



Aggregation of local parametric candidates with exemplar-based occlusion handling for optical flow[☆]



Denis Fortun, Patrick Bouthemy^{*}, Charles Kervrann

Inria, Centre de Rennes - Bretagne Atlantique, Campus Universitaire de Beaulieu 35042 Rennes, France

ARTICLE INFO

Article history:

Received 27 June 2015

Accepted 28 November 2015

Available online 17 December 2015

Keywords:

Optical flow

Occlusion

Large displacement

Local parametric motion

Aggregation framework

ABSTRACT

Handling all together large displacements, motion details and occlusions remains an open issue for reliable computation of optical flow in a video sequence. We propose a two-step aggregation paradigm to address this problem. The idea is to supply local motion candidates at every pixel in a first step, and then to combine them to determine the global optical flow field in a second step. We exploit local parametric estimations combined with patch correspondences and we experimentally demonstrate that they are sufficient to produce highly accurate motion candidates. The aggregation step is designed as the discrete optimization of a global regularized energy. The occlusion map is estimated jointly with the flow field throughout the two steps. We propose a generic exemplar-based approach for occlusion filling with motion vectors. We achieve state-of-the-art results in the MPI-Sintel benchmark, with particularly significant improvements in the case of large displacements and occlusions.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Optical flow is a key information when addressing important problems in computer vision such as moving object segmentation, object tracking, egomotion computation, obstacle detection or action recognition. The challenge for an optical flow estimation method is to deal with a large variety of image contents and motion types. Optical flow has been historically evaluated on sequences exhibiting small displacements and smooth motion fields, like in the Yosemite sequence [7]. Once initial issues were solved, other challenges were addressed [54], and new situations have been proposed by more recent benchmarks [4,22]. Various and sometimes opposite scene conditions must be handled together, as illumination changes, large areas of smooth motion, motion details, large displacements, motion discontinuities, occluded regions (i.e., points disappearing in the next image).

Optical flow methods first rely on a data constancy assumption, e.g., applied to image intensity or spatial intensity gradient. Then, it is combined with a spatial, or sometimes space-time, coherency constraint on the expected velocity field. Existing approaches can be broadly classified into *local* and *global* methods.

Local spatial coherency arises when considering a parametric motion model, e.g., local translation [52], 4-parameter sub-affine model,

affine model, 8-parameter quadratic model [59], in a given neighborhood or an appropriate local region. Optimization requires that the neighborhood is sufficiently textured or contains interest points such as corners, to supply accurate and reliable velocity vectors.

In contrast, global methods express the flow field coherency by imposing a global smoothness constraint in addition to the data constancy term, known as the regularization term of the global energy as pioneered by Horn and Schunck [37]. Global methods overcome uncertainty yielded by local supports in uniform intensity regions by diffusing motion from informative to non informative regions via the global regularization constraint. The optimization problem of seminal model [37] was optimally solvable, but the estimation was affected by oversmoothing and was limited to small displacements.

Numerous modifications of this original model, starting with [10,36], have been designed to resolve these two crucial issues, namely, handling of large displacements and preservation of motion discontinuities. It was usually achieved by introducing a multi-resolution and incremental coarse-to-fine framework along with piecewise smoothing or robust estimation. A more recent attempt is to learn statistics of motion fields or motion bases as regularization means [42,57,65,66,82]. The data term of the global modeling has also received attention. Image features like image gradient [19], texture component [79], LDP (Local Directional Pattern) descriptor for illumination-robust data constancy [55], and matching criteria like Normalized Cross Correlation (NCC) [81] or Census transform [35], convey invariance properties to overcome limitations of the classical intensity constancy assumption. However, intricate optimization issues came with the increasing complexity of the modeling.

[☆] This paper has been recommended for acceptance by Shahriar Negahdaripour.

^{*} Corresponding author. Fax: +33 299847171.

E-mail address: Patrick.Bouthemy@inria.fr (P. Bouthemy).

Although existing local methods are far from being able to compete with global models in terms of accuracy in computer vision benchmarks, several works based on joint estimation and segmentation of the motion field have shown that when appropriate segmented regions are found, affine models can be very accurate representations [71,77]. Yet, the alternate optimization schemes involved are sensitive to the initialization of the region supports.

In this paper, we describe an occlusion-aware optical flow method that we name *AggregFlow*. It relies on an aggregation framework which explicitly separates the motion candidate computation and the global motion field recovery.

Firstly, we advocate the systematic computation of affine motion models over a set of size-varying square patches, without segmentation step. To this end, we introduce a pre-defined collection of estimation windows to compute motion candidates, which allows us to seamlessly handle any configuration of piecewise continuous motions and a variety of motion scales. To handle large displacements, we combine affine estimation with patch-based matching. Differently from other methods exploiting feature matching as additional constraints [20,80] or coarse initialization [24,85], patch-based matching directly contributes to the computation of real-valued motion vector candidates at every pixel. We experimentally demonstrate that these sets of candidate motion vectors can potentially yield an accurate global flow field.

Secondly, we handle occlusion detection and occlusion filling with motion vectors in the two steps of *AggregFlow*. The set of motion candidates is in fact extended in two ways: exemplar-based search in occluded areas and use of the estimated parametric dominant motion. The local motion candidates are also exploited to build an occlusion confidence map which intervenes in the global aggregation model. We introduce a novel generic exemplar-based model for occlusion filling. It takes the form of an additional term in the global aggregation energy imposing non-local and image-based constraint on the motion of occluded pixels.

Thirdly, we resort to a discrete aggregation scheme. This kind of optimization approach has been little explored for optical flow computation so far [50], but it appears very promising. In coherence with the above observation about candidates accuracy, we define the aggregation as the selection of one motion candidate at each pixel, while ensuring smoothing of the resulting flow field and preservation of motion discontinuities. The aggregation is achieved with a discrete optimization algorithm, since motion candidates can be seen as labels. The occlusion confidence map enables to guide the joint occlusion and motion estimation, while decoupling the estimation of the two sets of unknown variables.

The main contributions of our method *AggregFlow* can be summarized as follows:

- An accurate parametric patch-based scheme for the motion candidate computation step with an efficient integration of feature matching,
- A generic exemplar-based approach for recovering motion in occluded regions,
- A joint motion field and occlusion map estimation guided by a local occlusion confidence map obtained from motion candidates.

We have carried out a comprehensive experimental evaluation. Specifically, state-of-the-art results have been obtained for large displacements and occlusions on the challenging MPI Sintel dataset. A preliminary approach without any occlusion handling and dedicated to a specific application was presented in [30].

The paper is organized as follows. [Section 2](#) describes related work. In [Section 3](#), we present the parametric computation of motion candidates and the local detection of occlusions. [Section 4](#) is devoted to the aggregation stage. In [Section 5](#), we report experimental results on three optic flow benchmarks, demonstrating the performance of *AggregFlow*. [Section 6](#) contains concluding remarks.

2. Related work

Hereunder, we briefly review the literature on optical flow computation while focusing on issues related to our contributions. A recent comprehensive survey can be found in [31].

2.1. Feature correspondences and large displacements

The integration of feature correspondences in dense motion estimation has been investigated in several recent works. A first class of methods integrates feature correspondences in a global energy model. Variational methods [18,20,80] include an additional term to a classical global energy to force the flow to be close to pre-computed correspondences. Giving a fixed weight to the correspondences, this approach is sensitive to matching errors. To overcome this problem, [18,64,75,80] focused on improving the matching step. Another class of methods use correspondences to reduce the search space for discrete optimization and provide a coarse initialization for subsequent refinement [24,56,85]. The main motivation of the attempts based on feature matching is to get rid of the drawbacks of the coarse-to-fine scheme imposed by variational optimization, in particular the loss of large displacements of small objects.

Our patch correspondence is related to [24,56,85] in the sense that it is used in the candidates generation process. However, our method does not produce coarse approximations to be refined in a continuous subsequent step and we do not perform any global variational optimization.

2.2. Occlusion handling

Occlusions play a crucial role for motion estimation [69], especially under large displacements, since no motion measurements are available in occluded areas. Therefore, a proper occlusion handling must distinguish between *occlusion detection*, segmenting the image into occluded and non-occluded regions, and *occlusion filling*, applying a specific treatment to motion estimation in occluded regions. Occlusion detection has been mostly undertaken as a subsequent operation to motion computation, by thresholding a consistency measure issued from the estimated motion field, like forward-backward motion mismatch [41], mapping unicity [85] or data constancy violation [83]. A distinctive geometric criterion is introduced in [44]. Occlusions can also be detected independently from motion estimation using image cues like spatiotemporal T-junctions [2]. In stereovision, other criteria like visibility [70] or ordering constraints [16] are also exploited. The main limitation of this sequential approach is that accuracy of occlusion detection is highly dependent on the quality of the initial motion estimation. Several flow and image criteria have been combined in a learning framework [40]. Other approaches estimate the occlusion map jointly with the motion or disparity field in an alternate optimization scheme, encoding one of the above-mentioned criterion in a global energy [3,41,61,70]. Our occlusion detection falls in the latter category.

The problem of filling occluded regions with estimated velocity vectors when the occlusion map is known is closely related to the image inpainting problem. Inpainting methods can be coarsely divided into two classes, diffusion-based methods [9,23] and exemplar-based methods [26,48]. A synthesis of these two approaches has been investigated in [21] in a variational framework. In exemplar-based image inpainting, the missing part is filled by copying pixels of the observed images. The framework is non local in the sense that similar pixels can be sought anywhere in the image. Occlusion filling is usually tackled by diffusion-based (or geometry-oriented) methods, propagating motion from non-occluded regions to occluded regions via partial derivative equation (PDE) resolution [3,8,41,49,61,85]. In stereovision, a weighted least-squares strategy exploits a local averaging of

Download English Version:

<https://daneshyari.com/en/article/525596>

Download Persian Version:

<https://daneshyari.com/article/525596>

[Daneshyari.com](https://daneshyari.com)