Crisp Weighted Support Vector Regression for robust single model estimation : application to object tracking in image sequences

Franck Dufrenois
Laboratoire d'Analyse des Systeme du Littoral
50 rue Ferdinand Buisson, 62228 CALAIS
franck.dufrenois@lasl.univ-littoral.fr

Johan Colliez, Denis Hamad
Laboratoire d'Analyse des Systeme du Littoral
50 rue Ferdinand Buisson, 62228 CALAIS
johan.colliez@lasl.univ-littoral.fr

Abstract

Support Vector Regression (SVR) is now a well-established method for estimating real-valued functions. However, the standard SVR is not effective to deal with outliers and structured outliers in training data sets commonly encountered in computer vision applications. In this paper, we present a weighted version of SVM for regression. The proposed approach introduces an adaptive binary function that allows a dominant model from a degraded training dataset to be extracted. This binary function progressively separates inliers from outliers following a one-against-all decomposition. Experimental tests show the high robustness of the proposed approach against outliers and residual structured outliers. Next, we validate our algorithm for object tracking and for optic flow estimation.

1. Introduction

Estimation of a real-valued function from finite set of samples so as to explain their structure is a central problem in statistics. Introduced by Vapnik [21], Support Vector Regression (SVR) has become an emerging technique to solve this kind of problem. However, the standard approach is sensitive to a weak error rate in the training set. This sensitivity is a central drawback in some real image applications where datasets are highly degraded. In such applications, the structure to be estimated is often corrupted by outliers and sometimes also by residual structured outliers. Outliers are bad measurements, which may arise from physical imperfections in sensors or/and from previous image pretreatments. Generally, outliers are considered as uniformly distributed [22], [3], etc.. Structured outliers are measurements from one or more additional (residual) structures. In this paper, we propose a new estimator being able to resist these two types of errors. The proposed method is based on a weighted version of Support Vector Regression : Crisp SVR (C-SVR). The principle of the weighted SVR is to assign each datum a different penalty coefficient according to a predefined criterion. This concept has been already introduced in SVM based classification problem [10]. We propose a new methodology to extract a single (dominant) model in a degraded dataset. A binary function progressively separates inliers from outliers of the training data set and updates the weight values after each SVR estimation. It is based on the comparison of the absolute value of residuals with an adaptive threshold and thus it realizes a crisp data partition in two subsets. We show experimentally that the proposed method can tolerate up to 80% randomly distributed outliers in the data set and is resistant to residual structured outliers. We also show that these robustness properties are also preserved in a nonlinear context.

This approach, although similar to the robust SVR based approach of Colliez *et al.*in [5] and the Multiple Model Estimation (MME) method proposed by Cherkassky and Ma in [4], differs from them by several points : **Firstly**, in [4, 5], outliers are removed from the data set with a rejection criterion based on the absolute value of residuals and the estimation is updated with the remaining data subset. C-SVR directly integrates the data partition in the SVR formulation by the way of weights as in the Reweighted Least Squares Estimator [17]. In this way, the data partition is naturally based on the concept of support vectors which is an essential property of SVM theory. **Secondly**, in [4], the threshold is proportional to the standard deviation of additive noise in the dominant structure and the noise level is known a priori. In C-SVR, the threshold is adaptive and no information about noise level is needed. **Thirdly**, in a very degraded training data set, an analytic estimation of the SVM parameters ε and C such as proposed in [4] seems incomplete. In C-SVR, no analytic estimation of the parameters ε and C is necessary. The only assumption is to consider a small fixed ε value in order to obtain an initial estimate close to the dominant structure.

The paper is organized as follows. Section 2 presents the proposed weighted version of SVR and enlightens the reader about the terms contributing to robustness. In section

1-4244-1180-7/07/$25.00 ©2007 IEEE

3, the performances of C-SVR are studied and compared with other standard robust regression approaches. Finally, we apply this approach for rigid object tracking in image sequence and for the optic flow computation problem .

2. Weighted Support Vector Regression

2.1. Principle

Consider a training data set $S = \{(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), ..., (\mathbf{x_n}, y_n)\}$, where each $\mathbf{x}_i \in R^d$ denotes an input value and has a corresponding target value $y_i \in R$. The standard SVR builds a nonlinear function : $f(\mathbf{x}) = (\mathbf{w}, \phi(\mathbf{x})) + b$, such as the regression vector $\mathbf{w} \in R^d$ and the bias term $b \in R$ minimize the following constrained optimization problem [21, 1]:

$$\min_{\mathbf{w},\xi_i,\xi_i^*} L = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i\in S}(\xi_i + \xi_i^*)$$

$$sc : \begin{cases} r_i = y_i - (\mathbf{w}, \phi(\mathbf{x_i})) - b \leq \varepsilon + \xi_i \\ -r_i = -y_i + (\mathbf{w}, \phi(\mathbf{x_i})) + b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (1)$$

where the function $\phi(.)$ is a nonlinear application from R^d to high dimensional space called *feature space.* This formulation adopts the ε linear insensitive loss function proposed by Vapnik [21], which does not penalyse errors below some $\varepsilon > 0$. The form of the loss function is one of the fundamental keys to ensure robustness properties of an estimator. Besides, the form of the loss function is closely connected to noise modelling [21, 1] and this correspondence is obtained by maximizing the a posteriori probability (MAP). The slower the loss function increases with the errors, the more robust the estimation is. Thus, the linear ϵ-insensitive loss function presents the best characteristics of robustness among other common loss functions. The presence of *errors* in the data set is measured by other internal parameters ξ_i and ξ_i^* 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 (1). 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.
Therefore, standard SVR is still not sufficiently robust. In fact, as we will show it later, a small percentage of outlier contamination (<50%) is sufficient to force SVR to produce arbitrarily large values (small breakdown point). We propose here to develop a weighted SVR-based technique which can tolerate more than 50% outliers in the data. Let $\chi_i^{(k)}$ denotes the weight factor associated with the pair $(\mathbf{x_i}, y_i)$ at step k. The principle of the proposed approach consists of refining the regression vector $\mathbf{w}^{(k)}$ (estimated at step k) with a modified SVR where each penalization term in the sum is weighted by χ_i.The constrained optimization problem of the weigthed SVR is formulated as follows :

$$\min_{\mathbf{w}^{(k)},\xi_i,\xi_i^*} L = \frac{1}{2}\|\mathbf{w}^{(k)}\|^2 + C\sum_{i\in S}\chi_i^{(k)}(\xi_i + \xi_i^*)$$

$$sc : \begin{cases} r_i^{(k)} = y_i - \left(\mathbf{w}^{(k)}, \phi(\mathbf{x_i})\right) - b^{(k)} \leq \varepsilon + \xi_i \\ -r_i^{(k)} = -y_i + \left(\mathbf{w}^{(k)}, \phi(\mathbf{x_i})\right) + b^{(k)} \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (2)$$

It turns out that the previous optimization problem can be solved more easily in its dual formulation. For that, we construct the corresponding Lagrangian function which is expressed by :

$$Lp = \frac{1}{2}\left\|w^{(k)}\right\|^2 + C\sum_{i\in S}\chi_i^{(k)}(\xi_i + \xi_i^*) \quad (3)$$

$$-\sum_{i\in S}\alpha_i\left(\varepsilon + \xi_i - r_i^{(k)}\right) - \sum_{i\in S}\alpha_i^*\left(\varepsilon + \xi_i^* + r_i^{(k)}\right)$$

$$-\sum_{i\in S}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$

where $\alpha_i^{(*)}$and $\eta_i^{(*)}$ are nonnegative Lagrange multipliers. First, we differentiate Lp with respect to w, b, and $\xi_i^{(*)}$. Then, the resulting conditions of optimality are substituting in (3) and the dual problem becomes :

$$\max_{\alpha_i,\alpha_i^*} L = -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*).(\phi(\mathbf{x_i}), \phi(\mathbf{x_j})) \quad (4)$$

$$-\varepsilon\sum_{i,j=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n}y_i(\alpha_i - \alpha_i^*)$$

subject to: $\sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0$ and $\alpha_i^*, \alpha_i \in [0, \chi_i^{(k)}C]$.

Where the dual variables α_i and α_i^* are determined by Quadratic Programming techniques [1]. Then, the vector solution $\widehat{\mathbf{w}}$ and the estimated function $\widehat{f}$ are obtained from the following expressions:

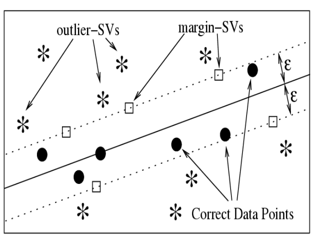
$$\widehat{\mathbf{w}} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(\mathbf{x_i}) \quad (5)$$

$$\widehat{f}(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\phi(\mathbf{x_i}), \phi(\mathbf{x})) + b \quad (6)$$

In expressions (4) and (6), the inner product in feature space $(\phi(\mathbf{x_i}), \phi(\mathbf{x}))$ can be favourably replaced by a kernel function $K(\mathbf{x_i}, \mathbf{x})$. Kernel functions enable dot product to be performed in high dimensional feature space using low dimensional input space without knowing an explicit expression of ϕ. Some common kernel functions are: the linear kernel function, the polynomial kernel function, the radial basis kernel function, etc.

2.2. Influence of the weight factors $\chi_i^{(k)}$

As we can see, the introduction of the weight factors $\chi_i^{(k)}$does not change the dual formulation problem of the standard SVR (see [21, 1]). However, the upper bounds of Lagrange multipliers $\alpha_i^{(*)}$are modified by dynamical boundaries (the upper bounds become $\chi_i^{(k)}C$ instead of C, see the constraints of Eq. 4). To explain the effect of the weight factors, the values of the dual variables $\alpha_i^{(*)}$ and their corresponding slack variables $\xi_i^{(*)}$ are analysed from Figure 1. Figre1 1-a shows a typical result of the linear SVR where the optimal line is illustrated by a solid line and the ε-margin by dotted lines. According to this figure, the data set can be divided into 3 subsets: data points inside the margin ('•'), data points on the margin ('□') and data points outside the margin ('*'). In practice, a simple test on the pair (α_i, α_i^*) (Fig. 1-b) makes it possible to classify the data in these three groups. As we can see from this table, only the subsets on ('□') and outside the margin ('*'), corresponding to $\alpha_i^{(*)} \neq 0$, give a non-zero contribution to SVM solution (see Eq. 5 and 6). The data inside the margin ('•') are considered as not support vectors and therefore have no contribution to SVM solution. This property is defined by $\alpha_i^{(*)} = 0$ and $\xi_i^{(*)} = 0$. Then, on the assumption that an outlier is identified, we can nullify its effect by forcing to 0 its corresponding slack variable $\xi_i^{(*)}$.



(a)

●	Not SV	$\xi_i^{(*)} = 0$	$\alpha_i^{(*)} = 0$
□	Is SV	$\xi_i^{(*)} = 0$	$\alpha_i^{(*)} \in]0,C[$
*	Is SV	$\xi_i^{(*)} > 0$	$\alpha_i^{(*)} = C$

(b)

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

2.3. Outlier rejection rule

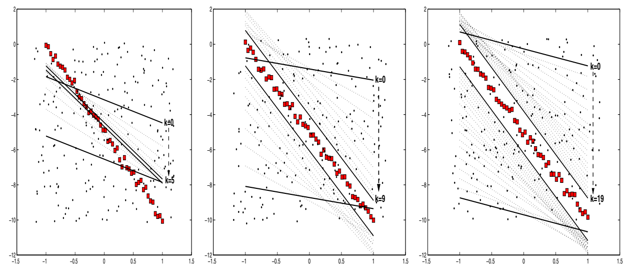
Following the previous idea, we propose a binary rule for outlier detection : The value 1(0) is assigned to the weight χ_i when the data point i has an absolute residual value lower (higher) than a given cut-off value M. So, the weight χ_i at time k+1 is determined by the following expression:

$$\chi_i^{(k+1)} = \frac{(1 + sign(M^{(k)} - |r_i|^{(k)}))}{2} \qquad \forall\, i \,\epsilon\, S \tag{7}$$

where the cut-off value M is proportional to the maximal absolute value of residuals, i.e. :

$$M^{(k)} = \beta \max_{i \in S} \left(\chi_i^{(k)}.\, |r_i|^{(k)} \right), \quad 0 < \beta < 1 \tag{8}$$

and $r_i = y_i - \widehat{f}(x_i)$ represents the residual value between the data and the estimated function. As for the parameter β,



(a) (b) (b)

Figure 2. Evolution of the binary margin according to the value of β. β=0.5 (a), β=0.7 (b), β=0.9 (c).

it manages the proportion of data which must be removed from the training data set at each step. To illustrate its influence, we consider a training data set composed of a dominant linear structure and randomly distributed outliers (Fig. 2). In this example, the outlier percentage is about 70%. Each figure depicts the evolution of the binary margin (i.e $\widehat{f}(x)^{(k)} \pm M^{(k)}$) according to the value of β. The data points outside the binary margin have zero weights. In figure 7.a, β is fixed at 0.5. In this example, the proportion of removed data is high and the convergence is not guaranteed. For the values of β =0.7 and 0.9, the proportion of removed data is enough to ensure the convergence (Fig. 2.b and 2.c). Of course, the higher the value of β will be, the longer the computation time will be (k = 9 for β = 0.7 and k = 19 for β = 0.9). Then, the crisp weighted SVR algorithm can be summarized below :

1. Initialize β and set $\chi_i^{(k=0)} = 1, \; 1 \leq i \leq n$.

2. Compute $\mathbf{w}^{(k)}$and $b^{(k)}$ from Eq. 9. Compute the residuals $r_i^{(k)}$ and the cut-off value $M^{(k)}$with Eq. 11.

3. Update $\chi_i^{(k+1)}$ with Eq. 10.

4. If the reconstruction error $\|\widehat{f}(x)^{(k)} - \widehat{f}(x)^{(k-1)}\| \leq \zeta(\ll 1)$, then stop. Otherwise, $k \longleftarrow k+1$ and go to step 2.

As we can point out to the reader, the proposed strategy is rather similar to M-estimators which use generally a weighting strategy to solve a regression problem. Usually, the weighting functions in M-estimators are real-valued, piecewisely defined and use a cut off point which is proportional to the scale of the data. The performance of the M-estimators lies in a robust scale estimate which is commonly solved with the MAD estimator. However, the MAD estimator is inaccurate when the outlier contamination (uniformly distributed outliers) is more than 50% or when other structured outliers are present. One of the contributions of our SVR based estimator is to propose an adaptive cut off point (Eq. 8) which involves no scale estimate. Of course, a real-valued weight function such as Huber's function or Tuckey's function could also replace the binary weight function defined by Equation 7. The robustness of the proposed approach depends on both the weak robustness of standard SVR and the proposed iterative weighting strategy. Thus, we obtain a very robust estimator and less sensitive to initialization than M-estimators.

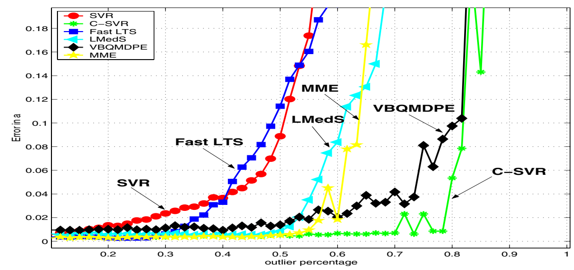
3. Performances and comparative results

3.1. Linear regression tests

First, the performances of our algorithm in line fitting will be demonstrated and its tolerance to large percentages of outliers will be compared with standard SVR, Cherkassky's method (MME) [4] and other robust popular estimators such as Least Median of Squares (LMedS) [17], Least Trimmed Squares (LTS) [16], and recently Variable Bandwidth Quick Maximum Density Power Estimator (VBQMDPE) [22]. In order to compare the abilities of the methods to resist different percentages of outliers, we will draw the "breakdown plot" [11]. This curve illustrates the evolution of the relative error between the estimated solution and the true solution according to the contamination rate. In the following tests, we fixed ε=0.001, C=10 for the standard SVR and other SVR-based techniques. For Cherkassky's method [4], we assume that the noise level σ for the domimant structure is known and the value of the threshold in [4] is constant (=2σ). For the proposed approach, we put β=0.7.

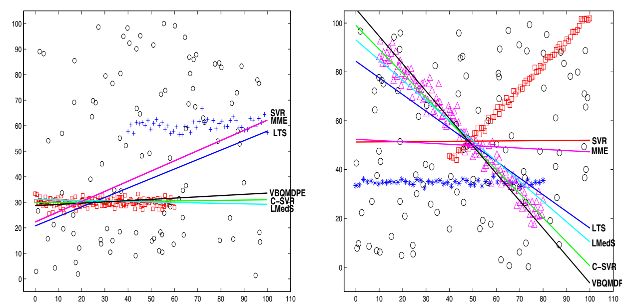
- **Experiment 1: Influence of outlier percentage on single linear model regression.** In this experiment, 300 data are composed of a noisy linear structure with outliers. n_1 data points define a single line y=-x+100 corrupted by a gaussian noise with zero mean and unit variance σ, and n_O data points are uniformly distributed in the range of (0,100). The percentage of outliers in the data points changes from 10% to 95%. We repeat this experiment 100 times and the averaged results are shown in Figures 3. Figures 3 illustrates the evolution of the relative error (breakdown plot) of the slope a of the estimated line for the six methods. This figure clearly shows three levels of robustness : a first group with the standard SVR and LTS which first break down between 30 and 40% outliers, a second group with LMedS and MME which break down between 50 and 60% outliers, and next VBQMDPE and C-SVR which break down between 70 and 80% outliers. In the third group, C-SVR (β=0.7) clearly outperforms VBQMDPE which is less stable in this interval and breaks down more quickly. These results show that C-SVR is very robust against uniformly distributed ouliers and outperforms the more standard robust estimators. In the next section, we illustrate the robustness behaviour of C-SVR in the presence of residual structured outliers in the data.

-**Experiment 2: Multiple linear model regression.** In



(a)

Figure 3. Breakdown plots of studied regression estimators. The x-axis represents percent data contamination and the y-axis characterizes the relative error for the slope.



(a) (b)

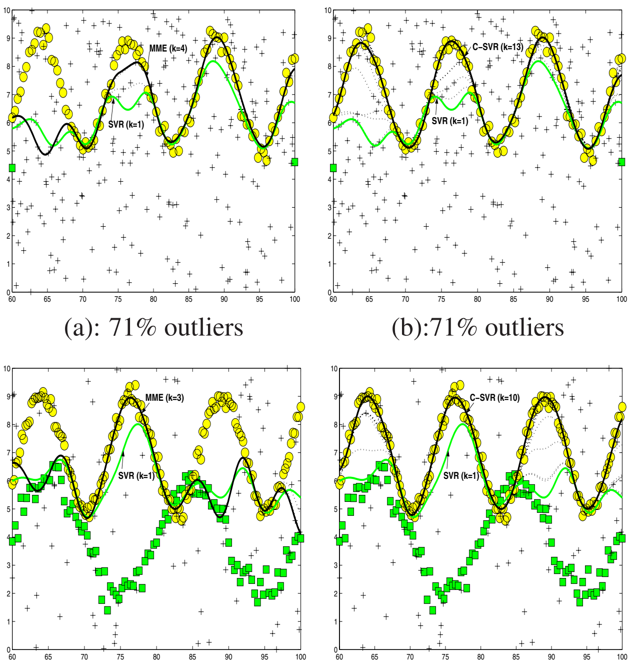
Figure 4. Illustration of line fitting with different robust approaches. Step edge (a). Three crossed lines (b)

this experiment, we investigate the characteristics of the six methods to fit lines in presence of outliers and residual linear structures. Every line is corrupted by gaussian noise with zero mean and different variance σ. The i^{th} line has n_i samples. We add n_O uniformly distributed outliers in the range of (0,100). The training data sets are defined as follows:

A step : x=(0 60), y=30, $\mathbf{n_1}$=150, σ=1.5 ('□'); x=(40,100), y=60, $\mathbf{n_2}$=80, σ=2 ('+'); $\mathbf{n_0}$=100 ('o') (Fig. 4-a)).

Three crossed linear structures : x=(40,100), y=x+3, $\mathbf{n_1}$=80, σ=1 ('□'); x=(0,80), y=35, $\mathbf{n_2}$=50, σ=1 ('*'); x=(10,80), y=-x+100, $\mathbf{n_3}$=150, σ=4 ('△'); $\mathbf{n_0}$=100 ('o') (Fig. 4-b)).

The line fitting results are shown on Figures 4.a and 4.b. Figure 4.a illustrates, for example, a pair of range surfaces forming a step discontinuity. Fitting a model to this kind of data is problematic because a standard fit (a least-squares fit) is skewed so much that it crosses (or "bridges") the point sets from both surfaces, placing the fit in close proximity to both point sets. It is besides what occurs for standard SVR, MME and LTS whereas VBQMDPE, LMedS and C-SVR seem less sensitive to this discontinuity. The second test (Fig. 4.b.) shows three crossed lines with dif-



(a): 71% outliers (b):71% outliers

(c): 57% outliers (d): 57% outliers

Figure 5. Extraction of a dominant nonlinear structure with SVR, MME and C-SVR. Noisy sinusoidal signal with outliers (a,b) . Noisy sinusoidal signal with outliers and a residual sinusoidal signal (c,d) .

ferent variances. Despite a high noise level, the dominant structure is correctly estimated by C-SVR. LTS, LMedS and VBQMDPE are close to the solution whereas standard SVR and MME fail. These examples show the abilities of the proposed approach to extract correctly a dominant structure from highly corrupted datasets. Of course, several complementary works as in [19] will have to be made in order to characterize the robustness limits of the proposed approach in multiple structure cases.

3.2. Nonlinear regression tests

The great advantage of SVR-based approaches compared to standard robust estimators, such as those studied here, is to be able to treat nonlinear structures by simply changing the kernel function. In this test, we experimentally demonstrate that the previous robustness properties are preserved in a nonlinear context.

In this experiment, we use the radial basis function kernel with a width fixed to 1. We generated two datasets, each with a total of 350 data points :

- The first example includes a noised sinusoidal signal (gaussian noise with σ=0.3, n1=100, 'o') corrupted with n_O =250 data points uniformly distributed in the range of (60,100) ('+') (Fig. 5-a and 5-b).

- The second example includes a noised sinusoidal signal (gaussian noise with σ=0.3, n1=150, 'o'), a noised sinusoidal signal (gaussian noise with σ=0.3, n2=100,'□') and n_O =100 data points uniformly distributed in the range of (60,100) ('+') (Fig. 5-c and 5-d).

The estimation of the dominant structures from both datasets is illustrated in Figure 5. In each figure, the grey curve represents the reconstruction result of standard SVR (first iteration of MME or C-SVR), the black curve illustrates the final result of the MME approach (Fig.5-a and 5-c) or the C-SVR approach (Fig.5-b and 5-d) and, the dashed curves the intermediate reconstructions. The results illustrated on Figures 5.b and 5.d confirm that C-SVR is able to extract nonlinear dominant structures even in a very disturbed context. On the other hand, the standard SVR and MME are not adapted to solve this kind of situation and fail to correctly extract the dominant structure (Figure 5.a and 5.c).

4. Real data tests

4.1. Application to video object tracking

In this section, we illustrate the performance of the proposed approach for rigid object tracking in image sequences. Here, the tracking problem is formulated as discovering the geometric transforms of object images between frames according to the extracted feature correspondences. In order to obtain a valid estimate of the transforms, a correct detection of feature correspondences is essential, which is, however, not easy in practice due to three factors : (1) the similarities, such as similar intensities and shapes, shared among features; (2) the occlusions and (3) the noise, which might also drive the data away from where they should be. Therefore, developing a robust feature matching technique is especially important.

In this study, the matching problem is solved with a two-step algorithm. The first step computes the "corners" on the object of interest in two consecutive frames. These feature points are extracted according to the Harris operator [8]. Then, the best correspondences are selected in the sense of both shape similar and intensity agreeing. The intensity based similarity measure is defined by the similarity of the histograms computed in the neighborhood of feature points. We use Bhattacharyya coefficient as a measure of similarity between two histograms [6]. The shape based similarity measure is based on the maximum-likelihood edge template matching technique of [14]. For simplicity, we assume the object motions to be locally affine. The affine transform of object images is formaluted as follows:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a1 & b1 \\ a2 & b2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c1 \\ c2 \end{pmatrix} \quad (9)$$

where the coefficients (a_1, b_1, c_1, a_2, b_2, c_2) are the affine parameters and (x',y')/(x,y) are the points in the target tem-

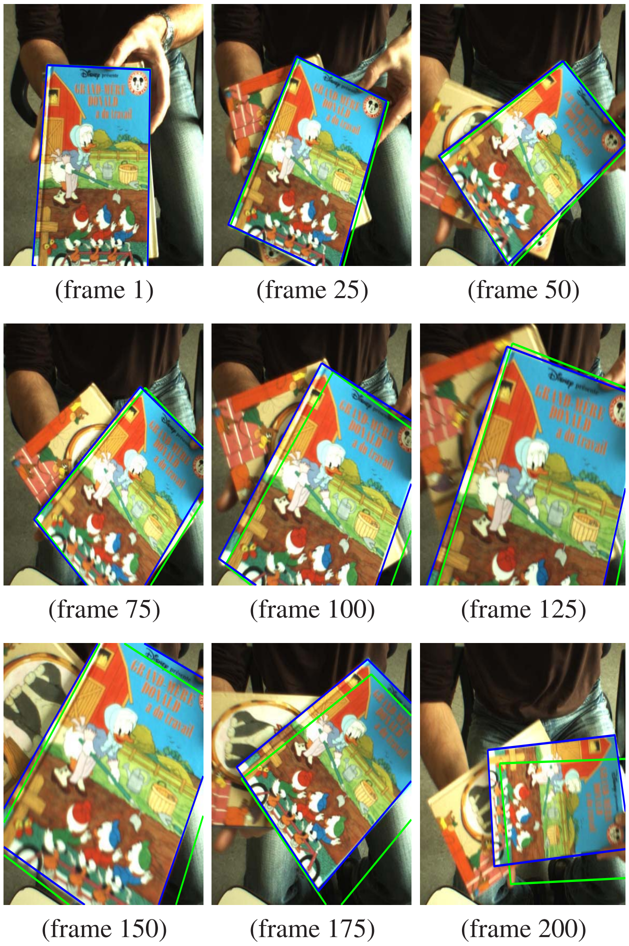


Figure 6. Tracking of a target book in a color image sequence (200 frames). Black rectangles are the results of the C-SVR based tracking and grey rectangles are the results of the MME based tracking.

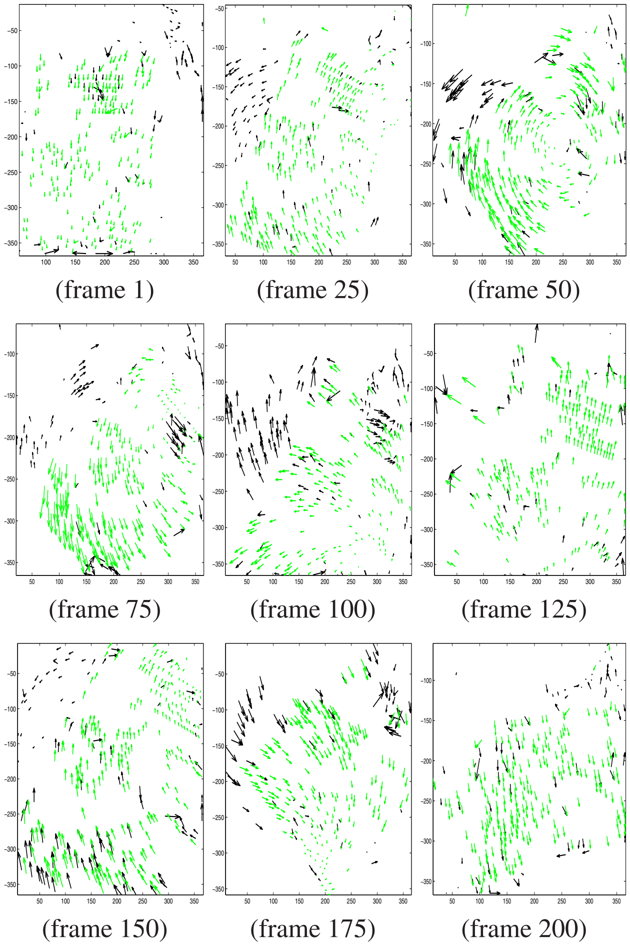


Figure 7. The grey vectors correspond to the dominant motion extracted ($\chi_i = 1$) and black vectors correspond to outlier motions ($\chi_i = 0$).

plate images. The estimation of the affine parameters can be reformulated as a regression problem. Under SVR formalism, if we put w=(a_1, b_1, c_1, a_2, b_2, c_2), we obtain two constraint sets :

one constraint over x' :

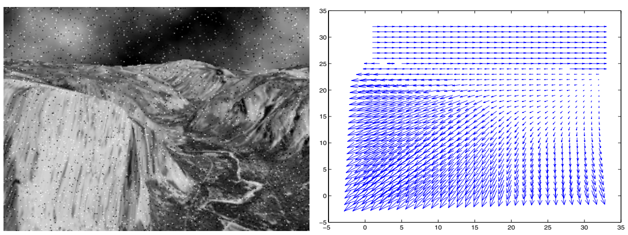
$$\begin{cases} r_{x'_i} = x'_i - \Big(\mathbf{w}, \phi_{x'_i}(x_i, y_i)\Big) - b \leq \varepsilon + \xi_{x'_i}, \\ -r_{x'_i} = -x'_i + \Big(\mathbf{w}, \phi_{x'_i}(x_i, y_i)\Big) + b \leq \varepsilon + \xi^*_{x'_i} \end{cases},$$

and one constraint over y' :

$$\begin{cases} r_{y'_i} = y'_i - \Big(\mathbf{w}, \phi_{y'_i}(x_i, y_i)\Big) - b \leq \varepsilon + \xi_{y'_i}, \\ -r_{y'_i} = -y'_i + \Big(\mathbf{w}, \phi_{y'_i}(x_i, y_i)\Big) + b \leq \varepsilon + \xi^*_{y'_i} \end{cases}$$

where $\phi_{x'_i}(x_i, y_i) = (x_i, y_i, 1, 0, 0, 0)$ and $\phi_{y'_i}(x_i, y_i) = (0, 0, 0, x_i, y_i, 1)$. Thus, the problem of the estimation means extracting two independent planes in the data set. Next, $\mathbf{w}$ and b are estimated with the Crisp-weighted SVR methodology as previously presented. It should be noted that the values of the parameters c_1 ($\mathbf{w}(\mathbf{3})$) and c_2 ($\mathbf{w}(\mathbf{6})$) must be updated with the value of b ($c_1 = c_1 + b$, $c_2 = c_2 + b$). The application concerns the estimation of the dominant motion in a color image sequence and the tracking of the corresponding object. The dominant motion corresponds to the displacement of a target book in the foreground. During the sequence, a residual motion of a second book appears. On the first frame, we manually select a window surrounding the target book. Feature correspondences are computed in the whole image. Figure 6 displays 9 selected frames of the sequence. In each frame, the black rectangle corresponds to the tracking result of the C-SVR method whereas the grey rectangle illustrates the tracking result of the MME method. Figure 7 shows vector fields which illustrate corner matches computed from the frames of figure 6. We clearly distinguish a dominant vector field (target book) corrupted by the presence of a residual motion (second book) as well as outlier motions (false matches). As we can see on this figure, the proposed approach accurately separates the dominant motion (grey vectors corresponding to $\chi_i = 1$ for a stopping criterion ζ fixed to $1e-3$) from the others (black vectors, $\chi_i = 0$). At the beginning of the sequence, the target book is successfully tracked by the two approaches because the book behind the target book is mainly occulted. Next, the second book appears more and more and then its leverage influence becomes more and more dominating. The MME rectangle (grey window on Fig. 6) then quickly shrinks too much and the tracking is lost. On the other hand, the target book is consistenly tracked throughout the sequence with the C-SVR based tracking (black rectangle).

4.2. Optic flow estimation



(a) (b)

Figure 8. (a) One frame of the Yosemite sequence severely degraded by 6% noise added. (b) Correct velocity field.

Table 1. Comparative results from Yosemite sequence (AAE)

Noise(%)	0	2	4	6	8
FAR [7]	7.56	12.12	16.37	20.19	22.62
LS [12]	8.14	48.10	49.72	49.94	49.97
MM [24]	8.08	9.81	15.11	22.12	34.57
SVR [20]	8.20	27.98	39.72	43.79	46.83
C - SVR	**7.58**	**8.25**	**8.99**	**9.99**	**12.75**
FastLTS [23]	7.54	7.49	10.42	20.9	34.90
LMedS [15]	8.18	8.42	9.32	11.11	16.65
vbQMDPE [22]	8.15	8.48	9.71	11.52	16.32

In this section, we illustrate the performance of the proposed approach in the optic flow computation problem. Many differential methods for optic flow estimation are based on the fundamental assumption of the image brightness conservation. Then, under this assumption, the optic flow constraint equation (OFCE) is approximated by a first order linear equation[9]:

$$I_x.u + I_y.v + I_t = 0 \qquad (10)$$

where (u, v) is the optic flow vector and (I_x, I_y, I_t) is the spatio-temporal 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 informations.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 [9], is a global approach and it is based on the definition of a functional derived from the OFCE and a smoothness penalty term. The second one is based on the assumption that motion is locally homogeneous [12]. It consists in forming a set of OFCEs in a small neighborhood around each pixel in order to improve the optic flow vector estimation. Thus, the optic flow problem is formulated as a set of over-determinated simultaneous linear equations. In this work, we adopt this solution and consider a constant motion.

Under SVR formalism, if we put w=$(u$,$v)$, for each pixel in the block, we obtain the following constraint set :

$$\begin{cases} r_{x'_i} = -I_t - \big(\mathbf{w}, \phi_{I_t}(I_x, I_y)\big) - b \leq \varepsilon + \xi_{I_t}, \\ -r_{x'_i} = I_t + \big(\mathbf{w}, \phi_{I_t}(I_x, I_y)\big) + b \leq \varepsilon + \xi^*_{I_t} \end{cases},$$

where $\phi_{I_t}(I_x, I_y) = (Ix, Iy)$. Next, $\mathbf{w}$ and b are estimated with the Crisp-weighted SVR methodology as previously presented.

Our results have been compared with five local approaches : (1) the Lucas Kanade's LS based method [12] - (2) the Yohai's MM based algorithm [24],[18] - (3) the Ye's LTS based algorithm [23]- (4) the Ong's LMedS based algorithm [15] - (5) the Wang's MeanShift based approach [22] - (6) the Farneback's approach (FAR)[7]. In the following tests, we fixed $C = 10$ and $\varepsilon = 0.001$ for the standard SVR and our approach (β=0.7).

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 $\frac{1}{12}(-1,\, 8,\, 0,\, -8,\, 1)$.

The comparative test is based on the well-known Yosemite sequence which depicts a simulated flight through to the Yosemite National Park (Fig. 8-a). This sequence combines two divergent motions and a translational motion of the sky (Fig. 8-b). To illustrate the robust behaviour of the proposed approach, the original sequence has been degraded by randomly distributed outliers with an increasing density in the range [0, 2, 4, 6, 8] (in percent). This impulse noise locally simulates a high motion discontinuity. Table 1 summarizes the comparative results of the different approaches. The Average Angular Error measure (AAE) used by Baron et al [2] 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

A weighted version of Support Vector Regression has been presented. The proposed weighting strategy effectively reduces the effect of outliers and structured outliers. Empirical results demonstrate that this approach can tolerate up to 80% uniformly distributed outliers and then out-performs most traditional robust estimators. Several experiments have also shown the effectiveness of the method to isolate a dominant structure from outliers and residual structured outliers. The application of the method to video object tracking and optic flow computation has given promising results.

Several issues are still under investigation. Firstly, the proposed approach can be extended to multiple model estimation as in [4]. Secondly, if the proposed approach is very resistant to vertical outliers (randomly distributed and structured vertical outliers), its robustness decreases more and more when the presence of bad leverage points or horizontal outliers increases. It is the case of the step as illustrated on Figure 4.a when the weight of the right edge compared to the left edge increases. We are working on reliable leverage diagnostics for removing these horizontal outliers. Thirdly, the object tracking experiment is based on the use of a specialized quadratic programming code to solve the dual formulation (Eq. 9) which is time consumming. In order to solve the problem considerably faster, we can use an active set strategy as proposed in [13].

References

[1] R. W. B. Schölkopf, A.J. Smola and P. Bartlett. New support vector algorithms. *Neural Computation*, 12:1083–1121, 2000. 2, 3

[2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optic flow techniques. In *Int. J. of Computer Vision*, volume 12, pages 43–77, 1994. 7

[3] H. Chen and P. Meer. Robust computer vision through kernel density estimation. *European Conference on Computer Vision*, pages 236–250, 2002. 1

[4] V. Cherkassky and Y. Ma. Multiple model regression estimation. *IEEE Transactions on Neural Networks*, 16:785–798, 2005. 1, 4, 8

[5] J. Colliez, F. Dufrenois, and D. Hamad. Robust regression and outlier detection with svr : application to object tracking in image sequences. *British Machine Vision Conference*, 3:1229–1239, 2006. 1

[6] V. R. D. Comaniciu and P. Meer. Real-time tracking of non rigid objects using mean shift. *IEEE Inter. Conf. On Computer Vision and Pattern Recognition*, 2:142–149, 2000. 5

[7] 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. 7

[8] C. Harris and M. Stephens. A combined corner and edge detector. *Proceedings Alvey conference*, pages 189–192, 1988. 5

[9] B. Horn and B. Schunck. *Deteniming optic flow*, chapter 17, pages 185–203. 1981. 7

[10] H. Huang and Y. Liu. Fuzzy support vector machines for pattern recognition and data mining. *International Journal of Fuzzy Systems*, 4(3):826–835, 2002. 1

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

[12] B. 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. 7

[13] D. Musicant and A. Feinberg. Active set support vector regression. *IEEE Transactions on Neural Networks*, 15(2):268–275, 2004. 8

[14] C. Olson. Maximum-likelihood template matching. *Proceedings of the International Conference on Computer Vision*, 2:52–57, 2000. 5

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

[16] P. Rousseeuw and V. Katrien. *Computing LTS regression for large data sets*, volume 12, chapter 1, pages 29–45. 2006. 4

[17] P. Rousseeuw and A. Leroy. *Robust regression and outlier detection*. John Wiley, 1987. 1, 4

[18] M. Salibian-Barrera and V. Yohai. A fast algorithm for s-regression estimators. *To appear in the Journal of Computational and Graphical Statistics*, 2005. 7

[19] C. Stewart. Robust parameter estimation in computer vision. *Society for Industrial and Applied Mathematics Reviews*, 41:513–537, 1999. 5

[20] S. G. V. Vapnik and A. Smola. *Support vector method for function approximation, regression estimation and signal processing*. MIT press, 1997. 7

[21] V. Vapnik. *The Nature Of Statistical Learning Theory*. 2nd edition. Springer Verlag, 1999. 1, 2, 3

[22] H. Wang and D. Suter. Variable bandwidth qmdpe and its application in robust optical flow estimation. In *ICCV*, pages 178–183, Nice, France, 2003. 1, 4, 7

[23] M. Ye and R. Haralick. Optical flow from a least-trimmed squares based adaptive approach. In *ICPR*, pages 7064–7067, 2000. 7

[24] V. Yohai. *High breakdown-point and high efficiency robust estimates for regression*, volume 15, pages 642–656. 1987. 7