QUALITY EVALUATION OF OPTIC FLOW ESTIMATORS : A NEED FOR VISION-BASED ROBOTICS

Etienne Grossmann, José Santos-Victor

Instituto de Sistemas e Robótica,
Instituto Superior Técnico, Torre Norte - Piso 7
Av. Rovisco Pais, 1
P 1096 Lisboa Codex - PORTUGAL
{etienne,jasv}@isr.ist.utl.pt

This paper describes the work done on performance characterization of vision algorithms at ISR's Vision Laboratory. We addressed this problem through the study of an "affine optical flow estimator". This estimator relies on a modified robust regression algorithm. After briefly defining the affine flow and explaining how the estimator works, we discuss the choice of a criterion for assessing the quality of an estimator. We then show in detail how to set some of the tuning parameters of the implementation. We also give a quick description of other parts of our work on performance characterization.

1 Introduction

The optic flow, a vector velocity field defined from a sequence of images, is commonly used in computer vision. We present the implementation of an algorithm that estimates the parameters of the first-order development of the optic flow at the center of the image. Many problems in computer vision, like the one we study here, can be closely linked to problems of estimation. We use a method derived from robust regression, which is not described in detail. An effort has been made to be able to assess the quality of the estimator we implemented. The possibility of assessing the performance of a vision algorithm is determinant for the successful application of vision to problems in Robotics. "Quality" has many possible mathematical definitions, and choosing one is non-trivial. We will detail how we proceeded in Sec. 4. Our quality is based on an error measure (a norm) defined on the quantities we want to estimate —the affine flow— which we define in Sec. 2. The estimator we developed has many parameters; we will detail the choice of two of them in Sec. 4. Finally, in Sec. 5, we cover briefly other parts of the work, which are more fully detailed in another paper [1] ; we also discuss possible extensions of our approach, both on the specific subject of optical flow, and more generally, about the methodology we used.

2 Optical Flow

The definition of the optical flow and the ill-posed nature of the problem of its estimation are well known. We refer to the book "Robot Vision" [2] for a richer definition, and to Barron et al [3] for a review of some available techniques for computing it.

The optical flow (u, v), defined over an image sequence I[a], is such that for every point (x, y) in the image, at any time instant t, the following equation is fulfilled :

$$\mathcal{D}_{(u,v,1)} I = 0 \tag{1}$$

where $\mathcal{D}_{(u,v,1)} I$ is the directional derivative of I in the direction $(u, v, 1)$. This means that along the lines of the 3D vector field $(u, v, 1)$, the brightness is constant. In computer vision literature, Eq. 1 is most often written as :

$$I_x u + I_y v + I_t = 0,$$

where I_x, I_y and I_t are the partial derivatives of the image sequence. If we consider the image as having continuous support, the above definitions are equivalent. We transpose the definition of Eq. 1 to discrete images :

$$\mathcal{D}^*_{(u,v,1)} I = 0, \tag{2}$$

where

$$\mathcal{D}^*_{(u,v,1)} I(x, y, t) = I(x + u, y + v, t + 1) - I(x, y, t)$$

It is widely recognized that the model of the optic flow is only *approximately* valid. The optical flow is a rich source of information about structure and motion, as shown by Koenderinck and Doorn [4] or, more extensively, Subbarao [5].

2.1 Affine Optical Flow

When watching a flat surface, the optic flow at the center of the image can be expressed as a degree-2 polynomial of the image coordinates, and one can show that the amplitude of the constant and linear terms is somewhat greater than that of the others. In many cases [7], we can approximate the flow by a first order model, which is easier to estimate, and often sufficient. The affine flow model results from a first-order expansion of the flow $(u(x, y, t), v(x, y, t))$

[a] An image sequence is a function taking as arguments a point (pixel) in the image plane p (also written (x, y)), and a time instant t. $(u(p, t), v(p, t))$ is then a "vector in the 2D image plane", or, similarly, $(u, v, 1)$ is a "vector in the 3D support of the image sequence" (we often omit the arguments (p, t)) for brevity.

around a point (x_0, y_0), with flow (u_0, v_0). The affine flow model can then be expressed as :

$$\begin{bmatrix} u(x,y,t) \\ v(x,y,t) \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

where u_x, u_y (resp. v_x, v_y) are the x and y-derivatives of u (resp. v). This equation can equivalently be written as

$$\begin{bmatrix} u(x,y,t) \\ v(x,y,t) \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} D + S_1 & -R + S_2 \\ R + S_2 & D - S_1 \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}, \quad (3)$$

where $D = (u_x + v_y)/2$ is called "Dilation", $R = (-u_y + v_x)/2$, "Rotation", and $S_1 = (u_x - v_y)/2$, $S_2 = (u_y + v_x)/2$ are the "Shear" terms. We prefer these notations, because the terms D, R, S_1 and S_2 appear more naturally in analytical expressions linking 3D motion to affine flow.

The purpose of the system we present here is to compute the parameters $(u_0, v_0, D, R, S_1, S_2)$ when the point (x_0, y_0) is $(0, 0)$, the center of the image. The generalization to any other point is trivial. A point worth mentioning is that solving for the six parameters that define the affine flow can be seen as finding the affine deformation[b] of the image plane that "best fits" the image sequence.

3 Estimation of Affine Optical Flow using a Modified Robust Regression Algorithm

3.1 Deriving the Observation Equation :

The observation equation we use to estimate the optical flow is —to our knowledge— different from the usual approaches. We further assume that x, y, t are integer numbers, and I has discrete support. I represents a —possibly filtered— image. A first-order approximation of $I(x + u, y + v, t + 1)$ in a neighborhood of $I(x + \tilde{u}, y + \tilde{v}, t + 1)$, for any (x, y) and $(\tilde{u}, \tilde{v})$ can be written as :

$$I(x+u, y+v, t+1) = I(x+\tilde{u}, y+\tilde{v}, t+1) + [I_x, I_y]\,[u - \tilde{u}, v - \tilde{v}]^T + h(u-\tilde{u}, v-\tilde{v}), \quad (4)$$

where h is the approximation "error" [c].

[b] Two images I_1 and I_2 are said to be affine transformations one of another if there are parameters $(u_0, u_x, D, R, S_1, S_2)$ such that $I_1(x + u, y + v) = I_2(x, y)$, where u, v are defined as in Equ. 3.

[c] A "h" function is defined for every pixel in the image (and every image), and it is thus more rigorous to call it $h_{I,x,y}$.

We assume that $h(0,0) = 0$, and that the expectancy of $|h(a,b)|$ (taken over all possible images) is an increasing function of $||(a,b)||$. This property can easily be verified in practice, for example, by plotting the empirical expectancy of "$|h|$". We may further assume that the functions $E|h(.,v)|$ and $E|h(u,.)|$ have their minimum in 0 (for all v and u).

The problem now is to derive an observation equation depending on the unknown flow values (u,v), by using Eq. 2 and keeping the error as small as possible. The flow vector (u,v) is a random variable, and we assume that we have an unbiased estimate $(\hat{u},\hat{v})$ of it. For a *chosen* vector $(\tilde{u},\tilde{v})$, we may write for any point (x,y), the following equations, by using Eq. 2 and Eq. 4 :

$$-\mathcal{D}^*_{(\tilde{u},\tilde{v},1)}I(x,y,t) + [I_x, I_y]\,[\tilde{u},\tilde{v}]^T + h(u-\tilde{u}, v-\tilde{v}) = [I_x, I_y]\,[u,v]^T \qquad (5)$$

where, apart from the error term $h(u-\tilde{u}, v-\tilde{v})$, all the terms in the left hand side can be computed from image data.

The choice of $(\tilde{u},\tilde{v})$, is then important , as it determines the magnitude of the approximation error. We took $(\tilde{u},\tilde{v}) = (\hat{u},\hat{v})$. This choice is based on the fact that : If we assume that Eq. 2 is exact; the density of (u,v) is such that the conditional densities of u and v are unimodal (for any value of v and u), with modes $\hat{u}$ and $\hat{v}$; the functions $E|h(.,v)|$ and $E|h(u,.)|$ have their minimum in 0 (for all v and u), then this choice can be shown to minimize the expectation of $|h(u-\tilde{u}, v-\tilde{v})|$. We believe this choice can be justified under weaker conditions. Then, the (noiseless) observation equation becomes :

$$-\mathcal{D}^*_{(\hat{u},\hat{v},1)}I(x,y,t) + [I_x, I_y]\,[\hat{u},\hat{v}]^T = [I_x, I_y]\,[u,v]^T \qquad (6)$$

3.2 Solving for the Affine Flow Parameters

Here we discuss how to solve for the six parameters that define the affine flow. We have a sequence of images, and will successively estimate the value of the parameters at all instants. By using the observation equation Eq. 6, together with the affine flow model Eq. 3, we get the final observation equation :

$$-\mathcal{D}^*_{(\hat{u},\hat{v},1)}I + [I_x\ I_y]\begin{bmatrix}\hat{u}\\ \hat{v}\end{bmatrix} = [I_x\ I_y\ xI_x{+}yI_y\ xI_y{-}yI_x\ xI_x{-}yI_y\ yI_x{+}xI_y]\cdot\boldsymbol{\theta}, \qquad (7)$$

where $\boldsymbol{\theta} = [u_0\ v_0\ D\ R\ S_1\ S_2]^T$ is the 6 dimensional vector characterizing the affine flow. We will solve for $\boldsymbol{\theta}$ by using many observations.

A natural choice for solving this equation would be a maximum-likelihood approach, which would lead us to a non-linear minimization problem. However, each noise term in Eq.7 would have to be characterized separately, which in practice is difficult.

3.3 Robust Regression Estimation

In the work presented here, we have simplified the problem by summing up all the noise terms : We then obtain a linear regression problem in which the left hand side of Eq. 7 are the observed values. Figure 1 shows an example of the residues in Eq. 7. It shows that the noise term cannot be assumed to have a Gaussian distribution, or even a "slightly contaminated" Gaussian. Despite

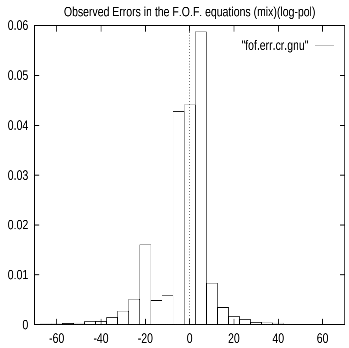

Figure 1: Histogram of the residues in Eq. 1, using the affine flow approximation, computed from real images. The unknown flow parameters are replaced by a good estimate (computed off-line, allowing big computation resource). The residue is the sum of the noise terms in Eq.7 and a term due to the error in the estimate.

this, we used robust regression, because we obtained fairly good numerical results. The robust regression estimate is computed iteratively, through a sequence of estimates which -hopefully- converges to the solution. Like many iterative numerical algorithms, the stopping point is an important issue, as will be discussed in Sec. 4. An important change we made in the robust regression algorithm, is that between iterations, we recompute all the observations (the left hand sides of Eq. 7) with our latest estimate, $(\hat{u}, \hat{v})$. In another paper [1], we show experimentally that this modification has positive effect.

For a complete description of the algorithm[d], we refer to the book[6] "Robust statistics", where it is called "robust regression using modified weights". We consider our estimator to be, like the robust regression it is derived from, asymptotically unbiased.

[d]Details of our implementation are: We took as "ψ-function", $\frac{2x}{1+x^2}$, after having observed that it performed better than a trimmed mean, and slightly better than the "Huber" estimator. We took a trimmed mean as scale estimate. The initial estimate is the least squares solution.

4 Experimental Results and Study of Influencing Factors

4.1 Error Measure of an Affine Flow Estimate

Here, we define the "error" of an estimated affine flow parameter vector $\hat{\boldsymbol{\theta}}$, when the ground truth $\boldsymbol{\theta}$ is available. We have used the norm defined by the inverse Σ of the covariance matrix of $\boldsymbol{\theta}$ (which is then seen as a random variable) :

$$||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}||^2_{OF} \stackrel{\Delta}{=} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\Sigma^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \tag{8}$$

The covariance matrix Σ is unknown, and it is estimated by the empirical covariance matrix $\hat{\Sigma}$ obtained from a population of estimates[e] , computed from "representative" sequences. One can see analytically that the approximation of the norm (i.e. of Σ) will improve as more images are used. We used a total of 40 images, although the use of more could be advantageous.

Arguing that in general, the different components $(u_0, v_0, D, R, S_1, S_2)$ of an estimate have no reason to be correlated, we removed all off-diagonal terms. The diagonal of the obtained inverse covariance matrix is :

$$\mathrm{diag}(\hat{\Sigma}^{-1}) = [3.89 \;\; 8.95 \;\; 8.35e5 \;\; 5.81e5 \;\; 6.73e5 \;\; 9.19e5]\,,$$

and is used from now on to compute the various error measures.

4.2 Measure of Performance of an Estimator

We will define the "performance" ("quality") of the implementation of an estimator, and discuss about setting the "tuning parameters". We took as the performance measure of an estimator, the mean squared norm of the error of the estimates it produces on a given set of sequences, as written in Eq. 9 :

$$\mathcal{E} \;=\; \hat{E}(||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2_{OF}) \;=\; \frac{1}{N_e}\sum_{\hat{\theta}} ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2_{OF}, \tag{9}$$

where N_e is the number of estimates, $\hat{E}$ is the empirical mean and $\boldsymbol{\theta}$ is the ground truth. The sum is taken over all the estimates $\hat{\boldsymbol{\theta}}$ produced by the studied estimator, over the considered sequences. Clearly, this measure is dependent on the used sequences. Unfortunately, it is non-robust, and a few far-lying estimates may degrade significantly the performance of an estimator. A trimmed mean may be a better measure. In any case, it may only be

[e] This implies that one already has a trustworthy estimator. We assert its trustworthiness with an approximate norm

calculated provided that the ground truth values are known, which is hardly ever the case in practise.

Transposing this performance measure to the case where the ground-truth is unknown is non-trivial : A natural idea is to substitute the true (but unknown) flow value, $\boldsymbol{\theta}$, by the empirical mean of the available flow estimate $\bar{\boldsymbol{\theta}}$ in Eq. 9.

$$\bar{\mathcal{E}} \;=\; \hat{E}(||\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}||^2_{OF}) \;=\; \frac{1}{N_e}\sum_{\hat{\theta}} ||\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}||^2_{OF} \tag{10}$$

When doing so, we want that the relative performance of two estimators remains unchanged, i.e. that the outcome of the comparison of two estimators will be unchanged when we use this error measure instead of the error "with ground truth" $\mathcal{E}$. Unfortunately, as shown in Table 1, this is not always the case : We can see that the estimator with the *lowest* error, using $\bar{\mathcal{E}}$, has the *highest* error $\mathcal{E}$.

Table 1: Experimental Errors $\mathcal{E}$ and $\bar{\mathcal{E}}$ of Estimators using from 0 to 5 iterations of the robust regression algorithm. The error measures using ground-truth, $\mathcal{E}$, and that using the mean of estimates, $\hat{\mathcal{E}}$, would give nearly opposite conclusions about the quality of the different estimators. The error measure using "good" estimates, $\mathcal{E}^*$ appears to give good results.

Number of Iterations	0	1	2	3	4	5
Sequence : Synthetically Deformed Outdoors						
$\mathcal{E}$	110.30	65.26	53.60	46.70	42.52	49.53
$\bar{\mathcal{E}}$	5.51	10.66	12.76	14.66	17.48	21.18
$\mathcal{E}^*$	30.00	14.27	13.55	15.77	20.56	24.27

To overcome this inconsistency, we can define still a different criterion, replacing $\boldsymbol{\theta}$ in Eq. 9, by a "good" estimate $\boldsymbol{\theta}^*$:

$$\mathcal{E}^* \;=\; \hat{E}(||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*||^2_{OF}) \;=\; \frac{1}{N_e}\sum_{\hat{\theta}} ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*||^2_{OF} \tag{11}$$

The computation of the "good" estimate $\boldsymbol{\theta}^*$ is a delicate choice : A large error in an estimate $\boldsymbol{\theta}^*$ will bias all error measures computed with it. We used a mean of 20 estimates produced by estimators that —to our knowledge[f]— were the best available[g] . The error measure $\mathcal{E}^*$ was used on real data only if it had

[f] Their quality being demonstrated on sequences with ground truth

[g] The characteristics of these estimators are : Spatial filter width : 7 or 9 , Number of iterations of robust regression algorithm : 2 or 3 (see below), Time-smoothing coefficient (see below) : 0.35. Number of observations : 2000

given acceptable results on synthetic data. Future work will of course address the problem of finding analytical results to support this quality criterion.

In the case of choosing the number of iterations of the estimator, this method is effective on synthetic data, as shown in Table 1, and gives coherent results with real-world data, as shown in Table 2.

Table 2: Effectiveness of $\mathcal{E}^*$ as Error Measure for the same estimators as in Table 1

Number of Iterations	0	1	2	3	4	5
	Sequence : Outdoors					
$\bar{\mathcal{E}}$	0.77	2.15	3.34	3.81	5.28	5.40
$\mathcal{E}^*$	6.17	3.31	3.90	4.17	6.02	6.58

We explain the increase of the error $\mathcal{E}^*$, when more iterations are done, by a (possibly) large error in the "good" estimates $\boldsymbol{\theta}^*$. The increase of $\mathcal{E}$ when 5 iterations are used, in Table 1, cannot be easily explained otherwise than by the unusual characteristics of the noise terms, as Figure 1 suggests.

4.3 Parameter Tuning

Once a criterion has been established to assess the quality of an estimator, it can be used to tune some of the parameters existing in almost every numerical algorithm. We have studied separately the tuning of some of the parameters that define the estimator : Space and time-filtering, number of observations used, etc. We here detail the tuning of the temporal smoothing.

We adopted an autoregression approach; more precisely, we used $I_t(x,y) = aI_{t-1}(x,y)+(1-a)I^o_t(x,y)$, where I^o is the original (non-smoothed) image,and I is the smoothed image. The subscript t is the considered instant. We will call a as the "time smoothing coefficient". Results are shown Table 3.

The error measure $\bar{\mathcal{E}}$ seems to fail on real-world sequences. On the contrary, $\mathcal{E}^*$ displays a good behavior (the result is what we expected, after seeing the results with synthetic data).

All the presented results were obtained in the following conditions :

- Time-smoothing coefficient (see below) : 0.35.
- Filter width : 7. (this value gave the best results)
- Number of iterations of the robust regression algorithm : 3.
- Number of observations : 1000.

Table 3: Error as a Function of Time Smoothing Coefficient. Error measure $\bar{\mathcal{E}}$ behaves badly, while $\mathcal{E}^*$ gives similar results to $\mathcal{E}$.

	0.0	0.1	0.2	0.3	0.4	0.5	0.7	0.9
	Sequence : Synthetically Deformed Outdoors							
$\mathcal{E}$	26.3	24.3	21	19.6	19.1	20.8	23.3	45.5
$\bar{\mathcal{E}}$	7.71	6.26	5.2	4.49	4.05	3.17	3.2	13.04
$\mathcal{E}^*$	10.4	8.13	5.82	4.93	5.00	6.36	9.47	53.2
	Sequence : Outdoors							
$\bar{\mathcal{E}}$	8.23	6.98	6.51	5.90	4.88	4.17	4.20	8.64
$\mathcal{E}^*$	10.8	8.62	8.02	6.56	6.16	6.96	9.44	27.7

When measuring errors, we used 20 estimates of each value, for sequences of 10 to 26 images. Then each error measure is the mean of 200 to 520 values. Two real-world sequences of 20 (310x232)) images were used ("indoors" and "outdoors"). The first image of each (shown in Fig. 2) was used in 4 synthetically deformed sequences of lengths varying from 10 to 26.

Figure 2: First Images of the Used Sequences

It is important to note that these two examples show that $\bar{\mathcal{E}}$ is —in our case— not a useful error measure, although we can theoretically justify its use by saying that it is an asymptotically unbiased estimate of $\mathcal{E}$.

5 Discussion

Various aspects related to quality evaluation of affine flow estimators have been presented. In another paper[1], we have also discussed some related research, dealing both with the justification of our observation and modified robust regression approaches, and with parameter tuning when the computational cost of the estimator is fixed. This work is currently being extended to different image geometries, such as log-polar images. In the future, we plan to address the following problems :

- Give firmer statistical foundations to the principles we used in assessing the quality of estimators : More rigorous performance assessment criteria would certainly increase the value of the assessment itself.
- Characterizing and modeling of the noise terms in Eq. 7 : These include measurement errors, but also the validity of the definition of optic flow, and that of the affine approximation on which the present work is based.

Working in these direction could be applied to ameliorating the numerical methods we used (e.g. presenting an estimator based on another principle than robust regression), but also to more general problems : We believe that many difficulties in estimation problems in computer vision are linked to insufficient specifications of both the input data, and the desired estimate.

Acknowledgments

This work is partially funded by the HCM-European Robotics Network (ERNET), and by a previous fellowship from the European Commission HCM Programme. Special thanks to César Silva for his help in troubleshooting.

References

1. Etienne Grossmann, José Santos Victor, "Robust affine flow estimation", 4th International Symposium on Intelligent Robotic Systems-SIRS96, Lisbon, Portugal, July 96.
2. Berthold Klaus Paul Horn, "Robot vision", M.I.T. Press, 1985.
3. J.L.Barron , D.J Fleet, S.S.Beauchemin "Performance of Optical Flow Techniques" ,Proc CVAP 1992 or Robotics and Perception Laboratory (Queen's University Kingston, Ontario, Canada , 1992
4. Jan J. Koenderinck and A.J. Doorn "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer" Optica Ccta 22, pp 773-791, 1975.
5. Muralidhara Subbarao "interpretation of visual motion: a computational approach", Morgan Kaufmann publishers,Inc, San Mateo, California
6. Peter J. Huber, "Robust statistics", John Wiley and sons 1981.
7. Shahriar Negahdaripour and Shinhak Lee, "Motion recovery using only first order optical flow information" , IJCV 9:3, pp 163-184 , 1992.