



# Sparse Aggregation Framework for Optical Flow Estimation

Denis Fortun, Patrick Bouthemy, Charles Kervrann

## ► To cite this version:

Denis Fortun, Patrick Bouthemy, Charles Kervrann. Sparse Aggregation Framework for Optical Flow Estimation. Scale Space and Variational Methods in Computer Vision, May 2015, Lège Cap Ferret, France. hal-01138012

**HAL Id: hal-01138012**

**<https://inria.hal.science/hal-01138012>**

Submitted on 31 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sparse Aggregation Framework for Optical Flow Estimation

Denis Fortun, Patrick Bouthemy, and Charles Kervrann

Inria, Centre de Rennes - Bretagne Atlantique, Rennes, France

`denis.fortun@epfl.ch`

`{patrick.bouthemy,charles.kervrann}@inria.fr`

**Abstract.** We propose a sparse aggregation framework for optical flow estimation to overcome the limitations of variational methods introduced by coarse-to-fine strategies. The idea is to compute parametric motion candidates estimated in overlapping square windows of variable size taken in the semi-local neighborhood of a given point. In the second step, a sparse representation and an optimization procedure in the continuous setting are proposed to compute a motion vector close to motion candidates for each pixel. We demonstrate the feasibility and performance of our two-step approach on image pairs and compare its performances with competitive methods on the Middlebury benchmark.

**Keywords:** motion estimation, optical flow, sparse representation, optimization

## 1 Introduction

Optical flow estimation is based on a conservation assumption of image features able to capture the real motion (intensity image feature, gradient, feature descriptor ...). The so-called brightness constancy assumption is the mostly used one. It provides a single equation and is consequently insufficient to recover the two components of the motion vector. A usual way to overcome this under-determination is to impose a spatial coherency constraint for the flow field. Existing methods can then be classified into two main approaches: *i*/ the spatial coherency is ensured at pixel  $x \in \Omega \subset \mathbb{R}^2$  ( $\Omega$  is the image domain) by introducing parametric motion models in a neighborhood  $V(x) \subset \Omega$  [17] ; *ii*/ the flow field is assumed to be piecewise smooth and the strategy is to minimize a global energy that explicitly combines a potential  $\rho_{data}(\cdot)$  which penalizes deviation from the brightness constancy equation with a regularization potential  $\phi(\cdot)$  which penalizes high values of the norm of the gradient  $\nabla \mathbf{w}$  of the velocity field  $\mathbf{w} : \Omega \rightarrow \mathbb{R}^2$  [13]:

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

where  $I : \Omega \times [0, T] \rightarrow \mathbb{R}$  is an image sequence and  $\lambda$  is a balance parameter.

The best state-of-the-art results are achieved by minimizing an energy of the form (1). An over-smoothing phenomenon was particularly visible in the seminal work of [13] which uses quadratic penalty function for the regularization potential. This shortcoming has been greatly reduced by the introduction of robust penalty functions [3, 18], the adaptation of the regularization along image discontinuities [24] or non-local regularization strategies [26]. However, this family of methods is still limited by undesirable effects coming from the necessity to resort to coarse-to-fine schemes to handle large displacements. The motion of small objects is discarded at coarse scales, and the error is often propagated in the incremental updates at finer scales. As a result, motion details are often smoothed in the final estimated flow field.

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 transferred to the segmentation step [23]. 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 [4]. For more details about optical flow literature, see the recent survey [12].

In this paper, we present a method combining local parametric estimations and continuous optimization to preserve motion discontinuities and details while capturing large displacements, in a variational framework without coarse-to-fine scheme. It is composed of two stages: first, local parametric estimations are performed on overlapping square windows (Section 2); second, the resulting local motion vectors are used as candidates for a global continuous optimization (Section 3). It is worth noting that this two-step approach has already been successfully investigated in the discrete setting in [11]. We propose hereunder a sparse representation approach in the continuous setting which is faster and is able to produce competitive results (Section 4) on several sequences of the Middlebury database [1] when compared to [5, 6, 9, 11].

## 2 Computation of local motion candidates

To compute motion candidates, we follow the idea of [11] and perform local estimations in overlapping square patches of different sizes, so that each pixel is contained in several patches. This approach can be viewed as an alternative way to address the problem of the choice of the local neighborhood for parametric estimation. Rather than adapting the regions *a priori* or jointly with the motion field, we operate in two steps: 1) estimation of motion candidates on several supports at every pixel, 2) selection of the optimal candidate at each pixel within the aggregation step. In this section, we describe a combination of parametric estimation and patch correspondences for computing motion candidates. Nevertheless, the proposed framework allows for any types of local estimation.

### 2.1 Set of overlapping patches

The local supports for computing motion candidates are overlapping square patches of different sizes. Let us denote  $\mathcal{P}(x) = \{P_{\nu,\tau} : x \in P_{\nu,\tau}\}$  the set of patches with patch size  $\nu \in \{\nu_1, \dots, \nu_{max}\}$ , an overlapping ratio  $\tau \in [0, 1]$  indicating the proportion of surface shared by neighboring patches, and the set  $P_{\nu,\tau}$  of patches covering  $\Omega$  with sizes in  $\nu$  and overlapping ratio  $\tau$ . To capture different motion scales, the patch sizes must cover a large range of values. In all our experiments, we set  $\nu \in \{15, 45, 115\}$ . See [11] for a more detailed description of the set of patches. The motion vectors are estimated independently in each patch in two sub-steps including patch correspondences and affine motion estimations, as described in the following.

### 2.2 Patch correspondences

Let us consider two successive images  $I_1$  and  $I_2$ . For each patch  $P_1 \in \mathcal{P}(x)$  in image  $I_1$ , we first determine the set  $\mathcal{X}_N(P_1)$  of the  $N$  most similar patches to  $P_1$  in  $I_2$ . In our experiments, we use a combination of the saturation and value channels of the HSV color space to cope with illumination changes [27]. We consider the Sum of Absolute Distances to compare patches and we impose a minimal distance between the correspondences of a given pixel to ensure the diversity of motion candidates. Multiple distances to compare patches could be jointly considered to enlarge the collection of candidates at each location. The final set of correspondences is denoted  $\mathcal{X}_N(P_1)$ . The correspondences are found with the PatchMatch algorithm [2]. Finally, for each established pair of corresponding patches  $(P_1, P_2)$  with  $P_2 \in \mathcal{X}_N(P_1)$ , we denote  $\mathbf{d}_{1,2} \in \mathbb{Z}^2$  the translation vector shifting  $P_1$  onto  $P_2$ .

### 2.3 Affine motion refinement

The displacements  $\mathbf{d}_{1,2}$  estimated by patch correspondences are integer-pixel translational approximations. We refine the coarse translation computation with the estimation of a local affine motion model denoted  $\delta\mathbf{w}_{1,2} : \Omega_{P_1} \rightarrow \mathbb{R}^2$  in every pair  $(P_1, P_2)$  where  $\Omega_{P_1} \subset \Omega$  is the pixel domain of  $P_1$ . At pixel  $x = (x_1, x_2)^\top$ , we define:

$$\delta\mathbf{w}_{1,2}(x) = (a_1 + a_2x_1 + a_3x_2, a_4 + a_5x_1 + a_6x_2)^\top \quad (2)$$

and the parameter vector  $\boldsymbol{\theta}_{1,2} = (a_1, a_2, a_3, a_4, a_5, a_6)^\top$  is estimated using the brightness constancy constraint:

$$\hat{\boldsymbol{\theta}}_{1,2} = \arg \min_{\boldsymbol{\theta}_{1,2}} \int_{\Omega_{P_1}} \psi(P_2(x + \delta\mathbf{w}_{1,2}(x)) - P_1(x)) dx. \quad (3)$$

The penalty function  $\psi(\cdot)$  is chosen as the robust Tukey's function. The problem (3) is solved with the publicly available Motion2D software<sup>1</sup> [21], which implements a multi-resolution incremental minimization scheme involving an IRLS

<sup>1</sup> <http://www.irisa.fr/vista/Motion2D/>

(Iteratively Reweighted Least Squares) technique for solving the successive linearizations of the penalty function in (3).

## 2.4 Definition of motion candidates

The above described two-step estimation is repeated for every patch of  $\mathcal{P}(x)$  and generates a set of candidate motion vectors  $\mathcal{C}(x)$  at each pixel  $x \in \Omega$  defined as follows:

$$\mathcal{C}(x) = \{\mathbf{d}_{1,2} + \delta \mathbf{w}_{1,2}(x) : P_1 \in \mathcal{P}(x), P_2 \in \mathcal{X}_N(P_1)\}. \quad (4)$$

In what follows, we will denote  $\mathbf{w}_{\mathcal{C}}(x) = (\mathbf{w}_1(x), \dots, \mathbf{w}_M(x))^\top$  the vector of  $M$  motion candidates computed from the  $M$  overlapping patches in  $\mathcal{P}(x)$  and  $\mathcal{C}(x) = \{\mathbf{w}_1(x), \dots, \mathbf{w}_M(x)\}$  the corresponding set of candidate motion vectors.

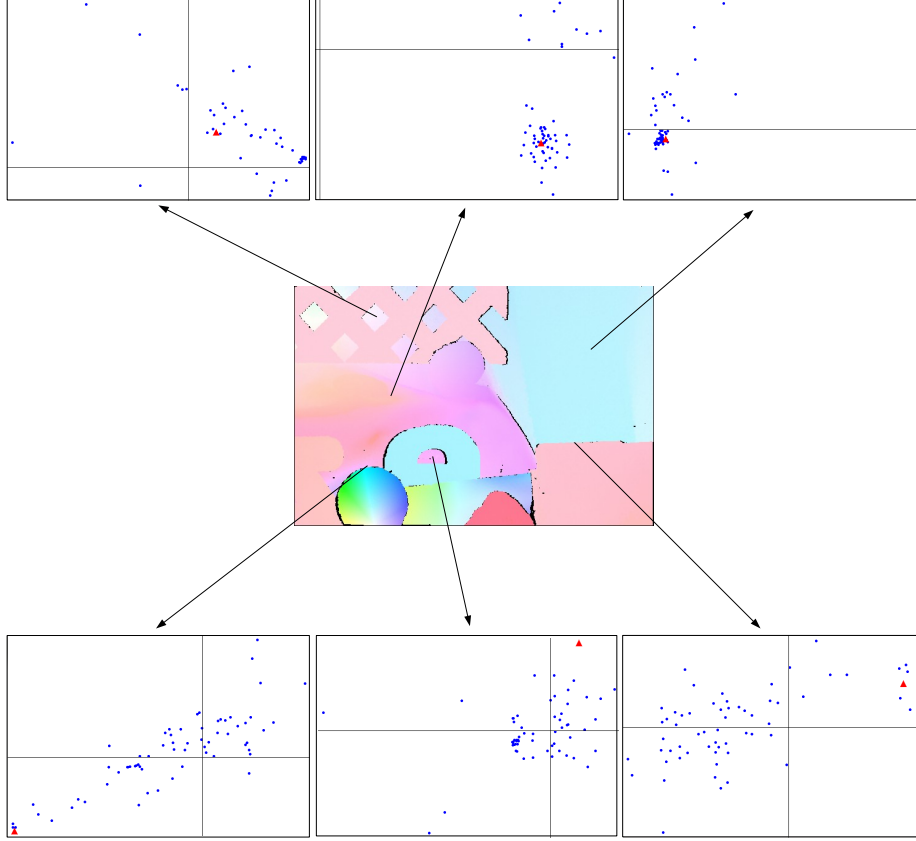
Combining a coarse motion estimation with a refinement step has already been investigated in [10, 16, 20] but essentially for global dense motion field estimation. This strategy is applied to semi-local neighborhoods and patches in our approach. Unlike [17], each pixel gets several motion vectors computed from overlapping patches. Finally, in contrast to several other methods based on feature matching [6, 10, 25], we select the  $N > 1$  best correspondences. Note that the correspondence sub-step enables to capture large displacements even for small patch sizes and allows us to deal with small structures undergoing large displacements unlike coarse-to-fine schemes. Generally, local parametric motion estimation needs an appropriate selection of spatial neighborhoods [17]. In our framework, we consider square patches with several sizes to tackle adaptively motion of different amplitudes. The selection of patches/neighborhoods is postponed in the aggregation step.

## 3 Sparse aggregation and continuous optimization

In this section, we present an aggregation strategy in a continuous setting which can be considered as an alternative to the discrete aggregation method described in [11]. The sparse aggregation we propose enables significantly lower computational time than [11], it is less dependent on the quality of candidates estimation, and it recovers more accurately smooth regions of the motion field. In a continuous setting, we minimize an energy of the form

$$E(\mathbf{w}) = \int_{\Omega} \rho(\mathbf{w}(x), \mathbf{w}_{\mathcal{C}}(x)) + \lambda_1 \phi(\nabla \mathbf{w}(x)) dx, \quad (5)$$

where  $\rho(\mathbf{w}(x), \mathbf{w}_{\mathcal{C}}(x))$  is a fidelity term and the second term imposes smoothness of  $\mathbf{w}$  while preserving motion discontinuities, balanced by the parameter  $\lambda_1$ . In the following, we consider a Total Variation regularization:  $\phi(\nabla \mathbf{w}(x)) = \|\nabla \mathbf{w}(x)\|_1$ . Unlike usual approaches for optical flow, the image intensities are not used as input of the data potential  $\rho(\mathbf{w}(x), \mathbf{w}_{\mathcal{C}}(x))$ , but are replaced by the motion candidate vector  $\mathbf{w}_{\mathcal{C}}(x)$ .



**Fig. 1.** Visualization of the distribution of the motion candidates at several locations in the image. The central image is the ground-truth motion field of the *RubberWhale* sequence of the Middlebury benchmark. The six plots represent the motion vector candidates and the motion vector ground-truth at each corresponding pixel. The horizontal and vertical axes are respectively the horizontal and vertical components of the motion vectors. Blue points are motion candidates and red triangles are ground-truth motion vectors.

Minimizing in the continuous domain w.r.t.  $\mathbf{w}$  implies that the estimated motion field is allowed to deviate from the motion candidate vectors whereas in [11], the fixed motion vectors are necessarily vectors selected among the motion candidates by the discrete optimization scheme. From a practical point of view, this deviation allows us to achieve good results even when the sets of candidates are less accurate. Thus, critical parameters for the computational cost of the method, such as the overlapping ratio  $\alpha$ , could be adapted to speed up candidates computation.

### 3.1 Candidate distribution

If the motion candidates set is considered as the input data for the aggregation stage, we have to study the distribution of the candidates. Figure 1 illustrates the 2D distribution of candidates at several locations in an image (blue points), while also plotting the ground truth motion vector among them (red triangle). We can first observe that it is not always possible to identify modes of the distribution of motion candidates. While in regions of constant or smoothly varying motion, most motion vectors are clustered around the same mode, the distribution at motion discontinuities is unpredictable. Secondly, when a mode exists, the ground truth motion vector does not always correspond to this mode. The best motion candidate is sometimes isolated from the rest of the candidates. As a matter of fact, the two cases (absence of modes and isolated best candidate) frequently occur in all types of sequences. We can conclude that the candidates distribution is not a relevant information for modeling the data term of the energy function in the continuous aggregation. Options like dense linear combination of candidates, fitting of a statistical distribution or clustering are then excluded.

### 3.2 Continuous aggregation

We propose two versions for the data potential of (5). We define  $\rho(\mathbf{w}(x), \mathbf{w}_{\mathcal{C}}(x))$  as a measure of proximity of  $\mathbf{w}(x)$  to components of the motion candidate vector  $\mathbf{w}_{\mathcal{C}}(x) = (\mathbf{w}_1(x), \dots, \mathbf{w}_M(x))^{\top}$ . However, the potential must not be a distance measure to a mode of  $\mathcal{C}(x)$  or a weighted average of candidates, as pointed out in the previous section. We rather define it as the distance to a single appropriately selected candidate from  $\mathcal{C}$ . Therefore, differently from [11], we exploit the selected candidate as a constraint in the data potential.

**Minimum distance** A first natural idea is to define  $\rho(\mathbf{w}(x), \mathbf{w}_{\mathcal{C}}(x))$  as the distance to the closest component of  $\mathbf{w}_{\mathcal{C}}(x)$ :

$$\rho(\mathbf{w}(x), \mathbf{w}_{\mathcal{C}}(x)) = \min_{m \in \{1, \dots, M\}} \|\mathbf{w}(x) - \mathbf{w}_m(x)\|_1. \quad (6)$$

The min function naturally selects one candidate used for distance measure. The proximal operator of  $\rho(\mathbf{w}(x), \mathbf{w}_{\mathcal{C}}(x))$  can be computed exactly and the resulting energy can then be minimized in a proximal splitting framework [9]. However, the problem of the potential (6) lies in high non convexity, leading inevitably to local minima. In practice, we experimentally observe a convergence of the algorithm, but it stays trapped in a local minimum very dependent on the initialization. In the next subsection, we introduce another model to relax the selection of a unique candidate and achieve more efficient minimization.

**Sparse constraint and dictionaries** In what follows, we propose an alternative potential composed of two terms to select a linear combination of a few

candidates:

$$\rho(\mathbf{w}(x), \mathbf{w}_C(x), \boldsymbol{\alpha}(x)) = \|\mathbf{w}(x) - \boldsymbol{\alpha}(x)^\top \mathbf{w}_C(x)\|_1 + \lambda_2 \|\boldsymbol{\alpha}(x)\|_1 \quad (7)$$

where  $\boldsymbol{\alpha}(x) = (\alpha_1(x), \dots, \alpha_M(x))^\top$  is a sparse coefficient vector associated to the  $M$  candidates at pixel  $x$  and  $\lambda_2$  balances the influence of the two terms. The first term is a reconstruction term which approximates the unknown motion vector  $\mathbf{w}(x)$  by a linear combination of motion candidates  $\mathbf{w}_m(x)$  in the set  $\mathcal{C}(x)$  viewed as a local motion dictionary. The second term imposes sparsity of the coefficients  $\boldsymbol{\alpha}(x)$ . A probabilistic interpretation could be given to  $\boldsymbol{\alpha}(x)$  as in [22] it was not considered in the proposed framework. If the balance coefficient  $\lambda_2$  is high enough, a few components of  $\boldsymbol{\alpha}(x)$  will be non null, which amounts to selecting almost one single candidate in (7). Besides, potential (7) is convex and minimization is more tractable compared to the optimization problem described in the previous subsection.

In an alternate optimization scheme, the tight coupling between  $\mathbf{w}$  and  $\boldsymbol{\alpha}$  could imply that in practice  $\boldsymbol{\alpha}(x)$  stays trapped in the local minimum at the first iteration. We overcome this problem by replacing the pure sparsity constraint of (7) by a weighted sparsity constraint defined by:

$$\|\boldsymbol{\alpha}(x)\|_{1,\beta(x)} = \sum_{m=1}^M \beta_m(x) |\alpha_m(x)| \quad (8)$$

where  $\beta_m(x)$  is a confidence measure associated to  $\mathbf{w}_m(x)$ . Apart from [14, 15], existing confidence measures are dedicated to specific motion estimation methods. For a variational approach, [7] uses the inverse of the global energy. For local approaches like [17], eigenvalues of the structure tensor are usually exploited [19]. For parametric estimations in general, the variance of the estimate is also a possible confidence measure. To keep the generality and simplicity of our method, we consider the following general weights:

$$\beta_m(x) = \exp - \left( \sigma^{-2} \int_{\Omega} g(x, y, I_1) \rho_0(\mathbf{w}_m(x), I_1(y), I_2(y)) dy \right) \quad (9)$$

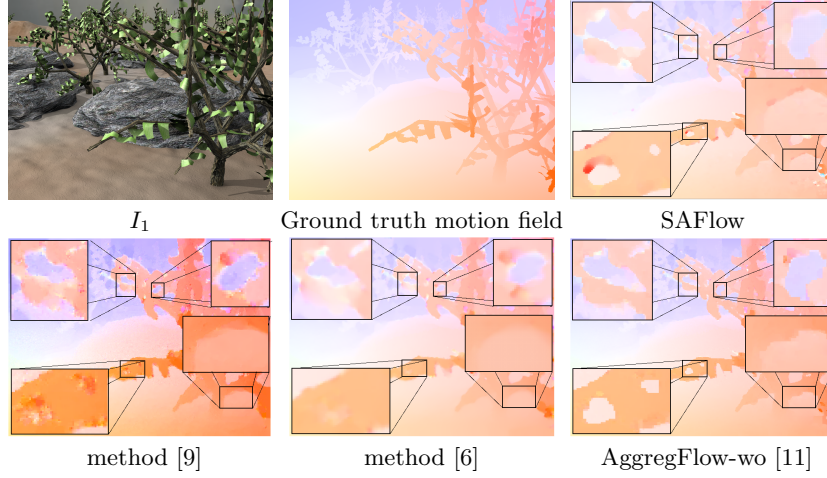
where  $g(x, y, I_1)$  are bilateral weights defined as:

$$g(x, y, I_1) = \exp - \left( \frac{\|x - y\|_2^2}{\sigma_s^2} + \frac{(I_1(x) - I_1(y))^2}{\sigma_r^2} \right) \quad (10)$$

and  $\rho_0(\mathbf{w}_m(x), I_1(y), I_2(y)) = |I_2(y + \mathbf{w}_m(x)) - I_1(y)|$  is a classical data potential penalizing deviation from the brightness constancy equation. The weight  $\beta_m(x)$  is then a measure of the local coherency of brightness constancy through bilateral filtering. The final energy is:

$$E(\mathbf{w}, \boldsymbol{\alpha}) = \int_{\Omega} \|\mathbf{w}(x) - \boldsymbol{\alpha}(x)^\top \mathbf{w}_C(x)\|_1 + \lambda_2 \|\boldsymbol{\alpha}(x)\|_{1,\beta(x)} + \lambda_1 \|\nabla \mathbf{w}(x)\|_1 dx. \quad (11)$$

We minimize  $E(\mathbf{w}, \boldsymbol{\alpha})$  alternatively on  $\mathbf{w}$  and  $\boldsymbol{\alpha}$ . Minimization w.r.t.  $\mathbf{w}$  is performed by solving the Euler-Lagrange equations with fixed point iterations [5].



**Fig. 2.** Preservation of small motion details and discontinuities on the *Grove3* sequence of the Middlebury benchmark. Top row: first frame, ground truth motion field and motion field estimated with SAFlow. Bottom row: motion field estimated with [9], [6] and [11]. Zooms on regions of interest overlay the images.

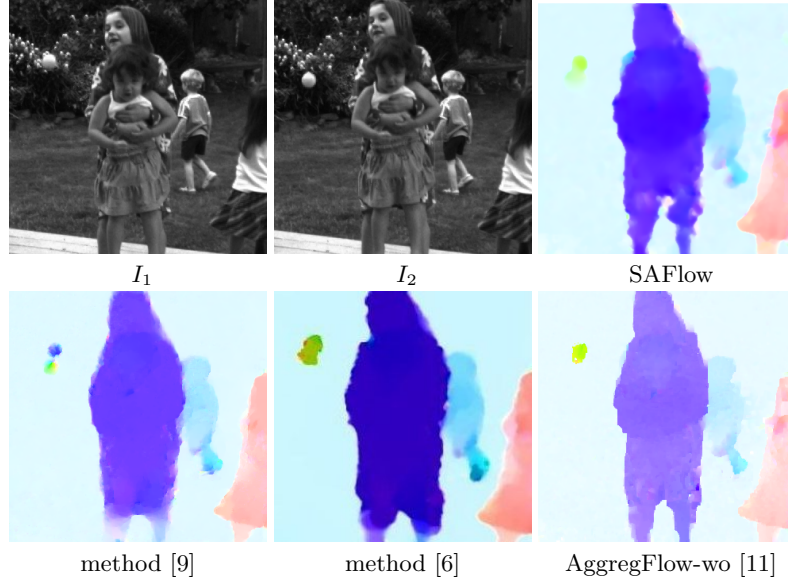
To minimize w.r.t.  $\alpha$ , we resort to a greedy algorithm. From an initial configuration of  $\alpha$ , we search for possible configurations of  $\alpha$ , and a configuration is kept if it leads to a decreasing of the energy. The search strategy consists in iteratively adding non null components ordered by decreasing value of the confidence measure.

## 4 Experimental results

We have evaluated our method on sequences of the Middlebury benchmark [1]. We experimentally compare our method named SAFlow (Sparse Aggregation for optical Flow) to the AggregFlow method (but without occlusion handling and designated by AggregFlow-wo) [11] and the variational methods described in [6, 9]. Local improvements related to discontinuity preservation are illustrated visually. The candidates sets were obtained with parameters  $\nu \in \{15, 45, 115\}$ ,  $\alpha = 0.8$ ,  $N = 2$ . Other parameters are set as follows:  $\sigma = 0.1$ ,  $\sigma_s = 5$  and  $\sigma_r = 20$ .

Figure 2 illustrates the ability of SAFlow to capture motion discontinuities and small details. Motion fields are less sharp than with AggregFlow-wo but, they are significantly better than [6, 9].

In Fig. 3, the displacement of the small ball is typically badly handled by variational methods using coarse-to-fine schemes, as [9]. In contrast, SAFlow satisfyingly recovers the displacement. The method of [6] integrating feature matching in a variational framework also captures the motion of the ball, but



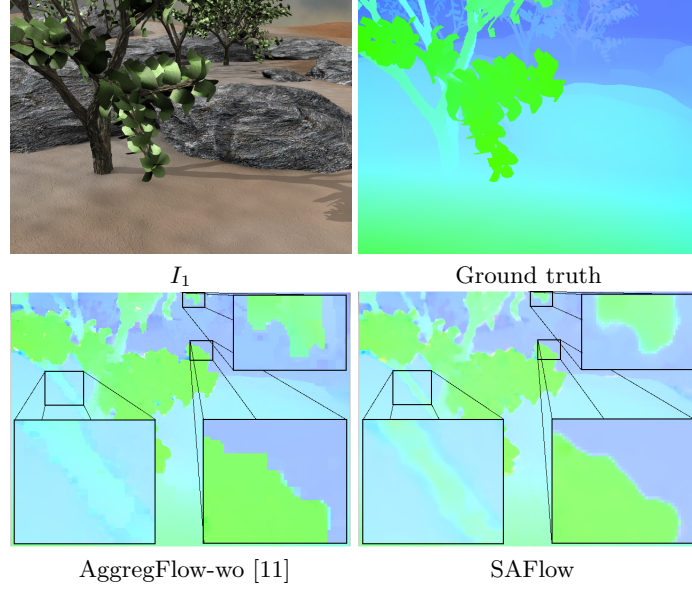
**Fig. 3.** Results on the *Backyard* sequence of the Middlebury benchmark. Top row: first and second frames (ground truth is not available for this sequence) and motion field estimated with SAFlow. Second row: motion field estimated with [9] , [6] and [11].

the shape of the ball is less preserved. It is also more impacted by the associated occlusion region.

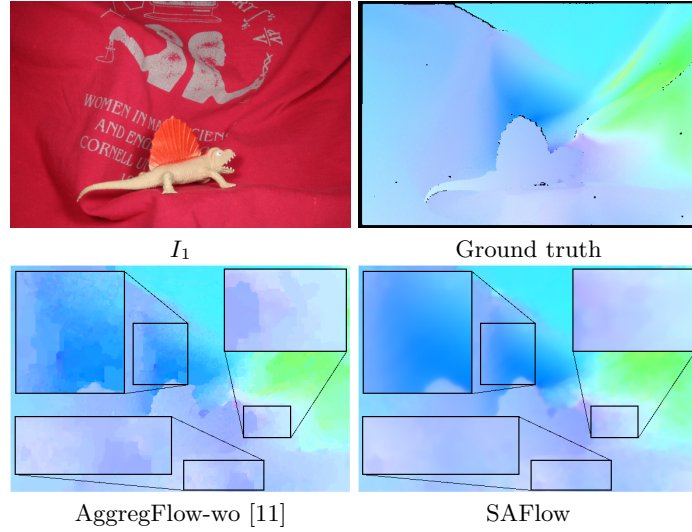
Figure 4 illustrates the ability of SAFlow to deal with less accurate motion candidates. In this experiment, we set the overlap ratio, i.e. the proportion of area shared by two neighbor patches, to  $\alpha = 0.5$ . This parameter is essential to deliver good candidates. In Fig. 4, typical artifacts of AggregFlow-wo can be observed. At motion discontinuities, the patches are not overlapping enough to produce accurate candidates, which implies block artifacts for AggregFlow-wo, due to the hard selection of one candidate. In contrast, SAFlow can deviate from the set of candidates to preserve clean discontinuities.

Discrete optimization in [11] tends to produce block artifacts for complex smooth deformations, as illustrated in Fig. 5. The variational optimization of SAFlow does not have this problem and estimate more accurately smooth flow fields. Finally, the computational time of SAFlow is around 5 minutes, mostly spent in the candidates computation step, while AggregFlow-wo requires 20 minutes.

Table 1 contains Angular Errors obtained with SAFlow, AggregFlow-wo [11] and the variational methods [6,9] for sequences of the Middlebury benchmark. The results of SAFlow are generally less accurate than those produced by AggregFlow-wo, especially sequences showing small motion details or sharp motion discontinuities. The performance of SAFlow can be affected by the confidence measures  $\beta_m(x)$ . Large errors of confidence measures can significantly de-



**Fig. 4.** Comparison of discrete and continuous aggregation for a small set of candidates on the *Grove2* sequence of the Middlebury benchmark. The candidates were computed with  $\alpha = 0.5$ . From left to right: first frame, ground truth, motion field estimated with AggregFlow-wo [11] and SAFlow. Zooms on regions of interest overlay the images.



**Fig. 5.** Comparison of discrete and continuous aggregation for complex and smooth flow fields on the *Dimetrodon* sequence of the Middlebury benchmark. From left to right: first frame and ground truth motion field, motion field estimated

**Table 1.** Angular errors obtained with SAFlow, AggregFlow-wo [11] and the methods [6] and [9] on sequences of the Middlebury benchmark.

	Grove2	Grove3	Hydrangea	Urban2	Urban3
SAFlow	2.43	5.92	2.29	2.53	4.12
AggregFlow-wo	2.19	5.43	2.47	2.47	3.42
method [6]	2.38	5.97	2.10	2.50	3.91
method [9]	2.92	6.72	2.29	2.63	6.10

crease the accuracy of SAFlow. Nevertheless SAFlow yields better performance than the method [9] and is competitive with the method [6].

## 5 Conclusion

We have proposed a continuous aggregation strategy minimizing a global energy for optical flow computation as an alternative to the discrete aggregation presented in [11]. A first version uses the min function but is limited because of severe non-convexity of the energy. A more attractive convex formulation exploits a sparse dictionary model. Experiments show that the overall quantitative performance remains lower than with discrete aggregation. Additional experiments on the Sintel MPI database will be performed as in [8]. However results are still competitive with standard variational methods. Moreover, the ability to reconstruct motion vectors beyond the candidates set makes the continuous aggregation more robust to smaller and suboptimal candidate sets. It also better behaves in case of complex and smooth motion fields. The computational cost of continuous aggregation is significantly lower than for discrete aggregation.

**Acknowledgments:** This work was carried out as part of the Quaero program, funded by OSEO, French State agency for innovation. It was also partly supported by the France-BioImaging project (ANR-10-INBS-04-07, "Investments for the future"). We thank also Dr. O. Demetz, Dr. S. Setzer and Prof. J. Weickert for fruitful discussions, and the reviewers for suggestions.

## References

1. S. Baker, D. Scharstein, JP Lewis, S. Roth, M.J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *IJCV*, 92(1):1–31, 2011.
2. C. Barnes, E. Shechtman, D.B. Goldman, and A. Finkelstein. The generalized patchmatch correspondence algorithm. In *ECCV*, pages 29–43, 2010.
3. M.J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVIU*, 63(1):75–104, 1996.
4. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.
5. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, pages 25–36, Prague, Czech Republic, 2004.

6. T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *PAMI*, 33(3):500–513, 2011.
7. A. Bruhn and W. Weickert. A confidence measure for variational optic flow methods. *Geometric Properties for Incomplete Data*, pages 283–298, 2006.
8. D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, pages 611–625. Springer-Verlag, 2012.
9. A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
10. Z. Chen, H. Jin, Z. Lin, S. Cohen, and Y. Wu. Large displacement optical flow from nearest neighbor fields. In *CVPR*, pages 2443–2450, 2013.
11. D. Fortun, P. Bouthemy, and C. Kervrann. Aggregation of local parametric candidates with exemplar-based occlusion handling for optical flow. Preprint arXiv:1407.5759.
12. D. Fortun, P. Bouthemy, and C. Kervrann. Optical flow modeling and computation: a survey. *CVIU*, 2015.
13. B.K.P. Horn and B.G. Schunck. Determining optical flow. *Art. Intel.*, 17(1-3):185–203, 1981.
14. C. Kondermann, M. Rudolf, and C. Garbe. A statistical confidence measure for optical flows. In *ECCV*, pages 290–301, Marseille, France, 2008.
15. J. Kybic and C. Nieuwenhuis. Bootstrap optical flow confidence and uncertainty measure. *CVIU*, 115(10):1449–1462, 2011.
16. M. Leordeanu, A. Zafir, and C. Sminchisescu. Locally affine sparse-to-dense matching for motion and occlusion estimation. In *ICCV*, pages 1221–1228, Sydney, Australia, 2013.
17. B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Int. Joint Conf. Art. Intel.*, pages 674–679, 1981.
18. E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *TIP*, 7(5):703–719, 1998.
19. C. Mota, L. Stuke, and E. Barth. Analytic solutions for multiple motions. In *ICIP*, pages 917–920, Thessaloniki, Greece, 2001.
20. M. Mozerov. Constrained optical flow estimation as a matching problem. *TIP*, 22(5):2044–2055, 2013.
21. J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *JVCIR*, 6(4):348–365, 1995.
22. Kenneth Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, 1998.
23. D. Sun, E.B. Sudderth, and M.J. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In *NIPS*, pages 2226–2234, Vancouver, Canada, 2010.
24. A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure-and motion-adaptive regularization for high accuracy optic flow. In *ICCV*, pages 1663–1668, Kyoto, Japan, October 2009.
25. P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, et al. Deepflow: Large displacement optical flow with deep matching. In *ICCV*, pages 1385–1392, Sydney, Australia, 2013.
26. M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *CVPR*, pages 2464–2471, San-Fransisco, 2010.
27. H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *IJCV*, 93(3):1–21, 2011.