A Coarse To Fine Multiscale Approach For Linear Least Squares Optical Flow Estimation

F. Lauze[1], P. Kornprobst[2], E. Mémin[3]

1. IT University of Copenhagen, Denmark,
francois@itu.dk

2. INRIA Sophia–Antipolis, France,
Pierre.Kornprobst@sophia.inria.fr

3. INRIA/IRISA University of Rennes, France,
Etienne.Memin@irisa.fr

Abstract

Tensor-based approaches for optical flow have been often criticized because of their limitations in handling large motions. This paper shows how to adapt them in a multiscale coarse-to-fine strategy. We show how the same ideas used in the variational framework can be adapted by working with both a multiscale image sequence as well as a multiscale, motion compensated tensor field. Several experiments are presented in order to compare it to some recent well known multiscale techniques. We demonstrate how this approach offers a good compromise between precision and computational efficiency.

1 Introduction

Computing Optical Flow or a deformation field from a sequence of images remains a crucial problem for many applications in computer vision. In the last twenty years, a lot of improvements were made since the pioneering work of Horn and Schunck and Lucas and Kanade (see [3, 2, 16] for some reviews).

The Lucas and Kanade approach, as well as its temporal extension, is based on a least squares fitting of a local translational flow on a small spatial support. This approach provides a very fast and local technique for motion estimation. It is in addition robust to noise, but presents nevertheless several important drawbacks such as delocalization of occlusion edges, difficulty to handle very smooth areas and important motion ranges. These difficulties have been quite well circumvented by hierarchical coarse-to-fine variational methods [1, 2, 14, 10] but at the expense of a heavy computational load.

In this paper we propose to adapt such a strategy to Lucas and Kanade-like approaches in order to handle some of the problems mentioned above. We propose to embed such an incremental scheme into a Gaussian scale space structure of the data together with a multiscale scheme which enables us to iteratively refine the spatio-temporal estimation support. This strategy enables us to rely on an appropriate hierarchy of scale-space data structure in order to estimate long displacement ranges, and to refine iteratively in a coarse to fine way the estimation support.

The paper is organized as follows: in section 2 we review some variational and Lucas and Kanade like approaches, with special focus on large scale motions. Our algorithm

BMVC 2004 doi:10.5244/C.18.79

is presented in section 3 and a link to the work of Nagel et al. is discussed. Section 4 presents a series of experimental evaluations, followed by conclusions in section 5.

2 OFC Based Approaches And Large Motions

Let $I : \Omega \subset \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}$ denotes an image sequence and $\mathbf{w} = (u, v, 1)^T$ be the displacement vector between the frame at time t and the frame at time $t+1$. Most of the apparent motion recovery algorithms start with the gray level constancy assumption, which means that the value of a pixel doesn't change while moving:

$$I(x_1 + u, x_2 + v, t+1) = I(x_1, x_2, t). \tag{1}$$

If we are considering small displacements, we can linearize (1) and obtain the optical flow constraint (OFC) equation

$$\mathbf{w} \cdot \nabla I = 0 \tag{2}$$

where ∇I is the 3D spatio-temporal gradient of I. This leads to the well known aperture problem and then additional constraints need to be considered.

2.1 Variational Approaches

One way to solve for the aperture problem is to add some *a priori* on the solution, such as smoothness properties. The standard approach is to minimize an energy incorporating the OFC and a penalty term (some function of the gradients of $\mathbf{w}$) for the smoothness:

$$\min_{\mathbf{w}} \int_\Omega (\mathbf{w} \cdot \nabla I)^2 \, dx_1' \, dx_2' + \int_\Omega \phi(\nabla u, \nabla v) \, dx_1' \, dx_2'. \tag{3}$$

Extensive research has been carried out in this direction, in particular to find the suitable penalty term (see [2] for a review).

Since the OFC becomes usually not valid for large displacements, such schemes have been embedded into hierarchical schemes to cope with long range motion. Several different approaches have been proposed. One may distinguish multiresolution techniques [14] which rely on a multiresolution pyramid of data, scale-space approaches [1] which use a scale-space decomposition of the data and multigrid techniques [10] which explore nested configuration spaces of solutions. All of these schemes rely on an incremental principle which consists in refining rough solutions iteratively along the levels of the considered hierarchical structure. More precisely, assuming a known coarse estimation $\mathbf{w}$ of the motion at a given pixel location $X = (x_1, x_2, t)$, one searches for an update $\mathbf{d}w = (du, dv, 0)^T$ such that

$$I(X + \mathbf{w} + \mathbf{d}w) = I(X). \tag{4}$$

The linearization of the above equation provides the so-called *shifted* OFC equation:

$$\mathbf{d}w \cdot \nabla^{\mathbf{w}} I = 0 \tag{5}$$

where $\nabla^{\mathbf{w}} I$ is the *shifted* or *motion compensated* gradient

$$\nabla^{\mathbf{w}} I = \begin{pmatrix} I_{x_1}(X + \mathbf{w}) \\ I_{x_2}(X + \mathbf{w}) \\ I(X + \mathbf{w}) - I(X) \end{pmatrix}. \tag{6}$$

2.2 The Lucas and Kanade Approach

The solution proposed by Lucas and Kanade [9] consists in assuming the optical flow vector to be constant within a some neighborhood, usually a spatial or spatio-temporal Gaussian window, if enough temporal information is available ([4]), defined through a Gaussian kernel h, centered at the location $X = (x_1, x_2, t)$ of estimation and solve for it in a least square sense, i.e. by finding the minimizers of the local energy function

$$E_X(\mathbf{w}) = h * (\nabla I \cdot \mathbf{w})^2(X) = \mathbf{w}^T h * \left(\nabla I \nabla I^T\right)(X)\mathbf{w}. \tag{7}$$

The quantity $S = S(X) = h * \left(\nabla I \nabla I^T\right)(X)$ is by definition the *Structure Tensor*. The problem to solve is then

$$\underset{\mathbf{w}=(u,v,1)^T}{\operatorname{Arg\,min}} \; E_X(\mathbf{w}) = \mathbf{w}^T S \mathbf{w}. \tag{8}$$

A minimizer $\mathbf{w}$ of this energy must satisfy the equation $\nabla E_X(\mathbf{w}) = 0$ which is carried out by solving the 2×2 linear system

$$\nabla E_X(\mathbf{w}) = \begin{pmatrix} h * I_{x_1}^2 & h * I_{x_1} I_{x_2} \\ h * I_{x_1} I_{x_2} & h * I_{x_2}^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} h * I_{x_1} I_t \\ h * I_{x_2} I_t \end{pmatrix} = 0. \tag{9}$$

These kind of approaches extract the motion from the analysis of the 2D+t structure of the volume which is coded in the structure tensor. The shape of the window of integration will determine which 2D+t structures can be successfully coded and so which motions can be recovered. One needs to find a compromise between having a sufficiently large kernel to disambiguate the OFC and a small enough one that does not average too much structures. When multiple range motions have to be estimated, a common kernel for every speed is not satisfactory (see figure 1).

2.3 Nagel et al. Approach [13, 12, 11]

Nagel *et al.* have introduced a very interesting algorithm, derived of the Lucas and Kanade approach, that aims at segmenting the motion field in different categories. This approach combines both image data smoothing and structure tensor smoothing. The algorithm is summarized in table 1. The pseudo-inverse operation in table 1, step 2-(c), consists in

```
1. $\Sigma_G^0(X) = \mathrm{diag}(4,4,1)$.                    *Nagel etal approach* (Sec. 2.3)
2. For $n = 0$ to $N$
    (a) $\nabla I^n(X) = (I * \nabla G_{\Sigma_G^n})(X)$
    (b) $T^n(X) = \left((\nabla I^n)(\nabla I^n)^T * G_{\Sigma_{M^n}}\right)(X)$, $\Sigma_M^n \approx 2\Sigma_G^n$
    (c) $\Sigma_G^{n+1}(X) =$ a pseudo-inverse of $T^n(X)$
3. Output motion segmentation from analysis of the tensor field $T^N$
```

Table 1: Nagel et al. algorithm

remapping in a strictly positive range the eigenvalues of $T^n(X)$, using a bounded decreasing function. A major drawback of this method is its high running time due mainly to all the convolutions with non-separable Gaussian kernels. We will see in section 3.2 that our proposed algorithm uses, in a rough sense, a simplification of this one.

3 Using Shifted Structure Tensors

3.1 A Coarse To Fine Strategy

Within an incremental coarse-to-fine strategy (4-5), assuming the update $\mathbf{d}w$ constant on a Gaussian neighborhood of X, we can built a least squares solution at X exactly as in (8)-(9). The incremental field $\mathbf{d}w$ is given at each point as the minimizer of the quadratic energy

$$E_{\mathbf{w}}(\mathbf{d}w) = \mathbf{d}w^T S^{\mathbf{w}} \mathbf{d}w \qquad (10)$$

where $S^{\mathbf{w}} = \langle (\nabla^{\mathbf{w}} I) \left(\nabla^{\mathbf{w}} I^T\right) \rangle$ is a *shifted* or *motion compensated* structure tensor. The final motion field $\mathbf{w}$ is a sum of piecewise constant approximations of the flow with the first estimates constituting roughly the principal components of the motion field whereas the last increments figure detail corrections. Such a description of this scheme acting by successive refinements needs to use decreasing estimation support. Sufficiently large Hessian supports are used to estimate rough average motion components whereas final detail corrections are estimated on a much smaller support. Low frequency space solutions are first exhibited and are then completed by high frequency details.

This finally suggests the following coarse to fine strategy: choose an integer $N \geq 0$ and a sequence (τ_s^n, τ_t^n), $n = 0 \dots N$, such that $\tau_s^{k+1} > \tau_s^k$, $\tau_t^{k+1} > \tau_t^k$ and let h_k denote the spatio-temporal Gaussian kernel with spatial standard deviation τ_s^k and temporal standard deviation τ_t^k. We then propose the algorithm in table 3 (upper part) which is somewhat close with the work presented in section 2.3. Indeed, step 2-(a) of their algorithm (in table 1) computes gradients giving higher weights in the spatio-temporal direction of minimal variations of grey-levels, i.e. in the motion direction, therefore performing an implicit motion compensation. While we do it explicitly, the loop 2 of their algorithm corresponds to an iterative refining approach.

```
1. Set $\mathbf{w}^{N+1} \equiv 0$.                    *Coarse-to-fine approach* (Sec. 3.1)
2. For $n = N$ down to 0
    • Form the shifted structure tensor $S^{\mathbf{w}^{n+1}}(X) = h_k * \left(\nabla^{\mathbf{w}^{n+1}} I\right) \left(\nabla^{\mathbf{w}^{n+1}} I^T\right)(X)$
    • Compute the update $\mathbf{d}w^n$ as the minimum of the quadratic energy $E_{\mathbf{w}^{n+1}}(\mathbf{d}w)$
    • Set $\mathbf{w}^n = \mathbf{w}^{n+1} + \mathbf{d}w^n$
3. Output $\mathbf{w}^0$.
```

Table 2: Synopsis of the coarse to fine approach

```
1. Set $\mathbf{w}^{K+1,0} = 0$.                    *Multiscale coarse-to-fine approach* (Sec. 3.2)
2. For $k = K$ downto 0
    (a) Compute $I_k = g_k * I$
    (b) $\mathbf{w}^{k,N+1} = \mathbf{w}^{k+1,0}$
    (c) For $n = N$ downto 0
        • Form the shifted structure tensor
          $S^{\mathbf{w}^{k,n+1}}(X) = h_k * \left(\nabla^{\mathbf{w}^{k,n+1}} I_k\right) \left(\nabla^{\mathbf{w}^{k,n+1}} I_k^T\right)(X)$
        • Compute $\mathbf{d}w^{k,n}$ as the minimum of the quadratic energy $E_{\mathbf{w}^{k,n+1}}(\mathbf{d}w)$
        • Set $\mathbf{w}^{k,n} = \mathbf{w}^{k,n+1} + \mathbf{d}w^{k,n}$
3. Output $\mathbf{w}^{0,0}$.
```

Table 3: Coarse-to-fine multiscale tensor-based motion estimation

3.2 Incorporating a Multiscale Framework

We describe here a multiscale algorithm coupling these two strategies: let K a positive integer, we define two family of Gaussian kernels:

- g_k, $k = K \dots 0$, with spatio-temporal standard deviations (σ_s^k, σ_t^k) so that $(\sigma_s^k, \sigma_t^k) > (\sigma_s^{k-1}, \sigma_t^{k-1})$. This will be used to smooth of the image data.

- h_{kn}, $k = K \dots 0$, $n = N \dots 0$, with standard spatio-temporal deviations $(\tau_s^{k,n}, \tau_t^{k,n})$ such that $(\tau_s^{k,n}, \tau_t^{k,n}) > (\tau_s^{k,n-1}, \tau_t^{k,n-1})$. This will be used for the least squares estimates of the motion update vectors.

We also require that $(\tau_s^{k,0}, \tau_t^{k,0}) \gg (\sigma_s^k, \sigma_t^k)$ which states that the minimal aperture at a given image scale for forming the structure tensors must be sufficiently larger than the one used to smooth the image data, otherwise the extracted gradients will be too well aligned in the chosen neighborhood, and we might be in presence of the aperture problem.

This algorithm is shown in table 3. Standard central differences are used for the spatial derivatives, and Gaussian filtering is approximated by Deriche recursive filters [7]. In order to compute the values $I(X + \mathbf{w})$ we use bilinear interpolation. The standard deviation sequences used are exponential: given $\alpha > 1$, $\beta_k > 1$, we set

$$(\sigma_s^k, \sigma_t^k) = \alpha^k (\sigma_s^0, \sigma_t^0), \quad (\tau_s^{k,n}, \tau_t^{k,n}) = \beta_k^n (\tau_s^{k,0}, \tau_t^{k,0}).$$

4 Experiments

Our algorithm has been implemented in C++ on a 1.6Ghz Pentium IV processor. In all experiments but the last, we have used $K = N = 3, \sigma_s^0 = 1.0, \sigma_s^0 = 0.5, \tau_s^{k,0} = 3.0, \tau_t^{k,0} = 1.5, \alpha = \beta_k = \sqrt{2}$. Interestingly the approach is not so sensitive to changes in the pyramids levels (as soon as we start from coarse enough scale) which is an advantage with respect to multiresolution schemes that can lead to quite different results with different scales.

The proposed method, denoted CFLS (Coarse to Fine Least Squares) is compared with the new variational algorithm by Brox *et. al.* [5] (denoted BBPW), which provides some of the best results known so far, and with the multiscale variational approach by Aubert *et. al.* [2] (denoted ADK).

We first comment some experiments using several synthetic image sequences where the ground truth is known. The following measurements have been estimated (see [3]): AAE (the average angular error), AESTD (the angular error standard deviation), ANE (the average norm error) and NESTD (norm error standard deviation), RT (running time).

Different aspects have been evaluated. In addition, we have computed orientation images, for some of the sequences, the orientation represented by a color, and the used color code is indicated by a small band at the image boundary.

(a) The ability to recover a large range of motions (Figures 1 and 2). This sequence consists of five stacked translating patterns at different speeds (20, 13, 7, 4 and 2 pixels per frame leftward). Figure 1-(a) shows one frame of the sequence, 1-(b) the true flow and 1-(c) shows that using the original Lucas and Kanade approach with several homogeneous Gaussian kernels doesn't allow to recover the different ranges of motions and discontinuities. The graph in figure 2-(a) and the table 2-(c) allow to compare the different methods. BBPW is clearly better as shown in figure 2-(a), which represents a vertical cut of the velocities for $x = 34$. Figure 2-(b) illustrates the refining behavior of estimations in our algorithm (locations indicated by the spots in figure 1-(a)).

(b) Robustness to noise (Figure 3). This sequence is of dimensions 170×255×20 pixels, has a leftward 10 pixels per frames motion, but is heavily contaminated by non stationnary noise: the frames are divided in 3 rectangles with Gaussian noise of standard deviation 35% for the upper one, 58% for the middle one and 86% for the bottom one. Figure 3)-(b) is the true flow, and figure 3)-(c) shows very good computed motion orientations, although severe errors (overestimates) are present at the left boundary.

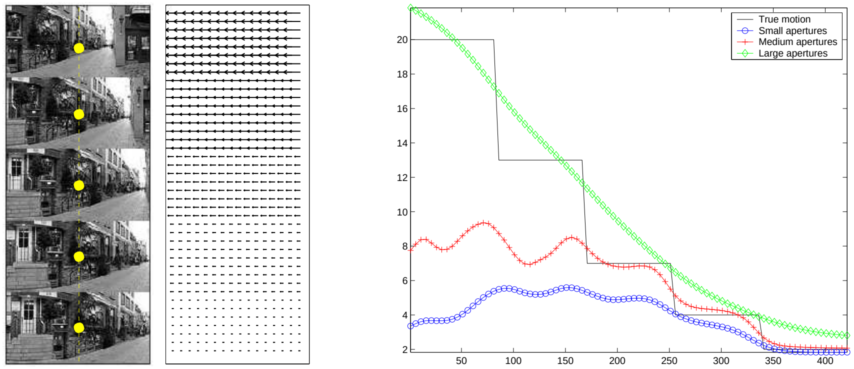
(c) Dealing with diverging motions (Figure 4). This sequence is a zoom in, with respect to the image center. Figure 4-(b) represents the true flow and figure 4-(c) represents velocity norms cuts for the red diagonal of figure 4-(b). Figure 4-(d),(e),(f) compare the true flow to the ones computed using respectively BBWP, ADK and our method. We capture reasonably the motion range and runs much faster than BBPW and ADK.

(d) Dealing with rotational motions (Figure 6). This sequence is a 4° counterclockwise rotation, with respect to the image center. Figure 6-(b) represents the true flow. Figure 6-(c),(d),(e) compare the true flow to the ones computed using respectively BBWP, ADK and our method. BBPW returns the best results, ADK gives also very good results, and our algorithm is a bit less accurate but much faster.

(e) Recovering motion with occlusions (Figure 5). This real sequence shows a boat moving slighly to the right while the background is moving leftward. Although the two variational methods BBWP and ADK give slightly better edges, we are able capture correctly the two motions.

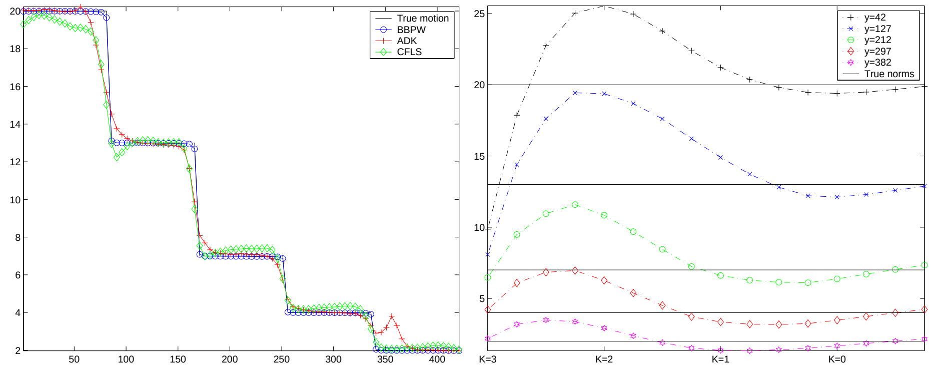
5 Conclusion

This paper demonstrates that a linear tensor-based approach can be successfully extended in a multiscale framework and estimate high range velocities robustly. The computational



(a) One image (b) $\mathbf{w}_{\text{True}}$ (c) Lucas and Kanade with different σ

Figure 1: Multispeed sequence (21 frames of size 170×425)

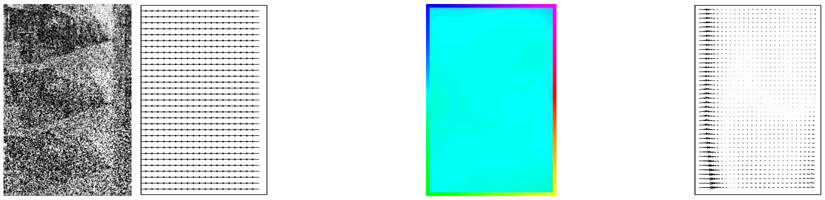


(a) (b)

(c)

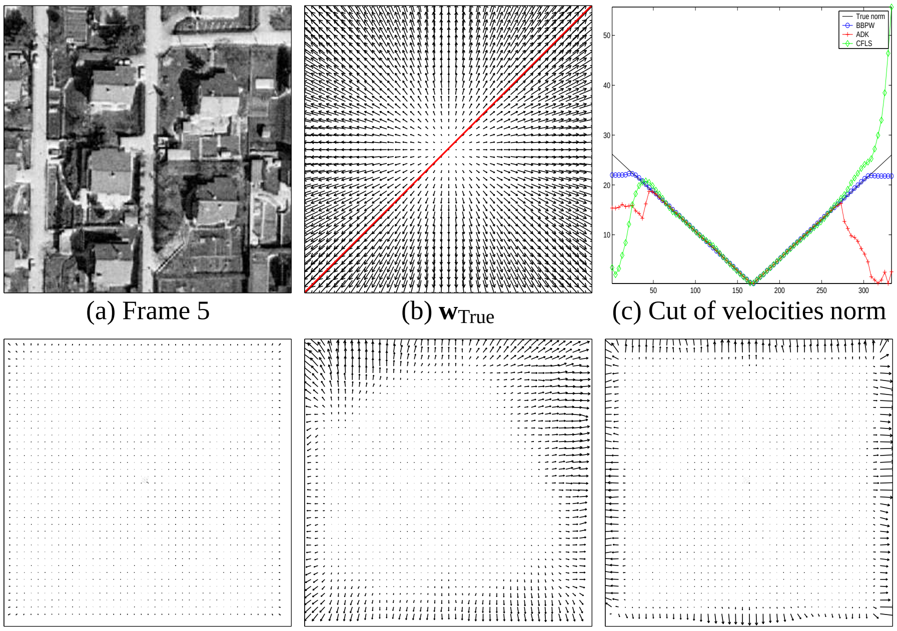
Algorithm	AAE	AESTD	ANE	NESTD	RT
BBPW [5]	0.67°	0.81°	0.30	0.14	17mn
ADK [2]	1.37°	1.02°	0.51	0.15	11mn40s
CFLS	1.43°	0.62°	1.60	2.90	1mn35s

Figure 2: Multispeed sequence: (a) Vertical cuts of the velocitiy norms, (b) evolution of velocities across scales, at position $x = 85$, $t = 4$ and y as shown in figure 1, (c) error measurements.



(a) frame 4 (b) $\mathbf{w}_{\text{True}}$ (c) Orientations computed (d) $\mathbf{w}_{\text{True}} - \mathbf{w}_{\text{CFLS}}$

Figure 3: Robustness to noise (21 frames of size 170×255). Error measurements are respectively AAE = 1.65°, AESTD = 0.62°, ANE = 2.20, NESTD = 0.62.



(a) Frame 5 (b) $\mathbf{w}_{\text{True}}$ (c) Cut of velocities norm

(d) $\mathbf{w}_{\text{True}} - \mathbf{w}_{\text{BBPW}}$ (e) $\mathbf{w}_{\text{True}} - \mathbf{w}_{\text{ADK}}$ (f) $\mathbf{w}_{\text{True}} - \mathbf{w}_{\text{CFLS}}$

Algorithm	AAE	AESTD	ANE	NESTD	RT
BBPW [5]	0.69°	0.23°	0.46	0.09	13mn
ADK [2]	6.74°	3.47°	3.65	0.81	$>$ 40mn
CFLS	3.01°	1.58°	3.94	1.59	2mn30s

Figure 4: Zooming sequence (21 frames of size 334×334). Error measurements are reported above.

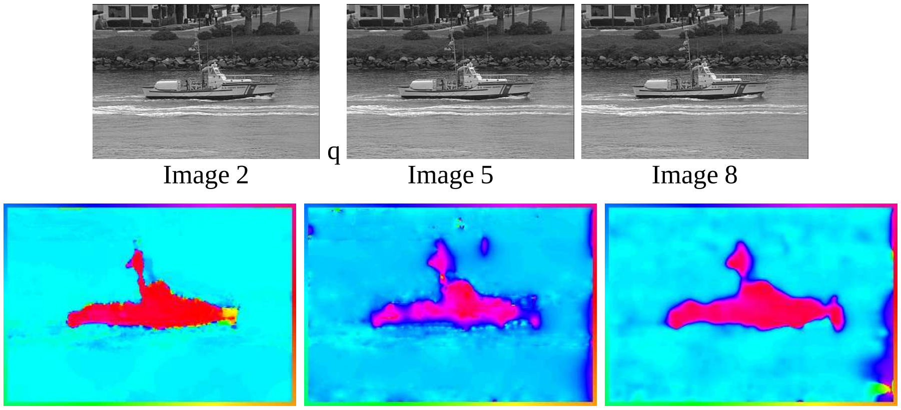
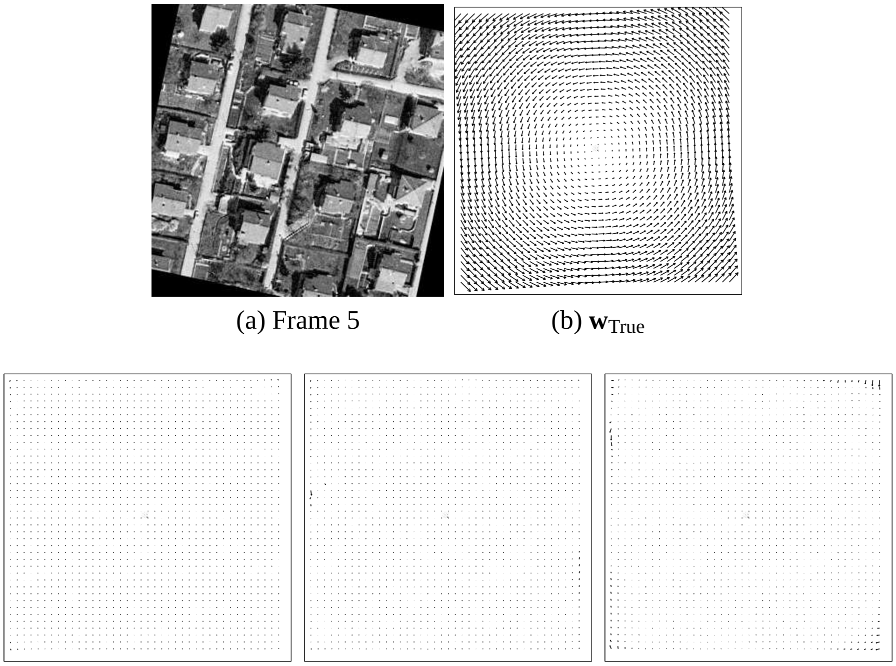


Image 2 Image 5 Image 8

$\mathbf{w}_{\text{BBPW}}$(9mn) $\mathbf{w}_{\text{ADK}}$(13mn) $\mathbf{w}_{\text{CFLS}}$(1mn30)

Figure 5: Real sequence with occlusions (10 frames of size 352×240). Flow orientation is shown below for image 5 which gives a basic object segmentation. Computaional time is indicated between parentheses.



(a) Frame 5

(b) $\mathbf{w}_{\text{True}}$

(d) $\mathbf{w}_{\text{True}} - \mathbf{w}_{\text{BBPW}}$

(e) $\mathbf{w}_{\text{True}} - \mathbf{w}_{\text{ADK}}$

(f) $\mathbf{w}_{\text{True}} - \mathbf{w}_{\text{CFLS}}$

Algorithm	AAE	AESTD	ANE	NESTD	RT
BBPW [5]	0.22°	0.21°	0.08	0.12	9mn
ADK [2]	0.56°	0.94°	0.18	0.27	10mn45s
CFLS	1.38°	1.40°	0.44	0.59	1mn45s

Figure 6: Rotation sequence (15 frames of size 334×334). Error measurements are reported above.

coast is also quite low when compared to classical variational approaches since there is no need to ”iterate until convergence”.

One main drawback for this kind of approaches remains the recovery of accurate occluding boundaries. More accurate techniques could surely be proposed incorporating some kind of motion discontinuity detection at coarsest scales. Such process could be implicitly included considering a robust cost function instead of the quadratic norm. Otherwise, one advantage that hasn’t been taken fully into acount yet is that these approaches introduce naturally some error measurements (for exampe the system condition number) that could be used as a confidence measure for the velocities. These issues will be considered next.

6 Aknowlegdments

We wish to thank T. Brox, N. Papenberg and J. Weickert for providing us mathematical as well as implementation details about their optical flow algorithm, and especially T. Brox who kindly reran all the experiments involving his algorithm.

References

[1] L. Alvarez, J. Weickert, and J. Sànchez. Reliable estimation of dense optical flow fields with large displacements. *The International Journal of Computer Vision*, 39(1):41–56, August 2000.

[2] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal of Applied Mathematics*, 60(1):156–182, 1999.

[3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *The International Journal of Computer Vision*, 12(1):43–77, 1994.

[4] J. Bigun, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, August 1991. Report LiTH-ISY-I-0828 1986 and Report LiTH-ISY-I-1148 1990, both at Computer Vision Laboratory, Linköping University, Sweden.

[5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision*, Prague, Czech Republic, 2004. Springer–Verlag.

[6] T. Brox and J. Weickert. Nonlinear matrix diffusion for optic flow estimation. In *DAGM-Symposium*, pages 446–453, 2002.

[7] R. Deriche. Fast algorithms for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(12):78–88, January 1990.

[8] F. Lauze, P. Kornprobst, C. Lenglet, R. Deriche, and M. Nielsen. Sur quelques méthodes de calcul de flot optique à partir du tenseur de structure : Synthèse et contribution. In *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, 2004.

[9] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[10] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *The International Journal of Computer Vision*, 46(2):129–155, 2002.

[11] M. Middendorf and H.-H. Nagel. Estimation and interpretation of discontinuities in optical flow fields. In *Proceedings of the 8th International Conference on Computer Vision*, pages 178–183.

[12] M. Middendorf and H.-H. Nagel. Empirically convergent adaptive estimation of grayvalue structure tensors. In *DAGM-Symposium*, pages 66–74, 2002.

[13] H.-H. Nagel and A. Gehrke. Spatiotemporally adaptive estimation and segmentation of OF-fields. In Hans Burkhardt and Bernd Neumann, editors, *Proceedings of the 5th European Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 86–102, Freiburg, Germany, June 1998. Springer–Verlag.

[14] H.H. Nagel and W. Enkelmann. An investigation of smoothness constraint for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[15] H. Spies and H. Scharr. Accurate optical flow in noisy image sequences. In *Proceedings of the 8th International Conference on Computer Vision*, pages 587–592.

[16] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in pde-based computation of image motion. *The International Journal of Computer Vision*, 45(3):245–264, December 2001.