the robustness against outliers contamination and multiple structures than traditional SVR.

Our paper is organized as follows : In Section 2, we briefly recall the principle of the optic flow constraint. Section 3 presents the robust linear SVR technique. Section 4 is devoted to performance evaluation on synthetic image sequence and Section 5 concludes the paper.

2   Optic flow estimation as a regression problem

In computer vision, the motion estimation may be formulated as a linear regression problem based on the optic flow constraint equation (OFCE). This equation is based on the fundamental assumption of the brightness conservation, i.e.,

$$I(x,y,t) = I(x + u.dt, y + v.dt, t + dt) \tag{1}$$

where (u,v) is the optic flow vector. The gradient-based optic flow estimation is obtained thanks to the linear approximation of Eq.1:

$$I_x.u + I_y.v + I_t = 0 \tag{2}$$

Equation 2 represents the OFCE and (I_x, I_y, I_t) is the spatiotemporal image intensity gradient. Since there is only one equation with two unknown variables, this equation cannot be solved for both horizontal and vertical components of the optic flow without additional assumptions or information. This is the well-known aperture problem. To make the problem well-posed, various alternative regularization strategies have been suggested. The first one, introducing by Horn and Schunck [3], is based on the definition of a functional derived from the OFCE and a smoothness penalty term. Many variations have been proposed to improve and robustify this approach. The second one is to form a set of OFCEs in order to improve the optic flow vector estimation. This set may be obtained in different ways: Lucas and Kanade suppose a constant motion and a group of n OFCEs in a small neighborhood around each pixel [5]. Nagel proposes to differentiate the OFCE to obtain more than one constraint [6]. Other constraints may be added from multispectral images, multiple illumination sources, multi-band decomposition of the image sequence

or from a particular type of motion.

Regardless of the strategy for overcoming the aperture problem, one usually arrives at a set of n linear equations:

$$\mathbf{y} = X\theta + e \tag{3}$$

where $\mathbf{y} = (y_1, y_2, ..., y_n)^T$ is a given output n-vector, $X = (X_1, X_2, ..., X_n)^T = (x_{ij})$ is a $n \times p$ input matrix and $\theta = (\theta_1, \theta_2, ..., \theta_p)^T$ the unknown p-vector of coefficients which must be estimated. $\mathbf{e} = (e_1, e_2, ..., e_n)^T$ is a $n-$vector of unknown errors.

For example, in the first-order case and for a constant motion model, the OFCE problem is defined by the output vector $\mathbf{y} = -(I_{t1}, I_{t2}, ..., I_{tn})^T$ which groups temporal derivatives of the image, a $n \times 2$ input matrix X which groups spatial derivatives in each component $X_i = (I_{xi}, I_{yi})^T$ and by the unknown vector $\theta = (u, v)$.

Next, θ is estimated as

$$\widehat{\theta} = \arg\min_{\theta} f(\frac{r(\theta)}{\sigma}) \tag{4}$$

Where $r(\theta) = \mathbf{y} - \widehat{\mathbf{y}}$ is the residual error $r(\theta) = \mathbf{y} - X\widehat{\theta}$ and σ is a scale parameter (usually unknown). The criterion function f differs among estimators depending on what error model is assumed.

In classical theory, the Ordinary Least Square estimator (OLS) assumes the errors e_i to be additive in Gaussian with zero mean and variance σ^2, thus $f_{OLS}(r) = \sum_{i=1}^{n} r_i^2$. Most optic flow estimators solve the regression problem in the OLS sense. However as it is well-known, a LS estimator might fail at the presence of a single gross error either in y (*y-outlier*) and/or in X (*leverage points*), i.e., it has a breakdown point of 0%. In the optic flow problem, errors are present both in the input matrix X and the output vector **y**, i.e., in spatial and temporal derivatives. Indeed, most popular derivative estimators are neighborhood masks, which essentially come from polynomial models. Coefficients are then usually estimated in an OLS sense, that is why the image derivatives are sensitive to noise and gross errors.

In the next section, we propose a new robust estimator based on the support vector machines theory. We will study its performances with respect to outliers and multiple structures.

3   Linear support vector regression

The linear SVR builds a function $f(x) = \theta^T.x + b$ minimizing the functional

$$f_{SVM}(\theta) = \frac{1}{2}\|\theta\|^2 + \lambda \sum_{i=1}^{n} \rho(y_i - f(x_i)) \tag{5}$$

This functional may be interpreted as a classical regularization approach to solving an ill-posed inverse problem, with its first term a *regularization term* and its second part a *data fidelity term*. The constant λ controls the relative influence of these two terms. In the standard SVR, the loss function $\rho(.)$ takes the following form [10]:

$$\rho(r_i) = \left\{ \begin{array}{ll} 0 & \text{for } |r_i| < \varepsilon \\ |r_i| - \varepsilon & \text{otherwise} \end{array} \right\} \tag{6}$$

This ε-insensitive loss functions is convex and the corresponding SVR optimization problem can be transformed to a quadratic programming problem:

$$minL(\theta, \xi_i, \xi_i^*) = \frac{1}{2}\|\theta\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{7}$$

Subject to:

$$\left\{ \begin{array}{l} r_i = y_i - X_i\theta \leq \varepsilon + \xi_i \\ -r_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{array} \right. \qquad \forall\, i \in 1,...,n \tag{8}$$

The constant $C > 0$ determines the trade-off between the flatness of the output **y** and the amount up to which deviations larger than ε are tolerated. Variables ξ_i, ξ_i^*, called "slack variables", are upper and lower constraints of outputs of the system. It turns out that this optimization problem can be solved more easily in its dual formulation:

$$\left\{ \begin{array}{ll} maxL(\alpha_i, \alpha_i^*) = & -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)X_iX_j \\ & -\varepsilon\sum_{i,j=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n}y_i(\alpha_i - \alpha_i^*) \\ subject\ to & \left\{ \begin{array}{l} \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{array} \right. \end{array} \right. \tag{9}$$

Where the dual variables α_i and α_i^* are determinated by Quadratic Programming (QP) method. We are going to use the Sequential Minimal Optimization (SMO) of Platt [8] which breaks a large QP problem into a serie of smallest possible QP problems. These small QP problems solved analytically avoids using a time-consuming numerical QP optimization as an inner loop. The amount of memory required for SMO is linear in the training set size, which allows SMO to handle very large training sets.

Then, the vector solution θ is obtained from the following expression:

$$\theta = \Sigma_{i=1}^{N}(\alpha_i - \alpha_i^*)X_i \tag{10}$$

Thus, the vector θ is easily determinated independently from the number of input-dimensions and with only a linear complexity for reconstruction.

3.1 Robustness of the traditional SVR

In the traditional SVR method, the management of errors in data depends on several elements such as *the form of the loss function, the value of the hyperparameters (ε margin and the regularization parameter C)*, and *the concept of support vectors*.

- The form of the loss function is a one of the fundamental keys to ensure robustness properties of an estimator. The slower the loss function increases with the errors, the more robust the estimation is. For instance, the linear ε-insensitive loss function presents the best characteristics of robustness among other common loss functions such as Gaussian loss function, Laplacian loss function, Huber's loss function, Quadratic ε-insensitive loss function [11, 9], etc.

- The two parameters of SVR, e.g. the *ε-margin and the regularization parameter C*, play an important role in the regularity and the complexity of the solution. The parameter ε controls the "flatness" of the solution and so is related to the accuracy of the approximation. To obtain an optimal estimation, the value of ε must be adjusted to handle the *noise level* in the data set. The presence of *errors* in the data set are measured by other internal parameters called "*slack variables*", which characterize the deviation of training samples outside the ε-margin. The control of the global deviation is managed by the parameter C in (7). The larger C is, the more sensitive to errors the solution is and the more complex the model is. In standard SVR, the values of ε and C must be specified beforehand.

- Support vectors, which are directly dependent on the values of the dual variables (α_i, α_i^*), allows to identify outliers from the data set. As shown in Figure 1-a, the data set is divided into 3 subsets: data points inside the margin ('•'), data points on the margin ('□') and data points outside the margin ('*'). The solid line represents the optimal hyperplane, and the ε-insensitive margin is delineated by dashed lines. In practice, a simple test on the pair (α_i, α_i^*) (Figure 1-b) makes it possible to classify the data in these three groups.

After all, standard SVR is still not robust enough. In fact, as we will show it later, a small percentage of outlier contamination is sufficient to force SVR to produce arbitrarly large values (small breakdown point). In the next section, we propose a robust SVR-based method.
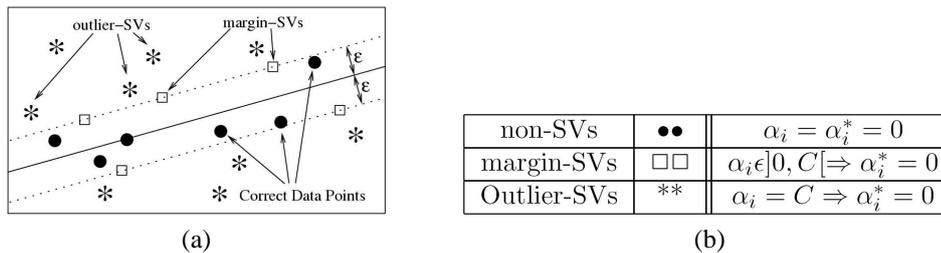

non-SVs	••	$\alpha_i = \alpha_i^* = 0$
margin-SVs	□□	$\alpha_i \epsilon]0, C[\Rightarrow \alpha_i^* = 0$
Outlier-SVs	**	$\alpha_i = C \Rightarrow \alpha_i^* = 0$

(a)   (b)

Figure 1: (a) Location of SVs and non SVs with respect to the margin. (b) Values of dual variables for different subsets.

3.2 Robust SVR: RSVR

We propose here to develop a SVR-based technique which can tolerate more than 50% outliers in the data. The main assumption, adopted here, is that the structure of interest occupies the higher proportion of the different structures present in the data set. The principle of our approach consists in refining the solution obtained with the traditional SVR by iteratively removing data with large absolute residuals from the data set. In a more formal way, the estimation of the unknown vector θ is obtained by re-iterating the two following stages:

(a) Apply SVR with a small ε:

$$min \, L(\theta, \xi_i, \xi_i^*) = \frac{1}{2} \|\theta^{(k)}\|^2 + C \sum_{i \in S^{(k)}} (\xi_i + \xi_i^*) \qquad (11)$$

subject to:

$$\begin{cases} r_i^{(k)} = y_i - X_i\theta^{(k)} \leq \varepsilon + \xi_i \\ -r_i^{(k)} \leq \varepsilon + \xi_i^* \qquad \forall\, i \in S^{(k)} \\ \xi_i, \xi_i^* \geq 0 \end{cases} \qquad (12)$$

(b) Preserve a subset of data which verifies:

$$S^{(k)} = \left\{ i \in S^{(k-1)} \,/\, \left| r_i^{(k-1)} \right| < M^{(k-1)} \right\} \qquad (13)$$

$S^{(k)}$ is the data subset at step k where each data has an absolute residual lower than a given value M. For k=0, $S^{(0)}$ is initialized with all training data samples. The cut-off value M is also evaluated at each iteration. Simulation studies showed that if we put $M^{(k-1)} = \frac{1}{2} \max_{i \in S^{(k-1)}} \left(\left| r_i^{(k-1)} \right| \right)$, we obtain stable and efficient results. The procedure is stopped when $\| \theta^{(k)} - \theta^{(k-1)} \|$ reaches a fixed threshold.

In the following, the performance of the procedure in line fitting will be demonstrated and its tolerance to large percentages of outliers will be compared with other popular methods: LMedS, LTS, SVR and vbQMDPE. In order to compare these methods, we will draw the "breakdown point" [4]. This curve illustrates the evolution of the relative error between the estimated solution and the true solution according to the contamination rate.

- Experiment 1

In this experiment, we generate $n = 300$ data points. Among these data points, n_1 data points define a single line $y = -x + 100$ corrupted by gaussian noise with zero mean and unit variance σ, and n_O data points were randomly distributed in the range of (0,100).The percentage of outliers in the data points changes from 10% to 95%. Figures 2-a and 2-b illustrate the evolution of the relative error (breakdown plot) of the slope a and the bias b of the estimated line for the five methods. SVR first breaks down (at about 20% of outliers), then LTS (35%), LMedS (50%), RSVR (75%) and next VBQMDPE (90%). As we can see, even if VBQMDPE outperforms all the approaches, RSVR is also effective and much more robust than standard SVR.

- Experiment 2

In this experiment, we investigate the characteristics of the five methods to fit data with multiple structures and outliers. We generate three kinds of data set (one step, two crossed lines and three crossed lines). Every line is corrupted by gaussian noise with zero mean and different variance σ. The ith line has n_i samples. n_O data points are randomly distributed in the range of (0,100).The three signals are defined as follows:

(1) **One step** : x=(0,55), y=30, $\mathbf{n_1}$=120, σ=1.5 ('□'); x=(20,100), y=60, $\mathbf{n_2}$=50, σ=2 ('+'); $\mathbf{n_0}$=50 ('o') (fig. 3-b)).

(2) **Two lines** : x=(20,70), y=x+10, $\mathbf{n_1}$=150, σ=3 ('□'); x=(35,85), y=-x+115, $\mathbf{n_2}$=100, σ=1 ('△'); $\mathbf{n_0}$=120 ('+')(fig. 3-c)).

(3) **Three lines** : x=(60,100), y=x+3, $\mathbf{n_1}$=80, σ=3 ('△'); x=(0,100), y=35, $\mathbf{n_2}$=80, σ=1 ('+'); x=(10,80), y=-x+100, $\mathbf{n_3}$=150, σ=4 ('□'); $\mathbf{n_0}$=100 ('o') (fig. 3-d)).
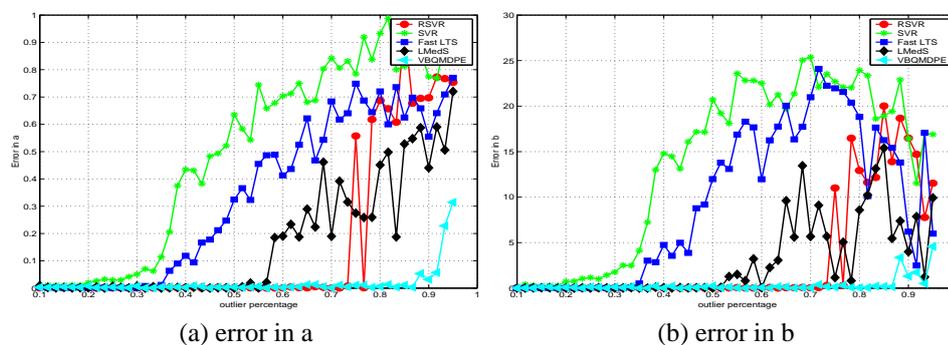
(a) error in a   (b) error in b

Figure 2: Breakdown plots of studied regression estimators.The x-axis represents percent data contamination and the y-axis characterizes the relative error for the slope (fig.2-a) and for the bias (fig.2-b).

(a) Experiment 1   (b) Experiment 2-1

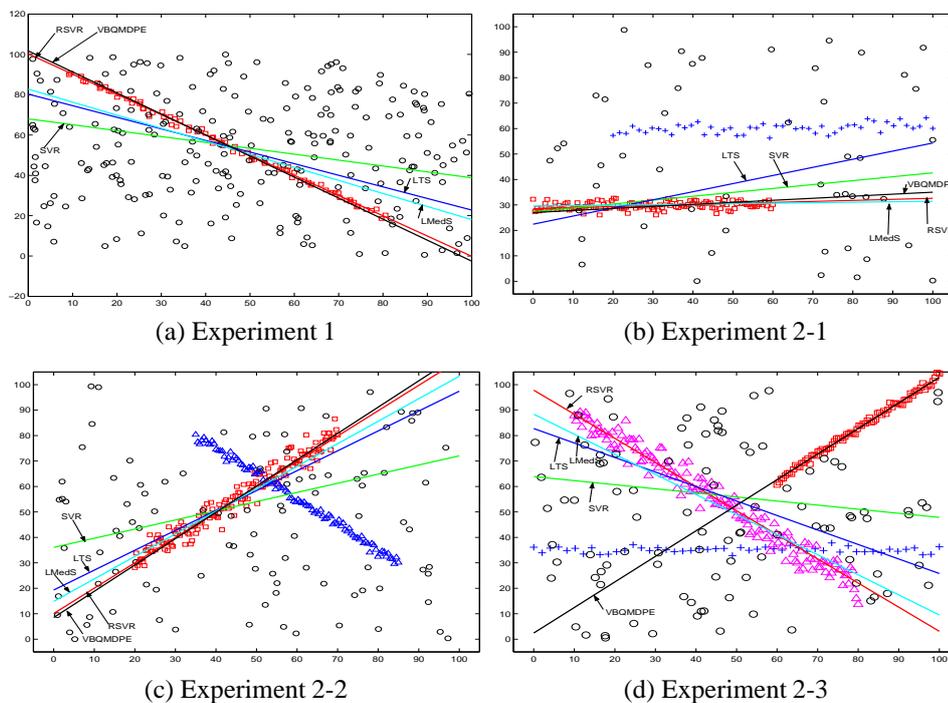(c) Experiment 2-2   (d) Experiment 2-3

Figure 3: comparative results

Results of experiments 2.1, 2.2 and 2.3 are illustrated on Figure 3. As we can note, when multiple structures are present in the data set, SVR and LTS fail to extract a dominant model, LMedS gives better results but it is less efficient and accurate than RSVR and VBQMDPE. In particular, the experiment 2.3 shows the robustness of RSVR in a very degraded training data set involving outliers and multiple structures with different

additive noise levels.

4 Application to the optic flow computation

In our experimentation, the OFCE is based on the first order Taylor approximation and we consider a constant motion model (see section 2). Our results have been compared with six published methods:
(1)- The Lucas Kanade's LS based method [5].
(2)- The Yohai's MM based algorithm [15].
(3)- The Ye's LTS based algorithm [14].
(4)- The Ong's LMedS based algorithm [7].
(5)- The Wang's MeanShift based approach [13].
(6)- The Farneback's approach (FAR)[2].

4.1 Parameters setting

Our algorithm has two significant parameters: the ε-insensitive margin and the regularization parameter C. We set $C = 10$ and $\varepsilon = 0.001$ for the tests.

In order to provide better estimates for objects moving with large displacements, all these estimators have been embedded in a multiresolution scheme which uses a coarse to fine strategy. In our tests, the image pyramid is sampled on three levels of resolution and a 7×7 pixel block is used to compute local optic flow. Derivatives were computed with four point central differences with mask coefficients $\frac{1}{12}(-1, 8, 0, -8, 1)$. Computations have been performed on AMD ATHLON XP 2400+ processor executing Matlab code.

4.2 Simulated sequence

The studied sequence depicts a simulated flight through to the Yosemite National Park. This sequence, which is available from *ftp://scd.uwo.ca* under the directory *pub/vision*, is one of the most complicated synthetic sequences that is widely used in the research community. We have added 6% randomly distributed outliers in the original sequence. This impulse noise causes locally a high motion discontinuity. Figure 4 shows some examples of estimated velocity fields. As we can note on these figures, RSVR, VBQMDPE and LMedS are effective in a degraded context and coincide well with the ground truth flow field in figure 4-a. The standard SVR, although better than the LS-based method, globally fails to recover the truth optic flow. These results are confirmed by the quantitative evaluation in table 1, where we have summarized the performance of the studied approaches for different noise levels. The Average Angular Error measure (AAE) used by Baron et al [1] is adopted as our performance measure and is a basis for comparison. As expected, we note that our method clearly improves the results of the classical SVR. The proposed algorithm also out-performs the most traditional estimators.

5 Conclusion

In this paper, we have presented a robust SVM regression-based method for the optic flow computation from an image sequence. We have experimentally shown that the proposed

Table 1: Comparative results from Yosemite sequence (AAE)

Noise (%)	FAR [2]	LS [5]	MM [15]	SVR [10]	**RSVR**	FastLTS [14]	LMedS [7]	vbQMDPE [13]
0	7.56	8.14	8.08	8.20	**7.56**	7.54	8.18	8.15
2	12.12	48.10	9.81	27.98	**8.25**	7.49	8.42	8.48
4	16.37	49.72	15.11	39.72	**8.99**	10.42	9.32	9.71
6	20.19	49.94	22.12	43.79	**9.99**	20.9	11.11	11.52
8	22.62	49.97	34.57	46.83	**12.75**	34.90	16.65	16.32
10	24.94	50.36	42.34	48.69	**17.91**	46.01	25.1	23.03

(a) Correct velocity field   (b) LMS method   (c) LTS method

(d) vbQMDPE method   (e) Standard SVR method   (f) RSVR method
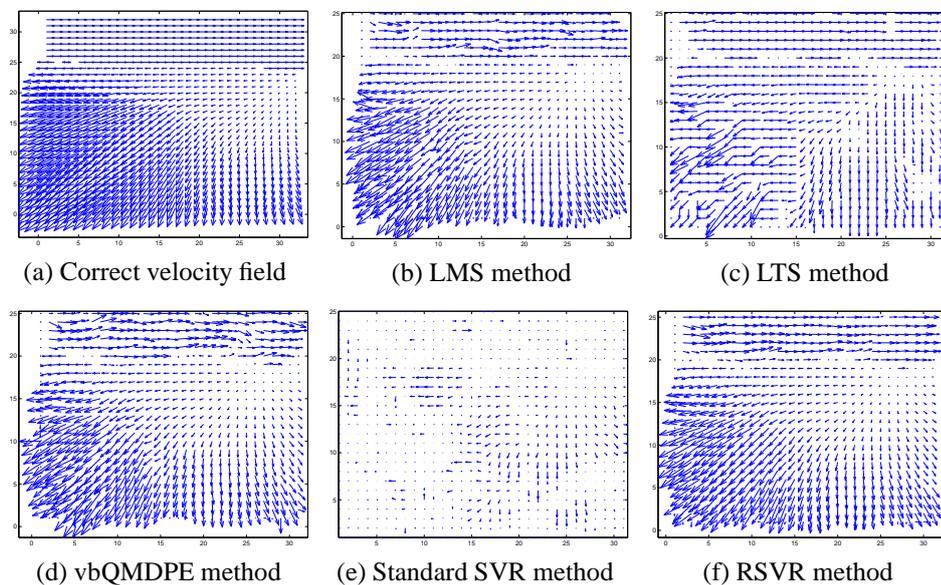
Figure 4: Optic Flows from Yosemite sequence with 6% noise added

method significantly improves the robustness against outliers contamination and multiple structures than traditional SVR. The principle of our approach consists in refining the solution obtained with the traditional SVR by iteratively removing outliers from the data set. The data partition consists of an adaptive cut-off value computed from the set of residuals instead of being constant. Evaluations on the optic flow computation have shown that the proposed algorithm may give promising results even when the image sequence is highly degraded. Nonlinear evaluations will have to be carried out to confirm these results.

6 Acknowledgments

References

[1] J.L. Barron, D.J. Fleet, and S. Beauchemin. Performance of optic flow techniques. In *Int. J. of Computer Vision*, volume 12, pages 43–77, 1994.

[2] G. Farneback. Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field. In *IEEE int. conf. on Computer Vision*, pages 171–177, 2001.

[3] B.K.P. Horn and B.G. Schunck. *Deterniming optic flow*, chapter 17, pages 185–203. 1981.

[4] P. Huber. *Robust Statistics*. Wiley, 1981.

[5] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *7th Int. Joint Conf. on Artificial Intelligence*, pages 674–679, 1981.

[6] H.H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 8, pages 565–593, 1986.

[7] E.P. Ong and M. Spann. Robust optical flow computation based on least-median of squares regression. *Int. J. of Computer Vision*, 31:51–82, 1999.

[8] J. Platt. *Fast Training of Support Vector Machines using Sequential Minimal Optimization*, chapter 12, pages 41–64. 1998.

[9] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT press, 2002.

[10] S. Golowich V. Vapnik and A. Smola. *Support vector method for function approximation, regression estimation and signal processing*. MIT press, 1997.

[11] V. Vapnik. The nature of statistical learning theory, 1995.

[12] H. Wang. Robust statistics for computer vision: Model fitting, image segmentation and visual analysis. *Ph.D. Thesis, Monash University, Department of Electrical and Computer Systems Engineering*, 2004.

[13] H. Wang and D. Suter. Variable bandwidth qmdpe and its application in robust optical flow estimation. In *ICCV*, pages 178–183, Nice, France, 2003.

[14] M. Ye and R.M. Haralick. Optical flow from a least-trimmed squares based adaptive approach. In *ICPR*, pages 7064–7067, 2000.

[15] V.J. Yohai. *High breakdown-point and high efficiency robust estimates for regression*, volume 15, pages 642–656. 1987.