Large Displacement Optical Flow Computation without Warping

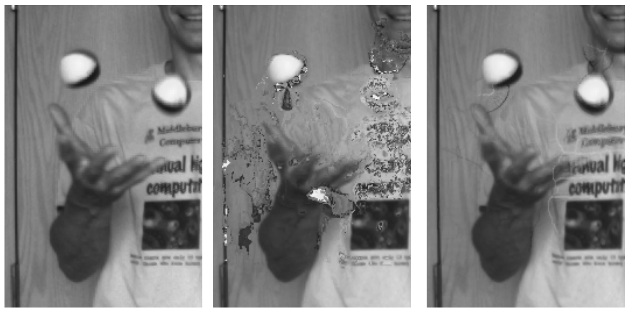
Frank Steinbrücker
Department of Computer Science
University of Bonn, Germany

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology, Austria

Daniel Cremers
Department of Computer Science
University of Bonn, Germany

Abstract

We propose an algorithm for large displacement optical flow estimation which does not require the commonly used coarse-to-fine warping strategy. It is based on a quadratic relaxation of the optical flow functional which decouples data term and regularizer in such a way that the non-linearized variational problem can be solved by an alternation of two globally optimal steps, one imposing optimal data consistency, the other imposing discontinuity-preserving regularity of the flow field. Experimental results confirm that the proposed algorithmic implementation outperforms the traditional warping strategy, in particular for the case of large displacements of small scale structures.



Original image Warping Proposed algorithm

Figure 1. Close-up of reconstructed second frame based on the first frame and the estimated motion. In contrast to existing coarse-to-fine warping schemes, the proposed algorithm allows to estimate large-displacement optical flow even for small scale structures.

1. Introduction

1.1. Variational Optical Flow Computation

Computing optimal correspondences between pairs of points remains one of the major challenges in Computer Vision. Applications include the computation of motion fields from videos, the registration of medical organs across different scans, the matching of facial images for the purpose of recognition and the tracking of deformable objects. The computational challenge is to determine for each point in one image an optimal corresponding point in the other image. To suppress meaningless correspondences and make the problem more well-posed one typically imposes spatial regularity of the correspondence function in an energy minimization framework. While the computation of one-dimensional correspondence functions – often referred to as *string matching* – can be solved in polynomial time using Dynamic Programming [4] approaches, for matching problems in two or more dimensions no efficient optimal solutions are known.

In 1981, Horn and Schunck introduced one of the first variational methods in Computer Vision in order to compute a dense motion field $v : \Omega \to \mathbb{R}^2$ on the image plane $\Omega \subset \mathbb{R}^2$ for matching a pair of consecutive images from a gray value sequence $I : \Omega \times [0, T] \to \mathbb{R}$. They proposed to minimize the functional

$$E(v) = \int_\Omega \underbrace{\left(\nabla I^\top v + I_t\right)^2}_{\text{data term}} + \lambda \underbrace{\left(|\nabla v_1|^2 + |\nabla v_2|^2\right)}_{\text{regularity term}} d^2x. \quad (1)$$

The data term aims at matching points of similar intensity by imposing the linearized brightness constancy constraint, while the regularity term (weighted by $\lambda > 0$) imposes spatial smoothness of the velocity field $v = (v_1, v_2)$.

In the wake of subsequent publications researchers successfully addressed numerous shortcomings of the above formulation. To avoid over-smoothing and preserve discontinuities in the computed flow field, researchers replaced the quadratic regularizer by image-adaptive anisotropic [9] or by robust non-quadratic ones [5]. Similarly, robust estimators were employed to account for outliers in the data term [5].

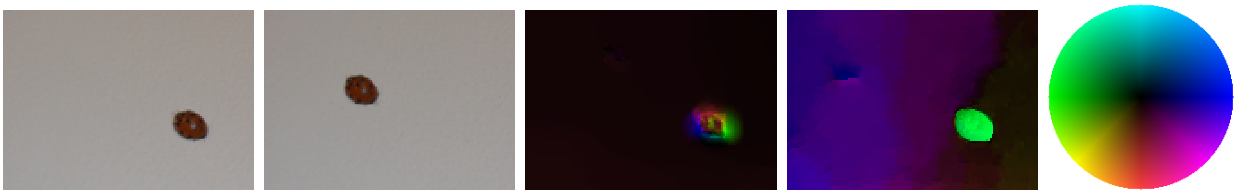


Figure 2. Large displacement of small-scale structures. For two images of a lady bug taken at very different times, in contrast to the coarse-to-fine warping schemes, the proposed approach allows to accurately estimate the correspondence.

1.2. The Problem of Large Displacements

One of the major practical limitations of the Horn and Schunck model is that it only applies to the case of *small* motion, the linearization in (1) only being valid for velocities of small magnitude. In the case of larger motion vectors that arise in most real-world applications, the computational challenge becomes substantially more cumbersome: The number of pixels that a given pixel can be matched to grows quadratically with the maximum permissible velocity magnitude.

To circumvent this combinatorial explosion of permissible solutions, researchers have reverted to coarse-to-fine strategies of estimation [1, 8, 10, 11]. The key idea is to iteratively compute the motion field from coarse to fine scales, always warping one of the two images according to the current motion estimate. As a consequence, the residual motion field on each level of refinement is expected to fulfill the small motion assumption and the motion estimates are successively refined.

Convergence properties of this technique were studied in greater detail in [7], a theoretical justification relating it to a fixed point iteration on the functional with non-linearized data term was developed in [10]. Coarse-to-fine warping is known to give excellent flow field estimates even for larger motions. To date it is the *only* competitive algorithmic approach to compute high quality dense flow fields from the established non-convex energy functionals. Nevertheless warping schemes have two important drawbacks:

- The numerical implementation of coarse-to-fine schemes is somewhat involved. The choice of coarsening pyramid and interpolation technique is known to substantially affect the quality of results [10].
- Coarse-to-fine warping strategies only provide reliable motion estimates for larger motion if the respective image structures are of a similar spatial scale. Fine scale image structure that are no longer visible in the coarsened version of the image clearly cannot be matched by the motion estimate on the coarse scale. As a result, motion estimates for image sequences containing large displacements of fine scale low contrast structures are likely to be inaccurate. Figure 1 provides an example of this limitation taken from an image of the Middlebury benchmark.[3]

1.3. Contribution

In this paper we propose a novel framework for motion estimation which allows to handle large motion without the need for warping. In contrast to warping strategies, the algorithmic implementation is much simpler. It does not require coarse-to-fine pyramid representations of the images and respective warping strategies. Moreover, experimental results demonstrate that it can handle large displacements even for small scale structures.

The key idea is to solve a quadratic relaxation scheme for minimization of the non-linearized optical flow functional by a sequence of globally optimal steps, each being computed on the full scale image. More specifically, by introducing an auxiliary vector field we decouple data term and regularizer in such a way that minimization can be done by alternating two globally optimizable problems: The first one aims at attracting the flow field to optimally match respective intensities (thus minimizing the data term), while the second one is a convex problem which aims at imposing discontinuity-preserving spatial regularity.

2. Alternating two Global Optimizations

In the following, we will propose a novel algorithm based on alternating global optimizations which allows to compute large displacement optical flow without warping.

Let $\Omega \subset \mathbb{R}^2$ denote the image plane[1] and $I_1, I_2 : \Omega \to \mathbb{R}$ denote two intensity images. Following [10], the problem of estimating a regularized motion field $v : \Omega \to \mathbb{R}^2$ which optimally matches intensities from one image to the other can be formulated as one of minimizing the functional

$$E(v) = \int_{\Omega} \lambda \rho(v, x) + \psi(v, \nabla v, \dots)\, d^2x. \qquad (2)$$

[1] In this paper, we are merely concerned with two-dimensional images. However, the model can be extended to higher dimensions.

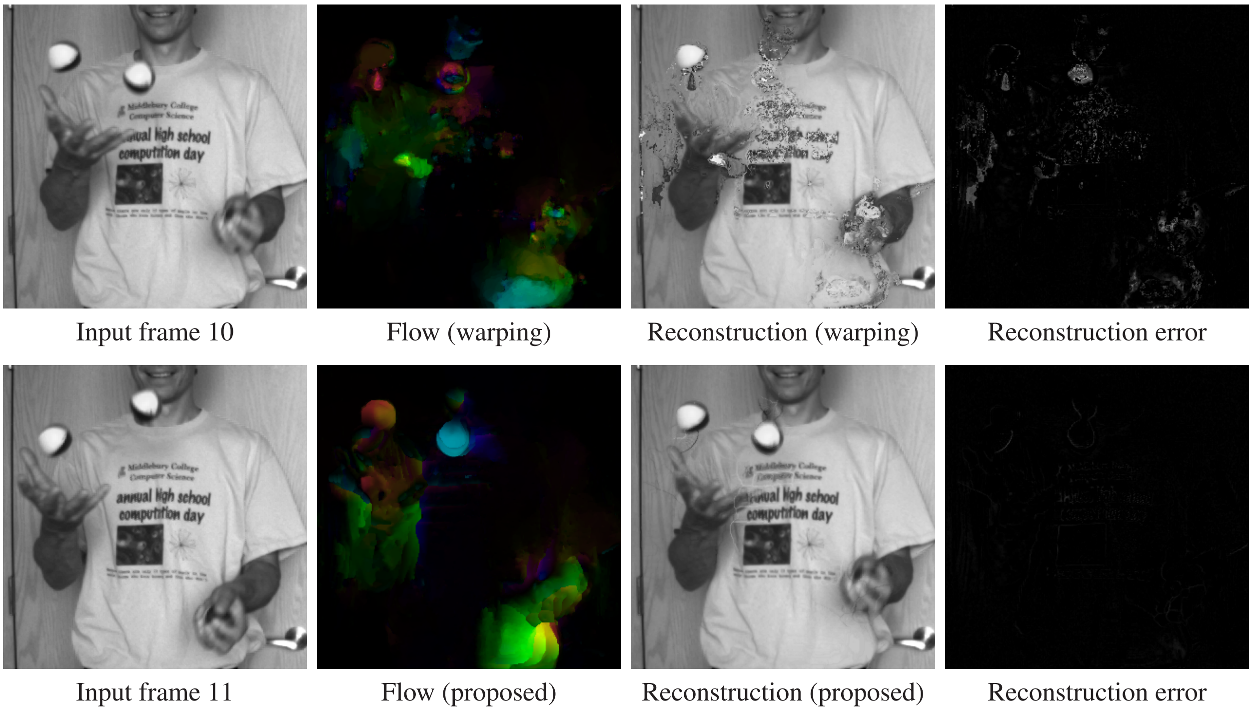


Figure 3. Comparison of reconstructed images from flow fields computed with and without warping. The experiments show the flow fields and reconstructions of frame 10 computed from frame 11 and the estimated flow field for consecutive images from the Beanbags sequence. While the warping scheme (above) clearly looses small scale structures such as the fast moving ball, these are appropriately preserved with the proposed algorithm (below). As a consequence, we obtain a substantially smaller reconstruction error.

In the following, we will consider a data term which favors the matching of similar intensities according to

$$\rho(v, x) = |I_1(x) - I_2(x + v(x))|. \tag{3}$$

To preserve discontinuities in the regularized flow field, we replace the quadratic penalty function of the Horn and Schunck model (1) with a TV-L1 penalizer, yielding the smoothness term

$$\psi(\nabla v) = |\nabla v_1| + |\nabla v_2|. \tag{4}$$

The contribution of the present paper is not a new functional for optical flow estimation, but rather a different algorithmic framework for computing minimizers. The major algorithmic challenge lies in the fact that the above functional is not convex in v. As a consequence, the quality of minimizers invariably depends on the strategy of minimization (initialization, coarse-to-fine warping) and on implementational aspects such as the choice of downsampling factor, of interpolation scheme etc.

In the following, we present a decoupling scheme which gives rise to a minimization algorithm that consists of two fractional steps each of which can be solved in globally optimal manner. Let us start by raising the question why and in what sense functional (2) is not convex. Firstly we observe that the regularity term (4) is indeed convex. Secondly, we observe that the data term data term (3) is non-convex. Thirdly – and this is the key observation – the data term is a *point-wise* term in the sense that optimal choices for v at different locations do not depend on one another (other than via the regularity term). Therefore, if we decouple data term and regularity term, we can decompose the optimization problem (2) into two subproblems each of which can be optimized globally. In particular, it turns out that this strategy removes the need for warping.

Following a series of papers on quadratic relaxation [6, 2, 11], we use an auxiliary vector field $u : \Omega \to \mathbb{R}^2$ in order to decouple data term and regularizer:

$$E(v, u) = \int_\Omega \lambda\rho(v, x) + \frac{1}{2\theta}(v - u)^2 + \psi(\nabla u)\, d^2x. \tag{5}$$

It can be shown [6] that for $\theta \to 0$ minimization of functional (2) is equivalent to minimization of (5). At a first glance, this decoupling seems to complicate things, because rather than one optimization problem in v we are now faced with two coupled optimization problems in v and u. Yet, both of these problems can be optimized globally:

- Functional (5) can be minimized globally with respect to u because it is convex in u. Therefore optimal solutions for u can be computed by gradient descent or alternative more efficient algorithms.

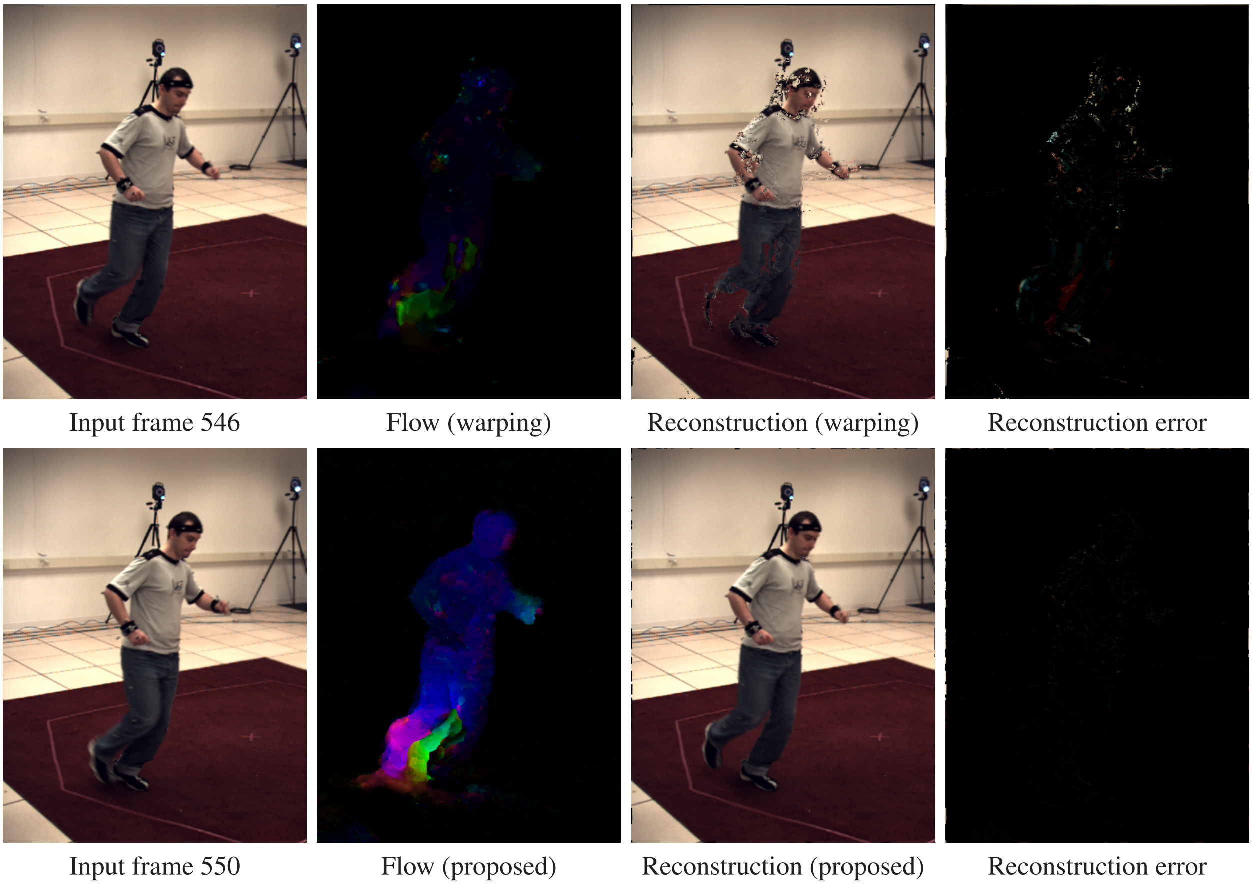


Figure 4. Performance of the proposed algorithm on color sequences. The experiments show the flow fields and reconstructions of frame 546 computed from frame 550 and the estimated flow field for two images from the HumanEva-II sequence. In contrast to the warping scheme, the proposed method finds correspondences for fast moving structures as well as for occluded areas.

- Functional (5) can be minimized globally with respect to v, because it merely exhibits a dependency on v. Optimal values for $v(x)$ for every x can be simply computed by a complete search. There is no spatial regularity term for v in functional (5). Without this coupling of solutions for v at different locations, the combinatorial explosion of possible solutions has vanished. While a complete search over possible values of $v(x)$ associated with each pixel x appears to be a computationally cumbersome problem, it can be efficiently parallelized on standard graphics hardware.

While with current graphics hardware a complete search is still slower than warping schemes such as the one employed in [11], the proposed algorithmic solution has two important advantages:

- Since the proposed algorithm does not rely on image coarsening, there is no issue with small-scale structures being lost on the coarser scales which warping schemes require for estimating larger motions. As a consequence, we can expect the resulting algorithm to provide better motion fields for small scale structures undergoing large displacements.

- Warping schemes ultimately require a linearization of the data term in (5). This is not the case for the proposed complete search. It can integrate arbitrary data terms including distance of local color values, patch comparisons or normalized cross-correlation. In addition, since there is no differentiation, the proposed approach naturally extends to truncated or other non-differentiable penalty functions.

In practice, we initialize with $u = v = 0$ and a large value for θ. Subsequently we alternatingly compute $v(x)$ as the minimizer of (5) with fixed $u(x)$ by a complete search for all pixels x, and minimize (5) with respect to u for fixed v. Between the iterations we continuously decrement θ forcing u and v to converge at the end. To save computation time the complete search for an optimal v is performed in a restricted search window chosen with respect to a user-specified upper bound on the velocity. Qualitatively, this user parameter corresponds to the number of pyramid levels specified in warping schemes.

While alternating two globally optimal algorithms for parts of the functional does not guarantee global optimality for the entire functional, our method significantly outperforms methods depending on local linearization as we will show in the next section.

3. Experimental Results

3.1. Large Displacements of Small Objects

Warping schemes require a coarsening of the input data in order to account for larger displacements. As a consequence, they cannot handle objects of a scale which is substantially smaller than their motion. If the object is no longer present on the coarse scale, the resulting motion estimates will be unreliable.

Since the proposed algorithm does not involve any image coarsening to account for larger motion, we expect that this algorithm should better estimate large motion, in particular for objects which are much smaller than the scale of motion. In the following, we will confirm this in several real-world experiments. Since these experiments typically do not have a ground truth, we evaluate the results in a two-fold manner:

- We verify qualitatively whether the computed and color-coded flow field is meaningful.
- We check the consistency of the flow-field by reconstructing the first of the two frames using the second frame and the estimated motion field v according to:

$$I_1^r(x) = I_2(x + v(x))$$

  If the motion field is correct, then the reconstructed first frame I_1^r should be identical to the observed one I_1. We can quantify this error by plotting the difference image $|I_1 - I_1^r|$.

Figures 2, 3 and 4 demonstrate that indeed the motion fields and the reconstructed frames obtained with the proposed approach are more convincing than those obtained with the warping scheme. In both cases, the motion is substantially larger than the size of the moving objects. A closer observation shows that the warping scheme gives rise to flow fields which tend to shrink the respective structures (to account for their disappearance) – see Figure 2. In more complex scenes, it incorrectly matches them to the most similar structures in the vicinity – see Figures 3 and 4. In contrast, the proposed method provides reliable motion estimates which give rise to faithful reconstructions of the first frame.

3.2. Handling More Sophisticated Data Terms

Since the proposed algorithm does not require differentiability of the data term or local linearization, we can make use of arbitrary data terms. In the following, we will demonstrate this for the case of color comparison and patch comparison.

Figure 4 shows an example of applying the proposed algorithm to match two color images despite large displacement. The images are taken from the HumanEva-II benchmark on human tracking. While the warping strategy fails to correctly match corresponding regions, the proposed warping strategy determines a reliable flow field. While the reconstructed image of the warping strategy has flaws around the fast moving right leg and the small scale structures of the tripod in the background, the reconstructions obtained with the proposed scheme have no visible errors.

Besides the natural support for vector-valued images, the proposed algorithm also supports advanced penalty functions. It can be easily extended to compare a small patch around each sampled position instead of comparing the intensities at single positions alone. Other vector-valued data terms such as normalized cross correlations, SIFT features etc. are possible as well. Yet a more detailed study of these aspects is beyond the scope of this paper and will therefore be left for future work.

3.3. Subpixel Accuracy via Oversampling

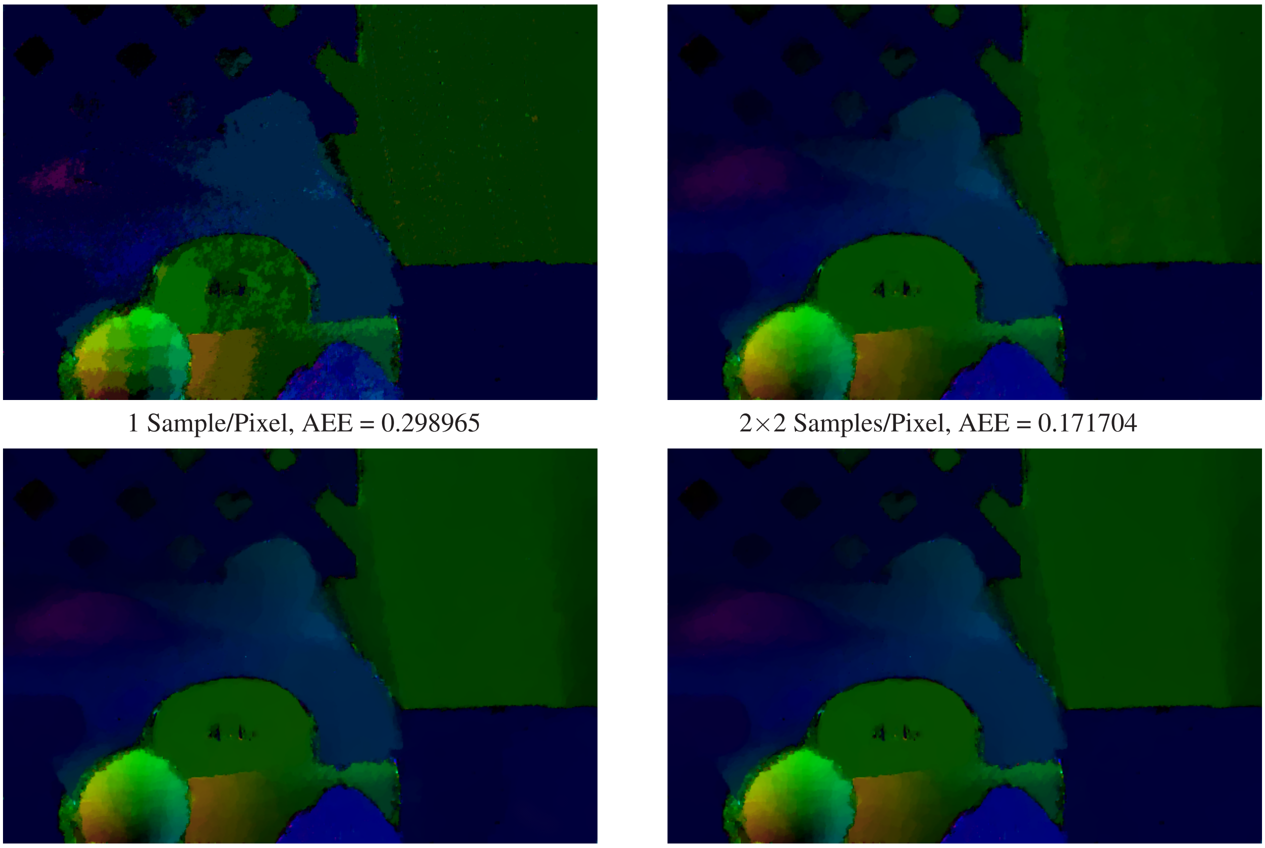
The strategy of complete search for globally minimizing in v is only good up to pixel precision. In order to increase the precision, one can simply revert to an oversampling strategy that considers a certain number of intermediate positions at the expense of additional computation time.

Figure 5 shows that while increasing the computation time, this oversampling strategy does improve the accuracy of estimated flow fields.

We tried an alternative subpixel strategy using an analytic solution based on linear interpolation. However,this strategy did not provide better experimental results.

4. Conclusion

We proposed a novel algorithm for estimating large-displacement optical flow which circumvents the need for warping schemes. By means of a quadratic relaxation scheme we decompose the original non-convex functional into a functional which can be minimized by alternating two globally optimal steps. The algorithm simply alternates a complete search with respect to the non-convex (but point-wise) data term and a convex optimization that takes into account the smoothness constraint. The flow estimation process is therefore decomposed in an alternation of searching for appropriate correspondents and discontinuity-preserving smoothing. In contrast to warping approaches, the proposed method can naturally make use of arbitrary data terms, including non-convex, non-differentiable terms and norms on color values or local patches. In numerous ex-



1 Sample/Pixel, AEE = 0.298965

2×2 Samples/Pixel, AEE = 0.171704

4×4 Samples/Pixel, AEE = 0.140528

10×10 Samples/Pixel, AEE = 0.130869

Figure 5. Effects of subpixel sampling rate on the quality of the computed flow field. An increasing number of samples per pixel improves the quality of the computed flow represented by the average end-point error (AEE) to the ground truth flow. However, it also increases the computation time from 7 seconds for the top left image over 18 and 63 seconds for the top right and bottom left images to 370 seconds for the bottom right image.

periments, we show that in contrast to state-of-the-art warping schemes, the proposed quadratic decoupling scheme allows to compute flow fields which accurately match small scale structures over large displacements.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Int. J. of Computer Vision*, 2:283–310, 1989. 2

[2] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition - modeling, algorithms, and parameter selection. *Int. J. of Computer Vision*, 67(1):111–136, 2006. 3

[3] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *IEEE Int. Conf. on Computer Vision*, 2007. 2

[4] R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, New Jersey, 1957. 1

[5] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *IEEE Int. Conf. on Computer Vision*, pages 231–236, 1993. 1

[6] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Im. Vis.*, 20(1-2):89–97, 2004. 3

[7] M. Lefébure and L. D. Cohen. Image registration, optical flow and local rigidity. *J. Math. Im. Vis.*, 14(2):131–147, March 2001. 2

[8] E. Memin and P. Perez. Hierarchical estimation and segmentation of dense motion fields. *Int. J. of Computer Vision*, 46(2):129–155, 2002. 2

[9] H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 8(5):565–593, 1986. 1

[10] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, April 2006. 2

[11] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Pattern Recognition (Proc. DAGM)*, LNCS, pages 214–223, Heidelberg, Germany, 2007. Springer. 2, 3, 4