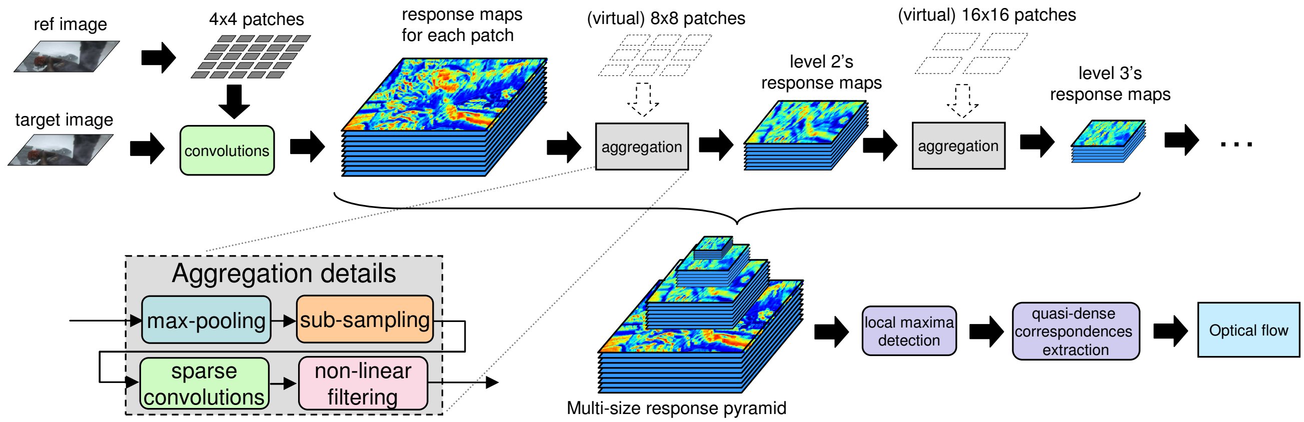

Figure 1. Outline of DeepFlow.

variational optical flow approach termed DeepFlow. Finally, we present experimental results in Section 5.

2. Related work

Large displacement in optical flow estimation. Variational methods are the state-of-the-art family of methods for optical flow estimation. Since the pioneering work of Horn and Schunck [1], research has focused on alleviating the drawbacks of this method. A series of improvements were proposed over the years [4, 31, 7, 21, 2, 25, 29]. Brox *et al*. [5] combine most of them into a variational approach. Energy minimization is performed by solving the Euler-Lagrange equations, then reducing the problem to solving a sequence of large and structured linear systems.

To handle large displacements, a descriptor matching component is incorporated in the variational approach in [6]. One major drawback of this method is that local descriptors are reliable only at salient locations and are locally rigid. Adding a matching component challenges the energy formulation as it could deteriorate performance at small displacement locations. Indeed, matching can give false matches, ambiguous matches, and has low precision (a pixel). In a different context, namely scene correspondence, descriptors or small patches were used in SIFT-flow [17] and PatchMatch [3] algorithms. Xu *et al*. [33] integrate matching of SIFT [26] and PatchMatch [3] to refine the flow initialization at each level. Excellent results were obtained, yet at the cost of expensive fusion steps. Leordeanu *et al*. [16] propose to extend sparse matching, with locally affine constraint, to dense matching before using a total variation algorithm to refine the flow estimation. We present here a computationally efficient, yet competitive approach for large displacement optical flow using a deep convolutional matching procedure.

Descriptor matching. Image matching consists of two steps: extraction of local descriptors and matching them. Initial image descriptors were extracted at sparse, scale-invariant or affine-invariant image locations [26, 20]. For the purpose of optical flow estimation, recent work showed that dense descriptor sampling improves performance [27, 6, 17]. In all cases, descriptors are extracted in rigid (generally square) local frames. Matching descriptors is then generally reduced to a nearest-neighbor problem [26, 3, 6]. Methods such as reciprocal nearest-neighbors allow to prune lots of false matches, but as a side effect also eliminate correct matches in weakly to moderately textured regions. We show here that (i) extraction of descriptors in non-rigid frames and (ii) dense matching in all image regions, yields a competitive approach, with a significant performance boost on MPI-Sintel [8] and KITTI [10] datasets.

Non-rigid matching. Our proposed matching algorithm, called *deep matching*, is strongly inspired by non-rigid 2D warping and deep convolutional networks [15, 28, 12]. It also bears similarity with non-rigid matching approaches developed in different contexts. In [9], Ecker and Ullman proposed a similar pipeline to ours (albeit more complex) to measure the similarity of small images. However, their method lacks a way of merging correspondences belonging to objects with contradictory motions (*e.g.*, on different focal planes). In a different context, Wills *et al*. [32] estimated optical flow by robustly fitting smooth parametric models (homography and splines) to local descriptor matches. In contrast, our approach is non-parametric and model-free. More recently, Kim *et al*. [13] proposed a hierarchical matching to obtain dense correspondences, but their method works in a coarse-to-fine (top-down) fashion, whereas deep matching works bottom-up. In addition, their method requires inexact inference using loopy belief propagation.

3. Deep Matching

In this section, we present the matching algorithm, termed deep matching, and discuss its main features. The matching algorithm builds upon a multi-stage architecture