SEMI-LOCAL VARIATIONAL OPTICAL FLOW ESTIMATION

Denis Fortun, Charles Kervrann

Inria Rennes - Bretagne Atlantique
35 042 Rennes Cedex, France

ABSTRACT

Global variational methods for optical flow estimation usually suffer from an over-smoothing effect. We propose a semi-local estimation framework designed to integrate and improve any variational method. The idea is to implicitly segment the minimization domain into coherently moving windows. In a first time, local variational estimations are performed in overlapping candidate square regions. Then, a global discrete optimization, non subject to the over-smoothing introduced by variational approaches, selects the optimal window for each pixel. Experimental results show an increasing of the sharpness of discontinuities and a significant improvement of global registration errors compared to the results of the baseline global variational method.

Index Terms— Optical Flow, variational methods, minimization domain, motion segmentation, discrete optimization

1. INTRODUCTION

The optical flow approximates the projection of the motion of a 3D scene on the image plane. Any optical flow estimation have thus to be based on a conservation assumption of some optical properties of the image able to capture the real motion (intensity, gradient, image descriptor . . .). This data conservation constraint provides in general a single equation and is consequently insufficient to recover the two components of the motion field (*aperture problem*). The typical way to overcome this under-determination is to add to the data conservation constraint a spatial coherency constraint. Existing methods can be classified regarding their *local* or *global* strategy to impose such a constraint.

The spatial coherency of *local* approaches is ensured at a pixel $x \in \Omega \subset \mathbb{R}^2$ by the assumption of common parametric motion (translational in [1]) in a neighborhood $V(x) \subset \Omega$, where Ω is the image domain. The *global* approach allows to compute a dense motion field and explicitly adds to a data potential $\rho_{data}(\cdot)$ which penalizes deviations from the data conservation constraint, a regularization potential $\rho_{reg}(\cdot)$ which penalizes high values of the norm of the gradient ∇w

of the velocity field $w : \Omega \to \mathbb{R}^2$. A global energy combining these two potentials is minimized [2]:

$$E_{global}(w) = \int_{\Omega} \rho_{data}(x, I, w) + \lambda \rho_{reg}(x, \nabla w)dx \quad (1)$$

where $I : \Omega \times [0, T] \to \mathbb{R}$ is an image sequence and λ is a balance parameter between data fitting and regularization.

The best results of the state-of-the-art are achieved by global variational motion estimation methods. However, this kind of methods still suffer from an over-smoothing effect. This phenomenon is particularly visible in the seminal work of [2] which uses quadratic penalty function for the regularization potential. This shortcoming has been greatly reduced by the introduction of robust penalty functions [3, 4], adaptation of the regularization along image discontinuities [5] or non-local regularization strategies [6], but it still remains. Indeed these methods are limited by the coarse-to-fine scheme [4, 7], necessary to cope with large displacements. This approach avoids most local minima due to the non-convexity induced by the non-linearized data potential, at the price of an over-smoothing of the discontinuities.

We mention two non-variational approaches related to our method that have been investigated to reduce the over-smoothing effect of global variational methods: (i) parametric motion estimation based on motion field segmentation; (ii) discrete optimization of the energy (1). In the first case, a parametric model of the flow field is estimated inside coherently moving regions. The estimation of the discontinuities is thus transfered to the segmentation step [8]. In the second case, discrete optimization of the energy (1) is able to find strong minima for non-convex functionals without coarse-to-fine schemes, but is limited by the quantization of the flow field range [9].

In this paper, we present a method combining local estimations and discrete optimization to sharpen the discontinuities of a global variational method by implicitly segmenting the flow field. It is composed of two stages: first, local variational estimations are performed on a regular grid of overlapping windows; second, the resulting local motion vectors are used as candidates for a global discrete optimization. In this scheme, the discrete optimization module selects of the optimal spatial minimization domains, yielding an implicit segmentation of the flow field. It is worth noting that our

framework is general and can be used to improve any baseline variational method. The results with the popular and representative method [7] show significant improvements over the global variational approach when applied on several sequences of the Middlebury database [10].

2. VARIATIONAL OPTICAL FLOW ESTIMATION

2.1. Global variational approach

All global variational optical flow estimation methods are based on the minimization of the energy (1). The method described in [7] contains most of the basis concepts still used in the most recent methods. Therefore we used this method to evaluate our semi-local framework, which can integrate any other variational optical flow estimation methods. The main features of the method [7] are described in this section.

Global energy functional The data potential penalizes deviations from intensity and gradient conservation constraints with a L_1 penalty function:

$$\rho_{data}(x, I, w) = \phi\left(|I(x + w(x), t + 1) - I(x, t)|^2\right) \quad (2)$$
$$+ \gamma\phi\left(\|\nabla I(x + w(x), t + 1) - \nabla I(x, t)\|^2\right)$$

where $\gamma > 0$ is a balance parameter and $\phi\left(z^2\right) = \sqrt{z^2 + \epsilon^2}$ is a regularized form of the L_1 norm with ϵ a small constant.

The regularization potential penalizes high gradients with the same convex and discontinuity-preserving L_1 norm:

$$\rho_{reg}(x, \nabla w) = \phi\left(\|\nabla u(x)\|^2 + \|\nabla v(x)\|^2\right) \quad (3)$$

Energy minimization The minimization of (1) is performed by solving the Euler-Lagrange equations. To make these equations tractable, the data potential (2) must be linearized, which limits the estimation to small displacements. Therefore all recent variational methods adopted a coarse-to-fine scheme to handle large displacements [4, 7]. The coarse-to-fine levels are interpreted in [7] as fixed point iterations enabling the minimization of the initial non-linear energy. At each level, a second fixed point allows to cope with the remaining non-linearity of the Euler-Lagrange equations due to the L_1 norm. The resulting linear system is then solved with Successive Over Relaxation (SOR).

2.2. Restriction to local domains

The minimum reached by global variational methods is usually suboptimal. Indeed, variational optimization is proved to find the global minimum only for convex energies, which is not the case of (1) with the non-linearized data potential (2). Actually, the coarse-to-fine scheme transforms the problem into successive minimizations of convex approximations of (1) which tend to smooth the discontinuities of the flow field.

To reduce this over-smoothing effect we propose to minimize several energies of the form (1) over local regions,
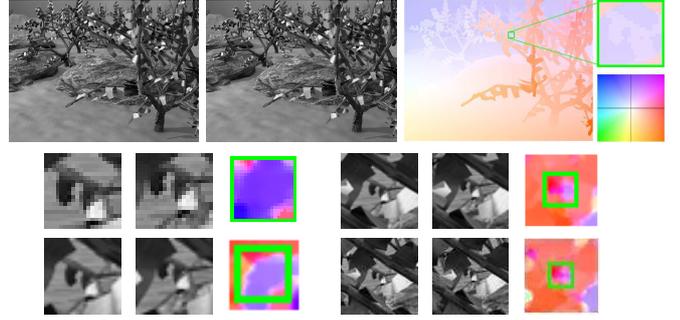
Fig. 1. Influence of the spatial minimization domain. 1st row: two frames and ground truth of the motion field with a zoom on a region of interest (green square). 2nd and 3^{d} rows: variational estimations over windows of different sizes centered on the same green square.

inspired by the localization of Total Variation for denoising [11]. The local motion field $w_V : V \subset \Omega \to \mathbb{R}^2$ is the minimizer of

$$E(w_V) = \int_V \rho_{data}(x, I, w_V) + \lambda\,\rho_{reg}(x, \nabla w_V)dx \quad (4)$$

obtained with the method described in the previous section. If V is a coherently moving region without strong discontinuities, the over-smoothing does not occur. We emphasize that contrary to other approaches computing optical flow in local regions [8, 12], where the motion is restricted to a parametric model, we prefer to compute a regularized flow field. Thus we overcome the difficulty of local parametric approaches to handle complex motion fields, where the region V must be small enough to ensure the validity of the parametric assumption, and large enough to avoid the *aperture problem*. Figure 1 shows how the localization of variational methods can improve the accuracy of the global approach when the region is suitably chosen: as the window size increases, the details of the flow tends to disappear. Our strategy to select the optimal region of each pixel is presented in the next section.

3. SEMI-LOCAL FRAMEWORK

Our aim is to estimate the motion at each pixel with a local variational model (4) in coherently moving regions. Existing similar approaches are restricted to parametric motion and are based either on an image gradient segmentation [12], subject to over-segmentation of the flow field, or on global variational minimization of non-convex functionals [8] for joint estimation of regions and motion, suffering from the drawbacks described in the previous section. Our semi-local framework performs an *implicit* segmentation of the flow field, without global variational estimation or image segmentation.

Local estimations Local variational estimations are performed on overlapping square windows of different sizes.

For a fixed size s, let $\mathcal{V}_{s,\alpha}$ be a set of regularly spaced windows covering the whole image, with an amount of overlap α between neighbors (see Fig. 2). For a set of varying sizes $\mathcal{S} = \{s_0, \ldots, s_n\}$, we define $\mathcal{V}_{\mathcal{S},\alpha} = \bigcup_{s \in \mathcal{S}} \mathcal{V}_{s,\alpha}$. For each window $V \in \mathcal{V}_{\mathcal{S},\alpha}$, a local motion field w_V is estimated by minimization of (4).

One pixel is contained in several overlapping windows with different locations and sizes. We denote $\mathcal{N}_V(x)$ the set of windows containing the pixel x (see Fig. 2, $\mathcal{N}_V(x) = V_1, .., V_4\}$). The computed flow fields over these windows provide a set $\{w_V(x)\}_{V \in \mathcal{N}_V(x)}$ of candidate motion vectors at each pixel x.

Global aggregation The aggregation step aims at combining the locally estimated flow fields to compute an optimal global flow field. The goal is to select at each pixel the most appropriate window. To this end, we consider the aggregation as a multi-label assignment problem, where local candidates $\{w_V(x)\}_{V \in \mathcal{N}_V(x)}$ constitute the discrete label space at pixel x. The global flow field w_Ω resulting from the aggregation is then the minimizer of a global objective energy:

$$w_\Omega = \arg\min_w E_\Omega(w) \text{ s.t. } w(x) \in \{w_V(x)\}_{V \in \mathcal{N}_V(x)}. \quad (5)$$

Thus, the solution is found by selecting the best motion vector among the small set of candidates $\{w(x)\}_{V \in \mathcal{N}_V(x)}$. We define E_Ω as:

$$E_\Omega(w) = \sum_{x \in \Omega} \rho_{data}(x, I, w) + \lambda \psi_{reg}(x, w) \quad (6)$$

where $\psi_{reg}(\cdot)$ is a Markov Random Field prior:

$$\psi_{reg}(x, w) = \sum_{y \in \Delta(x)} \phi\left(|u(x) - u(y)|^2 + |v(x) - v(y)|^2\right)$$

where $\Delta(x)$ is a 4 or 8 pixel neighborhood. The energy E_Ω is actually a discrete version of the energy (1). The difference with global variational minmization of (1) is that we impose the solution to belong to the set of local variational estimations $\{w_V(x)\}_{V \in \mathcal{N}_V(x)}$. Our discrete optimization scheme does not suffer from the over-smoothing effect of the global variational approach. Consequently, it selects the candidates coming from coherently moving regions, which are not affected by this over-smoothing (see the smallest window in Fig. 1). This selection of the optimal window at each pixel can be seen as an implicit segmentation of the spatial minimization domain used for the baseline variational method.

Discrete optimization In this subsection we detail our approach for the discrete optimization problem (5). The formulation of (5) differs from the classical multi-label assignment scheme because the regularization is not applied to the discrete labels but to the continuous-valued motion vectors assigned to the labels. This problem has been addressed in
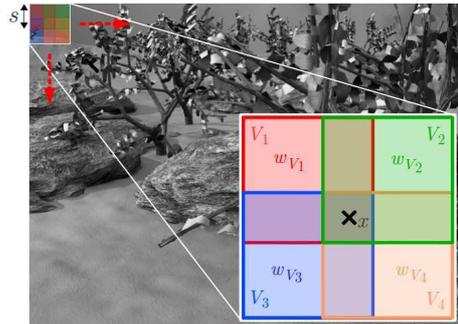
Fig. 2. Illustration of the patches distribution for a given size s and overlapping ratio $\alpha = 0.3$. The pixel x is contained in four patches $V_1, \ldots, V_4$ providing four candidates for x: $w_{V_1}(x), \ldots, w_{V_4}(x)$.

the context of optical flow estimation in [13] with candidates estimated by several global methods. The authors achieved the multi-label optimization with the *fusion-move* algorithm, which operates by successive fusions of proposal labellings (see [13] for more details about *fusion-move* and its applications). We apply this technique for solving (5).

The *fusion-move* algorithm requires a set of global flow fields $\{w_{\Omega_1}, \ldots, w_{\Omega_N}\}$ to be fused. Let us consider that the set of candidates at each pixel $\{w_V(x)\}_{V \in \mathcal{N}_V(x)}$ is composed by a maximum number of N candidates. Then for $i \in \{1 \ldots N\}$, we assign to $w_{\Omega_i}(x)$ one arbitrarily chosen candidate in $\{w_V(x)\}_{V \in \mathcal{N}_V(x)}$, so that each candidate is assigned to at least one global flow field $w_{\Omega_i}(x)$. Our experiments showed that the assignment strategy of the local candidates in the global flow fields has negligible influence on the final result.

4. RESULTS AND DISCUSSION

We use the Average Angular Error (AAE) to evaluate the performance of our method on sequences of the Middlebury benchmark [10]. Two aggregation procedures are considered: *SL-fusion* performs the discrete optimization described in Section 3.2 and 3.3; *SL-Mean* performs a weighted mean of the local candidates, favoring central pixels with a gaussian filter centered on each window. In all our experiments we fix the parameters of [7] $\lambda = 40$ and $\gamma = 5$.

Table 1 compares the AAE of the baseline variational method [7] with *SL-fusion* and *SL-Mean*, for several sets of window sizes $\mathcal{S}$. The superior performance of *SL-fusion* over *SL-Mean* highlights that the selection of the best window is crucial to prevent the global flow field from being influenced by outliers coming from inappropriate regions. For *SL-fusion*, the errors obtained with single sizes are always higher than those obtained with their combination. This result shows that a single window size is not able to capture all types of coherently moving regions and that the aggregation procedure successfully combines the advantages of each size

	Grove2	Grove3	Rubber-whale	Dime-trodon
Variational [7]	**2.38**	**5.97**	**3.92**	**1.83**
SL-Mean				
$\mathcal{S} = \{15\}$	4.91	17.7	5.42	4.40
$\mathcal{S} = \{49\}$	2.43	6.11	4.04	1.91
$\mathcal{S} = \{129\}$	2.38	6.01	3.98	1.83
$\mathcal{S} = \{15, 49, 129\}$	4.21	16.2	5.05	3.54
SL-fusion				
$\mathcal{S} = \{15\}$	2.95	12.8	4.47	3.30
$\mathcal{S} = \{49\}$	2.27	5.85	3.69	1.87
$\mathcal{S} = \{129\}$	2.30	5.83	3.71	1.81
$\mathcal{S} = \{15, 49, 129\}$	**2.10**	**5.60**	**3.34**	**1.79**

Table 1. Comparison of the results (AAE) obtained with our implementation of [7], *SL-Mean* and *SL-fusion* for $\alpha = 0.75$.

by selecting the best region. The results of *SL-fusion* with $\mathcal{S} = \{15, 49, 129\}$ are significantly better than those of the baseline variational method for the sequences *Grove3*, *Rubberwhale* and *Grove2*. This is due to the better preservation of the discontinuities illustrated in Fig. 3. We mention that even for very smooth sequences like *Dimetrodon*, large window sizes (here 129) ensure that the result cannot be worse than the global variational method, and is even slightly improved.

The influence of the amount of overlap α is shown in Fig. 4. As it can be intuitively expected, the error decreases when the overlap increase. Indeed, the overlap determines the number of candidate regions, and thus the probability that the windows fall in appropriate regions.

5. CONCLUSION AND FUTURE WORK

We proposed a new approach for optical flow estimation, combining global methods with local minimization regions. Our experiments showed that our semi-local framework improves the estimation accuracy of a baseline variational method along discontinuities of the motion field, by performing an implicit segmentation of the spatial minimization domain. In the future we plan to extend this general framework to non-variational baseline methods and adaptive region shapes.
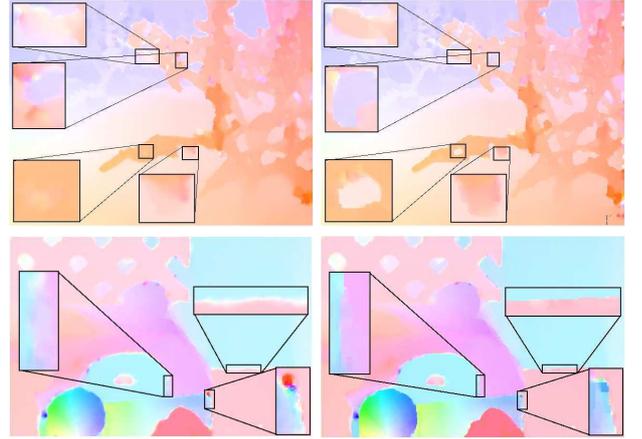
6. REFERENCES

[1] B.D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. Seventh Int. Joint Conf. on Artificial Intelligence*, pp. 674–679, 1981.

[2] B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[3] M.J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Comp. Vis. and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[4] E. Mémin and P. Pérez, "Dense estimation and object-based segmentation of the optical flow with robust techniques," *IEEE Trans. Image Process.*, vol. 7, no. 5, pp. 703–719, 1998.

[5] A. Wedel, D. Cremers, T. Pock, and H. Bischof, "Structure-and motion-adaptive regularization for high accuracy optic flow," in *Proc. Int. Conf. Comp. Vis.*, Kyoto,Japan, October 2009, IEEE, pp. 1663–1668.

[6] Werlberger M., Pock T., and Bischof H., "Motion estimation with non-local total variation regularization," in *IEEE Proc. of Comp. Vis. Pat. Rec.*, San Fransisco, June 2010, pp. 2464–2471.

[7] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Eur. Conf. Comp. Vis.*, Prague, Czech Republic, May 2004, pp. 25–36, Springer.

[8] D. Sun, E.B. Sudderth, and M.J. Black, "Layered image motion with explicit occlusions, temporal consistency, and depth ordering," in *Adv. Neur. Inf. Proc. Sys.*, Canada, December 2010, pp. 2226–2234.

[9] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Patt Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[10] S. Baker, D. Scharstein, JP Lewis, S. Roth, M.J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comp. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.

[11] Louchet C. and Moisan L., "Total variation as a local filter," *SIAM J. Imaging Sci.*, pp. 651–694, 2011.

[12] CW Zitnick, N. Jojic, and S.B. Kang, "Consistent segmentation for optical flow estimation," in *Int. Conf. Comp. Vis.*, Beijing, China, October 2005, IEEE, pp. 1308–1315.

[13] V. Lempitsky, C. Rother, S. Roth, and A. Blake, "Fusion moves for markov random field optimization," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1392–1405, 2010.

Fig. 3. Visual comparison between global variational estimation (1st column) and its integration in *SL-fusion* with $\mathcal{S} = \{15, 49, 129\}$ and $\alpha = 0.75$ (2nd column) on *Grove3* and *Rubberwhale*.
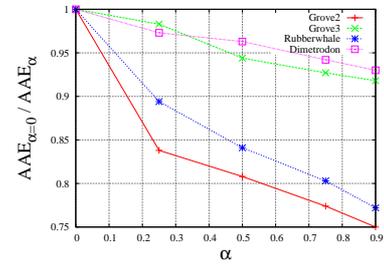
Fig. 4. Influence of α on the AAE. We plot $\mathrm{AAE}_{\alpha=0}/\mathrm{AAE}_{\alpha}$.