Randomized Global Transformation Approach for Dense Correspondence

Kihong Park
khpark7727@gmail.com
Seungryong Kim
srkim89@yonsei.ac.kr
Seungchul Ryu
ryus01@yonsei.ac.kr
Kwanghoon Sohn
khsohn@yonsei.ac.kr

School of Electrical and Electronic Engineering
Yonsei University
Seoul, Korea

Abstract

This paper describes a randomized global transformation approach to estimate dense correspondence for image pairs taken under challengingly different photometric and geometric conditions. Our approach assumes that a correspondence field consists of piecewise parametric transformation model. While conventional approaches consider large search space including flow and geometric fields exhaustively, our approach is based on an inference of optimal global transformation model from transformation candidates. To build a reliable global transformation hypothesis, we build optimal global transformation candidates with a randomized manner from an initial sparse feature correspondence, followed by a transformation clustering. Furthermore, the optimal global transformation is estimated as a cost filtering scheme with fast edge-aware filtering to provide a geometrical smoothness. Experiments demonstrate outstanding performance of our approach in terms of correspondence accuracy and computational complexity.

1 Introduction

Establishing dense visual correspondence between multiple images has been an active research area of computer vision and computational photography. Conventionally, dense correspondence researches for stereo matching [27, 40] and optical flow [7, 28], which aim to estimate dense matching fields for images adjacent in viewpoint or time, have been dramatically advanced in past decades [9, 26]. Recently, many researchers have begun to attempt to solve dense correspondence problem for more challenging images which have high variability in terms of photometric and/or geometric conditions [4, 11, 20, 25, 27, 37].

In these challenging scenarios, there exist two principal bottlenecks which make conventional methods provide limited performances; (1) photometric variations derived from different camera specifications, illumination or exposure conditions [37], and (2) geometric variations derived from viewpoint changes, object pose changes, and non-rigid deformation for objects [14]. For the first bottleneck, many robust feature descriptors have been proposed to alleviate photometric variations [21, 30]. As a pioneering work, the SIFT flow [20]

Pages 153.1-153.12
DOI: https://dx.doi.org/10.5244/C.29.153

(a) Source and target (b) SIFT flow [20] (c) SegSIFT flow [32] (d) DFF [37] (e) Proposed

Figure 1: Dense visual correspondence estimation for challenging image pairs in terms of photometric and geometric deformation. Unlike conventional methods, our approach estimates reliable correspondence from source to target image.

has shown satisfactory results on image pairs which have different but semantically similar property by employing robust SIFT descriptor [21]. However, geometric variations for the second bottleneck still remain unsolved due to its large search space, including translation (or flow), rotation, and scale, which induces extremely large computational complexity. As shown in Fig. 1(a) which has different scale and rotation deformations, the SIFT flow [20] failed to estimate reliable correspondence fields in Fig. 1(b) because the SIFT is extracted with fixed-patch. To alleviate these limitations, many researches have been popularly focused on proposing geometrically robust descriptors [18], re-designing conventional SIFT flow [16, 25, 29, 32], and employing PatchMatch strategy [3, 4, 37]. However, under the challenging geometrical variations, any feature descriptors cannot describe local patch reliably. The SIFT flow-based approaches need very high computational time since they should consider all possible label search spaces. Even with additional geometrical information such as segmentation [32], they also have limitation as Fig. 1(c). Furthermore, the PatchMatch-based approaches [37], which reduce large search space with a randomized concept, produce a plausible reconstructed image, but incorrect flow fields due to its randomized sampling property. To summarize, main bottlenecks of conventional methods come from tremendously large search space as the geometric variation for these tasks.

In this paper, we propose a randomized global transformation approach to estimate reliable correspondence between challenging image pairs having photometric and geometric deformation. Our approach starts from an intuition that geometric variations between two images can be formulated as a piecewise transformation model. Finding the optimal global transformation for each pixel enables us to infer correspondence fields without computing geometric fields. To build reliable candidates of correspondence fields, we choose a flow hypothesis by leveraging a randomized sampling scheme from initial sparse matching fields. In order to reduce repeated transformation and provide robustness to initial outliers, we employ a clustering scheme on initial global transformation candidates. By using these global transformation candidates, dense descriptor is built on target image and transformed source images respectively, which will be contributed to compute a cost volume between two image pairs. Furthermore, with fast edge-aware filtering scheme, we employ a cost filter scheme to provide a geometric smoothness. Experiments show that our algorithm outperforms the state-of-art algorithms in correspondence accuracy while computational complexity is kept very low. Our paper can be found in our project page [2].

Contributions The contribution of this paper can be summarized as follows. First of all, unlike conventional approaches that consider a large search space extensively, our approach reduces search space labels by leveraging optimal global transformation, which provides accurate correspondence results while producing very low complexity. Secondly, in order to find reliable correspondences, we employ a cost filter scheme by leveraging fast edge-aware filtering. Thirdly, our approach is very simple but provides satisfactory results, and it further

can be combined with any other feature descriptors and filtering approaches. Finally, we provide an intensive comparative study with state-of-the-art methods using various datasets.

2 Related Works

In this section, we introduce related works for our approach, by focusing on methods which have same goal with ours. Conventional optical flow methods based on variational methods [7, 28] are not tailored for the tasks, thus providing unsatisfactory performance definitely. Although some approaches were proposed based on parametric affine models [6, 15, 35] to handle large scale variations in optical flow, they inherently provide limitations for these tasks. The scale-invariant optical flow (SIOF) [36] further cannot be applied to general image deformation.

Based on the SIFT flow [20], many methods have been proposed to alleviate scale variation problems [11, 16, 25]. A scale-less SIFT flow (SLS) [11] constructs multiple scale SIFT descriptors, and projects into scale-invariant space. A scale-space SIFT flow (SSF) [25] adds scale smoothness term to the SIFT flow [20]. However, they have a critical limitation as huge computational complexity. A deformable spatial pyramid (DSP) [16] employs multiple grid graph to provide robust regularization, but it just uses the SIFT with a fixed scale and rotation, thus also producing limitations. More recently, by extending the DSP, generalized deformable spatial pyramid (GDSP) [14] was proposed to a scale and rotation invariance. It effectively preserves the meaningful inherent geometry, but it also has dramatically high complexity. On the other hands, the scale and rotation-invariant descriptor (SID) [18] was also proposed to provide geometrical deformable description. However, it is providing reliable solution for identical images only varying geometric conditions. Segmentation-aware description [32] could impose robustness to descriptors such as SIFT or SID in terms of geometric variations, but it has also limitation due to low robustness.

To reduce computational time for that tasks, a generalized PatchMatch (GPM) [4] scheme finds nearest-neighbor fields between two images by leveraging a randomized scheme in translation fields as well as geometric fields such as scale and rotation. With the PatchMatch (PM) [3], fast speed and high scene reproducibility properties are preserved. However, since it has no explicit smoothness term, it cannot estimate reliable correspondence fields although it provides plausible warping results. Recently, the DAISY Filter Flow (DFF) [37], which exploits the DAISY descriptor [30] with PatchMatch Filter (PMF) [22] based regularization, was proposed to provide scale and rotation invariance on their matching and showed well reconstructed results. However, their weak spatial smoothness, which is derived from the PM, induces local minima problem and mismatching results.

3 Randomized Global Transformation with Cost Filter

3.1 Problem Statement and Overview

Given image pairs $I_1, I_2 : \mathcal{I} \rightarrow \mathbb{R}$ or $\mathbb{R}^3$ taken under different photometrical and geometrical conditions, where $\mathcal{I} = \{p = [x_p, y_p]^T\} \subset \mathbb{N}^2$ is a discrete image domain, our goal is to estimate correspondence as a spatially coherent, discontinuity-preserving field $\mathcal{M} = \{\mu_p = [u_p, v_p]\}$. Under geometrical deformation between two images, the search space for estimating correspondence might have infinitely large possible candidates. That is, not only translation fields μ_p but also geometrically variational fields such as scale ρ_p and rotation θ_p should be considered, which can be defined as $\mathcal{F} = \{f_p = [u_p, v_p, \rho_p, \theta_p]^T\}$. To find reliable

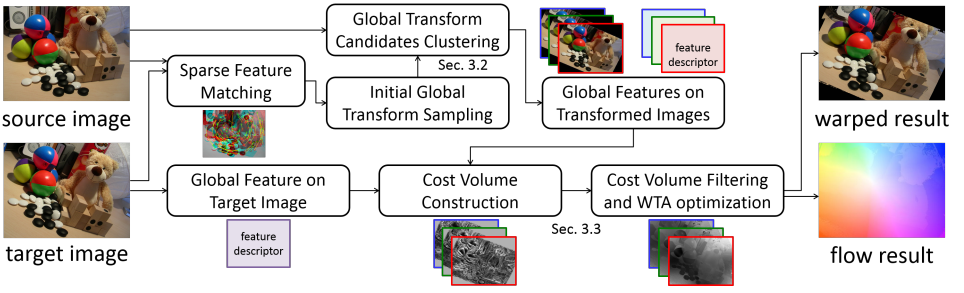


Figure 2: Framework of our randomized global transformation approach. It estimates initial global transformation candidates from sparse feature correspondences (Sec. 3.2). The cost volume between global feature descriptors are then constructed. To infer reliable correspondence, it employs a cost volume filtering using fast edge-aware filtering (Sec. 3.3).

dense matching, our approach employs dense feature descriptor $\mathcal{D}^d(p,\rho_p,\theta_p)$ defined on a local support patch centered at pixel p with scale ρ_p and rotation θ_p. The correspondence then can be derived as satisfying that $\mathcal{D}_2^d(p+\mu_p,\rho_p,\theta_p)=\mathcal{D}_1^d(p)$[1]. To estimate correspondence fields denoted as $\mathcal{F}$, the general energy function for correspondence fields $\mathbf{E}(\mathcal{F})$ can be formulated as follows:

$$\mathbf{E}(\mathcal{F})=\mathbf{E}_{data}(\mathcal{F})+\mathbf{E}_{smooth}(\mathcal{F})=\sum_{p\in\mathcal{I}}E_{data}(f_p)+\sum_{p,q\in\mathcal{N}_p}E_{smooth}(f_p,f_q), \tag{1}$$

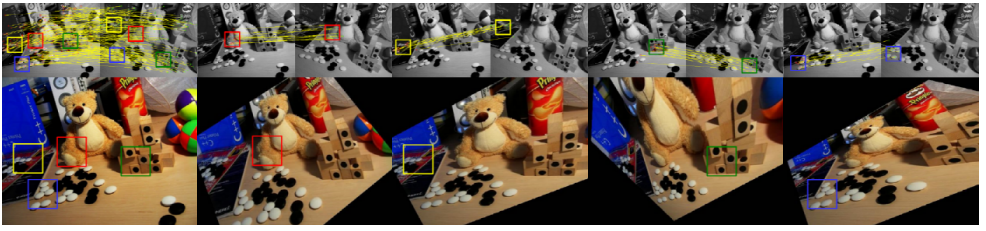
where $\mathbf{E}_{data}(\mathcal{F})$ encodes the penalty for dissimilarity for each pixel, as $E_{data}(f_p)$ is a similarity cost between $\mathcal{D}_2^d(p+\mu_p,\rho_p,\theta_p)$ and $\mathcal{D}_1^d(p)$. $\mathbf{E}_{smooth}(\mathcal{F})$ imposes the constraint that two adjacent pixels in local neighborhood $\mathcal{N}$ have similar correspondence fields. Compared to conventional methods such as stereo matching [27, 40] or optical flow [7, 28], the search space for these tasks is extended to not only flow fields u_p, v_p but also scale ρ_p and rotation θ_p fields, which makes conventional approaches have limited performance.

To overcome these limitations, unlike conventional methods which consider fully large search space fields, our approach tries to find optimal transformation candidates and infers the best transformation on each pixel. That is, we infer the best global transformation label from discrete label set $\mathcal{L}=\{1,...,N_t\}$ representing the set of N_t transformation models, *i.e.*, $\mathbf{\Phi_T}=\{\mathbf{T}_1,\cdots,\mathbf{T}_{N_t}\}$. Fig. 2 shows overall framework of our approach. To build optimal global transformation candidates, our approach employs randomized sampling in initial sparse feature matching and transformation clustering, which will be described in Sec. 3.2. Furthermore, to provide a geometrical smoothness, we reformulate smoothness term as a cost filter approach by leveraging fast edge-aware filtering, which will be described in Sec. 3.3. Finally, dense correspondence field is estimated as winner-takes-all (WTA) optimization.

3.2 Global Transformation Candidates

Our initial intuition is that geometric variations of same planes located in source and target image can be formulated as one piecewise transformation model as can be seen in Fig. 3. By finding all the planes of source image and estimating optimal transformations of their deformations, accurate correspondence fields could be obtained [38]. However, in these geometric challenging situations, it needs huge computational complexity. Thus, we limit the number of transformations to reduce the search space and the complexity. To build

[1] for the sake of simplicity $\mathcal{D}^d(p)$ is defined as a descriptor from fixed-patch to $\mathcal{D}^d(p,0,0)$.



(a) Initial matching (b) $\mathbf{T}_1$ (c) $\mathbf{T}_2$ (d) $\mathbf{T}_3$ (e) $\mathbf{T}_4$

Figure 3: Intuitions of our approach. With sparse features in (a), there exist sparse correspondence sets for global transformation which provide reliable flow on each sub-region.

optimal global transformation candidates, we first estimate sparse feature correspondences between input images as an initializer. First of all, with detected sparse features denoted as Π_1 for image I_1 and Π_2 for image I_2, sparse feature correspondence set $\Omega(I_1,I_2)$ is defined as

$$\Omega(I_1,I_2) = \{(p,\overline{p}) | \overline{p} = \text{argmin}_{q\in\Pi_2} \|\mathcal{D}_1^s(p) - \mathcal{D}_2^s(q)\|_1, p \in \Pi_1\}, \tag{2}$$

where $\mathcal{D}^s(p)$ is a sparse feature descriptor on pixel p. Note that compared to dense descriptor $\mathcal{D}^d(p)$, the $\mathcal{D}^s(p)$ is inherently described with corresponding scale and rotations. In this paper, we utilize multiple sparse features $\mathcal{D}^s(p)$ on pixel p including SIFT [21], SURF [5], and BRISK [19] with the assumption that optimal feature descriptor typically varies from pixel to pixel [13]. Within the sparse correspondence set $\Omega(I_1,I_2)$, we choose correspondence sub-set $\Lambda_i \subset \Omega(I_1,I_2)$ by a randomized sampling strategy for $i \in \{1,...,N_i\}$, where N_i is the number of initial candidate sampling. In order to reduce erroneous local minima, geometric proximity features are sampled with geometric constraints in a way that the distances among all pixels of Λ_i are lower than fixed threshold λ.

Global Transformation With sub-set Λ_i of sparse correspondence, global transformation $\mathbf{T}_i$ can be computed by a parametric model. In this paper, although other types of parametric models are also possible, we choose "6-dof homography", also called by "affine transform" [8]. This model is more stable than "8-dof homography" which is another widely used homography model because it needs the smaller matched pairs. It is formulated as follows :

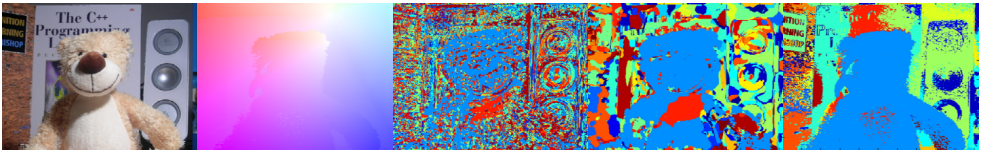
$$\mathbf{T}_i = \begin{bmatrix} \mathcal{R}(i) & t(i) \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}, \tag{3}$$

where $\mathcal{R}(i)$ is a 2×2 rotation matrix, $t(i)$ is a 2×1 translation vector, and $\mathbf{0}_{1\times 2}$ is a 1×2 zero matrix. We employ RANSAC [31] to estimate transformation $\mathbf{T}_i$. In an each iteration, by sampling 3 correspondences in Λ_i, the global transformation $\mathbf{T}_i$ is computed from $\mathbf{T}_i\mathbf{P}_1 = \mathbf{P}_2$, where $\mathbf{P}_1 = [\mathbf{p}_\varepsilon; \mathbf{1}_{1\times 3}]$ with $\mathbf{p}_\varepsilon = [p_1^T, p_2^T, p_3^T]$ from Λ_i and $\mathbf{1}_{1\times 3}$ is a 1×3 one vector. $\mathbf{P}_2$ also can be constructed in a similar way. The transformation matrix is iteratively estimated to maximize the inliers [31] with $\mathbf{T}_i = \mathbf{P}_2\mathbf{P}_1^T(\mathbf{P}_1\mathbf{P}_1^T)^{-1}$.

Clustering of Global Transformations Our approach clusters a large number of initial global transformations into optimal transformation candidates. It should be noted that any sparse feature matching scheme cannot estimate a fully reliable correspondence hypothesis. Thus, in order to reduce the effects of outliers in initial transformation from sparse correspondences, we employ the K-means clustering [33] to find optimal transformation sets as

$$\underset{\Phi_\mathbf{T}}{\text{argmin}} \sum_l \sum_i \|\text{vec}\{\mathbf{T}_l\} - \text{vec}\{\mathbf{T}_i^{\text{init}}\}\|_2, \tag{4}$$

(a) Source and target (b) WTA (c) WTA flow (d) WLS Filter [24] (e) WLS flow [24]



(f) GF Filter [12] (g) GF flow [12] (h) Index map of (b) (i) Index map of (d) (j) Index map of (f)

Figure 4: Effects of the cost volume filtering. As expected, just WTA optimization on cost volume cannot provide reliable solution with outliers. With cost filtering scheme such as WLS [24] or GF [12], we can estimate reliable correspondence with a context information.

where vec$\{\cdot\}$ is a vectorization of a matrix, and $\mathbf{T}_i^{\text{init}}$ is i-th initial transformation candidates. Because these transformation vectors contain a displacement, scale, and rotation information, we can obtain reliable transformation candidates by clustering them in this transformation domain. With the clustering scheme, we finally build optimal global transformation candidates $\Phi_{\mathbf{T}} = \{\mathbf{T}_1, \cdots, \mathbf{T}_{N_t}\}$.

3.3 Flow Inference with Cost Volume Filtering

With optimal global transformation candidates, we first build dense feature descriptors on target image and transformed source images. In this paper, we employ the DASC descriptor [17] due to its robustness and efficiency, but any dense descriptors can be used. We build $\mathcal{D}_1^d(\mathbf{T}_l \odot p)$ for l-th transformation $\mathbf{T}_l$ on source image I_1, where $\mathbf{T}_l \odot p$ is a matrix from 1st and 2nd components of $\mathbf{T}_l * [p,1]^T$, and $\mathcal{D}_2^d(p)$ for target image I_2.

With these descriptors, we construct a cost volume $\mathcal{C}(p,l)$ which encodes a similarity cost on pixel p between $\mathcal{D}_2^d(p)$ and $\mathcal{D}_1^d(\mathbf{T}_l \odot p)$. To provide the robustness, we compute truncated $L1$ distance between two descriptors as

$$\mathcal{C}(p,l) = \min(\|\mathcal{D}_1^d(\mathbf{T}_l \odot p) - \mathcal{D}_2^d(p)\|_1, \tau_c), \tag{5}$$

where τ_c is a truncated threshold to account for matching outliers and occlusions. However, the flow fields cannot be estimated reliably when computing only with WTA optimization on raw cost volume $\mathcal{C}(p,l)$ especially on erroneous local pixels or homogeneous regions.

To alleviate these limitations, we employ a cost volume filtering scheme by leveraging a fast edge-aware filtering, which was popularly used to enforce the spatial smoothness on the descriptor-based raw label cost [26]. It is worth noting that any other cost aggregation scheme can be applied to our framework. Given raw cost volume $\mathcal{C}(p,l)$, the aggregated cost $\overline{\mathcal{C}}(p,l)$ using edge-aware filtering can be computed as

$$\overline{\mathcal{C}}(p,l) = \sum\nolimits_{q \in \mathcal{W}_p} \omega_{q,p}(I_2)\mathcal{C}(q,l), \tag{6}$$

where $\mathcal{W}_p$ is the local aggregation window centered at pixel p. The weight $\omega_{q,p}(I_2)$ represents how similar two pixels p and $q \in \mathcal{W}_p$ are in terms of guidance image I_2, and is normalized, *i.e.*, $\sum_{q \in \mathcal{W}_p} \omega_{q,p}(I_2) = 1$. It can be defined with any kind of edge-aware weights [10, 12, 39]. This weighted sum better can handle outliers and local variations.

Figure 5: Comparison of dense correspondence for challenging non-rigid image pairs. (from top to bottom, from left to right) Source and target images, warping results from SIFT flow [20], SSF [25], SID [18], SegSID [32], SegSF [32], GPM [4], DFF [37], and proposed method. Compared to conventional methods, our method provides reliable performance.

With aggregated cost volume $\overline{\mathcal{C}}(p,l)$, optimal global transformation $\mathbf{T}_l$ is selected with WTA optimization by selecting an optimal labeling as follows:

$$\bar{l}_p = \text{argmin}_{l\in\{1,\ldots,N_t\}}\overline{\mathcal{C}}(p,l). \tag{7}$$

These candidate indices $\bar{l}_p$ are used to find a final flow field such that $\mu_p = p - \mathbf{T}_{\bar{l}_p} \odot p$ Fig. 4 shows the effects of cost filtering. Compared to WTA on raw cost volume $\mathcal{C}(p,l)$ in Fig. 4(b), cost filter based optimization $\overline{\mathcal{C}}(p,l)$ such as weight least square (WLS) [24] in Fig. 4(d) or guided filtering (GF) [12] in Fig. 4(f) show more accurate correspondence results. In this paper, we used GF aggregation scheme due to its robustness and low computational time. As shown in Fig. 4(e), our index map for global transformation shows edge-preserved and piecewise continuous results.

3.4 Computational Complexity Analysis

Conventional approaches, such as SSF [25] and SLS [11], have been tried to build data cost for all possible geometrical search space including scale and rotation variations, which inducing huge complexity as $O(N_\rho^2 N_\theta^2)$, where N_ρ and N_θ are the possible number of scale and rotation field, respectively. Furthermore, even though PM-based inference [3], such as

Figure 6: Comparison of dense correspondence for challenging non-rigid image pairs for a quantitative evaluation [2]. (from top to bottom, from left to right) Source and target images, warping results from SIFT flow [20], SSF [25], SID [18], SegSID [32], SegSF [32], GPM [4], DFF [37], and proposed method.

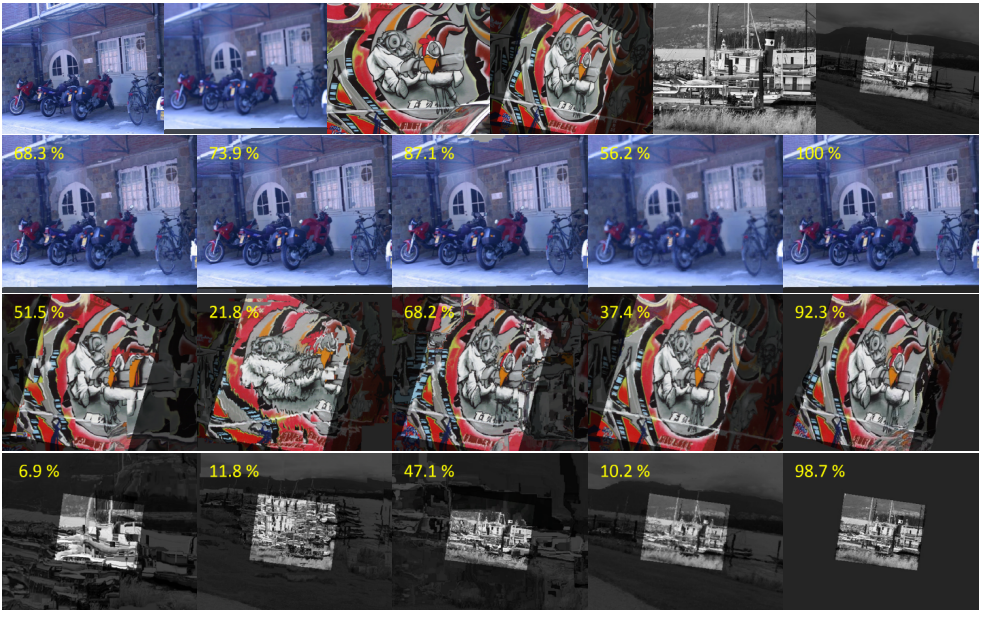
GPM [4] and DFF [37], might reduce large search space, they still have large computational time due to its iteration scheme as $O(N_\rho N_\theta \log N_\rho N_\theta)$. Compared to these conventional approaches, computational time of our approach only depends on the number of global transformation candidates N_t, which will be noted as $O(N_t)$. Since $N_t \ll N_\rho^2 N_\theta^2$, our approach enables us to reduce the complexity dramatically.

4 Experimental Results

In experiments, our approach is implemented with following same parameter settings for all datasets: $\{N_t, \lambda, \tau_c\} = \{20, 30, 10\}$. Note that any other descriptors or edge-aware filtering schemes can be applied in our framework. As described, we employed the DASC descriptor [17] and the GF filter [12]. It was also implemented in C++ on Intel Core i7-3770 CPU at 3.40 GHz, and measured the runtime on a single CPU core. In experiments, our approach was evaluated in various image datasets, such as challenging non-rigid images in Sec. 4.1, Mikolajczyk benchmark [23] in Sec. 4.2, and scaled-Middlebery benchmark [1] in Sec. 4.3. Discussions for the effects of N_t and computational times were analyzed in Sec. 4.4. Our method was evaluated with state-of-the-art methods, *e.g.*, SIFT flow [20], SSF [25], SID [18], SegSID [32], SegSF [32], GPM [4], and DFF [37].

4.1 Results on Challenging Non-rigid Image Pairs

Fig. 5 presents image warping results from estimated flow fields by computing dense correspondences from source to target images. These image pairs are non-rigid because their relationships cannot be expressed by just one global transformation [34]. We collected these image pairs among the geometric distorted images which were used in many previous researches, and thus can explain robustness property of our algorithm against various geometric challenging situation. As expected, the SIFT flow [20] cannot estimate flow fields properly on geometrical variations. Although the SSF [25] provides some improved performance with scale factors, it still showed limited performance, even worse than SIFT flow [20]. The SID[18] also cannot estimate properly because it does not describe the similarly for different scene images. Although the segmentation prior might improve matching performance, in SegSID [32] or SegSF [32], they still cannot provide reliable correspondence due to its low robustness. Furthermore, the warping from GPM [4] show plausible visual results, but it provides incorrect flow fields due to its weak smoothness. The results from DFF [37] were also unsatisfactory, which was described in details in [14]. We could evaluate quantitative performance as shown in Fig. 6. Accuracy of estimated flow was



(a) SIFT flow [20] (b) SSF [25] (c) SID [18] (d) GPM [4] Proposed

Figure 7: Comparison of dense correspondence (from top to bottom) for *Bikes*, *Graffiti*, and *Boats* on Mikolajczyk benchmark [23]. Compared to conventional methods, our approach provides reliable matching performance, even providing 100% matching rates.

measured by overlapped object labels of the warped source image and target image except background. Our approach represents 98.8% accuracy, while other approaches show low accuracy as SIFT flow [20], SSF [25], SID [18], SegSID [32], SegSIFT [32], and GPM [4] as 59.7%, 91.5%, 63.2%, 70.8%, 72.8% and 54.1%. Compared to conventional methods, our approach estimated reliably correspondence for challenging image pairs since even for challenging pairs the correspondence field can be modeled by piecewise parametric transformation models.

4.2 Results on Mikolajczyk Datasets [23]

We also evaluated our approach on the Mikolajczyk benchmark [23] as shown in Fig. 7, including various deformation such as scale, rotation, sharpness, and viewpoint. We evaluate quantitative matching performances similar to [37], in which estimated displacement errors that are lower than 15 pixels are considered as a correct. As described in above section, the SIFT flow [20] and the SSF [25] showed limited performance. For these tasks, the SID [18] showed improved performance since these datasets consist of similar image contents, but it still produced limited performance for challenging geometrical variations. However, our approach provides dramatically reliable correspondence thanks to its randomized global transformation.

4.3 Results on Scaled-Middlebury Benchmark [1]

Following qualitative experiments reported in [29], we measure the accuracy of optical flow in the presence of scale changes in Table 1. We use the Middlebury optical flow benchmark with measurements as Angular Error (AE) and Endpoint Error (EE) rates [1]. Since these

	Angular Error (AE)				Endpoint Error (EE)			
Datasets	SF[20]	SSF[25]	SID[18]	Prop.	SF[20]	SSF[25]	SID[18]	Prop.
Grove2	3.73±11	1.24±1.6	0.63±4.6	**0.16±0.3**	12.2±10	8.97±8.1	0.91±3.5	**0.66±0.3**
Grove3	3.12±2.5	1.64±1.5	1.69±5.7	**0.21±0.5**	12.6±7.8	14.0±9.6	4.81±8.4	**1.15±0.7**
Hydrangea	7.65±16	3.05±3.5	0.42±0.7	**0.13±0.2**	23.5±15	11.0±7.5	1.82±3.2	**0.83±0.6**
Urban2	3.72±8.3	1.60±2.3	**0.35±3.3**	1.50±0.9	15.6±17	8.53±7.3	**1.41±3.5**	3.31±0.3
Urban3	3.82±4.2	1.61±2.4	**0.34±0.8**	0.44±0.8	18.6±17	8.88±8.9	1.61±3.9	**1.29±0.8**
Average	4.40±8.4	1.83±2.2	0.67±3.0	**0.48±0.5**	16.5±13	10.3±8.2	2.11±4.5	**1.44±0.5**

Table 1: Comparison of quantitative evaluation on scaled-Middlebury datasets [1].

(a) Source and target (b) $N_t = 1$ (c) $N_t = 5$ (d) $N_t = 10$ (e) $N_t = 20$

Figure 8: Evaluation of our approach as varying the number of transformation candidate N_t.

images do not include scale changes, they were re-sized on both images in each pair, on to 0.7 and 0.2, respectively. Compared to conventional methods, our approach shows the best performance with the lowest average error rates.

4.4 Discussion

Fig. 8 evaluates the performance of our approach as varying N_t. In our approach, we assume that correspondence fields consist of piecewise parametric transformation model. As shown in Fig. 8(b), when estimating correspondence as only one transformation ($N_t = 1$), we cannot find reliable matching well. As increasing the N_t the correspondence performance was improved. However, with large enough N_t, the matching performance converges, which means that correspondence fields are fully covered. In terms of the computational complexity, for handling an image size 436×370, our approach averagely takes $43.21s$, while other approaches need high computational time for SIFT flow [20], SID [18], SLS [11], SegSID [32], and DFF [37] as $80.24s, 502.1s, 688.42s, 792.73s$, and $320.71s$.

5 Conclusion

This paper proposed simple but powerful correspondence estimation approach by leveraging randomized global transformation with a cost filtering for matching image pairs taken under different photometric and geometric conditions. Our approach was based on assumption that the correspondence field consists of piecewise continuous transformation model. Our approach built optimal global transformation candidates from initial sparse correspondence, followed by the transformation clustering. By using the global transformation candidates, dense descriptor was built on the transformed images. With fast edge-aware filtering scheme, we employed the cost filtering approach to find geometrical smooth correspondence fields. In experiments, our approach demonstrated its robustness in establishing dense correspondence between challenging image pairs taken under severely photometric and geometric different conditions. In further works, this performance can be improved by an occlusion handling which is based on forward-backward consistency check like [38]. Furthermore, our approach will be applied many other computer vision and computational photography as fundamental tools.

References

[1] http://vision.middlebury.edu/stereo/.

[2] http://diml.yonsei.ac.kr/RGTA/.

[3] C. Barnes, E. Shechtman, D. A. Finkelstei, and B. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. *In Proc. ACM SIGGRAPH*, 2009.

[4] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized patchmatch correspondence algorithm. *In Proc. of ECCV*, 2010.

[5] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *In Proc. of ECCV*, 2006.

[6] M. J. Black and P Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVIU*, 63:75–104, 1996.

[7] D. Butler, J. Wulff, G. Stanley, and M. Black. A naturalistic open source movie for optical flow evaluation. *In Proc. of ECCV*, 2012.

[8] H. Y. Chen, Y. Y. Lin, and B. Y. Chen. Robust feature matching with alternate hough and inverted hough transforms. *In Proc. of CVPR*, 2013.

[9] L. De-Maeztu, A. Villanueva, and R. Cabeza. Near real-time stereo matching using geodesic diffusion. *IEEE Trans. PAMI*, 34(2):410–416, 2012.

[10] E. Gastal and M. Oliveira. Domain transform for edge-aware image and video processing. *In Proc. of ACM SIGGRAGH*, 2011.

[11] T. Hassner, V. Mayzels, and L. Zelnik-Manor. On sifts and their scales. *In Proc. of CVPR*, 2012.

[12] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Trans. PAMI*, 35(6):1397–1409, 2013.

[13] K. Hsu, Y. Lin, and Y. Chuang. Robust image alignment with multiple feature descriptors and matching-guided neighborhoods. *In Proc. of CVPR*, 2015.

[14] J. Hur, H. Lim, C. Park, and S. C. Ahn. Generalized deformable spatial pyramid: Geometry-preserving dense correspondence estimation. *In Proc. of CVPR*, 2015.

[15] S. X. Ju, M. J. Black, and A. D. Jepson. Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. *In Proc. of CVPR*, 1996.

[16] J. Kim, C. Liu, F. Sha, and K. Grauman. Deformable spatial pyramid matching for fast dense correspondences. *In Proc. of CVPR*, 2013.

[17] S. Kim, D. Min, B. Ham, S. Ryu, D. Minh, and K. Sohn. Dasc: Dense adaptive self-correlation descriptor for multi-modal and multi-spectral correspondence. *In Proc. of CVPR*, 2015.

[18] I. Kokkinos and A. Yuille. Scale invariance without scale selection. *In Proc. of CVPR*, 2008.

[19] S. Leutenegger, M. Chli, and R. Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. *In Proc. of ICCV*, 2011.

[20] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans. PAMI*, 33(5):978–994, 2011.

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[22] 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.

[23] K. Mikolajczyk, T. Tuytelaars, C. Schmid, J. Matas, and F. Schaffalitzky. Cross-based local multipoint filtering. *In Proc. of CVPR*, 2005.

[24] 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*, 23(12):5638–5653, 2014.

[25] W. Qiu, X. Wang, X. Bai, A. Yuille, and Z. Tu. Scale-space sift flow. *In Proc. of WACV*, 2014.

[26] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *In Proc. of CVPR*, 2011.

[27] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, 2002.

[28] D. Sun, S. Roth, and M. Black. Secrets of optical flow estimation and their principles. *In Proc. of CVPR*, 2010.

[29] M. Tau and T. Hassner. Dense correspondences across scenes and scales. *In Proc. of CVPR*, 2014.

[30] E. Tola, V. Lenpetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Trans. PAMI*, 32(5):815–830, 2010.

[31] 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.

[32] E. Trulls, I. Kokkinos, A. Sanfeliu, and F. M. Noguer. Dense segmentation-aware descriptors. *In Proc. of CVPR*, 2013.

[33] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. 2008.

[34] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. Deepflow: Large displacement optical flow with deep matching. *In Proc. of ICCV*, 2013.

[35] L. Xu, J. Chen, and J. Jia. A segmentation based variational model for accurate optical flow estimation. *In Proc. of ECCV*, 2008.

[36] L. Xu, Z. Dai, and J. Jia. Scale invariant optical flow. *In Proc. of ECCV*, 2012.

[37] H. Yang, W. Lin, and J. Lu. Daisy filter flow: A generalized discrete approach to dense correspondences. *In Proc. of CVPR*, 2014.

[38] J. Yang and H. Li. Dense, accurate optical flow estimation with piecewise parametric model. *In Proc. of CVPR*, 2015.

[39] Q. Yang, K. Tan, and N. Ahuja. Real-time o(1) bilateral filtering. *In Proc. of CVPR*, 2009.

[40] K. Yoon and I Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. PAMI*, 28(4):650–656, 2006.