A Variational Aggregation Framework for Patch-Based Optical Flow Estimation

Denis Fortun, Patrick Bouthemy, Charles Kervrann

Noname manuscript No.
(will be inserted by the editor)

A Variational Aggregation Framework for Patch-Based Optical Flow Estimation

Denis Fortun · Patrick Bouthemy · Charles Kervrann

Received: date / Accepted: date

Abstract We propose a variational aggregation method for optical flow estimation. It consists of a two-step framework, first estimating a collection of parametric motion models to generate motion candidates, and then reconstructing a global dense motion field. The aggregation step is designed as a motion reconstruction problem from spatially-varying sets of motion candidates given by parametric motion models. Our method is designed to capture large displacements in a variational framework without requiring any coarse-to-fine strategy. We handle occlusion with a motion inpainting approach in the candidates computation step. By performing parametric motion estimation, we combine the robustness to noise of local parametric methods with the accuracy yielded by global regularization. We demonstrate the performance of our aggregation approach by comparing it to standard variational methods and a discrete aggregation approach on the Middlebury and MPI Sintel datasets.

Keywords Optical flow · parametric motion · aggregation · variational optimization

D. Fortun
Inria - Centre de Rennes -Bretagne Atlantique, Campus Universitaire de Beaulieu, 35 042 Rennes, France
Center for Biomedical Imaging - Signal Processing core (CIBM-SP), EPFL, Switzerland
Biomedical Imaging Group, EPFL, Switzerland
E-mail: denis.fortun@epfl.ch (corresponding author)

P.Bouthemy, C. Kervrann
Inria - Centre de Rennes -Bretagne Atlantique, Campus Universitaire de Beaulieu, 35 042 Rennes, France
E-mail: patrick.bouthemy@inria.fr
E-mail: charles.kervrann@inria.fr

1 Introduction

1.1 General positioning

Optical flow estimation is based on the conservation assumption of image features such as image intensity, image gradient or texture descriptor. The so-called brightness constancy assumption is the most 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 be classified into two main categories:

- Local spatial coherency is exploited when considering a parametric motion model, e.g., local translation [54], affine model or quadratic model [63], in a given neighborhood or an appropriate local region. The neighborhoods must be sufficiently textured or contain interest points to supply reliable velocity vectors.
- Global coherency [44] imposes a regularization constraint to the motion field on the whole spatial domain. The flow field is generally assumed to be piecewise smooth and the strategy is to minimize a global energy of the form

$$E(\boldsymbol{w}) = \int_{\Omega} \rho(\boldsymbol{x}, u, v, \boldsymbol{w}) + \lambda\, \phi(\nabla \boldsymbol{w}(\boldsymbol{x}))\, d\boldsymbol{x}, \qquad (1)$$

that explicitly combines a potential $\rho(\cdot)$, which penalizes deviations from the brightness constancy equation, with a regularization potential $\phi(\cdot)$ which penalizes high values of the norm of the gradient $\nabla \boldsymbol{w}$ of the velocity field $\boldsymbol{w} : \Omega \to \mathbb{R}^2$, where $\Omega \subset \mathbb{R}^2$ denotes the image domain. The two consecutive images are denoted by $u, v : \Omega \to \mathbb{R}$, $\boldsymbol{x} \in \Omega$ denotes

one pixel of the grid Ω, and λ is a balance parameter.

The best state-of-the-art results are obtained with the global approach. Nevertheless, several open issues remain unsolved. One of the main limitations comes from the undesirable effects due to the coarse-to-fine strategy used to handle large displacements [33]. The motion of small objects is discarded at coarse scales, and the error is propagated in the incremental updates at finer scales when the displacement is larger than the object size. As a result, motion details are not correctly recovered in the final estimated flow field [86]. Large displacements are also associated to large occlusions, which are another major source of errors. Occlusion handling is often treated as a post-processing task. It is then very sensitive to errors in the initial motion estimation. Finally, noise sensitivity is usually ignored in standard optical flow evaluation benchmarks. However, if pixelwise data potentials provide best results in absence of noise, they are not adapted when noise is present in input images. To limit the impact of these failure cases, the solution of the global approach is often to increase regularization, producing oversmoothed results, loosing motion details and blurring discontinuities.

Existing purely local methods [11,49,54,70] are far from being able to compete with global methods in terms of accuracy in optical flow benchmarks. The main issue is to be able to select appropriate local regions. The most basic approaches considering square patches centered on each pixel [54] are unable to retrieve motion discontinuites. They are also prone to the same large displacement and occlusion problems as the global methods. Nevertheless, joint global motion estimation and segmentation approaches [56,77,80] have demonstrated that piecewise parametric representation of flow fields can yield excellent results when local regions are appropriately chosen. However, the required alternate optimization scheme is computationally demanding and sensitive to the initialization. On the other hand, local methods are also known to be less sensitive to noise than global approaches [20]. These observations suggest that the potential of local methods may still be under-exploited.

The goal of this paper is to design a new way to combine parametric models with a global variational approach through aggregation procedure, in order to both overcome the above mentioned limitations of global methods and exploit the potential of parametric estimation.

1.2 Our contributions

We propose a novel aggregation approach for optical flow estimation based on motion reconstruction from spatially varying candidates computed with parametric models.

Our method is composed of a first step estimating a collection of parametric motion models generating local motion candidates, followed by an aggregation step combining the candidates to create a global dense motion field. The main contribution of the present work is in the aggregation step. We formulate the problem as a motion reconstruction step selecting the best candidate while ensuring global smoothness of the motion field. This approach differs from other motion estimation techniques, since it decouples motion estimation and motion reconstruction. The main interest is that the reconstructed motion field is not involved in a brightness conservation constraint, and is thus not affected by its limitations. In particular, our method is able to handle large displacements without coarse-to-fine schemes, it provides a valid data constraint in occluded regions, and it is more robust to noise in input images than standard variational approaches.

To achieve this, we provide motion candidates in the first step of our method that also handle large displacements, occlusions and noise in input images, by following the idea of our previous work [38]. We rely on the computation of parametric motion models over a set of overlapping size-variable square patches, that allows us to deal with various configurations of piecewise affine motions. An exemplar-based candidates extension strategy finds relevant motion candidates in occluded regions.

We provide an extensive experimental evaluation of our aggregation framework insisting on the versatility of its performance. We demonstrate that it outperforms the standard variational approach in case of large displacements, large occlusions and noise in input images, but also in more common situations as they can be found in the classical Middlebury benchmark. We also compare our variational aggregation with the aggregation based on discrete optimization we described in [38], removing any other specific features of [38] for fair comparison. We show that the method presented in this paper is faster and more robust to suboptimal candidate sets, while being competitive in terms of quantitative error. A first shorter version of this work was described in [37]. Compared to [37], we have integrated an occlusion handling module in the candidates estimation stage, we have modified the aggregation model to enforce the selection of a single candidate, we have improved the optimization step of our method, and

we have extended the experimental validation of the method.

1.3 Related work

In this Section, we offer a brief overview of the main open issues in optical flow estimation. A recent comprehensive survey is available in [36].

Numerous modifications of the Horn & Schunck model [44], starting with [13,43,61], have been proposed over years, specifically to cope with large displacements and preservation of motion discontinuities [18,56,76,82,86, 89]. The most common response to face these two issues has been to design a multi-resolution and incremental coarse-to-fine framework along with piecewise smoothing or robust estimation. As for the data term of the global energy function other image features have been introduced like image gradient [18], texture component [82]. Besides, invariance properties have been sought to overcome limitations of the classical intensity constancy assumption by using Normalized Cross Correlation (NCC) [84], Census transform [41], or LDP (Local Directional Pattern) descriptor [58]. However, optimization complexity increases with the sophistication of the modeling.

Local and global methods may involve parametric motion models [12,29,30,39,45,53,54,56,63,77,87]. The most frequently adopted ones are polynomial motion models such as translation, affine, quadratic, but other models can be investigated as well [45]. When attached to local optimization, the parametric motion models are estimated on local regions usually defined as square patches centered on each pixel [12,54], possibly with an adaptation of the patch size [55,70], or its position [49]. This local optimization setting is easy to implement with a low computational cost, but it is clearly outperformed by sophisticated extensions of [44] in recent optical flow benchmarks [5,23]. The motion candidates produced by our method are composed of affine motion vectors estimated in square patches without any motion segmentation. Our method implicitly selects the best patch size and position when selecting motion candidates to recover the global flow field in the second step.

When dealing with large displacements, using discrete optimization is a way to avoid resorting to coarse-to-fine schemes [38,57,87]. Another common approach is to somehow integrate feature correspondences in dense motion estimation. A first category of variational methods [17,19,83] includes an additional term in the global energy. This term makes the estimated flow be close to pre-computed correspondences. However, this approach may be sensitive to matching errors by giving a fixed weight to the correspondence fitting. To overcome this problem, recent works [17,79,83] have deliberately focused on improving the matching step. Another class of methods use correspondences to provide a coarse initialization for subsequent refinement [4,6,27,60,86]. In that vein, recovering a dense flow from initial sparse correspondences is also currently investigated [68,79]. In [74], the variational refinement process is iterative and interpreted as the minimization of the orginal nonlinearized energy. The main motivation to incorporate feature matching in global optical flow methods is to alleviate the drawbacks of the coarse-to-fine scheme imposed by the classical variational optimization, in particular the loss of large displacements of small objects. Our patch correspondence substep is only involved in the motion candidates generation process and it does not drive the global optimization subsequent step.

Occlusion is a key issue in motion estimation [73], especially in case of large displacements, since no motion measurements are available in occluded areas. By definition, a point of the current image which is occluded in the consecutive image has no corresponding point. One has to distinguish between *occlusion detection*, and *occlusion filling* with motion vectors. The two tasks can be addressed jointly within an alternate optimization strategy [3,38,47,64,78,75]. Filling occluded regions with velocity vectors given the occlusion map (or in other words, motion inpainting in occluded regions) can be related to the image inpainting problem. Image inpainting methods can be coarsely divided into diffusion-based methods [10,25] and exemplar-based methods [31,50]. Exemplar-based image inpainting fills missing parts by copying pixels of the observed image. In motion estimation, occlusion filling is usually solved by diffusion-based (or geometry-oriented) schemes, propagating motion from non-occluded regions to occluded regions using partial derivative equation (PDE) resolution [3,9,47,51,64,86]. In contrast, we adopt an exemplar-based strategy for candidates computation in occluded regions.

Our method share similarities with dictionary-based methods, looking for sparse combination of candidate motion vectors. Sparse representations of motion fields have recently been exploited for the design of regularization terms [28,32,48,71], replacing classical spatial regularization by a proximity constraint to a sparse combination of learned patch flow fields. These strategies only act on the regularization term and are thus affected by all the above mentioned issues of global methods. Estimating directly the motion field as a linear combination of learned motion models in patches has been investigated in [14,35,62] with PCA decomposition on various types of training sets. However, this approach tends to produce blurry results, and has been

combined with a layered approach in [85] to yield sharper results. One limitation is that the coefficients are estimated with a standard data term based on brightness constancy assumption. Finally, in [1], a pixel wise dictionary is learned online with phase correlation and a constraint on the entropy of the weights is imposed. However, the estimation only provides pixelic accuracy, without global regularization of the motion field, which causes large errors.

Robustness to noise in input images has only received little attention in the optical flow literature. In the local parametric estimation framework, explicit modeling of noise has led to dedicated methods [34, 72]. Experimental comparisons between local and global approaches [8, 40] have demonstrated the highest sensitivity to noise of global approaches. Improving robustness to noise of global variational methods has been achieved in [20] by integrating the local parametric assumption in the data term. However, this improvement comes at the cost of a loss of accuracy in the absence of noise.

Finally, we mention that a similar combination of candidates has been explored in the domains of image colorization [22, 65] and image completion [2].

1.4 Paper organization

The rest of the paper is organized as follows. In Section 2, we present the parametric estimation of motion candidates. In Section 3, we propose an aggregation method in a variational setting. In Section 4, we demonstrate the performance of our estimation algorithms on sequences of the Middlebury and MPI Sintel datasets and other real images. Section 5 contains concluding remarks and future work.

Notations The Euclidean norm (ℓ_2 norm) of a vector $\boldsymbol{z} = (z_1, \cdots, z_d)^T \in \mathbb{R}^d$ is given by $\|\boldsymbol{z}\|_2 = (\sum_{i=1}^d z_i^2)^{1/2}$ and the ℓ_1 norm of $\boldsymbol{z}$ by $\|\boldsymbol{z}\|_1 = \sum_{i=1}^d |z_i|$. The supremum norm of $\boldsymbol{z}$ is $\|\boldsymbol{z}\|_\infty = \sup_{1\leq i\leq d} |z_i|$.

We denote two consecutive 2D image frames as $u, v : \Omega \to \mathbb{R}$, with $\Omega \in \mathbb{R}^2$ denoting the image domain. We denote $\boldsymbol{x}, \boldsymbol{x}'$ or $\boldsymbol{y}$ one pixel of the image grid Ω and $\text{card}(\Omega)$ is the number of pixels.

We denote $\boldsymbol{p}_u(\boldsymbol{x_p}, h) := \big(u(\boldsymbol{x_p} + \boldsymbol{\tau}), \boldsymbol{\tau} \in \{\frac{h-1}{2}, \cdots, \frac{h+1}{2}\}^2\big)$ a patch of u centered at location $\boldsymbol{x_p} \in \Omega$. The square window[1] $U_{\boldsymbol{p}}(\boldsymbol{x_p}, h) = \{\boldsymbol{x} \in \Omega : \|\boldsymbol{x} - \boldsymbol{x_p}\|_\infty \leq h\}$ is the patch support centered at pixel $\boldsymbol{x_p}$ and the number of pixels falling in $U_{\boldsymbol{p}}(\boldsymbol{x_p}, h) \subset \Omega$ is $h \times h$. We define $\boldsymbol{\mathcal{P}}_u := \{\boldsymbol{p}_u(\boldsymbol{x_p}, h) : \boldsymbol{x_p} \in \Omega, h \in \boldsymbol{\mathcal{H}}\}$ as the set of all overlapping patches and $\boldsymbol{\mathcal{H}} = \{h_1, \ldots, h_M\}$ is a finite set of M prescribed patch sizes $h_m \in \mathbb{Z}^+$.

We denote $\boldsymbol{w}(\boldsymbol{x}) = (v_1(\boldsymbol{x}), v_2(\boldsymbol{x}))^\top$ the motion vector at pixel $\boldsymbol{x}$ of the motion field $\boldsymbol{w}$.

The occlusion map $o : \Omega \to \{0, 1\}$ is defined such that $o(\boldsymbol{x}) = \mathbb{1}[\boldsymbol{x}$ is occluded] where $\mathbb{1}[\cdot]$ is the indicator function. The set of occluded pixels is denoted $O = \{\boldsymbol{x} \in \Omega : o(\boldsymbol{x}) = 1\}$.

Additional notations will be introduced in the text.

[1] Without loss of generality, isotropic circular patches could be considered as well.

2 Local motion candidates and occlusion cues

We describe in this section the first step of our aggregation method. It follows the approach of [38] but its presentation is partly revisited. It exploits local information to supply motion candidates at each pixel. A set of motion vector candidates is generated at every pixel by a combination of patch correspondences and local parametric motion model estimations. A specific treatment is applied to occluded regions by exemplar-based extension of the motion candidates set. Our approach can be viewed as a new way to address the problem of choosing the local neighborhood for parametric estimation.

2.1 Local parametric motion candidates

The local supports for motion candidates computation are overlapping square patches of different sizes. To capture different motion scales, the patch sizes must cover a range of values. Due to the overlap and the number of patch sizes, one given pixel $\boldsymbol{x} \in \Omega$ belongs to several patches. The candidate motion vectors at each pixel $\boldsymbol{x}$ are computed independently in each patch in two substeps described below: patch correspondences and affine motion refinement.

2.1.1 Patch correspondences for large displacements

We assign to each patch $\boldsymbol{p}_u(\boldsymbol{x_p}, h)$ in u the set $\{\boldsymbol{p}_v(\boldsymbol{y}_1, h), \cdots, \boldsymbol{p}_v(\boldsymbol{y}_K, h)\}$ of the K patches $\boldsymbol{p}_v(\cdot, h)$ in v most similar to $\boldsymbol{p}_u(\boldsymbol{x_p}, h)$. Hence, for each established pair of corresponding patches, we get the translation vector $\boldsymbol{t}_k(\boldsymbol{x_p}, h) \in \mathbb{Z}^2$, shifting $\boldsymbol{p}_u(\boldsymbol{x_p}, h)$ onto $\boldsymbol{p}_v(\boldsymbol{x_p} + \boldsymbol{t}_k(\boldsymbol{x_p}, h)\ , h), k \in \{1, \cdots, K\}$. Let us put forward that we do not aim at keeping at this stage the best correspondence only but at selecting K relevant correspondences to subsequently constitute motion candidates (K is assumed to be constant for all patches). The matching step is generic and could be achieved with any arbitrary feature matching algorithm (e.g., *PatchMatch* algorithm [7]).

2.1.2 Affine motion refinement

The displacements estimated by patch correspondences are integer-pixel translational approximations. To attain subpixel accuracy and to allow for more complex motion, we refine the first sub-step with the estimation of a local affine motion model in every pre-registered patch pair. Denoting $U_{\boldsymbol{p}}(\boldsymbol{x_p}, h)$ the pixel support of $\boldsymbol{p}_u(\boldsymbol{x_p}, h)$, we estimate the affine motion model between two corresponding patches $\boldsymbol{p}_u(\boldsymbol{x_p}, h)$ and $\boldsymbol{p}_v(\boldsymbol{x_p}+\boldsymbol{t}(\boldsymbol{x_p}, h), h)$ at pixel $\boldsymbol{x} = (x, y)^T \in U_{\boldsymbol{p}}(\boldsymbol{x_p}, h)$ defined as:

$$\delta \boldsymbol{w}_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{p}}(\boldsymbol{x_p}, h)) = \begin{pmatrix} \theta_1 + \theta_2 x + \theta_3 y \\ \theta_4 + \theta_5 x + \theta_6 y \end{pmatrix}. \tag{2}$$

Assuming brightness constancy, an estimation of the parameter vector $\boldsymbol{\theta}_{\boldsymbol{p}}(\boldsymbol{x_p}, h) = (\theta_1, \cdots, \theta_6)^T$ is the minimizer of

$$\int_{U_{\boldsymbol{p}}(\boldsymbol{x_p},h)} \varphi(v(\boldsymbol{x} + \delta \boldsymbol{w}_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{p}}(\boldsymbol{x_p}, h)) \ + \ \boldsymbol{t}(\boldsymbol{x_p}, h)) - u(\boldsymbol{x}))d\boldsymbol{x} \tag{3}$$

where the penalty function $\varphi(\cdot)$ is a robust function of the family of M-estimators (e.g., Tukey's function).

2.1.3 Final set of motion vector candidates

The above described two-step estimation is repeated for every patch $p_u(\boldsymbol{x_p}, h)$ and generates a set of candidate motion vectors $\boldsymbol{\mathcal{C}}(\boldsymbol{x})$ at each pixel $\boldsymbol{x} \in \Omega$. In this paper, we consider sets of regularly spaced patches, defined by a set of sizes $\boldsymbol{\mathcal{H}}$ and an overlap ratio $r \in [0, 1]$ defining the proportion of area shared by two neighbour patches of the same size. Denoting $\Omega_p \subset \Omega$ the set of center pixels of the previously defined patches, the candidates are defined as follows:

$$\boldsymbol{\mathcal{C}}(\boldsymbol{x}) = \{\boldsymbol{t}_k(\boldsymbol{x_p}, h) + \delta \boldsymbol{w}_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{\theta}_{\boldsymbol{p}}(\boldsymbol{x_p}, h)) : h \in \boldsymbol{\mathcal{H}}, \ \boldsymbol{x_p} \in \Omega_{\boldsymbol{p}} : \|\boldsymbol{x_p} - \boldsymbol{x}\|_\infty \leq h, k \in \{1, \cdots, K\}\}. \tag{4}$$

The interest of the local set of motion candidates is first that the correspondence sub-step efficiently copes with large displacements. Specifically, it allows us to correctly deal with small structures undergoing large displacements. Second, by considering a variety of patches, we override the predefined choice of the local neighborhood. The implicit selection of the proper patch via its corresponding motion candidate is transferred to the aggregation stage. Third, introducing patches of several sizes enables to tackle motion of different scales.

2.2 Exemplar-based candidates extension in occluded regions

The computation of motion candidates described in Section 2.1 does not distinguish occluded and non-occluded pixels. However, in large occluded regions where the patches contain mostly occlusions, there is no chance to estimate relevant candidates with this local approach. Therefore, the occluded pixels require a dedicated process to compute additional motions candidates. This computation could nevertheless be considered as optional for small displacements. Indeed, considering large patch sizes enables to cope with small occlusion areas and to generate relevant candidates at motion discontinuities or at occluded positions.

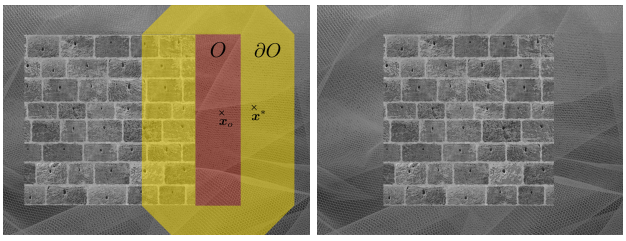


Fig. 1 Illustration of the exemplar-based inpainting of motion candidates. The foreground is shifting to the right over a static background. The candidate set of occluded pixel $\boldsymbol{x}_o \in O$ (in red) is extended by adding the candidates of its matched non-occluded pixel $\boldsymbol{x}^* \in \partial O$ (in yellow).

When the occluded regions are known or given by an occlusion detector [43,46,86], occlusion filling with motion vectors is conceptually closely related to image inpainting, since it recovers motion in regions where motion is by definition *not observable*. In order to deal with large occlusions produced by large displacements, we follow the inpainting analogy. In the first step of our aggregation method, the motion candidates set is thus augmented by "copy-paste" operations as described below.

We rely on the assumption that the motion at an occluded pixel $\boldsymbol{x}_o \in O$ is similar to the motion at a close non-occluded pixel in $\Omega \backslash O$ belonging to the same object or the same background part. The idea is to assign the set $\boldsymbol{\mathcal{C}}(\boldsymbol{x})$ of the most similar pixel $\boldsymbol{x}^* \in \Omega \backslash O$ to the occluded pixel $\boldsymbol{x}_o$. We limit the search for $\boldsymbol{x}^*$ in a band ∂O along the occlusion boundaries. Figure 1 illustrates the matching process and the definition of O and ∂O in a simple synthetic example. Searching for the most similar pixel denoted $\boldsymbol{x}^* \in \Omega \backslash O$ to $\boldsymbol{x}_o$ is actually easier for motion inpainting than for image inpainting. Indeed, the information supplied by image u is available even in O. Thus, as $\boldsymbol{x}_o$ is expected to belong to the same object as $\boldsymbol{x}^*$, we use patch similarity to find the best match in u.

An extended candidate set $\boldsymbol{\mathcal{C}}_+(\boldsymbol{x}_o)$ is created for occluded pixels $\boldsymbol{x}_o$ by adding to the initial set $\boldsymbol{\mathcal{C}}(\boldsymbol{x}_o)$ the motion candidates of their matched pixel $\boldsymbol{x}^*$:

$$\boldsymbol{\mathcal{C}}_F(\boldsymbol{x}_o) = \boldsymbol{\mathcal{C}}(\boldsymbol{x}_o) \cup \mathcal{C}(\boldsymbol{x}^*), \ \forall \boldsymbol{x}_o \in O. \tag{5}$$

Table 1 EPE-all scores of motion fields on sequences with ground truth from MPI Sintel and Middlebury datasets

	MPI Sintel	Middlebury
Full BCF	**0.792**	**0.0710**
BCF w/o extension	1.851	0.0833
DeepFlow [83]	4.691	0.386
MDP-Flow2 [86]	4.006	0.223

By convention, $\forall \boldsymbol{x} \in \Omega \backslash O, \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}) := \boldsymbol{\mathcal{C}}(\boldsymbol{x})$.

A particular class of occluded (or disappearing) regions occurs at image borders in the case of large camera motion. We cope with this issue by estimating the dominant image motion due to camera motion using the Motion2D software applied to the whole image [63], which provides additional motion candidates.

2.3 Best candidate flow

To validate our method for computing motion candidates, we have exploited sequences from MPI Sintel and Middlebury datasets [5,23] provided with ground truth. We introduce the so-called *Best Candidate Flow* (BCF) by selecting at each pixel $\boldsymbol{x}$ the candidate motion vector of $\boldsymbol{\mathcal{C}}_F(\boldsymbol{x})$ closest to the ground truth vector at $\boldsymbol{x}$. We distinguish between the BCF determined with the candidates extension described in the preceding section (or full BCF) and the BCF without it (or BCF w/o extension).

In Table 1, we report the objective evaluation given by the Endpoint Error (EPE) scores for the full BCF and BCF without candidate extensions, on the training sequences of the datasets MPI Sintel and Middlebury. Overall, the full BCF is very close to the ground truth motion field demonstrating the performance of the local parametric motion computation. We also compare these results with those of motion fields supplied by [83, 86], as obtained with publicly available code. Clearly, full BCF outperforms these state-of-the-art methods in the two benchmarks. Accuracy is especially improved with full BCF for the MPI Sintel sequences where large displacements and wide occluded regions are present. It demonstrates that the combination of local affine estimations in square patches with patch correspondences as described in Section 2.1, is quite relevant to recover very accurate motion vectors.

3 Variational motion reconstruction framework

We have now to recover the global dense motion field by aggregating motion candidates available at each pixel. We define an aggregation strategy in a variational setting, which consists in minimizing an energy of the form

$$E(\boldsymbol{w}) = \int_{\Omega} \rho(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x})) + \lambda_1 \phi(\nabla \boldsymbol{w}(\boldsymbol{x})) d\boldsymbol{x}, \qquad (6)$$

where $\rho(\boldsymbol{w}(x), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}))$ is a fidelity term and the second term imposes smoothness of $\boldsymbol{w}$, balanced by the parameter λ_1. In the following, we consider a total variation (TV) regularization: $\phi(\nabla \boldsymbol{w}(\boldsymbol{x})) = \|\nabla \boldsymbol{w}(\boldsymbol{x})\|_1$. Unlike usual approaches for optical flow, the image intensities are not used as input of the data potential $\rho(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}))$, but are replaced by the motion candidate set $\boldsymbol{\mathcal{C}}_F(\boldsymbol{x})$. We detail in this section the modeling and optimization issues related to this reconstruction term, and the solution we adopted.

3.1 Preliminary observations

3.1.1 Candidates distribution

As a first investigation, we explore the information carried by the distribution of the candidates of each pixel. This analysis is motivated by the analogy with practices in other domains like image denoising or completion, where distribution of candidate image patches is exploited [42,69]. We provide in Figure 2 six representative examples of the main forms of candidate distributions that occur in practice, and their relations with the ground truth motion vectors and the original image data. The motion vector candidates are represented by blue circles, the ground truth is the red rectangle, and the estimated motion vector is the green triangle (the full estimated motion field is given in Figure 7). In the background of the distributions, we display the value of the displaced frame difference (DFD) penalized by the ℓ_1 norm, which can be seen as a data fitting term: for the distribution at a given pixel $\boldsymbol{x}_i$ (one of the six pixels in Figure 2), the value displayed in background at coordinates $\boldsymbol{w} = (v_1, v_2)$ is $|v(\boldsymbol{x}_i + \boldsymbol{w}) - u(\boldsymbol{x}_i)|$. The question is then to identify characteristic patterns that can allow us to identify the ground truth, given the candidates and the DFD values.

Firstly, we observe that the form of the candidate distribution is highly variable. In some situations, e.g. in Figures 2(a) and 2(d), a unique mode can be clearly identified and gives a good estimate of the ground truth. However, other examples show that the main modes do not always correspond to the ground truth motion vector, and that the distribution can have highly multimodal and complex shape. In general, the form of the distribution cannot be accurately predicted from the input data. Thus, it turns out that the estimation cannot be only driven by local empirical distributions. Options

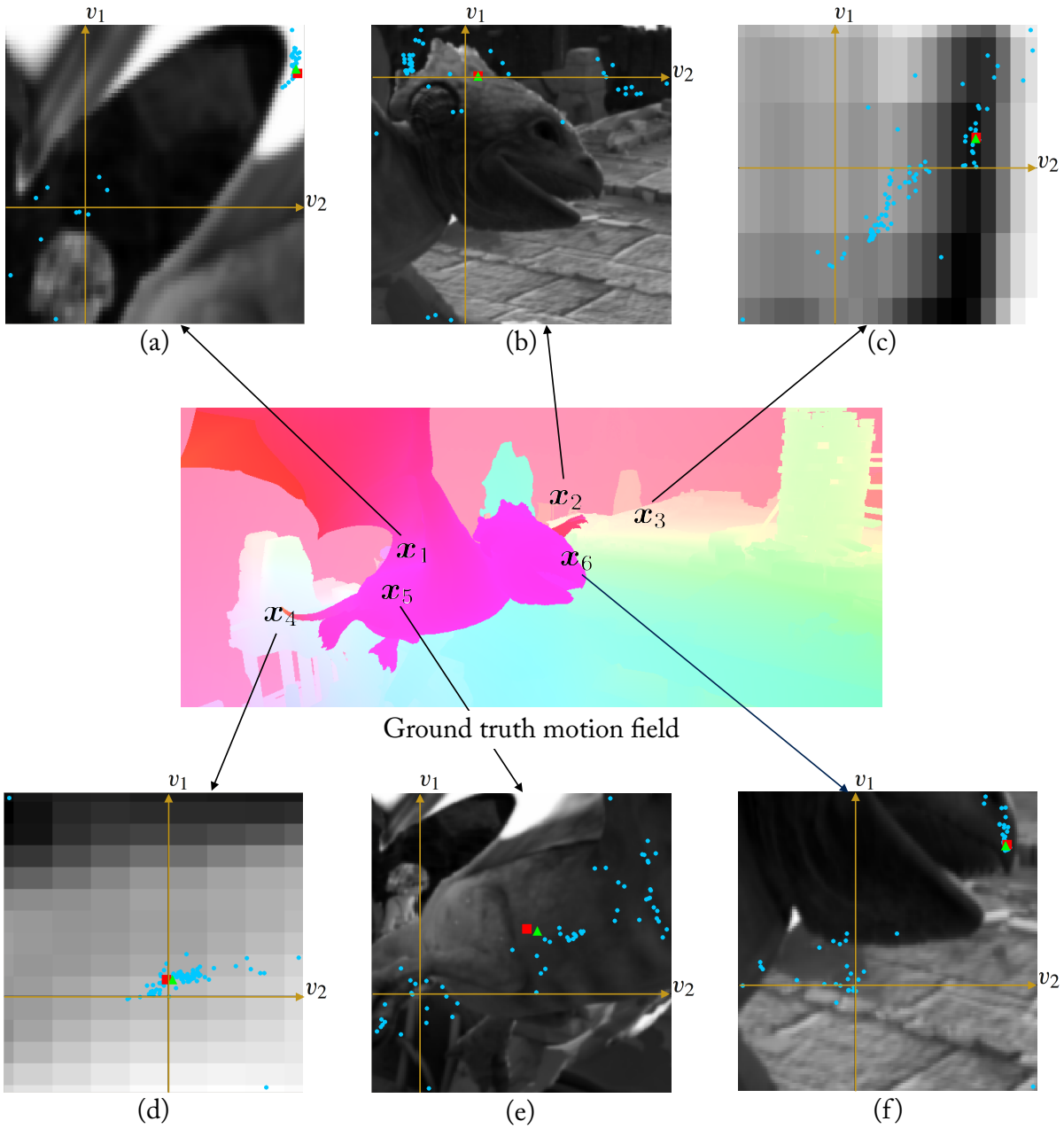


Fig. 2 Visualization of the distribution of the motion candidates at several pixels in the image. The central image is the ground truth motion field of the frame 23 of the *temple_2* sequence of the MPI Sintel data set. The six plots represent the motion vector candidates (blue circles), the motion vector ground truth (red rectangle) and the estimated motion vector with our method (green triangle)(the estimated motion field on the whole image is given in Figure 7) at each corresponding pixel. The horizontal and vertical axes are respectively the horizontal and vertical components of the motion vectors. The value of the displaced frame difference (DFD) penalized by the ℓ_1 norm is displayed in the background of the distributions.

like dense linear combination of candidates, fitting of a statistical distribution or clustering approaches are then not recommended.

Secondly, the relation between the DFD and the true motion vector also does not follow a general rule. It can constitute a relevant information to disambiguate complex distributions, as in Figures 2(c) and 2(f), where the ground-truth motion vector falls in regions with low values of the DFD (dark values in the background of the distribution). However, following the lowest DFD can sometimes be misleading, as it is in the case in the other figures, and it cannot be used as a unique estimation criterion.

Thirdly, cases where the true motion vector cannot be retrieved from the distribution or from low values of the DFD also occur in practice, and are illustrated

by Figures 2(b) and 2(e). To handle this situation, a third information must be introduced, which can take the form of an *a priori* smoothness assumption.

To summarize, we have identified three sources of information to guide the design of the aggregation model: the candidate distribution, a data fitting constraint, and a smoothness assumption. These constraints are complementary and only valid locally. They should be incorporated jointly in the aggregation model in a spatially adaptive way.

3.1.2 Minimum distance

In addition to the qualitative analysis of the modeling aspects of the aggregation given in the previous section, another requirement that we derived from the analysis of the BCF in Section 2.3 is the selection of a single candidate at each pixel. To achieve this goal, we could define $\rho(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}))$ as the distance to the closest element of $\boldsymbol{\mathcal{C}}_F(\boldsymbol{x})$:

$$\rho_{\min}(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x})) = \min_{i\in\{1,\cdots,M(\boldsymbol{x})\}} \|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p, \tag{7}$$

where $\boldsymbol{w}_i(\boldsymbol{x})$ is a motion candidate, $M(\boldsymbol{x})$ is the number of candidates at pixel $\boldsymbol{x}$, and $p \in \{1, 2\}$. The min function naturally selects one candidate used for distance measure. The proximal operator of $\rho(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}))$ can be computed exactly and the resulting energy can then be minimized in a proximal splitting framework [24]. However, the problem of potential (7) lies in its high non-convexity, leading inevitably to local minima. In practice, we experimentally observe that the algorithm converges but stays trapped in a local minimum which is very dependent on the initialization. Thus, we have to design a model that enforces the selection of a single candidate while relaxing the non-convexity of the min function (7) to facilitate minimization.

3.2 Aggregation model

To this end, we introduce an additional variable $\boldsymbol{\alpha}(\boldsymbol{x}) = \{\alpha_i(\boldsymbol{x})\}_{i=1,..,M(\boldsymbol{x})}$, weighting the contribution of each candidate. The fidelity term is then expressed as

$$\rho(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}), \boldsymbol{\alpha}(\boldsymbol{x})) = \sum_{i=1}^{M(\boldsymbol{x})} \alpha_i(\boldsymbol{x}) \|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p. \tag{8}$$

To ensure that only one candidate is selected, the weight vector $\boldsymbol{\alpha}(\boldsymbol{x})$ should be constrained to have binary values with only one non-zero element. To achieve this goal, we follow [66,65] and point the following property: if the problem $\arg\min_i \|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p$ has a unique solution $\hat{\imath}$, then the solution of the problem

$$\min_{\boldsymbol{\alpha}(\boldsymbol{x})} \sum_{i=1}^{M(\boldsymbol{x})} \alpha_i(\boldsymbol{x}) \|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p, \tag{9}$$
$$\text{s.t. } \begin{cases} \sum_{i=1}^{M(\boldsymbol{x})} \alpha_i(\boldsymbol{x}) = 1 \\ \forall i \in \{1, \dots, M(\boldsymbol{x})\}, \alpha_i(\boldsymbol{x}) \geq 0, \end{cases}$$

is $\rho_{\min}(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}))$ defined in (7), and is attained for $\alpha_{\hat{\imath}}(\boldsymbol{x}) = 1$ and $\alpha_j(\boldsymbol{x}) = 0, \forall j \neq \hat{\imath}$. The case where several coefficients are non-zero can only occur if the solution of $\arg\min_i \|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p$ is a non-singleton set $\mathcal{S}$. In that case, the non-zero coefficients are $\{\alpha_i\}_{i\in\mathcal{S}}$ and can take any configuration satisfying the constraints of (9). We observed that this situation rarely occurs in practice. The formulation (9) is convex w.r.t. to $\boldsymbol{w}$ and thus offers an algorithmically tractable alternative to the min function, while reproducing its behavior.

The fidelity term (8) relies only on the candidate distribution to guide the selection of a candidate. As mentioned in Section 3.1.1, purely distribution-driven estimation is insufficient to handle certain situations and should be complemented with a data-driven constraint. We exploit pre-computed confidence measures $\beta_i(\boldsymbol{x})$ associated to each candidate $\boldsymbol{w}_i(\boldsymbol{x})$. The fidelity term is then enriched by defining

$$\rho(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\mathcal{C}}_F(\boldsymbol{x}), \boldsymbol{\alpha}(\boldsymbol{x})) = \sum_{i=1}^{M(\boldsymbol{x})} \alpha_i(\boldsymbol{x}) \left(\|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p + \lambda_2 \beta_i(\boldsymbol{x}) \right). \tag{10}$$

where $\lambda_2 > 0$ is a balance parameter. The confidence measure reflects a feature constancy assumption, e.g. based on the DFD analysed in Figure 2. Low values of $\beta_i(\boldsymbol{x})$ correspond to high confidence and promote high value of $\alpha_i(\boldsymbol{x})$, such that the similarity to a distribution mode imposed by $\|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p$ is balanced with a data fitting constraint imposed by the confidence term. Apart from [51,52], existing confidence measures are dedicated to specific motion estimation methods. For a variational approach, [21] uses the inverse of the global energy. For local approaches like [54], eigenvalues of the structure tensor are usually exploited [59]. 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 simple weights based on a filtering of the DFD:

$$\beta_i(\boldsymbol{x}) = \frac{1}{Z} \int_\Omega g(\boldsymbol{x} - \boldsymbol{y}) |v(\boldsymbol{y} + \boldsymbol{w}_i(\boldsymbol{x})) - u(\boldsymbol{y})| d\boldsymbol{y}, \tag{11}$$

where g is a convolution kernel and $Z = \sum_{j=1}^{M(\boldsymbol{x})} \int_\Omega g(\boldsymbol{x} - \boldsymbol{y}) |v(\boldsymbol{y} + \boldsymbol{w}_j(\boldsymbol{x})) - u(\boldsymbol{y})| d\boldsymbol{y}$.

The analysis of Section 3.1.1 also revealed the necessity to introduce a smoothness assumption on the motion field. We complete the model with a standard Total Variation regularization to come up with the final optimisation problem

$$\min_{\boldsymbol{w},\boldsymbol{\alpha}} \left\{ \int_\Omega \sum_{i=1}^{N} \alpha_i(\boldsymbol{x}) \left(\|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_p^p + \lambda_2 \beta_i(\boldsymbol{x}) \right) + \lambda_1 \|\nabla \boldsymbol{w}(\boldsymbol{x})\|_1 d\boldsymbol{x} \right\}, \quad (12)$$
$$\text{s.t.} \begin{cases} \sum_{i=1}^{M(\boldsymbol{x})} \alpha_i(\boldsymbol{x}) = 1 \\ \forall i \in \{1,\dots,M(\boldsymbol{x})\}, \alpha_i(\boldsymbol{x}) \geq 0. \end{cases}$$

This model fulfills the modeling criteria identified in Section 3.1. Minimization w.r.t. $\boldsymbol{\alpha}$ enforces the selection of a single candidate at each pixel. The three terms of (12) combine similarity to the distribution, data-driven constraint and smoothness assumption. A key advantage of this formulation is that, differently from usual approaches based on non-linear feature conservation assumption, the optimization problem (12) can be solved without any linearization w.r.t $\boldsymbol{w}$. As a result, it does not impose coarse-to-fine optimization strategies with successive linearizations at each level. Moreover, if good motion candidates have been found at occluded pixels (see Section 2.2), this data term provides a valid measure even at occlusions. It is worth noting that $p = 1$ enables more deviations from the candidates in case of lack of good candidates or locally wrong candidate selection.

In [37], we proposed a related model in a sparse representation framework, where the number of selected candidates was controlled by a sparsity constraint on $\boldsymbol{\alpha}$. The confidence measures were associated to the sparsity constraint with a weighted ℓ_1 penalization function. The drawback of this approach comes from the coupling of the sparsity constraint and the confidence measures: to ensure the selection of a single candidate, the parameter weighting the sparsity constraint has to be very high, which also gives large weight to the confidence measures. As a result, the weighted ℓ_1 term becomes predominant and the estimation is mainly driven by the confidence measures. In the model (12), the selection of a candidate is decoupled from the influence of confidence measures.

3.2.1 Optimization

The minimization subproblems w.r.t. $\boldsymbol{w}$ and $\boldsymbol{\alpha}$ being both convex, we resort to a block-coordinate approach alternating updates of the two variables.

Minimization w.r.t. $\boldsymbol{w}$ The minimum of (12) w.r.t. $\boldsymbol{w}$ is obtained by solving the Euler-Lagrange equations. For simplicity, we consider $p = 1$ in this section. The algorithm is almost equivalent for $p = 2$. We approximate the vectorial ℓ_1 norm by a differentiable relaxation such that $\|\boldsymbol{z}\|_1 \approx \psi(\|\boldsymbol{z}\|_2^2) = \sqrt{\|\boldsymbol{z}\|_2^2 + \epsilon^2}$, with ϵ a small constant that we fix to 0.001. Under this assumption, the Euler-Lagrange equations at a given pixel $\boldsymbol{x}$ can be written:

$$\sum_{i=1}^{M(\boldsymbol{x})} \psi'(\|\boldsymbol{w}(\boldsymbol{x}) - \boldsymbol{w}_i(\boldsymbol{x})\|_2^2)\, \alpha_i \left(v_j(\boldsymbol{x}) - [\boldsymbol{w}_i(\boldsymbol{x})]_j\right) - \lambda_1 \mathrm{div}\left(\psi'(\|\nabla \boldsymbol{w}(\boldsymbol{x})\|_2^2)\, \nabla v_j(\boldsymbol{x})\right) = 0 \quad (13)$$

where $j = \{1, 2\}$ and $[\cdot]_j$ denotes the j^{th} component of a vector. Using standard forward finite differences for the discretization of the gradient operator, equations (13) yield a non-linear system of equations, where the nonlinearity is due to the terms in $\psi'(\cdot)$. We solve this system with the lagged nonlinearity method [18,81]. It consits in fixing in an inner loop the nonlinear parts of (13), and iterating linear system solving and nonlinearity update until convergence.

Minimization w.r.t. $\boldsymbol{\alpha}$ We solve the constrained optimization problem w.r.t. $\boldsymbol{\alpha}$ with an Augmented Lagrangian approach [16,88]. To facilitate readability, we omit the arguments in $\boldsymbol{x}$ in this section. The positivity constraint is handled with the indicator function $\iota_{\mathbb{R}_+^M}$ defined as

$$\iota_{\mathbb{R}_+^M}(\boldsymbol{z}) = \begin{cases} 0, & \text{if } \boldsymbol{z} \in \mathbb{R}_+^M \\ +\infty, & \text{else,} \end{cases} \quad (14)$$

which leads to the following problem:

$$\min_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i \left(\|\boldsymbol{w} - \boldsymbol{w}_i\|_p^p + \lambda_2 \beta_i \right) + \iota_{\mathbb{R}_+^M}(\boldsymbol{\alpha}), \quad \text{s.t. } \sum_{i=1}^{M} \alpha_i = 1. \quad (15)$$

We reformulate (15) by introducing a splitting variable $\boldsymbol{z}$ associated to the indicator function:

$$\min_{\boldsymbol{\alpha},\boldsymbol{z}} \sum_{i=1}^{N} \alpha_i \left(\|\boldsymbol{w} - \boldsymbol{w}_i\|_p^p + \lambda_2 \beta_i \right) + \iota_{\mathbb{R}_+^M}(\boldsymbol{z}), \quad \text{s.t.} \begin{cases} \boldsymbol{z} = \boldsymbol{\alpha} \\ \sum_{i=1}^{M} \alpha_i = 1. \end{cases} \quad (16)$$

The scaled form of the Augmented Lagrangian of problem (16) writes

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \sum_{i=1}^{N} \alpha_i \left(\|\boldsymbol{w} - \boldsymbol{w}_i\|_p^p + \lambda_2 \beta_i \right) + \iota_{\mathbb{R}_+^M}(\boldsymbol{z}) + \frac{\mu_1}{2} \left\| \sum_{i=1}^{M} \alpha_i - 1 + \frac{\boldsymbol{\rho}_1}{\mu_1} \right\|_2^2 + \frac{\mu_2}{2} \left\| -\boldsymbol{\alpha} + \boldsymbol{z} + \frac{\boldsymbol{\rho}_2}{\mu_2} \right\|_2^2, \quad (17)$$

where $\boldsymbol{\rho}_1 \in \mathbb{R}$, $\boldsymbol{\rho}_2 \in \mathbb{R}^M$ are Lagrange multipliers and μ_1, μ_2 are positive penalty parameters. We use the alternated direction method of multipliers (ADMM), which separates optimization subproblems w.r.t. each variable to converge to the solution of the original problem (15). Each iteration k is composed of the following steps:

$$\boldsymbol{\alpha}^{k+1} = \arg\min_{\boldsymbol{\alpha}} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{z}^k, \boldsymbol{\rho}_1^k, \boldsymbol{\rho}_2^k), \tag{18}$$

$$\boldsymbol{z}^{k+1} = \arg\min_{\boldsymbol{z}} \mathcal{L}(\boldsymbol{\alpha}^{k+1}, \boldsymbol{z}, \boldsymbol{\rho}_1^k, \boldsymbol{\rho}_2^k), \tag{19}$$

$$\boldsymbol{\rho}_1^{k+1} = \boldsymbol{\rho}_1^k + \mu_1 \left(\sum_{i=1}^{N} \alpha_i^{k+1} - 1 \right), \tag{20}$$

$$\boldsymbol{\rho}_2^{k+1} = \boldsymbol{\rho}_2^k + \mu_2(-\boldsymbol{\alpha}^{k+1} + \boldsymbol{z}^{k+1}). \tag{21}$$

Minimization problems (18) and (19) have analytical and efficiently computable solutions. The solution of (18) is given by

$$\boldsymbol{\alpha}^{k+1} = \left(\mu_2 \boldsymbol{I} + \mu_1 \mathbf{1}_M \mathbf{1}_M^\top\right)^{-1} \left(\mu_1 M \left(1 - \frac{\boldsymbol{\rho}_1^k}{\mu_1} \right) + \mu_2 \left(\boldsymbol{z}^k + \frac{\boldsymbol{\rho}_2^k}{\mu_2} \right) - \boldsymbol{b}_{\lambda_2} \right), \tag{22}$$

where we define the M-dimensional vector $\mathbf{1}_M = (1, \ldots, 1)^\top$ and the vector $\boldsymbol{b}_{\lambda_2} = (b_1, \ldots, b_M)^\top$ with $b_i = p\|\boldsymbol{w} - \boldsymbol{w}_i\|_p^p + \lambda_2 \beta_i$. The matrix inversion can be easily achieved with the Sherman-Morrison formula. The update of $\boldsymbol{z}$ is a simple projection onto the set $\mathbb{R}_+^M$ given by

$$\boldsymbol{z}^{k+1} = \max\left(\boldsymbol{\alpha}^{k+1} - \frac{\boldsymbol{\rho}_2^k}{\mu_2}, 0 \right). \tag{23}$$

We emphasize that the positivity and normalization constraints of the original problem (12) define a convex set for which efficient projection can be computed, e.g. using [26] as proposed in [22]. However, we experimentally observed that the decoupling of the constraints yielding faster minimization subproblems in the augmented Lagrangian framework described above yielded similar results with a significantly lower computational time.

4 Experimental results

In this section, we analyze the performance of our *VAFlow* (Variational Aggregation for optical Flow) aggregation method. We highlight the versatility of *VAFlow* by dealing with various issues: large displacements, occlusions, motion discontinuities, noise in input images, and suboptimal candidates set. We also quantitatively demonstrate its superiority over local parametric methods and classical variational approaches on the Middlebury benchmark.

4.1 Experimental protocol

Evaluation metric When ground truth is available, we use the standard error metric for optical flow evaluation, which is the averaged endpoint error (EPE). It is defined as the average of euclidean distances at each pixel between the estimated motion vector and the ground truth.

Implementation details The feature matching steps involved in the candidates computation (Sections 2.1.1 and 2.2) are implemented with the available code of the *PatchMatch* algorithm [7][2]. To achieve robustness to illumination changes, we consider a combination of saturation and value channels of the HSV color space, following [89]. The distance to minimize with *PatchMatch* is the sum of absolute differences (SAD) of patches. The distance between pixel $\boldsymbol{x}$ of image u and pixel $\boldsymbol{y}$ of image v is then defined for a patch size h by $SAD(\boldsymbol{x}, \boldsymbol{y}, h) = \|\boldsymbol{p}_u(\boldsymbol{x}, h) - \boldsymbol{p}_v(\boldsymbol{y}, h)\|_1$.

The affine motion estimation involved in the candidates computation step (3) is solved with the publicly available Motion2D software[3] [63], which implements a multi-resolution incremental minimization scheme based on the Iteratively Reweighted Least Squares (IRLS).

The occlusion detection required to extend the motion candidates in occluded regions (Section 2.2) is performed with a simple approach exploiting motion candidates computation. A coarse motion estimation is performed by block matching using *PatchMatch* with the smallest patch size ($h = 15$). The backward flow is then computed and a standard forward/backward consistency criterion [47,60] yields a coarse occlusion detection. More sophisticated methods could give more accurate occlusion regions and improve results.

In the optimization procedure described in Section 3.2.1, the motion field $\boldsymbol{w}$ is initialized by selecting at each pixel the motion candidate with best confidence measure (11). The weights $\boldsymbol{\alpha}$ are initialized by setting the weight corresponding to the best confidence measure to one and all the others to zero.

No post-processing is applied on the flow fields. The candidates sets were obtained with parameters $\boldsymbol{\mathcal{H}} = \{15, 35, 75\}$, $r = 0.75$, $K = 2$ (the typical number of candidates per pixel with these parameter values is around 100). The convolution filter g involved in the definition of the confidence measures in (11) is a rect function of size 5×5 pixels, which amounts to the SAD distance measure defined above and used in the patch matching step. The value of λ_2 is set to 15. Convergence of the ADMM optimization of $\boldsymbol{\alpha}$ has been observed to

[2] http://gfx.cs.princeton.edu/pubs/Barnes_2009_PAR/index.php
[3] http://www.irisa.fr/vista/Motion2D/

be reached for 500 iterations. To save computational cost, we set a maximum number of 100 iterations, which has a limited impact on the final results. The penalty parameters are set to $\mu_1 = 10$ and $\mu_2 = 10$. The number of alternate optimizations in the global minimization of (12) was 4, for which convergence has been experimentally observed. The number of outer iterations involved in the fixed point iteration scheme of the minimization w.r.t. $\boldsymbol{w}$ is 15, and the linear systems are solved with the successive over-relaxation method [18].

Methods exploited for comparison The candidates of *VAFlow* are obtained with parametric estimations. Thus, the comparison with local parametric methods [54,63] is informative about the efficiency of the aggregation step. In the following we will refer to "*multiscale* [54]" as the coarse-to-fine implementation of [54] described in [15], and to "*multiscale* [63]" as an extension of "*multiscale* [54]" performing the robust affine estimation described in [63] in each patch. The results of [54] are obtained with the publicly available implementation[4], and we use the Motion2D software[5] to apply the method [63]. The method we call "*block matching* [63]" mimics the candidates generation procedure of *VAFlow*. At each pixel, an initial block matching is performed and is followed by a parametric refinement between corresponding patches. Only the motion of the center pixel of the patch is kept.

As state-of-the-art results are achieved with global variational approaches, we also compare to the methods of [18] and [24] providing open access softwares[6,7], which implement TV-l_1 models with different optimization strategies. We also consider the method [19] and use the code made available by the authors[8]. It extends [18,24] with an additional energy term imposing similarity to pre-computed feature matches. It aims at reducing the undesirable effects of the coarse-to-fine scheme. Current top performing methods [67,68,83,86] rely on the baseline principles of [18,19,24], on which they elaborate more sophisticated modules like efficient feature matching, or non-local regularization. In this paper, we propose a baseline version of our continuous aggregation concept, with simple block matching and TV regularization. Therefore, we compare it with methods [18,19,24] using the same basic ingredients. More sophisticated features could be integrated as well in our method to still improve results in the future.

[4] http://www.mathworks.com/matlabcentral/fileexchange/23142-iterative-pyramidal-lk-optical-flow

[5] http://www.irisa.fr/vista/Motion2D/

[6] http://lmb.informatik.uni-freiburg.de/resources/software.php

[7] http://gpu4vision.icg.tugraz.at/index.php?content=downloads.php

[8] http://lmb.informatik.uni-freiburg.de/resources/software.php

Table 2 Average EPE results on the Middlebury benchmark for $p = 1$ and $p = 2$.

	$p = 2$	$p = 1$
Average EPE on Middlebury	0.415	0.284

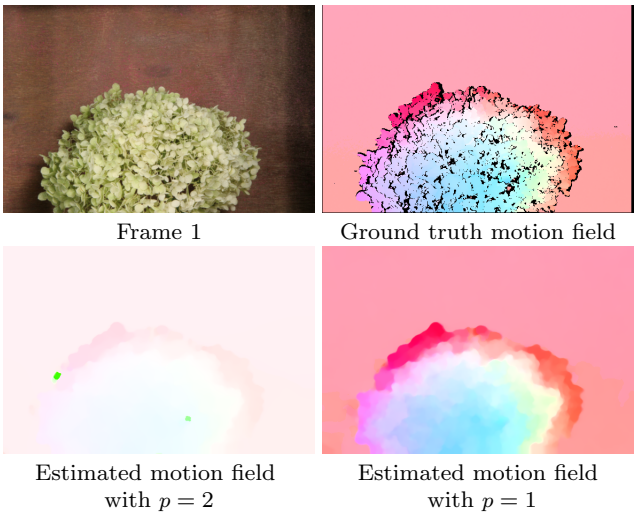


Fig. 3 Illustration of of the impact of the choice of ℓ_p for the data fidelity term.

Finally, we also compare *VAFlow* with the discrete optimization approach we introduced in [38]. We remove the exemplar-based aggregation term and post-processing of [38] to compare only the baseline aggregation methods. We refer to this method as *Discrete Aggregation*.

4.2 Results

Choice of the ℓ_p norm We first point out the importance of the choice of p in the ℓ_p norm promoting similarity to the selected candidate in (12). Table 2 gives the Average EPE obtained on the Middlebury dataset with ground truth for $p = 2$ and $p = 1$, and Figure 3 illustrates these results on an example. Choosing $p = 1$ yields robustness in the similarity constraint to the chosen candidate, such that few large differences between estimated motion vectors and the chosen candidate are allowed. This is a desirable property in case of locally wrong candidate selection. In Figure 3, the result with $p = 2$ contains two regions of large errors where the candidate selection was not optimal, whereas with $p = 1$, these outliers are properly handled. Few large errors could have a significant impact on the average EPE, as it can be seen in Table 2. In the light of these results, we will take $p = 1$ in the rest of the experiments.

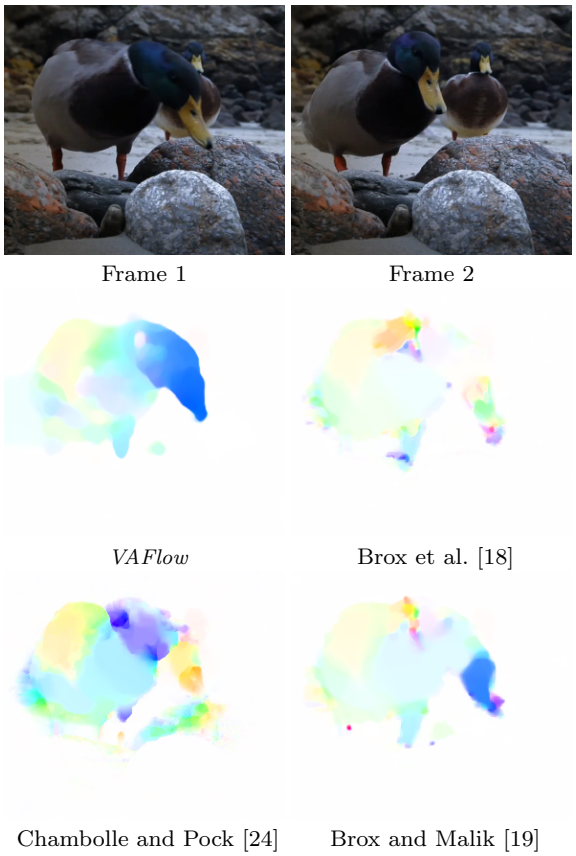


Fig. 4 Estimated motion fields with *VAFlow* and [18],[24],[19] on the *Bird* sequence.

Large displacements of small objects One of the main limitations of coarse-to-fine schemes arises in case of large displacements of small objects, as illustrated on real sequences without ground truth in Figures 4,5 and 6. The results supplied by [18,24] are typical examples of failures due to coarse-to-fine schemes, which prevent here from satisfyingly recovering the duck head in Fig. 4, the ball in Fig. 5 and the foot in Fig. 6. In contrast, *VAFlow* estimates correctly all these large displacements. In most cases, [19] also captures these movements, but at the same time, it generates large errors in other parts of the image. This is due to its high sensitivity to feature matching errors, which is better handled by VAFlow.

Motion details and discontinuities Motion details like the legs of the girl in Fig. 5 and the duck legs in Fig. 4 are better preserved by *VAFlow* compared to the three competing methods. In Fig. 6, the discontinuities of the motion field supplied by *VAFlow* are sharper and delineate better the leg and the foot of the football player.

Occlusion handling When the large displacements concern large parts of the image, occlusions become a prominent issue, as illustrated in the three image pairs of Fig. 7. To demonstrate the effect of our occlusion handling, we desactivate the occlusion handling module (*VAFlow w/o occlusions*) in the motion candidate generation step, and compare the results with those obtained by the full *VAFlow* method. In each case, *VAFlow w/o occlusion* still captures well large displacements, but it also exhibits large errors at occluded pixels, due to the absence of good candidates. When occlusion handling is activated, the result is visually greatly improved in these regions and is very close to the ground truth. This observation is confirmed by the large decrease of the EPE in each case (also reported in Fig. 7).

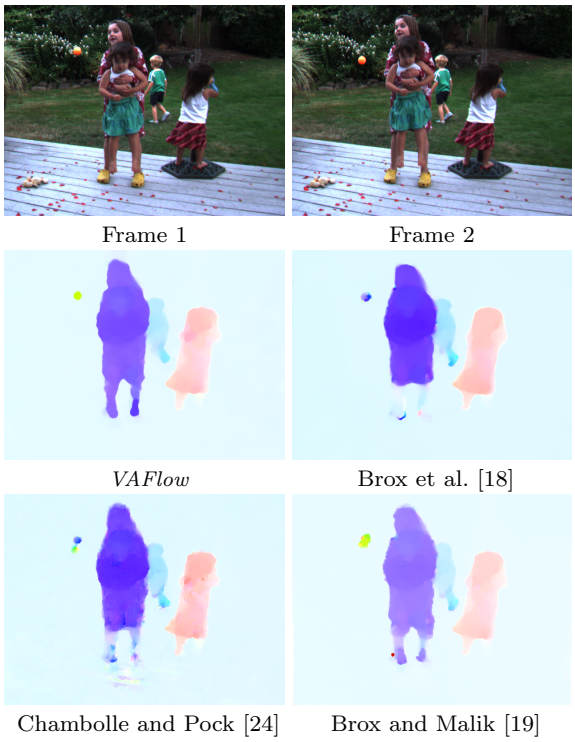


Fig. 5 Estimated motion fields with *VAFlow* and [18],[24],[19] on the *Backyard* sequence of the Middlebury dataset.

Quantitative evaluation We provide a quantitative evaluation in Table 3, reporting the EPE obtained with *VAFlow*, local approaches [54,63], and variational methods [18,19,24] for the sequences of the Middlebury dataset with ground truth. The candidates of *VAFlow* are computed by local methods. In particular, they are obtained with the same estimation procedure as *block matching* [63]. Therefore, the large improvement offered by *VAFlow* on these methods is due to the efficiency of the aggregation step, which is able to select the best motion candidate rather than just keeping the motion estimate at the central point of each patch. *VAFlow* also outperforms the global variational approaches [18, 19,24] on almost all the sequences.

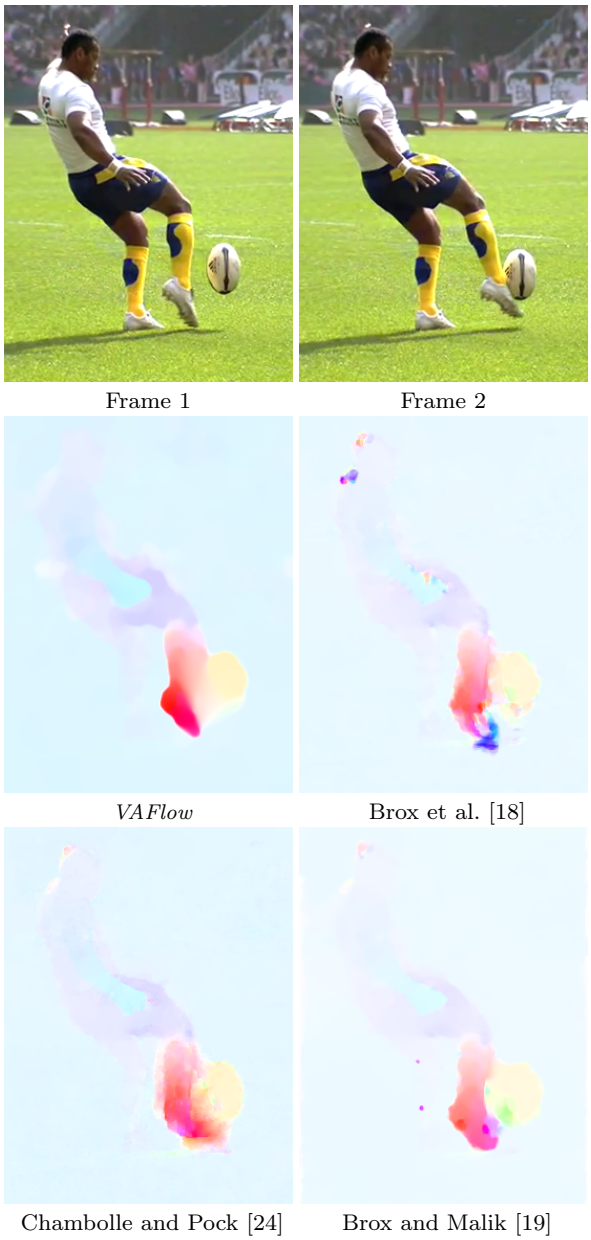


Fig. 6 Estimated motion fields with VAFlow and methods [18],[24],[19] on the *Football* sequence.

In Table 4, we report results obtained on the MPI Sintel training dataset [23], characterized by the presence of sequences with very large displacements. We give the average error on the whole benchmark, and we also give average errors obtained on the seven sequences with the largest displacements. The advantage between of *VAFlow* over the other methods is larger than in Table 3, which confirms the ability of our aggregation strategy to handle large displacements, in particular compared to the integration of feature matching as an additional constraint in a classical variational approach [19]. We mention some recent methods like

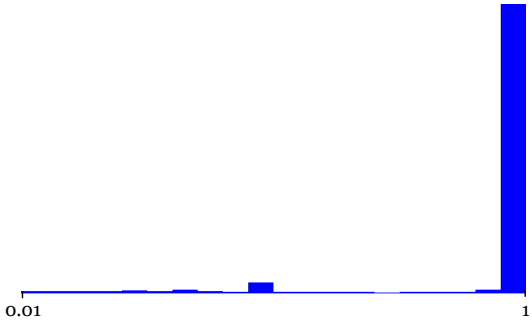


Fig. 8 Average distribution of the coefficients $\boldsymbol{\alpha}$ on the *temple_2* sequence of the MPI Sintel benchmark (50 frames). Only coefficients greater than 0.01 are displayed. 98.7% of the coefficients are below 0.01.

[4] and [57] are able to outperform these results with an average endpoint error of respectively 2.61 and 2.25 on the whole MPI Sintel training dataset. However, as explained in Section 4.1, these methods exploit sophisticated modules that could be integrated in our framework. For instance, the contributions of [4] and [57] are coarse feature matching methods that have to be refined with the variational method [68], which integrates a sophisticated edge detection to sharpen motion discontinuities. These ingredients could be incorporated in our aggregation model to improve results and compete on state-of-the-art computer vision benchmarks. Our primary aim is to propose a general aggregation framework for motion estimation.

We also report in Table 3 and Table 4 results obtained by selecting at each pixel the candidate with the best confidence measure (the lowest value of $\beta_i(\boldsymbol{x})$), which we refer to as "Best confidence flow". The results are always significantly worse than those of *VAFlow*. It demonstrates that the motion estimation with *VAFlow* is not over-guided by the confidence measures and can deviate from them to improve global accuracy of the motion field.

Values of the selection weights $\boldsymbol{\alpha}$ The averaged final estimation of $\boldsymbol{\alpha}$ obtained for the whole sequence *temple_2* of the MPI Sintel dataset (50 frames) is illustrated in Figure 8. 98.7% of the coefficients are lower than 0.01 and are considered to have no siginificant influence on the final results. Therefore, only coefficients superior to 0.01 are displayed in Figure 8. Among coefficients greater than 0.01, 96% are greater than 0.95, which confirms that the algorithm selects only one candidate for the reconstruction most of the time. In that sense, our method finds the sparsest solution in most cases.

Robustness to noise Existing optical flow benchmarks do not integrate robustness to noise as an evaluation criterion. However, it is common to deal with noisy images

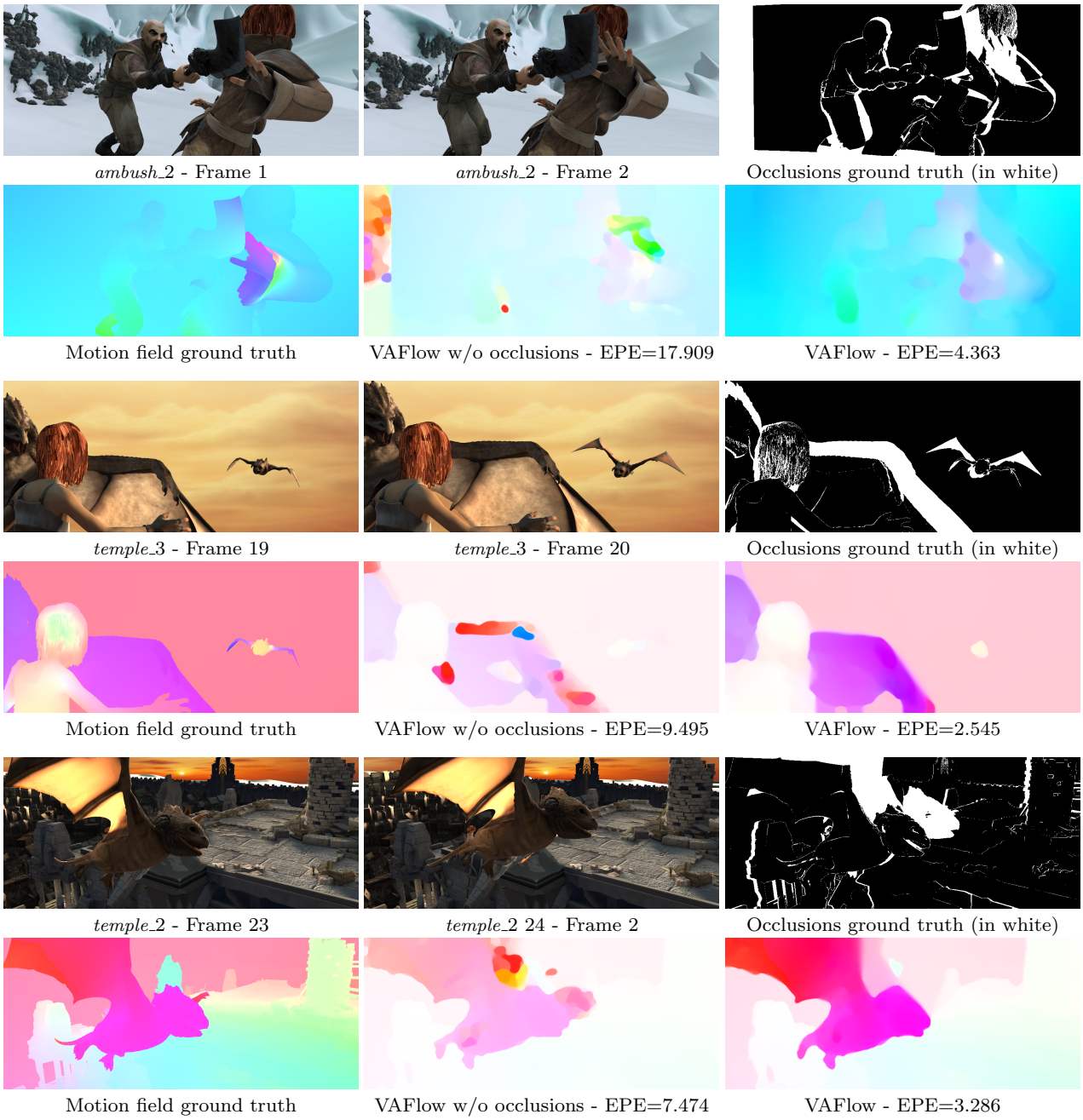


Fig. 7 Comparison of motion fields computed with *VAFlow* and with *VAFlow* without the occlusion handling module (w/o occlusions). Results are obtained on sequences of the MPI Sintel dataset [23] with large displacements. The EPE of each result is given in the captions attached to the motion fields.

when specific optical devices are used, as in microscopy or astronomy.

VAFlow performs patch-based parametric motion estimation, in the candidates generation step. The aggregation step (motion reconstruction) does not exploit any pixel wise feature conservation assumption, but only uses a patch-based confidence measure. Parametric estimations in patches [54,63] are known to be more robust to noise than global variational methods. Therefore, we expect *VAFlow* to provide with robustness to noise while ensuring of the accuracy global variational methods in the absence of noise, as previously demonstrated in Table 3.

In Fig. 9, we plot the average EPE for Middlebury sequences with ground truth after adding Gaussian noise to the input images with different standard deviations. The results supplied by *VAFlow* are compared with those of [18,24] in Fig. 9.a and with [19]

Table 3 Endpoints error obtained with *VAFlow*, the local methods [54,63] and the variational methods [18,24,19] on the Middlebury dataset with ground truth.

	Grove2	*Grove3*	*Urban2*	*Urban3*	*Venus*	*RubberWhale*	*Dimetrodon*	*Hydrangea*	Average
Best confidence flow	0.324	1.203	1.885	2.002	1.510	0.134	0.172	0.506	0.928
VAFlow	**0.161**	**0.630**	0.374	**0.395**	**0.298**	0.134	**0.090**	0.194	**0.284**
Local methods									
multiscale [54]	0.670	1.871	2.603	3.144	1.646	0.476	0.638	0.896	1.493
multiscale [63]	0.461	1.347	1.570	1.611	0.859	0.409	0.249	0.627	0.892
block matching [63]	0.437	1.362	1.512	1.766	1.678	0.448	0.241	0.571	1.002
Global methods									
Brox et al. [18]	0.184	0.724	0.420	1.044	0.484	0.138	0.175	**0.177**	0.358
Chambolle and Pock [24]	0.193	0.645	0.353	0.559	0.351	0.132	0.178	0.219	0.329
Brox and Malik [19]	0.176	0.680	**0.343**	0.586	0.402	**0.116**	0.100	0.198	0.325

Table 4 Endpoints error obtained with *VAFlow*, the local methods [54,63] and the variational methods [18,24,19] on the MPI Sintel dataset with ground truth. The last column gives the average result on the whole data set, and results on the seven sequences with the largest displacements are also given.

Sequences with large displacements	*cave_2*	*market_6*	*temple_3*	*ambush_5*	*ambush_6*	*ambush_2*	*market_5*	Average on whole benchmark
Best confidence flow	22.02	12.95	21.19	16.86	25.90	26.13	38.87	11.19
VAFlow	**7.99**	**4.82**	**8.74**	**6.34**	**7.86**	**10.17**	**11.79**	**3.90**
Brox et al. [18]	27.54	7.30	15.84	12.72	15.44	34.94	23.07	7.31
Chambolle and Pock [24]	25.01	8.55	21.43	12.22	16.07	35.67	23.74	7.91
Brox and Malik [19]	9.20	5.61	14.67	10.90	11.11	20.73	14.98	5.03

in Fig. 9.b. The impact of noise is significantly lower on the performance of *VAFlow* than on those of [18, 24]. The difference is even more pronounced between *VAFlow* and [19], which is due to the high sensitivity of [19] to wrong feature matches, as already emphasized in previous results.

Suboptimal candidates set The final output of *VAFlow* is dependent on the quality of the motion candidates. More patches should be considered to augment the variety of candidates. A crucial parameter is the overlap ratio $r \in [0,1]$, defining the amount of common area shared by two neighbor patches. When r is close to one, there are as many patches as pixels for a given patch size, and the number of candidates the is highest. However, the number of patches also increases the computation time, such that a trade-off has to be found between accuracy and complexity. The impact of the overlap ratio on these two aspects is reported in Table 5, which summarizes the evolution with r of the average EPE on the Middlebury benchmark sequences with ground truth on and the computational time on the *Urban_2* sequence of the Middlebury dataset. While the computation time increases slowly when r is small, it changes much faster when $r > 0.5$. In the same time the error increase remains relatively limited for $r > 0.5$.

Table 5 Evolution with the overlap ratio r of the Average EPE on the Middlebury dataset with ground truth and the computational time on the Urban_2 sequence of the Middlebury dataset.

Overlap ratio	0.75	0.5	0.25	0.1
Average EPE	0.296	0.310	0.329	0.354
Computation time (s)	305	142	117	111

This robustness to suboptimal candidates sets is further emphasized by the visual results of Figure 10, where we can observe that the results stay very similar when r decreases, in particular when $r > 0.5$. In practical scenarios where computational time matters, this robustness can allow us to make huge gains in complexity without loosing too much accuracy.

Comparision with discrete optimization [38] We focus now on the comparison between the variational aggregation scheme of *VAFlow* and the aggregation based on discrete optimization described in [38], that we call *Discrete Aggregation*. Table 6 reports the EPE obtained on sequences of the Middlebury and MPI Sintel datasets with ground truth by *VAFlow* and *Discrete Aggregation*. Results supplied by *Discrete Aggregation* are in

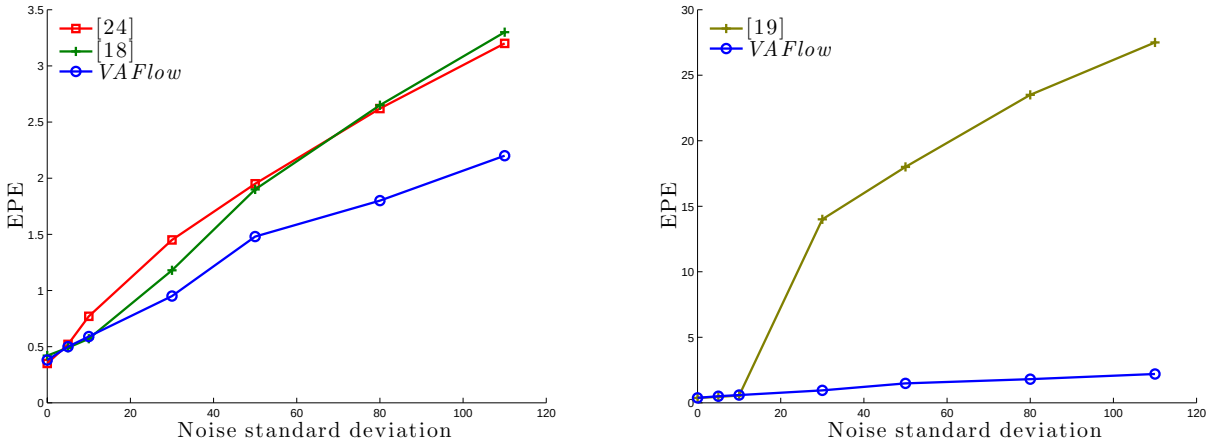


Fig. 9 Evolution of the EPE with the standard deviation of the added Gaussian noise in the input images. The reported EPE is the average EPE over all the sequences of the Middlebury dataset with ground truth. Fig. 9a compares *VAFlow* with [18, 24] and Fig. 9b compares *VAFlow* with [19].

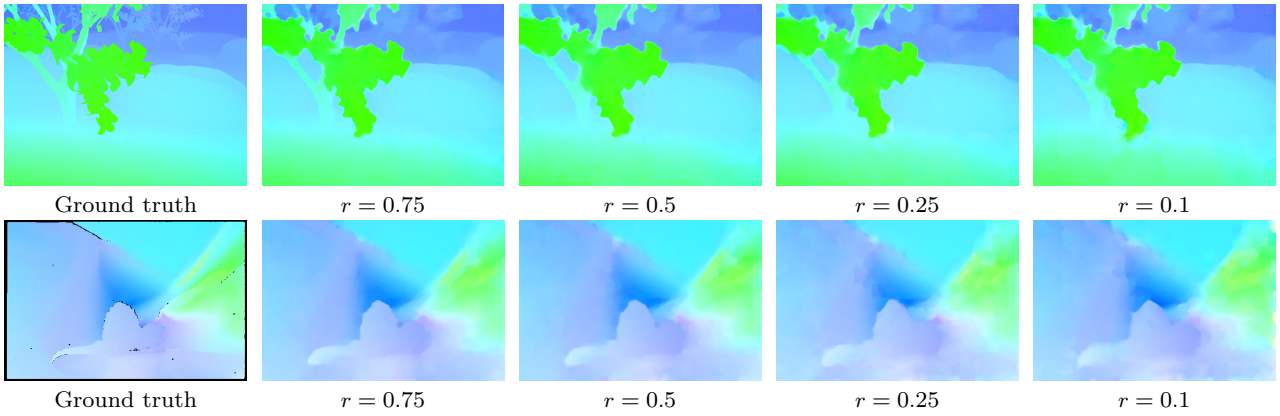


Fig. 10 Visual evaluation of the impact of the overlap ratio r on the results of *VAFlow*, for the *Grove*2 (top) and *Dimetrodon* (bottom) sequences of the Middlebury dataset.

general slightly more accurate than those of *VAFlow*. However, the advantage of *VAFlow* lies in its robustness to suboptimal candidate sets and its computation time. Figure 11 compares the impact of the overlap ratio on the EPE and the computation time. While the EPE of *Discrete Aggregation* is lower for a large overlap ratio, the results of *VAFlow* are less impacted by a lower quality of the candidates, and it gives lower EPE when the overlap ratio is approximately below 0.45. In the same time, the computation time of *Discrete Aggregation* increases faster than the one of *VAFlow* with r. For $r = 0.75$, *Discrete aggregation* is almost two times slower than *VAFlow*.

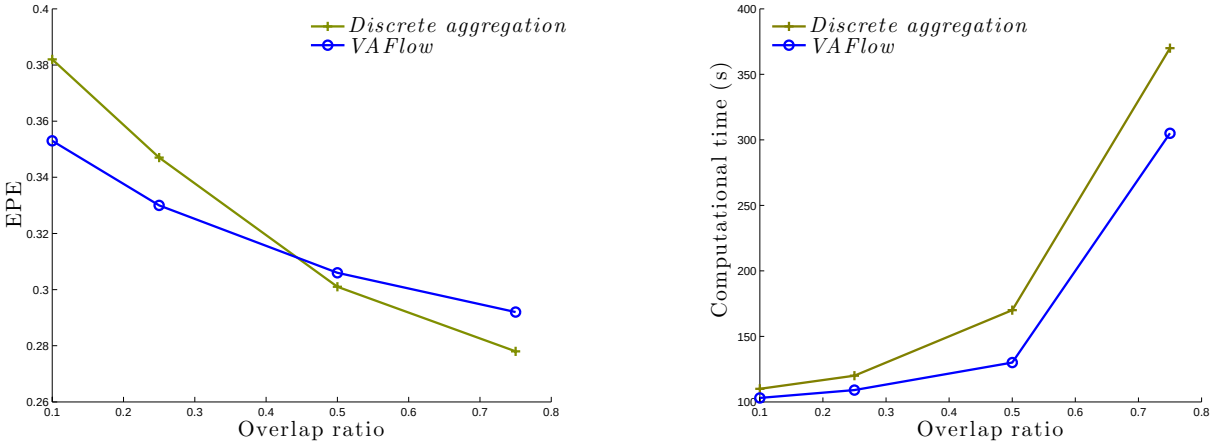
5 Conclusion

We have proposed a variational aggregation framework for optical flow estimation based on a sparse representation of the motion field. We combine in two successive steps local parametric estimation yielding motion candidates, and global aggregation supplying the global recovered flow. We formulated the aggregation step as a global energy minimization problem without coarse-to-fine strategy, combining the best motion candidates at every pixel while preserving motion discontinuities. We promoted sparse solutions, that is, the selection at each pixel of a few candidates in space-variant motion vector dictionaries. We handle occlusion with an exemplar-based motion inpainting approach in the candidates computation step. We demonstrated the improvements yielded by our method over standard variational approaches in various situations of large displacements of

Table 6 Endpoints error obtained with *VAFlow* and *Discrete Aggregation* on the sequences with ground truth of the Middlebury and MPI Sintel datasets.

Middleburry	*Grove2*	*Grove3*	*Urban2*	*Urban3*	*Venus*	*RubberWhale*	*Dimetrodon*	*Hydrangea*
VAFlow	**0.161**	0.630	0.374	0.395	0.298	0.134	**0.090**	0.194
Discrete aggregation	0.166	**0.621**	**0.337**	**0.381**	**0.287**	**0.121**	0.122	**0.179**

MPI Sintel	*cave_2*	*market_6*	*temple_3*	*ambush_5*	*ambush_6*	*ambush_2*	*market_5*
VAFlow	**7.99**	4.82	8.74	6.34	7.86	10.17	**11.79**
Discrete aggregation	8.228	**4.547**	**8.314**	**5.50**	**6.251**	**9.456**	11.958



(a) (b)

Fig. 11 Comparison of the behaviour of *VAFlow* and *Discrete Aggregation* w.r.t. the overlap ratio. Fig. (a) show the evolution of the average EPE on the sequences with ground truth of the Middlebury benchmark, and Fig. (b) shows the evolution of the computational time on the *Urban_2* sequence of the Middlebury benchmark.

small objects, occlusions, noise in input images and motion discontinuities. We also achieved a lower computational time and more robustness to suboptimal candidates set compared to the discrete aggregation approach introduced [38]. The framework is generic, and both the local and global steps could be adapted for specific purposes, especially using more sophisticated feature matching techniques.

Acknowledgements This work was realized as part of the Quaero program, funded by OSEO, French State agency for innovation. The authors acknowledge France-BioImaging infrastructure supported by the French National Research Agency (ANR-10-INBS-04-07, "Investments for the future"). They thank also the reviewers for useful comments helping improving the paper. Finally, they thank Ferreol Soulez, Martin Storath, Olivier Demetz, Simon Setzer and Joachim Weickert for inspiring discussions at different stages of this work.

References

1. Alba, A., Arce-Santana, E., Riviera, M.: Optical flow estimation with prior models obtained from phase correlation. Advances in Visual Computing pp. 417–426 (2010)
2. Arias, P., Facciolo, G., Caselles, V., Sapiro, G.: A variational framework for exemplar-based image inpainting. Int. J. of Computer Vision **93**(3), 319–347 (2011)
3. Ayvaci, A., Raptis, M., Soatto, S.: Sparse occlusion detection with optical flow. Int. J. of Computer Vision **97**(3), 322–338 (2012)
4. Bailer, C., Taetz, B., Stricker, D.: Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In: IEEE International Conference on Computer Vision, pp. 4015–4023 (2015)
5. Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., Szeliski, R.: A database and evaluation methodology for optical flow. Int. J. of Computer Vision **92**(1), 1–31 (2011)
6. Bao, L., Yang, Q., Jin, H.: Fast edge-preserving patchmatch for large displacement optical flow. In: Computer Vision and Pattern Recognition (CVPR). Columbus (2014)
7. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. ACM Trans. On Graphics **28**(3), 24 (2009)
8. Barron, J., Fleet, D., Beauchemin, S.: Evaluation of optical flow. Int. J. of Computer Vision **12**(1), 43–77 (1994)
9. Berkels, B., Kondermann, C., Garbe, C.S., Rumpf, M.: Reconstructing optical flow fields by motion inpainting. In: Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), pp. 388–400. Bonn, Germany (2009)

10. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th annual conference on Computer Graphics and Interactive Techniques, pp. 417–424 (2000)
11. Bigun, J., Granlund, G.H., Wiklund, J.: Multidimensional orientation estimation with applications to texture analysis and optical flow. IEEE Trans. Pattern Analysis and Machine Intelligence **13**(8), 775–790 (1991)
12. Black, M., Anandan, P.: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer Vision and Image Understanding **63**(1), 75–104 (1996)
13. Black, M.J., Anandan, P.: A framework for the robust estimation of optical flow. In: Int. Conf. on Computer Vision (ICCV), pp. 231–236 (1993)
14. Black, M.J., Yacoob, Y.: Recognizing facial expressions in image sequences using local parameterized models of image motion. Int. J. of Computer Vision **25**(1), 23–48 (1997)
15. Bouguet, J.Y.: Pyramidal implementation of the affine lucas-kanade feature tracker description of the algorithm. Intel Corporation **5**, 1–10
16. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning **3**(1), 1–122 (2011)
17. Braux-Zin, J., Dupont, R., Bartoli, A.: A general dense image matching framework combining direct and feature-based costs. In: International Conference on Computer Vision (ICCV) (2013)
18. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: European Conference on Computer Vision (ECCV), pp. 25–36. Prague, Czech Republic (2004)
19. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. IEEE Trans. Pattern Analysis and Machine Intelligence **33**(3), 500–513 (2011)
20. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/kanade meets horn/schunck: combining local and global optic flow methods. Int. J. of Computer Vision **61**(3), 211–231 (2005)
21. Bruhn, A., Weickert, W.: A confidence measure for variational optic flow methods. Geometric Properties for Incomplete Data pp. 283–298 (2006)
22. Bugeau, A., Ta, V., Papadakis, N.: Variational exemplar-based image colorization. IEEE Trans. on Image Processing **23**(1), 298–307 (2014)
23. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: European Conference on Computer Vision (ECCV), pp. 611–625. Springer-Verlag (2012)
24. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)
25. Chan, T.F., Kang, S.H., Shen, J.: Euler's elastica and curvature-based inpainting. SIAM Journal on Applied Mathematics pp. 564–592 (2002)
26. Chen, Y., Ye, X.: Projection onto a simplex. Ithaca, NY, USA: Cornell Univ. Press (2011)
27. Chen, Z., Jin, H., Lin, Z., Cohen, S., Wu, Y.: Large displacement optical flow from nearest neighbor fields. In: Computer Vision and Pattern Recognition (CVPR), pp. 2443–2450 (2013)
28. Chen, Z., Wang, J., Wu, Y.: Decomposing and regularizing sparse/non-sparse components for motion field estimation. In: Computer Vision and Pattern Recognition (CVPR), pp. 1776–1783 (2012)
29. Corpetti, T., Mémin, E.: Stochastic uncertainty models for the luminance consistency assumption. IEEE Trans. Image Processing **21**(2), 481–493 (2012)
30. Cremers, D., Soatto, S.: Motion competition: A variational approach to piecewise parametric motion segmentation. Int. J. of Computer Vision **62**(3), 249–265 (2005)
31. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Processing **13**(9), 1200–1212 (2004)
32. Dong, W., Shi, G., Hu, X., Ma, Y.: Nonlocal sparse and low-rank regularization for optical flow estimation. IEEE Trans. on Image Processing **23**(10), 4527–4538 (2014)
33. Enkelmann, W.: Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. Computer Vision, Graphics, and Image Processing **43**(2), 150–177 (1988)
34. Fermüller, C., Shulman, D., Aloimonos, Y.: The statistics of optical flow. Computer Vision and Image Understanding **82**(1), 1–32 (2001)
35. Fleet, D.J., Black, M.J., Yacoob, Y., Jepson, A.D.: Design and use of linear models for image motion analysis. Int. J. of Computer Vision **36**(3), 171–193 (2000)
36. Fortun, D., Bouthemy, P., Kervrann, C.: Optical flow modeling and computation: A survey. Computer Vision and Image Understanding **134**, 1–21 (2015)
37. Fortun, D., Bouthemy, P., Kervrann, C.: Sparse aggregation framework for optical flow estimation. In: Int Conf. Scale Space and Variational Methods in Computer Vision, pp. 323–334. Lège-Cap Ferret, France (2015)
38. Fortun, D., Bouthemy, P., Kervrann, C.: Aggregation of local parametric candidates with exemplar-based occlusion handling for optical flow. Computer Vision and Image Understanding (in press) (2016)
39. Fortun, D., Bouthemy, P., Paul-Gilloteaux, P., Kervrann, C.: Aggregation of patch-based estimations for illumination-invariant optical flow in live cell imaging. In: International Symposium on Biomedical Imaging (ISBI), pp. 660–663 (2013)
40. Galvin, B., McCane, B., Novins, K., Mason, D., Mills, S.: Recovering motion fields: An evaluation of eight optical flow algorithms. In: British Machine Vision Conference (1998)
41. Hafner, D., Demetz, O., Weickert, J.: Why is the census transform good for robust optic flow computation? In: Scale Space and Variational Methods in Computer Vision (SSVM), pp. 210–221 (2013)
42. He, K., Sun, J.: Image completion approaches using the statistics of similar patches. IEEE Trans. on Pattern Analysis and Machine Intelligence **36**(12), 2423–2435 (2014)
43. Heitz, F., Bouthemy, P.: Multimodal estimation of discontinuous optical flow using markov random fields. IEEE Trans. on Pattern Analysis and Machine Intelligence **15**(12), 1217–1232 (1993)
44. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence **17**(1-3), 185–203 (1981)
45. Hornacek, M., Besse, F., Kautz, J., Fitzgibbon, A.W., Rother, C.: Highly overparameterized optical flow using patchmatch belief propagation. In: European Conference on Computer Vision, Zurich,, pp. 220–234 (2014)
46. Humayun, A., Mac Aodha, O., Brostow, G.J.: Learning to find occlusion regions. In: Computer Vision and Pattern Recognition (CVPR), pp. 2161–2168 (2011)

47. Ince, S., Konrad, J.: Occlusion-aware optical flow estimation. IEEE Trans. Image Processing **17**(8), 1443–1451 (2008)
48. Jia, K., Wang, X., Tang, X.: Optical flow estimation using learned sparse model. In: Int. Conf. on Computer Vision (ICCV), pp. 2391–2398 (2011)
49. Jodoin, P.M., Mignotte, M.: Optical-flow based on an edge-avoidance procedure. Computer Vision and Image Understanding **113**(4), 511–531 (2009)
50. Komodakis, N., Tziritas, G.: Image completion using efficient belief propagation via priority scheduling and dynamic pruning. IEEE Trans. Image Processing **16**(11), 2649–2661 (2007)
51. Kondermann, C., Mester, R., Garbe, C.: A statistical confidence measure for optical flows. In: European Conference on Computer Vision (ECCV), pp. 290–301. Marseille, France (2008)
52. Kybic, J., Nieuwenhuis, C.: Bootstrap optical flow confidence and uncertainty measure. Computer Vision and Image Understanding **115**(10), 1449–1462 (2011)
53. Leordeanu, M., Zanfir, A., Sminchisescu, C.: Locally affine sparse-to-dense matching for motion and occlusion estimation. In: Int. Conf. on Computer Vision (ICCV), pp. 1221–1728. Sydney, Australia (2013)
54. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. International Joint Conference on Artificial Intelligence pp. 674–679 (1981)
55. Maurizot, M., Bouthemy, P., Delyon, B., Juditski, A., Odobez, J.M.: Determination of singular points in 2D deformable flow fields. In: Int. Conf. on Image Processing (ICIP), vol. 3, pp. 488–491. Washington D.C. (1995)
56. Mémin, E., Pérez, P.: Dense estimation and object-based segmentation of the optical flow with robust techniques. IEEE Trans. Image Processing **7**(5), 703–719 (1998)
57. Menze, M., Heipke, C., Geiger, A.: Discrete optimization for optical flow. In: Pattern Recognition, pp. 16–28. Springer (2015)
58. Mohamed, M., Rashwan, H., Mertsching, B., Garcia, M., Puig, D.: Illumination-robust optical flow approach using local directional pattern. IEEE Trans. on Circuits and Systems for Video Technology **24**(9), 1499–1508 (2014)
59. Mota, C., Stuke, L., Barth, E.: Analytic solutions for multiple motions. In: Int. Conf. on Image Processing (ICIP), pp. 917–920. Thessaloniki, Greece (2001)
60. Mozerov, M.: Constrained optical flow estimation as a matching problem. IEEE Trans. Image Processing **22**(5), 2044–2055 (2013)
61. Nagel, H., Enkelmann, W.: An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. IEEE Trans. Pattern Analysis and Machine Intelligence **8**(5), 565–593 (1986)
62. Nieuwenhuis, C., Kondermann, D., Garbe, C.S.: Complex motion models for simple optical flow estimation. Pattern Recognition pp. 141–150 (2010)
63. Odobez, J., Bouthemy, P.: Robust multiresolution estimation of parametric motion models. Journal of Visual Communication and Image Representation **6**(4), 348–365 (1995)
64. Papadakis, N., Yildizoglu, R., Aujol, J.F., Caselles, V.: High-dimension multilabel problems: Convex or nonconvex relaxation? SIAM Journal on Imaging Sciences **6**(4), 2603–2639 (2013)
65. Pierre, F., Aujol, J.F., Bugeau, A., Papadakis, N., Ta, V.T.: Luminance-chrominance model for image colorization. SIAM Journal on Imaging Sciences **8**(1), 536–563 (2015)
66. Pierre, F., Aujol, J.F., Bugeau, A., Ta, V.T.: Hue constrained image colorization in the rgb space. Preprint (2014)
67. Ranftl, R., Bredies, K., Pock, T.: Non-local total generalized variation of optical flow estimation. In: European Conference on Computer Vision, pp. 439–454. Zurich (2015)
68. Revaud, J., Weinzaepfel, P., Harchoui Z. Schmid, C.: Epicflow: Edge-preserving interpolation of correspondences for optical flow. In: IEEE Conf. Computer Vision and Pattern Recognition (CVPR'15). Boston, MA (2015)
69. Salmon, J., Strozecki, Y.: Patch reprojections for non-local methods. Signal Processing **92**(2), 477–489 (2012)
70. Senst, T., Eiselen, V., Sikora, T.: Robust local optical flow for feature tracking. IEEE Trans. Circuits and Systems for Video Technology **22**(9), 1377–1387 (2012)
71. Shen, X., Wu, Y.: Sparsity model for robust optical flow estimation at motion discontinuities. In: Computer Vision and Pattern Recognition (CVPR), pp. 2456–2463 (2010)
72. Simoncelli, E.P., Adelson, E.H., Heeger, D.J.: Probability distributions of optical flow. In: Computer Vision and Pattern Recognition (CVPR), pp. 310–315 (1991)
73. Stein, A.N., Hebert, M.: Occlusion boundaries from motion: Low-level detection and mid-level reasoning. Int. J. of Computer Vision **82**(3), 325–357 (2009)
74. Steinbrucker, F., Pock, T., Cremers, D.: Advanced data terms for variational optic flow estimation. In: Vision, Modeling, and Visualization Workshop (2009)
75. Sun, D., Liu, C., Pfister, H.: Local layering for joint motion estimation and occlusion detection. In: Computer Vision and Pattern Recognition (CVPR). Colombus (2014)
76. Sun, D., Roth, S., Black, M.J.: A quantitative analysis of current practices in optical flow estimation and the principles behind them. Int. J. of Computer Vision **106**(2), 115–137 (2014)
77. Sun, D., Sudderth, E.B., Black, M.J.: Layered segmentation and optical flow estimation over time. In: Computer Vision and Pattern Recognition (CVPR), pp. 1768–1775 (2012)
78. Sun, J., Li, Y., Kang, S.B.: Symmetric stereo matching for occlusion handling. In: IEEE Conf. Computer Vision and Pattern (CVPR'05), pp. 399–406. San Diego, CA (2005)
79. Timofte, R., Gool, L.V.: Sparse flow: Sparse matching for small to large displacement optical flow. In: IEEE Winter Conference on Applications of Computer Vision, WACV, Waikoloa, HI, pp. 1100–1106 (2015)
80. Unger, M., Werlberger, M., Pock, T., Bischof, H.: Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling. In: Computer Vision and Pattern Recognition (CVPR), pp. 1878–1885 (2012)
81. Vogel, C.R., Oman, M.E.: Iterative methods for total variation denoising. SIAM Journal on Scientific Computing **17**(1), 227–238 (1996)
82. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An improved algorithm for tv-l 1 optical flow. In: Statistical and Geometrical Approaches to Visual Motion Analysis, pp. 23–45 (2009)
83. Weinzaepfel, P., Revaud, J., Harchaoui, Z., Schmid, C., et al.: Deepflow: Large displacement optical flow with deep matching. In: Int. Conf. on Computer Vision (ICCV), pp. 1385–1392. Sydney (2013)

84. Werlberger, M., Pock, T., Bischof, H.: Motion estimation with non-local total variation regularization. In: Computer Vision and Pattern Recognition (CVPR'10), pp. 2464–2471. San-Fransisco (2010)
85. Wulff, J., Black, M.: Efficient sparse-to-dense optical flow estimation using a learned basis and layers. In: IEEE Conf. Computer Vision and Pattern Recognition (CVPR'15). Boston, MA (2015)
86. Xu, L., Jia, J., Matsushita, Y.: Motion detail preserving optical flow estimation. IEEE Trans. Pattern Analysis and Machine Intelligence **34**(9), 1744–1757 (2012)
87. Yang, J., Li, H.: Dense, accurate optical flow estimation with piecewise parametric model. In: IEEE Conf. Computer Vision and Pattern Recognition (CVPR'15). Boston, MA (2015)
88. Yang, J., Zhang, Y.: Alternating direction algorithms for \ell_1-problems in compressive sensing. SIAM journal on scientific computing **33**(1), 250–278 (2011)
89. Zimmer, H., Bruhn, A., Weickert, J.: Optic flow in harmony. Int. J. of Computer Vision **93**(3), 1–21 (2011)