A Complete Confidence Framework for Optical Flow*

Patricia Márquez-Valle, Debora Gil, and Aura Hernàndez-Sabaté

Computer Vision Center
Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain
{pmarquez,debora,aura}@cvc.uab.cat

Abstract. Assessing the performance of optical flow in the absence of ground truth is of prime importance for a correct interpretation and application. Thus, in recent years, the interest in developing confidence measures has increased. However, by its complexity, assessing the capability of such measures for detecting areas of poor performance of optical flow is still unsolved.

We define a confidence measure in the context of numerical stability of the optical flow scheme and also a protocol for assessing its capability to discard areas of non-reliable flows. Results on the Middlebury database validate our framework and show that, unlike existing measures, our measure is not biased towards any particular image feature.

Keywords: Optical flow, confidence measures, sparsification plots, error prediction plots.

1 Introduction

Analysis and interpretation of optical flow, plays a central role in several safety-critical applications as diverse as decision making in car driver assistance and pathology discrimination for medical diagnosis support. A good interpretation and application of flow fields requires a measure of the confidence on the accuracy of the computed flow field.

Two main approaches compute dense flow fields: local and global. Local approaches go back to the early 80's [1] and compute optical flow from the Brightness Constancy Constraint (BCC) in a neighborhood of each pixel. They produce sparse vector fields that are further interpolated to obtain dense flows [2]. In order to minimize the impact of erroneous vectors in the interpolation stage, plenty of confidence measures for local methods have been developed [2]. Global techniques produce dense flow fields by combining into a variational framework a data-term and a smoothness-term. The data-term puts into correspondence one frame with the following one using either the BCC [3–5] or local techniques [6–8]. The smoothness-term determines the global properties of the vector field across the image [9]. Given that current variational schemes are more stable under a local drop of the data-term performance, the use of confidence measures has decreased. However, in dense flow fields we still need a measure to determine in which points the estimation is reliable or not.

Current confidence measures are based on either local image structure, the energy of the computed flow or statistical patterns [10]. Given that data-terms depend on sequence

* This work was supported by the Spanish projects TIN2009-13618, TRA2011-29454-C03-01, CSD2007-00018 and the 2nd author by The Ramon y Cajal Program.

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 124–133, 2012.

derivatives, measures based on local image structures take into account either the image gradient or its structure tensor [2]. The local energy of computed flows introduced in [11] can be computed for any scheme, but its value is strongly linked to the assumptions made by the variational model. Finally, statistical confidence measures [10] are based on the estimation of the flow distribution from a training data-set. They are independent of the particular variational formulation, but require a database including any unusual motion pattern. The bootstrap method proposed in [12] computes the variability of the computed flow with respect to a perturbation of the variational model. Pixels with high variability are associated to model inconsistencies and, thus, discarded. A main concern is that none of the above measures have been defined taking into account the error sources of the numerical schemes.

Assessing the capability of a confidence measure for discarding areas of poor performance is as important as the definition of the measure itself. A reliable confidence measure should present a decreasing dependency on optical flow error in order to guarantee that a bound on its values also bounds the error. As far as we know, the only way of assessing the performance of confidence measures are the Sparsification Plots (SP) reported in [11]. These plots represent the average error for those points with a measure above a value against the measure sorted values. It follows that the more decreasing SP are, the better the confidence measure should be. Although they successfully reflect the overall performance of a given measure, they are unable to assess any dependency between confidence measures and flow errors. As a consequence, they can not be used to determine confidence measure ranges ensuring a bound on flow error.

We introduce a complete confidence framework to define and assess confidence measures. First, we propose defining a confidence measure using error analysis techniques [13]. In particular, we give a measure for Lucas-Kanade-based approaches: the classic local [1] and the Combined-Local-Global (CLG) method described in [11]. Second, we assess the performance of confidence measures by computing the probability density function of having a decreasing dependency between flow errors and confidence measures. The plot of this probability for confidence measure values is called Error Prediction Plot, EPP. Our framework is validated in the benchmark Middlebury database [14] and compared to state-of-art measures and SP. Results show the discriminative power of EPP for detecting bias in existing measures. Both, SP and EPP plots show the higher stability of our measure based on numerical errors compared to existing measures. Consequently it is better suited for defining ranges ensuring a bound on optical flow errors with a given confidence.

2 State of the Art

Most local approaches are based on the method of Lucas-Kanade (LK) [1]. This approach is based on the assumption that optical flow keeps constant in a neighborhood of each pixel. Under this assumption the flow (u, v) solves:

$$\underbrace{\begin{pmatrix} K_\sigma * (I_x^2) & K_\sigma * (I_x I_y) \\ K_\sigma * (I_x I_y) & K_\sigma * (I_y^2) \end{pmatrix}}_{A_{LK}} \begin{pmatrix} u \\ v \end{pmatrix} = \underbrace{\begin{pmatrix} -K_\sigma * (I_x I_t) \\ -K_\sigma * (I_y I_t) \end{pmatrix}}_{b_{LK}} \tag{1}$$

for $I(x, y, t)$ denoting the image sequence, the subscripts the partial derivatives (x and y for spatial derivatives and t for temporal ones), $*$ the convolution operator and K_σ a Gaussian kernel of standard deviation σ.

Variational approaches compute the flow field by minimizing an energy functional that combines a data and a smoothness term:

$$E(u, v) = \int \underbrace{D(u, v, \nabla I)}_{\text{Data Term}} + \alpha \underbrace{S(\nabla u, \nabla v)}_{\text{Smoothness Term}} \, \mathrm{d}x \, \mathrm{d}y \tag{2}$$

where the data-term is usually based on the Optical Flow Constraint (OFC) ($I_x u + I_y v + I_t = 0$) and the smoothness-term models the general properties of the flow field.

Confidence measures can be split in three main groups:

Local Structure-Based: Since the data-term is formulated using the image partial derivatives, several measures are defined in terms of the image local structure, either gradient or structure tensor. Gradient-based measures [2] is defined as the magnitude of the gradient. Note that for large values of the magnitude of the gradient we expect reliable motion vectors. However, large gradients usually denote occlusions or noise [11], and in those pixels the optical flow computation is not reliable. In order to minimize the impact of noise, structure tensor based measures use information about the local structure of the image. These measures are especially well suited for LK-based schemes.

A main concern about local structure based measures is that they only take into account the data-term, so that, the impact of the filling-in effect of the smoothness-term is not considered. This limits their applicability to variational approaches, which are based on an energy functional that includes assumptions about the computed flow. In order to account for all assumptions of the underlying energy functional, a confidence measure based on the flow model was introduced by [11].

Energy-Based: This measure takes into account that variational techniques compute optical flow by minimizing an energy functional (2), and thus, the confidence measure is computed evaluating the flow field over the functional. We will refer to it as c_e. A main advantage of c_e is that it can be computed for any variational scheme.

We note that c_e only measures that (u, v) minimizes the energy functional and, thus, that it fulfils the assumptions made in the model. However, this does not guarantee that (u, v) corresponds to the true flow field, since defining a model for variational flow is still an open problem. Measures based on pattern analysis of computed flows are an alternative for defining confidence measures regardless of the model assumptions [10].

Bootstrap-Based: The measure defined by [12] quantifies the uncertainty of the flow method, that is, in those points where the flow field varies, the computation is not reliable. They compute such measure using bootstrap resampling. We refer to this measure as c_b, and we will consider the inverse of ψ_{bootg} defined in [12] eq.(15). Like the energy-based measure, this measure assesses the consistency of the model assumptions.

Observe that none of the above measures take into account the numerical stability of the method itself and, thus, it is not straightforward to derive their link to flow error.

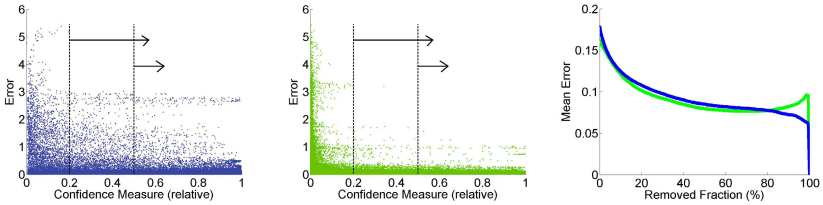

Fig. 1. Sparsification plots. On the left and on the middle the scatter plot of a confidence measure and the error. On the left a poor cm, on the middle a good one. The SP of both measures on the right.

2.1 Sparsification Plots

The most extended way to represent the performance of confidence measures is by means of the Sparsification Plots (SP) [11]. Such plots are given by the remaining mean error for fractions of removed flow vectors having increasing confidence measure values (cm). The scatter plots in fig. 1 illustrate the computation of SP for two representative cases selected from the Middlebury database. For a given removed percentage (vertical line in scatter plots and x-axis in SP below), arrows indicate the points that are considered for the computation of average errors (y-axis in SP).

Under the assumption that higher values of cm are associated to lower flow errors, SP should have decreasing profiles. An increase in their values for the higher removed fractions indicates artifacts in the decreasing dependency possibly due to a high error despite a high cm. However, the inverse does not always hold and random uniform dependencies could produce sensible plots. This is the case of the second representative sequence shown in fig. 1. Even if the dependency shown in the scatter plot is worse in the first sequence, its SP (blue line) indicates a better performance for high fractions.

Besides a poor power for assessing decreasing dependencies between confidence measures and flow error, SP are unable to properly detect if a measure is appropriate for giving a bound on flow accuracy. This is mainly due to the fact that its computation only considers confidence measure values for removing pixels regardless of optical flow error. Therefore, if the distribution of errors for high cm values concentrates around zero, the SP will be low even if we have some outliers with high errors.

3 A Complete Confidence Framework

From confidence measures we should expect a decreasing dependency between the measure and the accuracy, i.e., higher values of the measure are, higher accuracy we expect and viceversa.Thus, we state the dependency between confidence measure cm and error e using the following inequalities:

$$cm > \tau_c \Rightarrow e < \tau_e$$

for τ_c and τ_e chosen thresholds. That is, for a given probability, the ideal confidence measure should be able to guarantee that for a threshold τ_c on cm, the error, e, is

bounded. We call the above requirement the Condition of the Quality Threshold (CQT). We note that under CQT, the values of confidence measures would determine the accuracy of the flow field in the absence of ground-truth. Bearing the above requirements in mind, we propose the following confidence framework on the grounds of numerical stability analysis:

A Confidence Measure Based on OF Numerical Stability. There are two main sources of error: a deficient design of the algorithm (ill-conditioned) and round-off numerical propagation errors. The former can not be predicted from optical flow equations and requires a thorough analysis of the algorithm properties. The latter, can be analyzed using numerical stability concepts [13].

In regard to LK approaches [1, 7], the solution follows from a linear system. On the one hand, for local approaches [1] the system is given by (1). On the other hand, for the variational CLG method, the system (1) is the data-term of a variational framework:

$$E(u,v) = \int \psi_1\Big(\underbrace{K_\sigma * (I_x u + I_y v + I_t)^2}_{E_{LK}}\Big) + \alpha\, \psi_2\Big(|\nabla w|^2\Big)\, \mathrm{d}x\, \mathrm{d}y \tag{3}$$

for $\psi_i(s^2) = \sqrt{\beta_i^2 + s^2}$ a penalizing function, β_i a scaling parameter and $|\nabla w|^2 = |\nabla u|^2 + |\nabla v|^2$. The Euler-Lagrange equations of (3) are given in terms of the local LK system:

$$\frac{1}{\alpha}\psi_1'(E_{LK})\Big[A_{LK}\begin{pmatrix}u\\v\end{pmatrix} - b_{LK}\Big] = \begin{pmatrix}\mathrm{div}\big(\psi_2'(|\nabla w|^2)\nabla u\big)\\ \mathrm{div}\big(\psi_2'(|\nabla w|^2)\nabla v\big)\end{pmatrix} \tag{4}$$

for $\psi_i'(s^2) = 1/(\sqrt{1 + \frac{s^2}{\beta_i^2}})\ i = 1, 2$.

Therefore, we can determine the sources of errors of LK schemes by studying the properties and numerical stability of the system given by (1). Errors in the output data that come from errors in the input of the algorithm are called propagation errors. In our case, the error given by the input data is produced by the acquisition of the sequences.

The condition number [13] associated to a system of equations $Ax = b$, gives an upper bound of the error of the solution in relation with the error given in b. Given a square matrix A, if e is the error in b, then the error in the solution $x = A^{-1}b$ is $A^{-1}e$. The relative error in the solution to the relative error in b is the mentioned condition number and determines the error propagation. It is defined as follows:

$$\mathrm{K}(A) = \frac{\|A^{-1}e\|/\|A^{-1}b\|}{\|e\|/\|b\|} = \|A\|\|A^{-1}\| \tag{5}$$

for $\|\cdot\|$ a matrix norm. If we consider the L^2 norm and the matrix A is symmetric, the condition number simplifies to: $\kappa(A) = \lambda_{max}/\lambda_{min}$, where λ_{max} and λ_{min} are the maximum and minimum eigenvalues of A, respectively.

The condition number range is $[1, \infty)$. For large values the propagation of input errors is bad (ill-conditioned problem), whereas for low values (near to one) errors in the output compare to input errors and the problem is well-conditioned. In other words, if the condition number is large, then, the output error may not be bounded, that is, it can take any value (errors comparable with the input data errors or higher errors in the

output data). Meanwhile, if the condition number is close to one, the error of the output data is comparable to the error of the input data, and thus, the output data is reliable. It follows that low values of the confidence measure are associated with stability of the numeric solution, and, thus it can been used as a measure of the confidence of the computed OF.

Since the condition number is not bounded, we propose the following equivalent measure:

$$c_k = \frac{\lambda_{min}}{\lambda_{max}} \tag{6}$$

Notice that now, the range is $(0, 1]$. And thus, for small values the error propagation might be large, whereas for values near to 1 the error propagation will be small. We note that with this formulation, singularity of the system (1) cancels c_k, so that, one of the design errors of LK is also under control. In order to avoid indeterminate values such as $0/0$, formula (6) is computed setting to zero such cases. Since the LK matrix is symmetric, we propose c_k given by (6) as a measure that correlates with the accuracy.

The Capability of Confidence Measures for Predicting Errors. The CQT is fulfilled only if the scatter plots between a confidence measure cm and an error e show a decreasing pattern. Such pattern is difficult to measure using mathematical analysis tools because they are unable to properly handle point distributions. The best way to explore point distribution is by means of probability density functions. In probabilistic terms CQT can be stated as a conditional probability:

$$P_C(\tau_e, \tau_c) := P(e \geq \tau_e | cm \geq \tau_c) < \varepsilon \tag{7}$$

for $\varepsilon < 1$ the probability of having an error above τ_e provided that cm is above τ_c. The conditional probability can be computed by scanning the scatter plots given by cm-e. Taking into account that the condition $cm \geq \tau_c$ corresponds to a vertical line and $e \geq \tau_e$ to an horizontal one, the conditional probability is given by the fraction of points lying on the superior quadrant defined by the former lines. The scatter plots in fig.2 illustrate the computation of (7) for two representative cases selected from the Middlebury database. Arrows indicate the points that are considered for the computation of conditional probabilities.

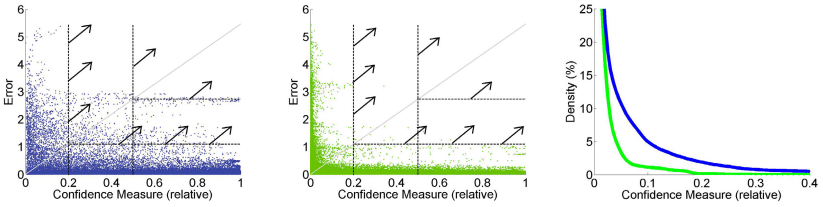

Fig. 2. Error Prediction plots. On the left and on the middle the scatter plot of a confidence measure and the error. On the left a poor measure, on the middle a good one. The EPP plot of both measures on the right.

The conditional probability (7) is a bi-dimensional graph, not easy to interpret. In order to get a simpler representation able to assess the capability of the measure for predicting the error, it suffices to consider the values for the diagonal of the square support of the variables e, cm. In this way, we ensure that unusual non-decreasing patterns (as the one shown in the green scatter point cloud in fig.2) are detected. We define our Error Prediction Plots, EPP, as the plot given by $(cm, P_C(e_{max} \cdot cm/cm_{max}, cm))$, for e_{max} and cm_{max} the maximum values of e and cm. Figure 2 shows scatter plots and their corresponding EPP. Unlike the SP shown in 1, we observe that EPP is worse for the non-decreasing case.

Besides their better potential for detecting poor confidence measures, EPP also serve to determine a threshold τ_c ensuring a bounded error, $e \leq \tau_e$. Given that points having $cm < \tau_c$ should be discarded, we will determine τ_c in terms of the percentage, $100 * \varepsilon$, of discarded points. In this context, τ_c is given by the intersection of the horizontal line $y = 100 * \varepsilon$ with EPP. That is, τ_c is given by the value of the confidence measure that satisfies:

$$P_C(e_{max} cm/cm_{max}, cm) = 100 * \varepsilon \tag{8}$$

The procedure described so far can only be computed for a representative sample of sequences with a ground truth. For the generalization to any sequence, statistical inference should be applied. In this framework, we should determine a confidence interval for τ_c ranges. In order to do so, the variability of EPP across the representative sequences should be as low as possible [17]. In this context, *the most relevant feature of confidence measures is not a highly decreasing pattern but a stable behavior across different sequences*. An homogeneous profile of EPP plots ensure small ranges for τ_c and, thus, reflect a higher capability for bounding e in terms of cm.

4 Experiments

Our framework has been validated on the well-known Middlebury database [14]. Motion has been computed using the CLG scheme implemented by [18]. In addition to c_k, the measures c_e, c_d and c_b have also been computed. The End-Point Error (EE) [14] is our accuracy score.

Two experiments have been carried out, one to validate the confidence framework, and a second one to test the error prediction capabilities of measures.

Validation of the Confidence Framework. In order to validate the presented framework, we firstly test the overall performance of c_k as well as the capability of EPP for assessing the decreasing dependency between confidence measures and errors. We assess c_k overall performance compared to other cm's by means of the gold-standard SP. Our EPP is validated by comparing their profiles to scatter plots.

We compare the different confidence measures in fig.3, which shows SP (left) and EPP (right) plots for two representative sequences of the Middlebury database. Below each quality plot we show the error scatter plots for each measure.

According to SP, none of the measures can be chosen as the best performer in global terms. We observe that c_k presents a more stable profile without any sudden increase for higher removed fractions, as Urban3 shows. This is consistent with the distribution

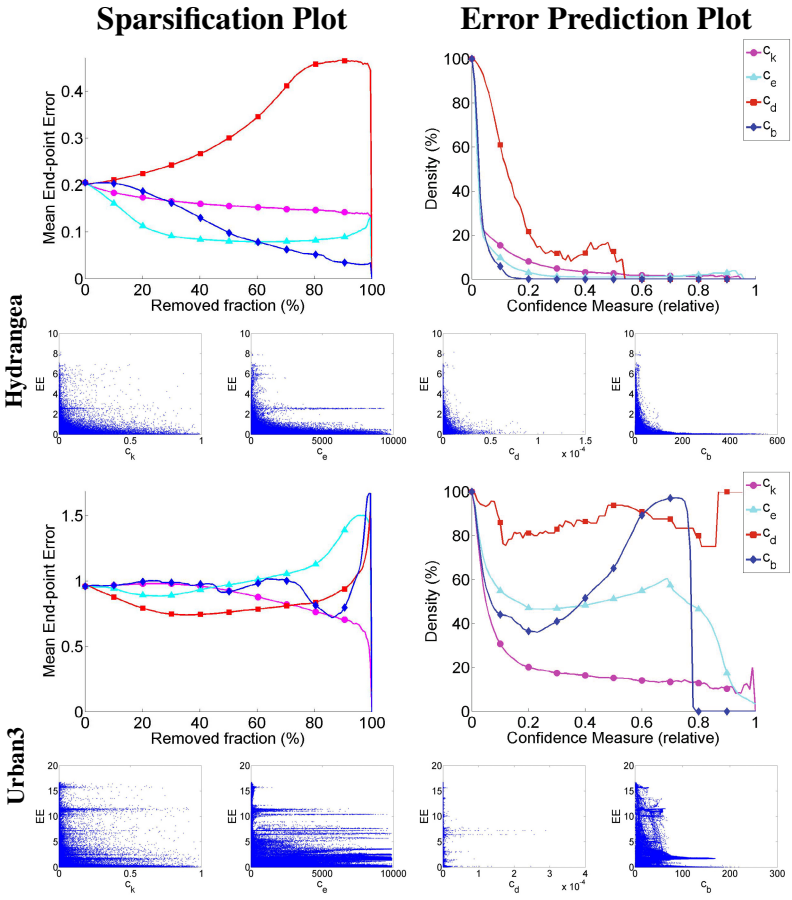

Fig. 3. On the left, the SP, on the right the EPP for the sequences Hydrangea and Urban3. Below the SP and EPP the scatter plots for the measures: c_k, c_e, c_d and c_b (from left to right).

of scatter plots for Urban3 shown below. However, in the case of Hydrangea, SP profile for c_d does not reflect the tendency observed in the scatter plots, which presents a clear decreasing dependency. This discrepancy between sparsification and scatter plots follow from few pixels having non-zero error (below 1 pixel) at the highest removed fractions. In comparison, SP for c_e has a lower profile, despite having pixels with an error above 2 for any removed fraction. Therefore, we must conclude that SP do not reflect dependencies between the confidence measure and error for all cases.

The ranking given by EPP plots agrees with SP for the Urban3 sequence, but clearly indicates a specific poor performance for the measures based on the model, c_e and c_b. Concerning Hydrangea, EPP profiles better reflect the dependency observed in scatter plots and indicate a slightly poorer performance for c_e. Another worthy point is the stability across the two sequences of the non-model based measures, c_k and c_d.

Error Prediction Capabilities of Confidence Measures. The goal of this experiment is two-fold. Firstly, detecting any grouping in cm behavior arising from a biased

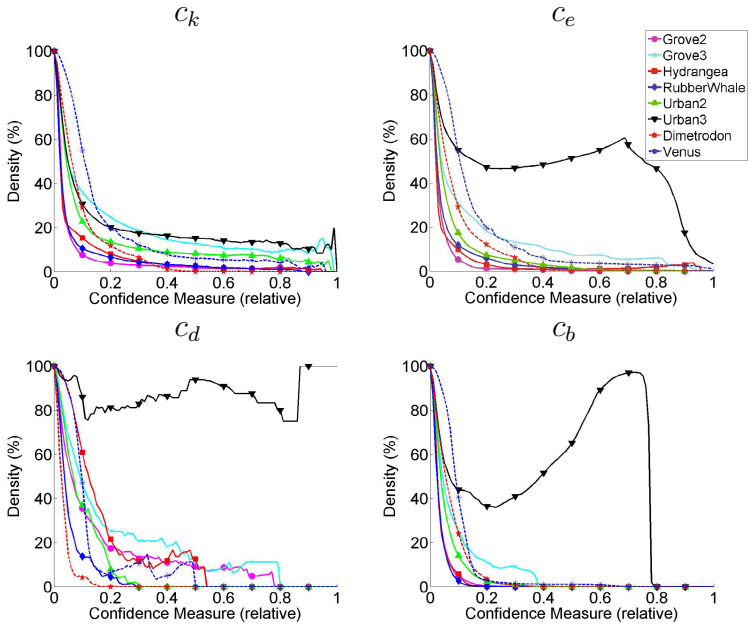

Fig. 4. EPP for c_k, c_e, c_d and c_b measures considering several sequences

definition. Secondly, assessing the variability across sequences and thus the accuracy of a confidence measure for controlling flow errors. For each cm we will consider the EPP for all sequences for the joint comparison of their profiles.

In order to assess the variability across sequences and detect any bias in confidence measures definition, for each measure we show its EPP for several sequences in fig.4. Observe that the measure that has stable behavior across different sequences is c_k. On the contrary, c_d, c_e and c_b are biased by the sequence. Firstly, the measure c_d has a different behavior for each sequence. Secondly, since the measures c_e and c_b are biased towards the assumptions of the flow model, both have similar profiles for the same sequences. Therefore, the measures c_d, c_e and c_b are less accurate for assessing bounds on optical flow error.

5 Conclusions and Future Work

We presented a confidence framework for assessing the performance of flow techniques by means of a numerical stability-based confidence measure and EPP.

As far as we know, none of the existing measures take into account the numerical stability of the method itself.

The SP can not give faithful comparisons across sequences because, the score they plot (mean EE) is not normalized across cases. On the contrary, the proposed EPP take into account both, the measure and the accuracy. That is, they provide a global vision of the capability of a measure to discard high errors. And thus, the problems that we

were facing with the SP are now solved and we can also determine a threshold for the confidence measure that gives a bound for the flow error.

Finally, in order to assess the flow field accuracy by means of a measure we need a stable behavior of the measure across different sequences.

In the near future we plan to generalize the proposed confidence measure based on the condition number for any kind of variational framework. In addition, we plan to explore the results considering different choices of evaluating EPP, that is, not only on the diagonal of the graph. And also, it would be interesting to do the same experiments considering different flow algorithms.

References

1. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: DARPA IU Workshop, pp. 121–130 (1981)
2. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. IJCV 12(1), 43–77 (1994)
3. Horn, B., Schunck, B.: Determining optical flow. AI 17, 185–203 (1981)
4. Nagel, H.H., Enkelmann, W.: An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. PAMI 8, 565–593 (1986)
5. Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: CVPR, pp. 2432–2439 (2010)
6. Bigün, J., Granlund, G.H., Wiklund, J.: Multidimensional orientation estimation with applications to texture analysis and optical flow. PAMI 13(8), 775–790 (1991)
7. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. IJCV 61(2), 221–231 (2005)
8. García-Barnés, J.: Variational framework for assessment of the left ventricle motion. Math. Mod. of Nat. Phen. 3(6), 76 (2008)
9. Weickert, J., Schnörr, C.: A theoretical framework for convex regularizers in pde-based computation of image motion. IJCV 45, 245–264 (2001)
10. Kondermann, C., Mester, R., Garbe, C.: A Statistical Confidence Measure for Optical Flows. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 290–301. Springer, Heidelberg (2008)
11. Bruhn, A., Weickert, J.: A confidence measure for variational optic flow methods. In: Geometric Properties for Incomplete Data, pp. 283–298 (2006)
12. Kybic, J., Nieuwenhuis, C.: Bootstrap optical flow confidence and uncertainty measure. Computer Vision and Image Understanding, 1449–1462 (2011)
13. Cheney, W., Kincaid, D.: Numerical Mathematics and Computing, 6th edn. Bob Pirtle, USA (2008)
14. Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. IJCV 92(1), 1–31 (2011)
15. Haussecker, H., Spies, H.: Handbook of Computer Vision and Applications, vol. 2. Academic Press (1999)
16. Shi, J., Tomasi, C.: Good features to track, pp. 593–600 (1994)
17. Newbold, P., Carlson, W., Thorne, B.: Statistics for Business and Economics. Pearson Education (2007)
18. Liu, C.: Beyond pixels: exploring new representations and applications for motion analysis. Ph.D. dissertation, Cambridge, MA, USA (2009)