Homography Flow for Dense Correspondences

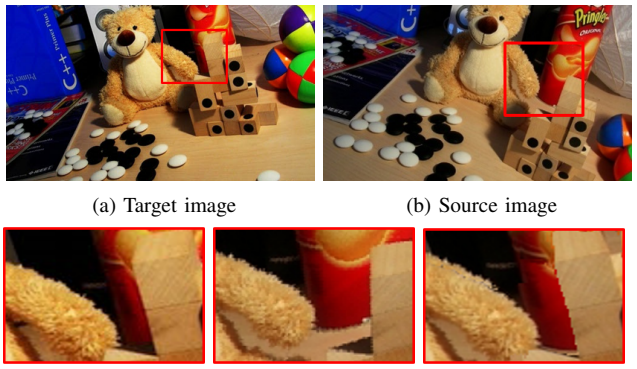
Kihong Park, Seungryoung Kim, and Kwanghoon Sohn
Yonsei University, Seoul, Korea
E-mail: khsohn@yonsei.ac.kr Tel/Fax: +82-2-2123-2879

***Abstract*—We present a unified framework for dense correspondence estimation, called Homography flow, to handle large photometric and geometric deformations in an efficient manner. Our algorithm is inspired by recent successes of the sparse to dense framework. The main intuition is that dense flows located in same plane can be represented as a single geometric transform. Tailored to dense correspondence task, the Homography flow differs from previous methods in the flow domain clustering and the trilateral interpolation. By estimating and propagating sparsely estimated transforms, dense flow field is estimated with very low computation time. The Homography flow highly improves dense correspondence estimations, especially in flow discontinuous area. Experimental results on challenging non-rigid image pairs and Moseg dataset, having large geometric deformations and multi-layered motions, show that our approach suppresses state-of-the-art algorithms in both accuracy and computation time.**

I. Introduction

Estimating dense correspondences between images has remained as an active research topic in many computer vision and computational photography society. It has been popularly studied due to its usefulness in many applications, such as stereo matching [1], optical flow [2], and image stitching [3]. Although significant advances have been made towards estimating dense stereo and optical flow fields for two images adjacent in viewpoint or time, establishing dense correspondences in a scene level still remains largely unsolved [4]. There exist two principal bottlenecks that make this task very challenging; (1) photometric deformations derivded from different camera specifications, different illumination or exposure conditions. (2) geometric deformations derived from viewpoint changes, object pose changes, and non-rigid deformations [5].

To solve the inherent limitations for that task, many approaches have been proposed. First of all, existing approaches have attempted to resolve the photometric deformations by adopting robust local descriptors, *e.g.*, scale invariant feature transform (SIFT) [6], and achieved a satisfactory performance. As a pioneering work, the SIFT flow [7] had shown satisfactory results by employing SIFT descriptor [6]. However, it cannot solve large geometric variation problems. Secondly, to overcome this limitation, many approaches have tried to extend a search space into not only displacement fields but also scale and rotation fields. Inspired by the SIFT flow [7], many geometric-invariant methods have been proposed, such as scale-space SIFT flow [8] and deformable spatial pyramid matching (DSP) [9]. However, they could consider limited geometric variations. DAISY filter flow (DFF) [5] reduces the search space by leveraging a randomized sampling scheme from PatchMatch filter optimization [10], but their correspondence fields are very noisy since it cannot preserve spatial smoothness. Generalized deformable spatial pyramid [4], an extension of the DSP [9], employed global optimization to deal with possible search range effectively, and thus, it naturally needs dramatically high computational complexity. To summarize, the geometric variations make the dense correspondence problems very challenging since they induce dramatically large search space. Thus, conventional methods inevitably have high complexity for that task.



(a) Target image (b) Source image

(c) Target patch (d) APAP [3] (e) Homograph flow

Fig. 1: Comparison of warping results between geometrically varying images in large occluded area. (a) source image, (b) target image, and for (c) target patch, enlarged warping results of (d) APAP [3] and (e) Homography flow. Unlike APAP [3], the Homograpy flow provides visually plausible warping results with accurately aligned flow discontinuity.

Unlike conventional methods which exhaustively consider all possible search spaces, our approach starts from an observation that sparse correspondences can be considered as evidences to estimate a dense correspondence field. We found that such a sparse-to-dense framework can be an inherent solution to alleviate the computational burden. In fact, in stereo matching and optical flow field, a number of methods have been already proposed by employing sparse-to-dense framework [2]. However, these conventional frameworks cannot be extended to deal with large geometric variations. On the other hands, in an image stitching field, as-projective-as-possible (APAP) [3] method, extending moving least square (MLS) [11] method with local adaptation, has been proposed, which can be considered as an alternative solution for that task. However, APAP [3] can not preserve flow discontinuities effectively as shown in Fig. 1. Because it only depends on spatial smoothness constraints, regions on the flow discontinuity are also degraded, thus providing erroneous flows.

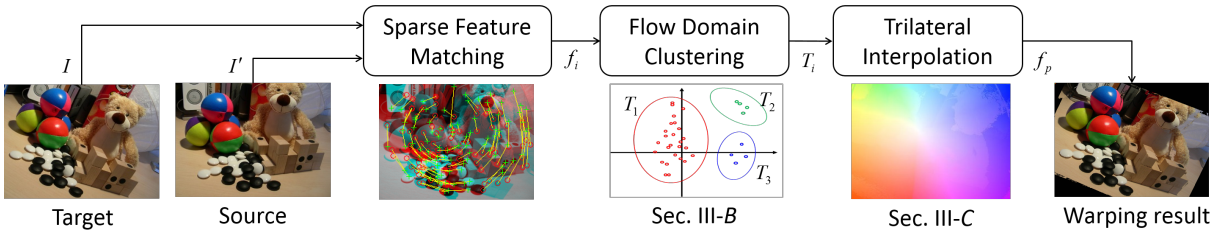


Fig. 2: Overall framework of the Homography flow.

In this paper, we propose a reliable sparse-to-dense framework, called Homography flow, to improve accuracy. Our algorithm formulates geometric variations of same planes as single plane-wise homography. Specifically, we employ clustering process on flow domain to divide the sparse matching set into plane-wise subsets, and extract a plane-wise transform from each subset. These transforms are then used as initial clues for the dense correspondence estimation. Unlike conventional method based on only spatial smoothness term, initial sparse transforms are interpolated with a trilateral weight function considering color, flow, and transform smoothness terms. Experiments show that our algorithm outperforms the state-of-the-art algorithms in terms of correspondence accuracy while computational complexity is kept very low.

II. Problem Statement and Backgrounds

Given a pair of images, source image I and target image I', our goal is to generate a spatially coherent and discontinuity-preserving flow field $f_p \in \mathbb{R}^2$ at pixel p with a very low computational complexity. When two images are taken under geometric deformations, to reliably estimate translation field f_p, additional search space for geometric variations should be considered, such as scale variation s_p and rotation variation r_p. Unlike conventional stereo or optical flow problems, considering only translation fields f_p, its search space can be defined in an infinite space, which prohibits its practical applications.

As described in above, a sparse-to-dense framework can be considered to an inherent solution for that problem. As a pioneering work, APAP [3] is also based on a propagation of sparse correspondence. It estimates a location dependent homography matrix T_p to minimize warping errors between ground control pixel q and its corresponding pixel q', as

$$x_{q'} = T_p \odot \widetilde{x}_q, \tag{1}$$

where $x_{q'}$ is the pixel coordinate and $\widetilde{x}_q = [x_q, 1]^T$. $T_p \in \mathbb{R}^{3\times3}$ defines an affine transform at the pixel p, which has 8 dimension of freedom. $T_p \odot \widetilde{x}_q$ is 1^{st} and 2^{nd} components devided by 2^{rd} components of $T_p * \widetilde{x}_q$.

Since the affine transform T_p can cover geometric variations such as scale and rotation, APAP [3] ideally can be used for dense correspondences between geometrically varying images.

However, it has inherent limitations to provide accurate displacement fields. First of all, when different flows meet at object boundaries, the local warp often fails to estimate an accurate alignment because this spatial smoothness term assumes only spatially smoothing flow field as shown in Fig. 1. It induces the necessity of additional smoothness that preserve a flow discontinuity by selecting highly related flows. Secondly, brute-force implementation of the method is very slow. Therefore, a grid version of APAP [3] is more frequently used to consider trade-off between accuracy and efficiency, where it warps all pixels within the same cell using one T_p. However, it loses some performance gain.

III. Homography Flow

A. Overview

Overall framework of the Homography flow is illustrated in Fig. 2. Our approach hypothesizes that the overall flow field consists of the optimal number of piecewise parametric models. To this end, we initialize affine transformations T_p by clustering sparse correspondence fields from sparse matching, and surface-fitting scheme (Sec. III-*B*). Furthermore, we define trilateral smoothness term for the sparse transform interpolation to preserve flow discontinuity more effectively (Sec. III-*C*). After that, dense correspondences are extracted as a simple matrix computation.

B. Sparse Transformation Candidate Generation

In the geometrically challenging situations, conventional algorithms need huge computational cost to directly estimate the dense correspondence field. We reduce the complexity by representing and estimating the entire flow field with piecewise transform models. Our algorithm initializes transform candidates from sparse correspondence $\Omega_f = \{f_1, \cdots, f_{N_s}\}$. For that initial matching, any sparse correspondence methods can be used, and in experiments, we utilize SIFT [6] to estimate scale and rotation invariant sparse correspondence.

Based on the observation that similar flows tend to be located in same plane, we employ clustering algorithm on the flow domain, and extract optimum transform candidates by using flows of each cluster. Specially, K-means clustering [13] is utilized by minimizing the enrgy function such that

$$\underset{\Phi_{\mathbf{F}}}{\operatorname{argmin}} \sum_{l=1}^{N_c} \sum_{i=1}^{N_s} \|\mathbf{F}_l - f_i\|^2, \tag{2}$$

where f_i is i^{th} initial sparse flow. We set that the number of subset clusters N_c is proportion to the number of the sparse matching points N_s since few sparse matching points of each cluster result in an unstable transform. We build optimal flow subsets $\Phi_{\mathbf{F}} = \{\mathbf{F}_1, \cdots, \mathbf{F}_{N_c}\}$, and finally cluster all sparse pixels into fixed clusters.

Now, we can estimate the transform model for each cluster. Specially, T_l can be formulated as a parametric model. For

(a) Target image (b) Source image (c) SIFT flow [7] (d) DFF [5] (e) APAP [3] (f) Homography flow

Fig. 3: Image alignment results under challenging photometric and geometric variations. Target and source images, and warping results from SIFTflow [7], DFF [5], APAP [3], and the Homography flow.

considering trade-off between stability and representation of geometric deformation, we choose "6-dof homography", also called by "affine transformation" [12]. RANSAC algorithm [14] is utilized to formulate the transform T_l of each cluster l

$$T_l = \begin{bmatrix} \mathcal{G}(l) & t(l) \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}, \tag{3}$$

where $\mathcal{G}(l)$ is a 2×2 geometric matrix, which describes scale, rotation, and shear changes. $t(l)$ is a 2×1 translation vector, and $\mathbf{0}_{1\times 2}$ is a 1×2 zero matrix. This RANSAC process is performed in homogeneous coordinate.

After these process steps, each T_l is allocated to the each cluster l. We call these points as control points. Their transforms T_p^* become initial clues for dense flow field.

C. Trilateral Interpolation

As described in above section, using only spatial smoothness term for the transform propagation provides very noisy output. In our approach, smoothness term is constructed to ensure high weights on pixels located in the same plane. For that task, energy function of the sparse transform interpolation is formulated as

$$\mathbf{E}(t^k) = \sum_p \left(h_p (t_p^k - t_p^{*,k})^2 + \lambda \sum_{q \in \mathcal{N}_4(p)} w_{p,q} (t_p^k - t_q^k)^2 \right), \tag{4}$$

where h_p is an index function, which is 1 for valid (constraint) pixels, and 0 otherwise. t_p^k and $t_p^{*,k}$ are k^{th} components of dense transform T_p and initial sparse transform T_p^*. We perform 7 iterations to optimize each component independently because two components are fixed to zero at the affine transformation. $\mathcal{N}_4(p)$ is 4-neighbors of p.

Furthermore, to construct a weight $w_{p,q}$, three features are defined. All of these features concentrate on which factors can distinguish different planes for preserving a flow discontinuity. First feature is color value c_p because pixel colors of same plane have similar values. However, this feature may divide a plane into more several parts according to their color patterns. This problem can be alleviate by leveraging flow value f_p, which is a second feature of the trilateral weight. It is based on the assumption that motion boundaries can be used to predict object boundaries. Third feature is transform value t_p, and it assumes that geometric deformations on same plane are similar. Since, however, there are no dense values of the second and last features, we first interpolate the sparse features, f_p and T_p^*, to get dense flow feature $\hat{f}_p$ and vectorized dense transform feature $\hat{t}_p$ by using the bilinear interpolation.

To summarize, the trilateral weight is extracted based on these three features as

$$w_{p,q} = \exp\left(-\frac{||c_p - c_q||^2}{{\sigma_c}^2} - \frac{||\hat{f}_p - \hat{f}_q||^2}{{\sigma_f}^2} - \frac{||\hat{t}_p - \hat{t}_q||^2}{{\sigma_t}^2} \right), \tag{5}$$

where the parameter σ_c, σ_f, and σ_t are set by user to adjust the color similarity term, flow similarity term, and transform similarity term respectively.

In this paper, we minimize the energy function $\mathbf{E}(t^k)$ with the trilateral weight by utilizing fast global image smoothing (FGS) filtering [15]. By solving the above linear system with $d\mathbf{E}(t^k)/dt^k$, the sparse transforms T^* are adaptively propagated to the same object plane. In a matrix form, minimizing the $\mathbf{E}(t^k)$ is represented such that

$$(\mathbf{H} + \lambda \mathbf{W})\mathbf{T}^k = \mathbf{H}\mathbf{T}^{*,k}, \tag{6}$$

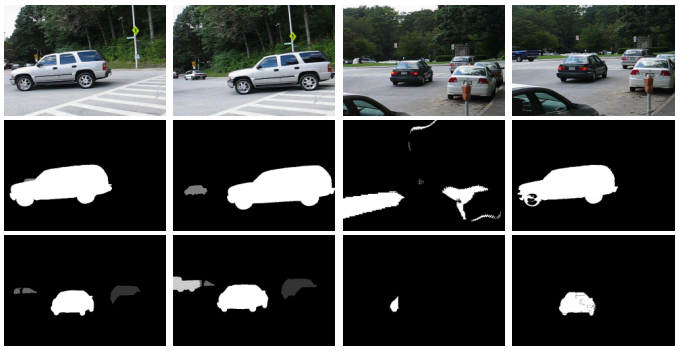


Fig. 4: Visual results of warping the source label to target on the Moseg dataset [16]. (from left to right) Target and source labels, results of APAP [3] and the Homography flow.

where $\mathbf{H}$ is a diagonal matrix whose elements are 1 for valid pixels and 0 otherwise. $\mathbf{W}$ means the matrix of the trilateral weight. $\mathbf{T}^k$ and $\mathbf{T}^{*,k}$ are k^{th} components of the final dense transform in a vector form.

Finally, we can extract dense correspondence field f_p from dense transformation field T_p such that

$$f_p = T_p \odot x_p - x_p. \tag{7}$$

IV. Experimental Results

In experiments, the proposed method is implemented with Matlab using the VLFeat [13] and the FGS [15] provided by author. We set $\sigma_c = \sigma_f = \sigma_t = 0.01$ and $\lambda = 100$ for the trilateral interpolation. N_c is set to $\lfloor N_s/10 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor operator. The proposed method is evaluated with other state-of-the-art methods, including SIFT flow [7], DFF [5], and APAP [3].

A. Visual Comparison on Non-rigid Image Pairs

We first evaluated the Homography flow with images taken under challenging non-rigid deformations. Visual comparisons with other methods are shown in Fig. 3. As expected, SIFT flow [7] is sensitive to scale and rotation variation by design, and fail to estimate accurate flow. DFF [5] cannot deal with asymmetric scale changes according to vertical and horizontal axis, as shown in the second row of Fig. 3. APAP [3] achieved quite competitive results, but showed limited performance in the occlusion (the third row of Fig. 3) since it considers only spatial smoothness. Compared to these methods, the Homography flow gives coherent alignment results.

B. Results on the Moseg Dataset [16]

We also evaluated the Homography flow on the Moseg dataset [16] including large displacements and multi-layered motions as shown in Fig. 4. On these scenarios, APAP [3] was influenced by the background flows, inducing a limited performance. In Fig. 5, the transfer accuracy was measured in image pairs of 10 frame difference. We followed the same evaluation protocol in [4], which measures overlap percentages between warping images and the ground truth using Dice coefficient. The proposed algorithm represented 91.5%, while other algorithms showed low accuracy as SIFT flow [7], DFF [5], and APAP [3] as 73.7%, 70.4%, and 25.6%.

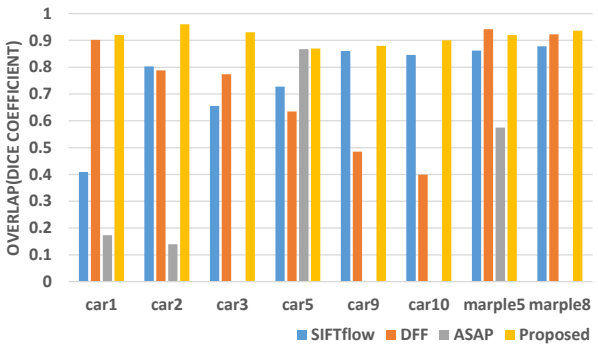


Fig. 5: Quantitative results on the Moseg dataset [16]

V. Conclusions

We proposed the Homography flow for efficiently matching image pairs taken under large photometric and geometric deformation. The Homography flow starts from the sparse to dense framework to estimate dense correspondence field in a very efficient manner. By leveraging flow domain clustering and trilateral interpolation, it extracted reliable flow field, especially in flow discontinuous area. In experiments, the Homography flow demonstrated its efficiency and robustness in various photometric and geometric challenging situations, and outperformed the state-of-the-art algorithms. In future work, the Homography flow can benefit from more powerful sparse matching algorithm.

References

[1] K. Yoon and I. Kweon, "Adaptive Support-Weight Approach for Correspondence Search," *IEEE Trans. PAMI*, vol. 28, no. 4, pp. 650-656, 2006.

[2] J. Wulff and M. J. Black, "Efficient Sparse-to-Dense Optical Flow Estimation using a Learned Basis and Layers," *In Proc. of CVPR*, 2015.

[3] J. Zaragoza, T. Chin, M. S. Brown, and D. Suter, "As-Projective-As-Possible Image Stitching with Moving DLT," *In Proc. of CVPR*, 2013.

[4] J. Hur, H. Lim, C. Park, and S. C. Ahn, "Generalized Deformable Spatial Pyramid: Geometry-Preserving Dense Correspondence Estimation," *In Proc. of CVPR*, 2015.

[5] H. Yang, W. Lin, and J. Lu, "DAISY Filter Flow: A Generalized Discrete Approach to Dense Correspondences," *In Proc. of CVPR*, 2014.

[6] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *IJCV*, vol. 60, no. 2, pp. 91-110, 2004.

[7] C. Liu and J. Yuen, "SIFT Flow: Dense Correspondence across Scenes and its Applications," *IEEE Trans. PAMI*, vol. 33, no. 5, pp. 978-994, 2011.

[8] W. Qiu, X. Wang, X. Bai, A. Yuille, and Z. Tu, "Scale-Space SIFT Flow," *In Proc. of WACV*, 2014.

[9] J. Kim, C. Liu, F. Sha, and K. Grauman, "Deformable Spatial Pyramid Matching for Fast Dense Correspondences," *In Proc. of CVPR*, 2013.

[10] J. Lu, H. Yang, D. Min, and M. N. Do, "PatchMatch Filter: Efficient Edge-Aware Filtering Meets Randomized Search for Fast Correspondence Field Estimation," *In Proc. of CVPR*, 2013.

[11] S. Schaefer, T. McPhail, and J. Warren, "Image Deformation Using Moving Least Squares," *ACM ToG*, vol. 25, no. 3, pp. 533-540, 2006.

[12] K. Park, S. Kim, S. Ryu, and K. Sohn, "Randomized Global Transformation Approach for Dense Correspondence," *In Proc. of BMVC*, 2015.

[13] A. Vedaldi and B. Fulkerson, "VLFeat: An Open and Portable Library of Computer Vision Algorithms," http://www.vlfeat.org/, 2008.

[14] Q. Tran, T. Chin, G. Carneiro, M. S. Brown, and D. Suter, "In Defence of RANSAC for Outlier Rejection in Deformable Registration," *In Proc. of ECCV*, 2012.

[15] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast Global Image Smoothing based on Weighted Least Squares," *IEEE Trans. IP*, vol. 23, no. 12, pp. 5638-5653, 2014.

[16] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," *In Proc. of ECCV*, 2010.