Regional Foremost Matching for Internet Scene Images

Xiaoyong Shen Xin Tao Chao Zhou Hongyun Gao Jiaya Jia*
The Chinese University of Hong Kong
http://www.cse.cuhk.edu.hk/leojia/projects/internetscenematch

Figure 1: *Our regional foremost matching for Internet images estimates accurate regional correspondence and enables several applications.*

Abstract

We analyze the dense matching problem for Internet scene images based on the fact that commonly only part of images can be matched due to the variation of view angle, motion, objects, etc. We thus propose *regional foremost matching* to reject outlier matching points while still producing dense high-quality correspondence in the remaining foremost regions. Our system initializes sparse correspondence, propagates matching with model fitting and optimization, and detects foremost regions robustly. We apply our method to several applications, including time-lapse sequence generation, Internet photo composition, automatic image morphing, and automatic rephotography.

Keywords: image matching, alignment, time-lapse images, Internet images, registration

Concepts: •**Computing methodologies** → **Image processing;**

*e-mail:{xyshen, xtao, zhouc, hygao, leojia}@cse.cuhk.edu.hk

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. © 2016 ACM.
SIGGRAPH ASIA 2016, December 5-8, 2016, MACAO
ISBN: 978-1-4503-4514-9/16/12
DOI: http://dx.doi.org/10.1145/2980179.2980249

ACM Reference Format
Shen, X., Tao, X., Zhou, C., Gao, H., Jia, J. 2016. Regional Foremost Matching for Internet Scene Images. ACM Trans. Graph. 35, 6, Article 178 (November 2016), 12 pages. DOI = 10.1145/2980179.2980249 http://doi.acm.org/10.1145/2980179.2980249.

1 Introduction

Establishing pixel correspondence between images is a long-standing problem due to its fundamental importance for many applications in computer graphics, computer vision, and image processing. Researchers have proposed a number of methods. For example, optical flow methods [Horn and Schunck 1981] estimate motion in consecutive frames for the same scene, and stereo matching [Rhemann et al. 2011] estimates image disparities. For rigid transform caused by camera position change, homography-based models [Liu et al. 2011b] can be applied.

In this paper, we aim differently to match complicated Internet scene images in pixel level to broadly benefit applications, such as scene reconstruction and visualization [Snavely et al. 2006], image/video composition [Chen et al. 2009; Martin-Brualla et al. 2015a], and image editing/enhancement [Zhang et al. 2014; Chia et al. 2011; HaCohen et al. 2013; HaCohen et al. 2011].

Dense matching on Internet scene images is by not easy for even state-of-the-art methods. The inherent difficulty stems from the large change in color, resolution, viewpoint, camera parameters, time to capture images, weather, lighting, moving objects, and possible post-processing.

1.1 Brief Evaluation of Existing Methods

We give one example in Figure 2(a)-(b) to demonstrate the difficulty of image matching by state-of-the-art methods, which include

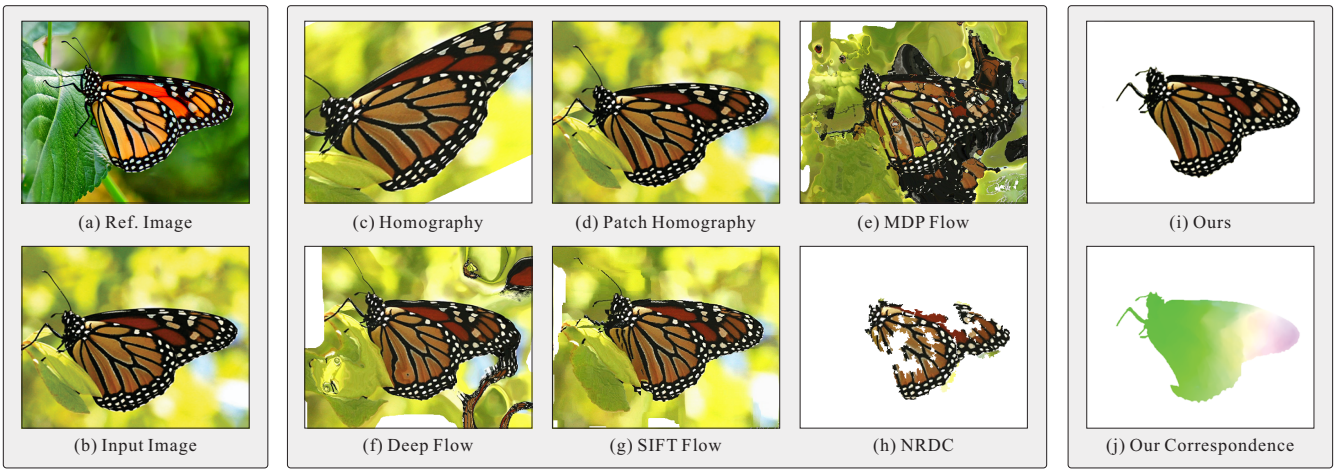


Figure 2: *Comparisons of state-of-the-art matching methods. (a) and (b) are the reference and input images, respectively. (c-h) are the warping results of (b) according to the correspondence to (a) by different methods. (i) shows our regional foremost warping result and (j) is the color-coded correspondence.*

homography, patch homography [Liu et al. 2013], MDP Flow [Xu et al. 2012], DeepFlow [Weinzaepfel et al. 2013], SIFT flow [Liu et al. 2011a] and NRDC [HaCohen et al. 2011]. The two input images are butterflies with similar patterns, which seem to be an easy case for matching.

SIFT features [Lowe 2004] can help find sparsely matched points. Although patch-based homography [Liu et al. 2011b] fits transform in grids, matching accuracy is still an issue due to the difficulty to segment images accurately, as shown in (d), especially when image motion violates the homography assumption. Optical flow methods [Xu et al. 2012; Weinzaepfel et al. 2013] process temporally-close video frames. They do not work similarly well on Internet images. The errors in (e) and (f) are also caused by aforementioned significant appearance and structure variation. SIFT flow [Liu et al. 2011a], on the other hand, may not produce high-quality results for non-SIFT-feature points.

Following another line, PatchMatch [Barnes et al. 2009] based methods can densely find similar patches with consideration of structure variation. We also evaluate the NRDC method [HaCohen et al. 2011], which includes random search in color transform space. The result in (h) shows that only a small portion of the image can find correspondence.

1.2 Problem Definition and Our System

Why do these methods not work well on these seemingly easy butterfly images? An intriguing finding is that Internet scene images generally contain reasonable-size objects, content, or regions, which are important for us to find correspondence. Meanwhile there are always pixels that cannot be matched, such as the different backgrounds in the butterfly images.

We thus advocate the concept of *regional foremost dense matching* with the goal to capture common contents and densely align them to form large-size confident area while eliminating other pixels – for example, different sky regions in individual images. Our system produces displacement maps as shown in Figure 2(i)-(j). Their properties are threefold.

- Displacement is in the form of large-size confident regions.
- Within the regions, dense accurate matching is yielded.
- Out of them, pixels matching is neither possible nor necessary.

Our system includes several steps to achieve these goals. The initial patch matching step produces a good number of features while eliminating outliers using an iterative update scheme. It quickly removes false matching and locates the object region for correspondence establishment. This initial estimate is then fed to propagation for reliable per-pixel displacement estimation, which benefits from model fitting and variational optimization. Finally, with the high quality displacement map, we employ a fully connected pairwise conditional random field (CRF) model to compute the foremost region. In each phase of our computation, we compare our results with others to demonstrate the effectiveness of our method.

In addition, we present applications on challenging image data. We demonstrate time-lapse sequence generation from Internet photos, Internet image composition, image morphing, view interpolation, difference spotting, and image enhancement.

2 Related Work

We review image correspondence estimation and Internet image based applications in this section.

2.1 Image Correspondence Estimation

Dense Matching For image pairs captured in the same scene with intensity or gradient consistency, the correspondence can be computed by optical flow methods following the variational framework of Horn and Schunck [1981]. In these models, data terms are employed to enforce color or gradient consistency [Brox et al. 2004; Bruhn and Weickert 2005]. Regularization terms are usually in the form of robust functions, such as L_1 norm and Charbonnier function [Brox et al. 2004; Black and Anandan 1996; Wedel et al. 2008; Zach et al. 2007], to achieve piece-wise smooth flow fields.

To handle large motion, Brox et al. [2011] and Weinzaepfel et al. [2013] enhanced the variational framework by incorporating the sparse descriptor matching term. Xu et al. [2012] improved the coarse-to-fine variational framework by fusing feature matches in each iteration to handle the large motion problem. Recently, Revaud et al. [2015] directly interpolated DeepMatching correspondence using geodesic distance. These methods are successful in optical flow estimation, but are still not suitable for images with intensity or gradient inconsistency.

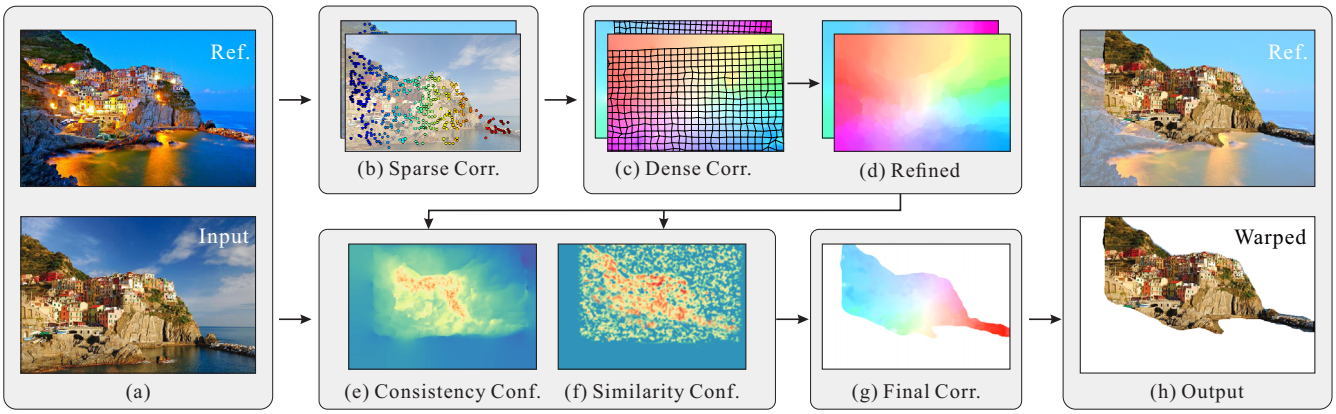


Figure 3: *Illustration of our framework. (a) Two input images. (b) Our initial correspondence. (c) Dense correspondence propagated from (b) (color encoded for visualization). (d) Refined correspondence based on (c). (e)-(f) Consistency and similarity confidence. (g) Final regional foremost correspondence. (h) Output foremost matching region.*

Based on the variational framework and SIFT features, SIFT Flow [Liu et al. 2011a] was proposed for general scene image matching. Tau et al. [2016] extended it by adding scale selection to handle object scale variation. By applying deformable spatial pyramids, Kim et al. [2013] handled multi-scale object matching. Besides SIFT features, Yang et al. [2014] applied DAISY descriptors and densely estimated correspondence by a discrete labeling process. Unlike our method, all pixels are matched globally.

Fast approximate nearest-neighbor field (NNF) strategies, such as PatchMatch [Barnes et al. 2009; Barnes et al. 2010], can be applied as well. Bao et al. [2014] proposed edge-preserving PatchMatch for optical flow estimation and recently Bailer et al. [2015] computed NNF to handle large motion and then rejected outliers via a coarse-to-fine scheme. HaCohen et al. [2011] proposed a framework where a coarse-to-fine scheme and a global nonlinear color transform model were used, together with local geometry consistency, to enforce region consistency. Sen et al. [2012] improved the framework for different-exposure image matching under the bidirectional similarity constraints. Matching outliers are still hard to remove, which are however common for Internet images.

Sparse Feature Based Schemes Feature based methods find sparse points and fit matching models like patch homography [Liu et al. 2013] and bounded distortion [Lipman et al. 2014]. To generate dense correspondence, smoothing regularization, such as thin-plate spline (TPS) and harmonic functions, is incorporated. Yücer et al. [2012] estimated homography by combining bounded biharmonic weights in an optimization framework. To reject feature matching outliers, Liu et al. [2013] divided images into regular grids and estimated global transform in each grid by RANSAC. To improve efficiency, a hierarchical estimation scheme was developed in [Liu et al. 2014]. The features adopted in these frameworks can be SIFT [Lowe 2004], SURF [Bay et al. 2006], learning based descriptors [Dosovitskiy et al. 2014; Simo-Serra et al. 2015; Simonyan et al. 2014], etc. We note this line of methods rely on the homography motion model, which may not be satisfied in Internet images.

2.2 Internet Image Applications

Internet provides a great amount of photos. They were used in scene reconstruction and visualization, such as photo tourism [Snavely et al. 2006]. A survey is provided in [Noah 2011]. Basic techniques are structure-from-motion and multi-view stereo, which were presented in [Snavely et al. 2008; Furukawa et al. 2010; Agarwal et al. 2009].

Using data-driven methods, Chen et al. [2009] proposed the skecth2photo system, which employs Internet images to composite semantic and sketch based photos. Martin-Brualla et al. [2015a; 2015b] computed video time-lapse by mining Internet photos. Image editing can be also achieved in enhancement [Johnson et al. 2011; Zhang et al. 2014; HaCohen et al. 2011; Shih et al. 2014; Joshi et al. 2010; Yan et al. 2016], recoloring [Chia et al. 2011], deblurring [HaCohen et al. 2013], completion [Hays and Efros 2007], and restoration [Dale et al. 2009]. Internet image photometric stereo was presented in [Shi et al. 2014; Shen and Tan 2009].

3 Our Framework

Our method is composed of three steps illustrated in Figure 3. We first detect a good number of corresponding points for initialization. Then we propagate them to other pixels and estimate the regional foremost displacement map.

In the following, we take two images I_1 and I_2 as inputs. More images can be progressively processed. Let p index image pixels. We estimate a displacement vector $w_p = (u_p, v_p)^T$ in I_1 for the foremost region pixel p as shown in Figure 3(g). The vector w_p makes pixel p in I_1 correspond to $p + w_p$ in I_2. $I_{1,p}$ and $I_{2,p}$ are intensities of pixel p in I_1 and I_2 respectively.

3.1 Initial Sparse Correspondence

Image sparse correspondence is often established by finding the nearest neighbors of feature descriptors. Then outliers can be rejected through RANSAC or bounded distortion matching [Lipman et al. 2014]. However, this process may produce a small number of matched points due to large image-structure diversity, as shown in 6(a) and (b). We instead adopt the nonrigid-matching method below to find more correspondences.

Region Similarity Measure Our initial nonrigid-matching is similar to DeepMatching [Weinzaepfel et al. 2013; Revaud et al. 2016] for its hierarchical matching structure as shown in Figure 4(a) and (b) where each rectangular patch in image I_1 is decomposed into four smaller ones in the lower level for finer matching in image I_2. Figure 4(d) shows the tree structure. We modify it by introducing a new matching measure based on confidence.

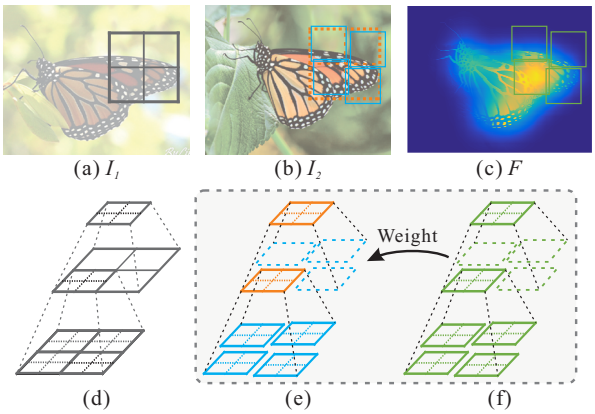


Figure 4: *Illustration of our first step for sparse matching. (a) and (b) are images I_1 and I_2 respectively. (c) is the regional foremost map of I_2. (d-f) are the matching quadtrees of (a)-(c), respectively.*

Algorithm 1 Iterative Confidence and Match Update

Input: Initial constant $\mathcal{F}$.
Procedure:
  for n iterations **do**
    Calculate matching map w.
    Update $\mathcal{F}$ based on the new w map.
  end for
Output: estimated w map in step 1.

We build a quadtree for each patch in the input two images by separating each patch to four smaller ones in the next level. When we calculate the matching score of two patches in level i, the four smaller child patches are considered to calculate finer matching scores in level $i+1$, as shown in Figure 4(d)-(f). So the overall procedure is to start from atomic patches in the leaves and go up to aggregate their measures by a *min* pooling operator until the root patch is reached [Weinzaepfel et al. 2013; Revaud et al. 2016].

In our system, we modify the aggregation term. For each patch $I^{(i)}_{1,p}$ in level i centered at p in image I_1, the aggregation process is expressed as

$$\mathcal{S}(I^{(i)}_{1,p}, I^{(i)}_{2,p+w_p}) = \varphi\Big(\sum \min_{w_p} \alpha^{(i+1)}_{p+w_p} \mathcal{S}(I^{(i+1)}_{1,p}, I^{(i+1)}_{2,p+w_p})\Big), \quad (1)$$

where $I^{(i)}_{2,p+w_p}$ is the found patch to match $I^{(i)}_{1,p}$ and w_p is the displacement. Their patch similarity score $\mathcal{S}$ comes from the $i+1$th level patches with similarity recursively defined in Eq. (1). The leaf-level patch similarity is calculated as the DAISY descriptor [Tola et al. 2008] L_2-norm distance. It is always positive and becomes small when the two patches are similar. α is the weight to indicate the importance of differently matched patches, which will be detailed later. $\varphi(x) = \sqrt{x^2+\epsilon^2}$ is a robust function to reject outliers where ϵ is a small positive value $1e-4$ in all our experiments. Since we estimate correspondence for each patch and discard low-score correspondence, the initial displacement map w is sparse. We still call each point in w a *feature* for simplicity.

Difference from DeepMatching Our similarity measure in Eq. (1) rejects outliers based on weight map α. α_p is set to $\exp(-\mathcal{F}_p)$ for every pixel p with the confidence map $\mathcal{F}$ shown in Figure 4(c). α is small when the patch is matched, which requires a good confidence. Contrarily, a large α is used when matching is unreliable.

The $\mathcal{F}$ map is thus important. We apply an iterative procedure to update it along with correspondence w, as outlined in Algorithm 1 and illustrated in Figure 5. Initial $\mathcal{F}$ is constant for estimating a uniform correspondence map w. It is then updated with the new w by simple but effective consistency check – two points are with a high confidence when matching from I_1 to I_2 is consistent with matching from I_2 to I_1.

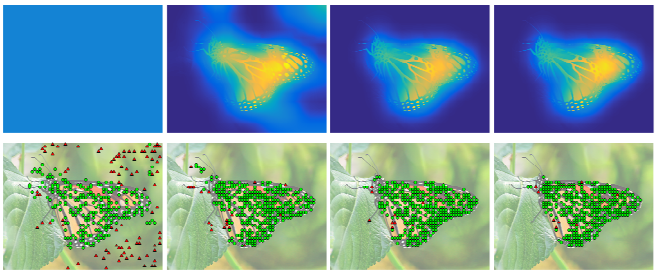


Figure 5: *Update of confidence $\mathcal{F}$ and correspondence w iteratively. The green dots in w are matched points while red ones are false matching. The input images are shown in Figure 2(a)-(b).*

To evaluate confidence for point p in I_1, we find its matching point $p+w'_p$ in I_2. Since $p+w'_p$ may not be a feature point computed in above process, we find its nearest neighbor p^* and match it back to I_1. Suppose now its displacement is w_{p^*}. The ideal situation is that $d_p = \|p - p^* - w_{p^*}\| = 0$. If it is not, the value indicates how reliable the match is. For distance d_p smaller than 15 pixels, we set $\mathcal{F}_p = 1$. Otherwise $\mathcal{F}_p = 0$. The confidence is propagated to other non-feature pixels by joint bilateral filtering [Tomasi and Manduchi 1998] with guidance I_2. Finally, we linearize values between 0 and 1. The new confidence map is then fed into correspondence calculation in Eq. (1) again in next iteration.

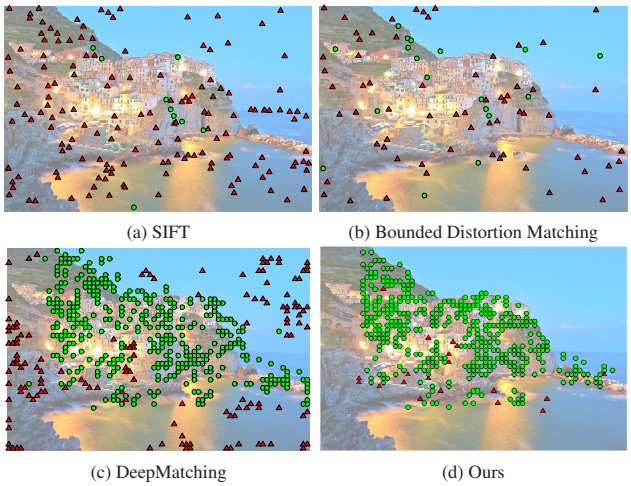
As illustrated in Figure 5, our iterative scheme quickly rejects incorrect matching and improves accuracy. The result in (b) shows that one iteration already suppresses outliers in the background. Results from later passes are with small change. Thus in our experiments, only two-pass computation is performed.

Simple Comparison with Other Methods We compare our initial feature matching with SIFT [Lowe 2004], bounded distortion matching[Lipman et al. 2014], and DeepMatching [Weinzaepfel et al. 2013; Revaud et al. 2016], as shown in Figure 6(a)-(c). The first two methods find much less points since the feature requirement is high. Original DeepMatching [Weinzaepfel et al. 2013; Revaud et al. 2016] produces more correspondences than these methods. But it does not consider the fact that many regions/pixels cannot be matched in the other image. Our result in (d) has a good number of correspondences and reliably reduces false matching. More results are presented later.

3.2 From Sparse to Per-pixel Correspondence

In order to densify the correspondence, we in this step propagate the initial sparse match w to all pixels. Instead of applying interpolation [Alexa et al. 2003; Wasserman 2013] or performing variational optical flow optimization [Brox and Malik 2011], we fit local regional models and then refine displacements. This strategy deals with large displacements and is not easily stuck in local minima. It also takes advantage of our initial correspondence.

Sparse Correspondence Propagation Given the good number of initial correspondence estimates as shown in Figure 6(d), we apply regional homography model fitting, as illustrated in Figure 3(c). The transform matrix H_p for each pixel p is related to the displace-



(a) SIFT (b) Bounded Distortion Matching

(c) DeepMatching (d) Ours

Figure 6: *Comparisons of sparse matching. (a) and (b) are the SIFT and bounded distortion matching results, respectively. (c) is the DeepMatching result and (d) is ours. The green dots indicate the matched points while the red ones mark false matching. The input images are shown in Figure 3(a).*

ment w as

$$H_p\hat{p} = \hat{p} + \hat{w}_p, \tag{2}$$

where $\hat{p}$ and $\hat{w}_p$ are the homogeneous coordinates of p and w_p respectively. To fit H_p, we apply the moving least square (MLS) method expressed as

$$\min_{H_p} \sum_q \mu(p,q)\|H_p\hat{q} - \hat{q} - \hat{w}_q\|^2, \tag{3}$$

where $\mu(p,q)$ is the weight to measure the confidence of $\hat{w}_q$ with respect to H_p. $\mu(p,q)$ is set to 0 when there is no sparse correspondence in q. Otherwise, we set $\mu(p,q) = \exp(-\beta_1\|I_{1,p} - I_{1,q}\| - \beta_2\|p-q\|)$, which is the bilateral weight measuring the confidence considering spatial and range distances. We set β_1 and β_2 to 10 and 0.01 respectively in our experiments.

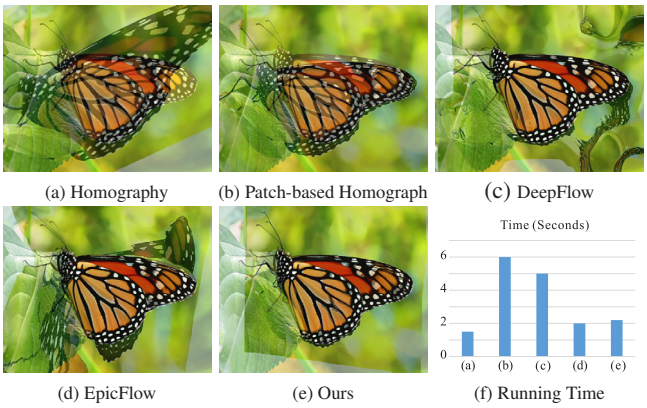
We quickly solve for H_p in the MLS system using the method suggested in [Zaragoza et al. 2014]. For further speed-up, we only compute H_p in rectangular grid nodes, as shown in Figure 3(c), and interpolate them for other pixels using the Shepard method [Shepard 1968] with Gaussian kernels. By converting H_p into displacement vectors, we get the updated displacement map w containing all pixels.

Dense Correspondence Refinement We then refine the propagated correspondence with even higher accuracy, as illustrated in Figure 3(d). We employ the well-studied variational framework to optimize

$$E(w) = \sum_p \Big(\varphi(\|\mathcal{D}(I_{1,p}) - \mathcal{D}(I_{2,p+w_p})\|) + \lambda\varphi(\|\nabla w_p\|)\Big), \tag{4}$$

where $\varphi(\|\mathcal{D}(I_{1,p}) - \mathcal{D}(I_{2,p+w_p})\|)$ is the data term to measure matching similarity and the other term is a regularizer. $\varphi(x)$ is the robust function also used before in Eq. (1) to reject outliers. λ balances the two terms.

$\mathcal{D}(I_{1,p})$ and $\mathcal{D}(I_{2,p})$ are the descriptors to measure image structure similarity. We use absolute gradient magnitudes $\mathcal{D}(I_{1,p}) = \|\nabla I_{1,p}\|$ for this measure since corresponding pixels can be with different gradient directions caused by variation of lighting and motion. Pixel-gradient based features can also be optimized efficiently compared with other patch-based ones.



(a) Homography (b) Patch-based Homograph (c) DeepFlow

(d) EpicFlow (e) Ours (f) Running Time

Figure 7: *Comparison of different propagation methods. The input images are shown in Figure 2(a) and (b). We warp the input image to reference and blend them by different methods. (a-d) are the matching results of global homograph, patch-based homograph, DeepFlow and EpicFlow respectively. (e) is our result. (f) shows running time of each method.*

We solve Eq. (4) on the original resolution because we have already good-quality correspondence w. Thus the coarse-to-fine model [Xu et al. 2012] that was commonly employed is not needed any more, saving much computation in our system. Also our system only requires a small number of iterations compared with general initialization. A brief evaluation is given below.

Brief Evaluation In evaluation, we warp input images and blend them with reference ones. The results are shown in Figure 7. It is notable that global and patch-based homography [Liu et al. 2011b] results shown in (a) and (b) are with errors caused by nonrigid body transform. Our method improves accuracy.

EpicFlow [Revaud et al. 2015] starts with segmentation and applies interpolation in segmented regions according to the geodesic distance of different regions. It relies on the segmentation quality. In contrast, we do not perform segmentation. Dense initial matching and robust propagation are two major factors to the success of our method. Our running time until this step is shown in (f), indicating the potential to develop efficient implementation.

3.3 Regional Foremost Correspondence

Our final step is to detect reliable dense correspondence and form the regional foremost map. Pixels that cannot be matched are removed due to the nature of Internet images with different content coverage. One example is shown in Figure 3(g). We denote the foremost correspondence map as o. $o_p = 1$ if pixel p is in the matchable region; otherwise $o_p = 0$. To produce this map, the following conditions are considered.

- **Structure Similarity** Corresponding pixels are generally with high structure similarity.
- **Matching Consistency** Estimates from I_1 to I_2 should be consistent with those from I_2 to I_1.
- **Coherence** The foremost-region size in o is reasonably large for matching of common content in scene images.

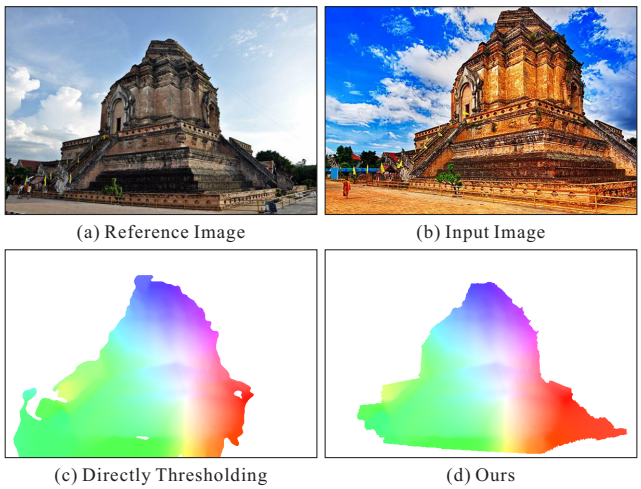


Figure 8: *Regional foremost labeling. (a) and (b) are the input images. (c) is the regional foremost correspondence by directly thresholding the confidence map $\mathcal{C}$. (d) is our result.*

Regional Foremost Labeling We model our final step as a labeling problem with a fully connected pairwise conditional random field (CRF) model [Krähenbühl and Koltun 2011] as

$$E(o) = \sum_p \phi_u(w_p, o_p) + \sum_{p<q} \phi_p(o_p, o_q), \tag{5}$$

where o is the binary mask to compute. The unary term ϕ_u models the confidence of regional foremost correspondence and the pairwise term ϕ_p is for extra constraints given below.

Unary Term $\phi_u(w_p, o_p)$ The unary term is based on image structure similarity and correspondence consistency. The confidence of matching is measured by structure similarity using normalized cross correlation (NCC) between patch p in I_1 and patch $p + w_p$ in I_2 as shown in Figure 3(f). It is expressed as

$$\mathcal{D}(w_p) = 1 - \left| \frac{\rho(I_{1,p}, I_{2,p+w_p})}{\sqrt{\sigma(I_{1,p})\sigma(I_{2,p+w_p}) + \epsilon}} \right|, \tag{6}$$

where $\rho(I_{1,p}, I_{2,p+w_p})$ is the covariance of intensities. $\sigma(I_{1,p})$ and $\sigma(I_{2,p+w_p})$ are the intensity variance. ϵ is a very small value to handle the case where the two patches both have no structure. $\mathcal{D}(w_p)$ is close to zero when the two patches are corresponding.

Another confidence measure is the bidirectional correspondence consistency similar to that used in our matching initialization (Section 3.1). Denoting the estimated correspondence from I_2 to I_1 as w', the confidence $\mathcal{C}(w_p)$ is denoted as the sum of consistency errors as $\mathcal{C}(w_p) = \|w_p + w'_{p+w_p}\|$.

Based on these two confidence measures, o_p approaches 1 when $\mathcal{S}(w_p)$ and $\mathcal{C}(w_p)$ are small. Otherwise, it goes to zero. We define the unary term $\phi_u(w_p, o_p)$ as

$$\phi_u(w_p, o_p) = g(w_p)o_p + (1 - g(w_p))(1 - o_p), \tag{7}$$

where $g(w_p)$ is the cost for the case that $o_p = 1$, and $1 - g(w_p)$ is the cost when $o_p = 0$. We define $g(w_p)$ using the logistic function $\ell(x, k) = \frac{1}{1+e^{-(x-k)}}$ according to the confidence measure

$$g(w_p) = \ell(\mathcal{D}(w_p) + \gamma\mathcal{C}(w_p), k), \tag{8}$$

where γ is the weight for the two confidence measures and k makes $g(w_p) = 0.5$ when $\mathcal{S}(w_p) + \gamma C(w_p)$ equals to k. We will discuss parameter setting more in our experiment section.

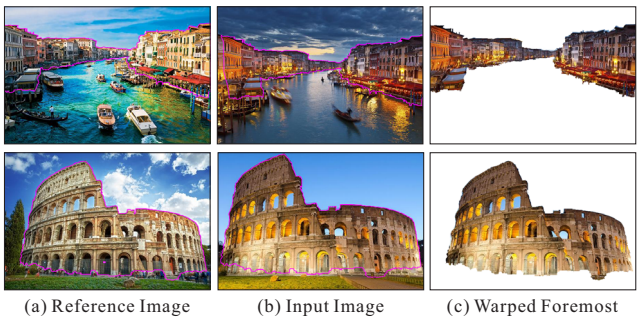


Figure 9: *Difficult cases. (a) and (b) are the input images. We highlight corresponding foremost region boundaries in purple. (c) shows the accurately warped foremost regions from (b) to (a).*

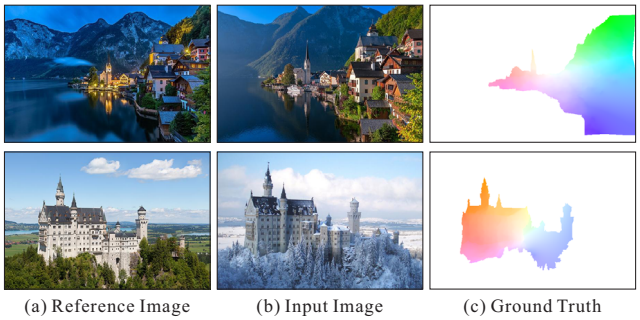


Figure 10: *Examples in our benchmark dataset. (a) and (b) are the reference and input images respectively. (c) is our color-coded ground truth regional foremost correspondence map.*

We note only using one of the measures degrades system performance. As can be observed from Figure 3(e)-(f), using either of the measures misses out a set of points. Our combination makes the system perform better.

Pair-wise Term $\phi_p(o_p, o_q)$ The pair-wise term models underlying structure of o. We apply the bilateral weights to set $\phi_p(o_p, o_q)$ as

$$\phi_p(o_p, o_q) = \exp(-\frac{\|p - q\|}{\sigma_s} - \frac{\|I_{1,p} - I_{1,q}\|}{\sigma_r})\|o_p - o_q\|, \tag{9}$$

where σ_s and σ_r are the spatial and range weights respectively. A large σ_s enforces strong spatial coherence of o and a small σ_r benefits o to be consistent with image structure.

With above terms, Eq. (5) forms the pair-wise CRF, which can be efficiently inferred by the method of [Krähenbühl and Koltun 2011]. One result is shown in Figure 3(g).

We show one more result in Figure 8 to illustrate the effectiveness of our regional foremost labeling. Starting from the same dense correspondence yielded from previous steps in our system, (c) is computed by thresholding consistency error $\mathcal{C}(w)$ with 3 pixels. Our result in (d) better shapes the foremost region, i.e., the tower.

3.4 More Analysis

The two examples in Figure 9 are with the reference and input images containing quite different appearance and details. Our results shown in (c) not only include reliable foremost regions, but also contain pixel-wise correspondence.

Methods	MD ($\times 10^{-4}$)	IR (%)	MD in FR ($\times 10^{-3}$)	IR in FR (%)
SIFT	1.99	39.82	0.53	64.12
SURF	0.98	56.37	0.24	65.72
BDM	0.60	59.70	0.51	81.74
DeepMatching	**18.0**	41.61	**4.90**	65.22
Ours	11.0	**63.65**	3.30	**89.12**

Table 1: *Evaluation of different sparse matching methods. "MD" and "IR" stand for the matching density and inlier ratio respectively. "FR" stands for evaluation in foremost regions.*

The decent performance stems from the development of aforementioned steps in our framework. First, initial matching finds a sufficient amount of correspondence points with reasonable accuracy. It has considered existence of outliers and possible areas that cannot be matched. The region similarity measure with the robust function helps match patches hierarchically. Second, homography-based estimation establishes dense correspondence. With this computation, the following variational refinement can be done in quite a small number of iterations in the original resolution. Finally, to form the foremost region, we apply CRF with effective similarity and consistency confidence. These steps guarantee the final result quality. We evaluate our system more thoroughly below.

4 Implementation and Evaluation

Our system is developed in C++ and all our experiments are conducted on a PC with an Intel i7 3.4GHz CPU and 16GB memory. We calculate the running time of each step with image size 800×600. The sparse matching step uses 30 seconds on a single thread of CPU and 0.6 second on an NVIDIA Titan X display card. The second propagation step takes 3 seconds where 2.2 seconds are for propagation and 0.8 second is for variational refinement on CPU. Our regional foremost labeling takes 1.6 seconds on CPU.

Our method is not sensitive to parameter setting and we apply the following default values. In Eq. (4), we set λ to 2.0. In the regional foremost labeling step, k is set to 2.0 and γ is set to 0.4 by default. σ_s and σ_r in Eq. (9) are 20.0 and 0.1 respectively. The default parameter setting is good enough in our experiments.

Benchmark Data Since existing optical flow benchmark data mostly satisfy the intensity constancy assumption and images for nonrigid matching [HaCohen et al. 2011] do not have per-pixel matching ground truth, they are not appropriate for our evaluation. We instead build a new dataset containing challenging Internet scene images. We use over 100 keywords for online image retrieval and randomly choose one pair from each keyword result. The keywords are related to famous scenic spots (e.g., Champs-Élysées) and objects (e.g., butterfly) which help narrow down search space and retrieve a good number of share-content images where variation in viewpoint, day/night, weather, season, lighting, style, etc. still exists.

Then we manually select the most diverse 30 image pairs to construct our dataset. Two examples are shown in Figure 10. All images are with nonrigid transform and only part of the pixels can be matched in each image pair. To label the regional foremost correspondence ground truth, we manually mark pixels. Inside these regions, we label correspondence for each 5×5 grid and apply Nadaraya-Watson estimation [Wasserman 2013] to generate dense correspondence. Two ground truth maps are shown in Figure 10(c). Other images are included in our project website.

Sparse Correspondence Evaluation We evaluate methods in terms of correspondence density and percentage of matched points.

Methods	EPE	MR by $\mathcal{C}(w_p)$ (IoU%)	MR by Our Labeling (IoU%)
LDOF	17.92	27.12	45.55
MDP	16.60	30.47	52.18
Flow Fields	10.33	41.56	52.68
EpicFlow	19.49	49.80	50.56
SIFT Flow	11.54	30.68	36.99
DAISY Flow	14.09	40.14	43.27
DeepFlow	7.58	47.94	59.87
NRDC	16.60	27.76	45.61
Ours	**5.28**	**77.35**	**81.39**

Table 2: *Evaluation of different dense correspondence estimation methods. "EPE" and "MR" stand for end point error and matched-region respectively.*

The methods we compare with are those finding the nearest neighbors in feature descriptors of SIFT [Lowe 2004] and SURF [Bay et al. 2006]. We also compare state-of-the-art sparse matching frameworks – bounded distortion matching (BDM) [Lipman et al. 2014] and DeepMatching [Weinzaepfel et al. 2013; Revaud et al. 2016]. We use author-released implementation with suggested parameter setting.

We calculate two evaluation scores – i.e., MD (short for matching density) to denote the percentage of matched points among all and IR (short for inlier ratio) to represent the ratio between the number of inliers and the total number of correspondences output from a method. We consider a pair of points as an inlier when the matching error is smaller than 5 pixels.

We compute the average of these two scores in the whole image and only the foremost regions on our benchmark data respectively. The results are reported in Table 1. It is not surprising that SIFT, SURF and bounded distortion matching yield smaller matching densities because of sparse feature points. Bounded distortion method rejects outliers and outputs higher inlier ratios. The DeepMatching method generates denser points; but it does not consider the case that many points cannot be matched. Our method produces decent results.

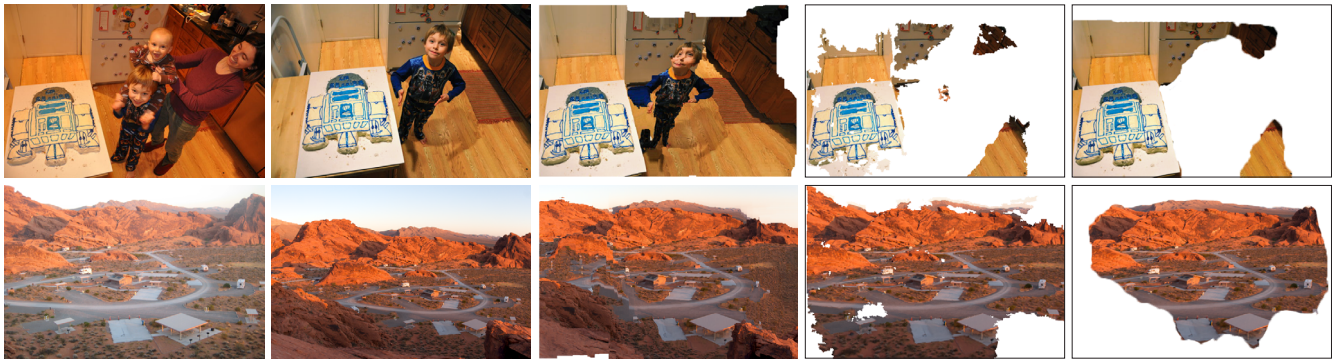
Foremost Dense Correspondence Evaluation We then evaluate dense matching accuracy. State-of-the-art dense image matching methods include large-displacement optical flow (LDOF) [Brox and Malik 2011], motion-detail-preserving optical flow (MDP) [Xu et al. 2012], flow fields [Bailer et al. 2015], EpicFlow [Revaud et al. 2015], SIFT flow [Liu et al. 2011a], DAISY flow [Yang et al. 2014], DeepFlow [Weinzaepfel et al. 2013; Revaud et al. 2016] and NRDC [HaCohen et al. 2011]. We measure the mean end point error (EPE), which is the correspondence L_2 distance with respect to ground truth. We only calculate EPE on our labeled foremost regions since there is no ground truth for other pixels. All evaluation is based on author-released implementation with the default setting.

In addition, matched-region quality is evaluated as intersection-over-union (IoU) of the estimated matching region to the ground truth foremost region. Since LDOF, MDP, flow fields, EpicFlow, SIFT flow, DAISY flow and DeepFlow do not output this type of results, we apply consistency confidence (elaborated on as $\mathcal{C}(w_p)$ in Section 3.3) by extracting matched ones with threshold of 5 pixels and our regional foremost labeling approach as illustrated in Section 3.3 to estimate the matched-region of each method. As reported in Table 2, our method yields accurate matching in terms of EPE because of the effective matching framework for Internet scene images. Our method also produces good-quality foremost regions compared with the simple threshold scheme as shown in the last column of Table 2.

Visual Comparison We visually compare results of our method, SIFT flow and NRDC since the latter two are state-of-the-art non-

(a) Reference (b) Input (c) SIFT Flow (d) NRDC (e) Ours

Figure 11: *Visual comparison with other methods.*



(a) Reference (b) Input (c) SIFT Flow (d) NRDC (e) Ours

Figure 12: *Visual comparison with other methods on the data from NRDC Real World Scenes.*

rigid matching approaches. As shown in Figure 11, images in (a) and (b) are with quite diversified appearance and content. We match (b) to (a) and show the warping results in (c)-(e) by different methods. SIFT flow may not be accurate enough for pixels that are not SIFT features. The NRDC method is also not designed to tackle our problem and results in small matched regions. Our results shown in (e) contain pixel-wise correspondence for foremost regions. More comparisons are provided in our project website.

We also evaluate our method on NRDC data [HaCohen et al. 2011], where images were taken with drastic human/object motion. As shown in Figure 12, our approach can still produce acceptable foremost regions and dense matching for a set of images. When the input images contain quite different content where common pixels cannot constitute foremost regions with reasonable sizes, our performance degrades. We will discuss it more as limitation later.

5 Applications

Our method benefits many applications. We in this paper present its employment in Internet time-lapse sequence generation, image composition, automatic image morphing, image manipulation, and spotting image difference. Since most of these applications need to perform whole-image warping or blending. We first explain the post-process in our system to quickly generate a full-image displacement map from the foremost regional results.

Post-processing to Interpolate Foremost Correspondence It is a simple interpolation process to minimize

$$E(w^*) = \sum_p \left(o_p \cdot \varphi(\|w_p^* - w_p\|) + \eta \cdot \|\nabla w_p^*\|^2 \right), \qquad (10)$$

where w^* is the whole-image correspondence. w is the regional foremost correspondence and o is the corresponding foremost 0-1 mask. A robust function $\varphi(x)$ is similarly used here. Weight η balances the influence and is set to 1.2 in all experiments. An example is shown in Figure 13 where the whole-image warping result in (e) maintains structure.

Internet Time-lapse Video Generation Our method is useful for time-lapse sequence generation from Internet scene images. Our method finds correspondence in main-object regions and interpolate it in other pixels with the above post-process. After warping, smooth transition between time-lapse frames can be produced.

Compared with the method of Martin-Brualla et al. [2015a], which used global transform to match images, our method handles more

Figure 13: *Extending displacements to the whole image. (a) and (b) are the reference and input images respectively. (c) is the regional foremost correspondence result and (d) is the whole image interpolation. (e) is the warping result based on (d).*

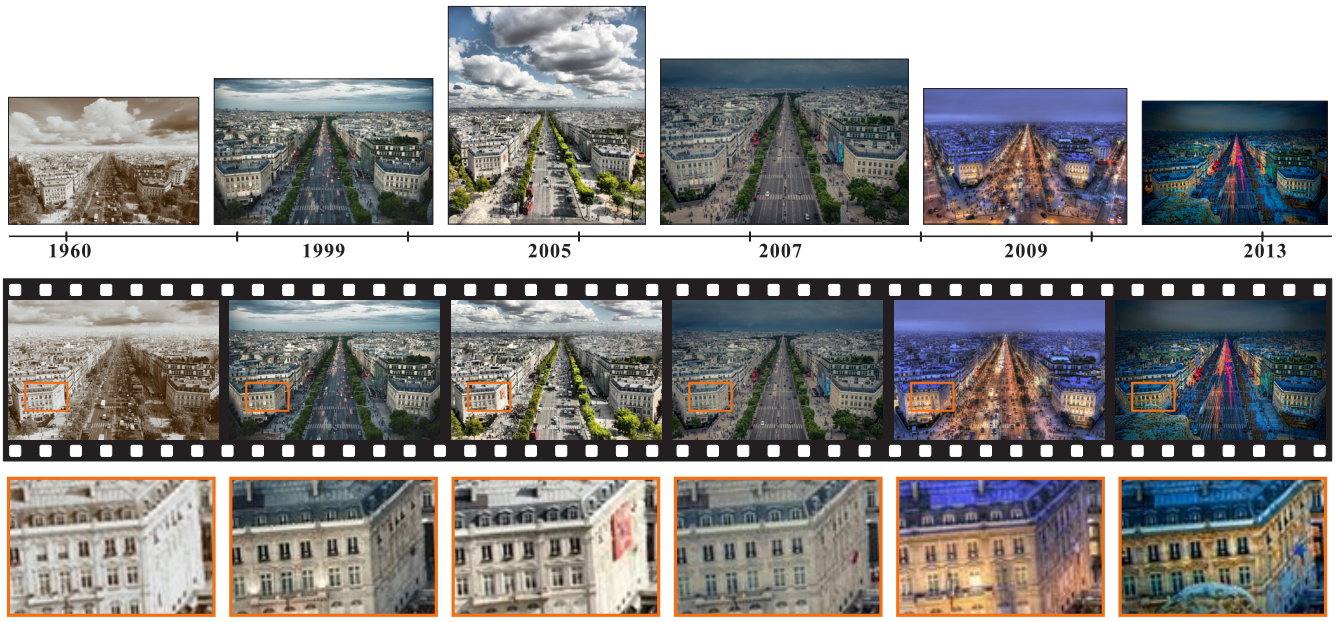


Figure 14: *An example of time-lapse sequence generation from Internet scene images.*

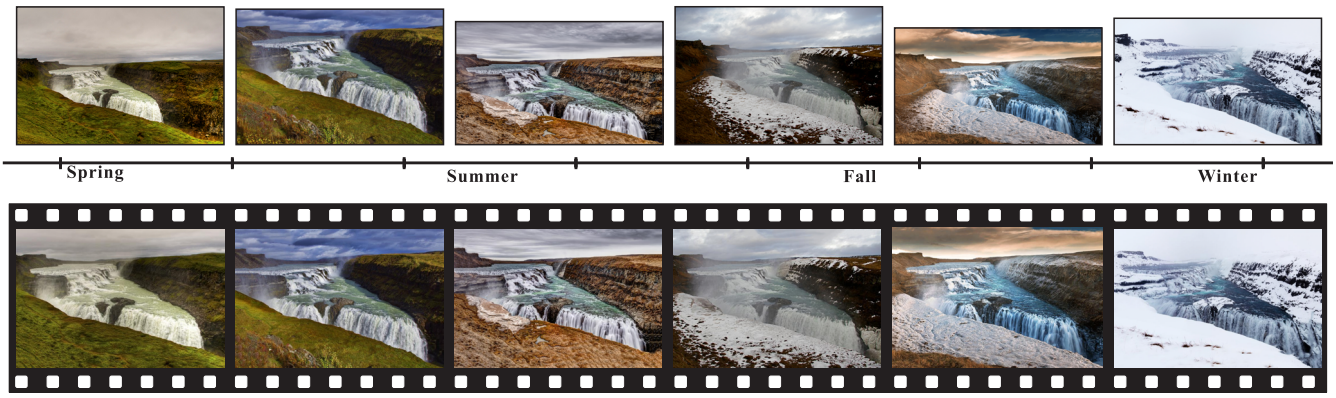
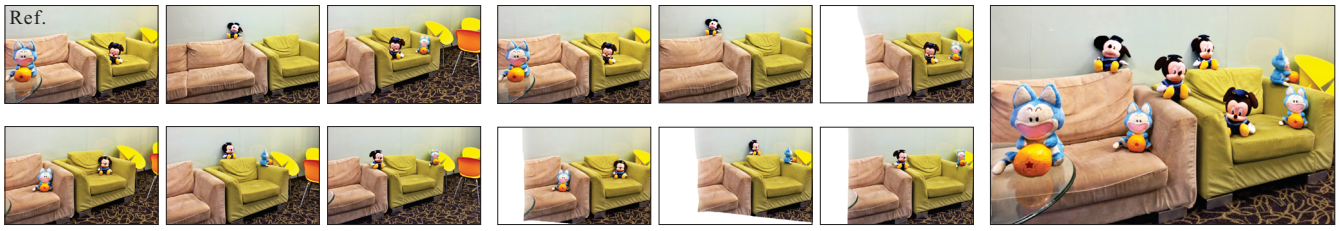


Figure 15: *Another example of time-lapse sequence generated from Internet scene images.*

difficult data with different appearance and nonrigid motion. As shown in Figures 14 and 15, our time-lapse sequences do not cause blurring since matching is done for each pixel.

Internet Image Composition Interactive digital photomontage [Agarwala et al. 2004] provides a way to put user selected parts from different source images into a single one. It requires aligned source images. Our method achieves similar composite without prior manual alignment. Figure 16 shows the process to first rectify the input images by our method automatically and then apply the process of Agarwala et al. [2004] to merge them.

Our algorithm works on more challenging images. As shown in Figure 17, the input five images were captured when the Sydney opera is in a light show. The largely varied salient edges and texture make accurate matching difficult. Our method still works for this example. We create the time-lapse mosaics [Agarwala et al. 2004] by merging images into a single one with time lapse from left to right and show it in (d). As illustrated in Figure 17(b) and (c), global matching based on SIFT descriptors and optical flow es-



(a) Input Images (b) Warped Images (c) Our Result

Figure 16: *Unaligned digital photomontage. In this example, we match and warp images regarding a reference as shown in (a) and (b) respectively and then compose objects in (c) using the digital interactive photomontage method.*

(a) Input Images

(b) Homography (c) DeepFlow (d) Ours

Figure 17: *Unaligned image fusion for time-lapse mosaics. (a) shows input images. (b) and (c) are the fusion results of global matching and DeepFlow estimation respectively. (d) is our ghost-free fusion result.*

timation produces less satisfying results.

Concurrent Image Editing For a group of images with similar content, our matching results can be used for concurrent interactive image editing. As shown in Figure 18, edit in (a) can be immediately applied to target in (d) without extra interaction. Compared with the method of Yücer et al. [2012], our matching and alignment can be performed on more challenging examples.

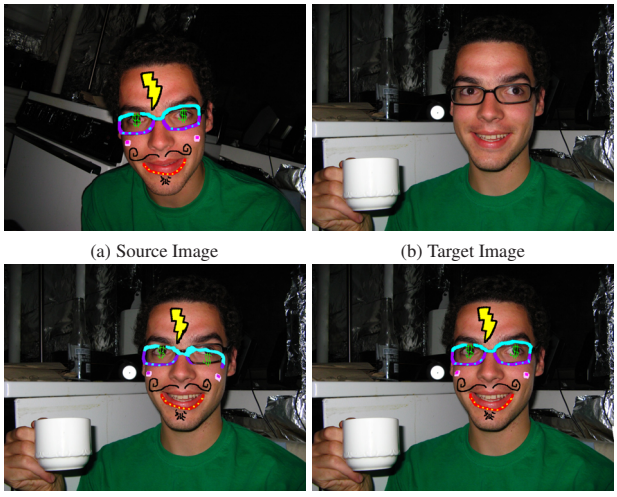
Automatic Internet Image Morphing The dense correspondence yielded by our method is also very helpful for automatic image morphing. In Figure 19, (a) and (b) are used to create morphing. The intermediate transition frame by the method of Liao et al. [2014] is shown in (c). The errors are caused by the complicated scene variation. Our result is shown in (d) where errors are suppressed.

Other Applications Our matching method also benefits spot difference, rephotography, image enhancement, view interpolation, image similarity measurement, image annotation. We show these applications in our website.

6 Concluding Remarks

We have presented a practical system for Internet scene image dense regional matching where accurate corresponding points are produced inside confident regions. It involves several steps to guarantee good initialization and robust displacement propagation.

Our method yet has the following limitations. First, it is hard to estimate regional correspondence when dynamic and featureless objects undergo large motion. One example is demonstrated in the first row of Figure 20 where people are with quite different poses and the clothes and skin do not contain enough feature. Addressing this problem may need other specific steps of pose estimation and human body segmentation. It is also our future work to incorporate patch-level information in the framework. Second, computing



(a) Source Image (b) Target Image

(c) Result of Yücer et al. 2012 (d) Ours

Figure 18: *Example of image editing. (a) and (b) are the source and target images respectively. (c) is the result of Yücer et al. [2012] and (d) is our result.*

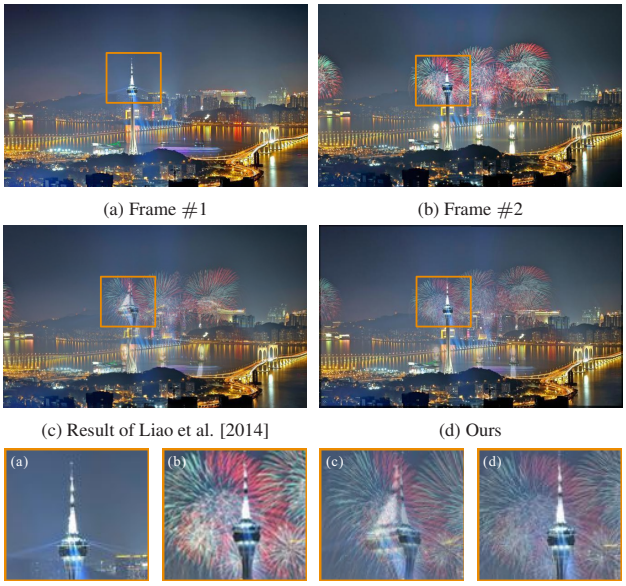


Figure 19: *Automatic image morphing. (a) and (b) are two input images. (c) and (d) are the results of Liao et al. [2014] and our method respectively.*

Figure 20: *Limitations. (a) and (b) are the reference and input images respectively. (c) shows our results. The first case represents dynamic objects with large nonrigid motion and the second one is with significant view and scale variation.*

reliable correspondence with very large view and scale variation in input images could be difficult as shown in the second row of Figure 20. So the images in data of [Cho et al. 2008] are difficult to handle by our method. Extra geometry information may be helpful.

Acknowledgements

We thank the anonymous reviewers for their suggestion, and Li Xu, Qi Zhang, Ziyang Ma and Ruiyu Li for helpful discussion. We also thank flickr users "Scott Cartwright Photography", "TomNC (Tom Charoensinphon)", "akwan.architect" for allowing us to use their photos in the paper. This work is supported by a grant from the Research Grants Council of the Hong Kong SAR (project No. 2150760) and by the National Science Foundation China, under Grant 61133009.

References

AGARWAL, S., SNAVELY, N., SIMON, I., SEITZ, S. M., AND SZELISKI, R. 2009. Building rome in a day. In *ICCV*, 72–79.

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S. M., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. F. 2004. Interactive digital photomontage. *ACM Trans. Graph. 23*, 3, 294–302.

ALEXA, M., BEHR, J., COHEN-OR, D., FLEISHMAN, S., LEVIN, D., AND SILVA, C. T. 2003. Computing and rendering point set surfaces. *IEEE Transactions on Visualization and Computer Graphics 9*, 1, 3–15.

BAILER, C., TAETZ, B., AND STRICKER, D. 2015. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In *ICCV*, 4015–4023.

BAO, L., YANG, Q., AND JIN, H. 2014. Fast edge-preserving patchmatch for large displacement optical flow. *IEEE Trans. Image Processing 23*, 12, 4996–5006.

BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. B. 2009. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph. 28*, 3.

BARNES, C., SHECHTMAN, E., GOLDMAN, D. B., AND FINKELSTEIN, A. 2010. The generalized patchmatch correspondence algorithm. In *ECCV*, 29–43.

BAY, H., TUYTELAARS, T., AND GOOL, L. J. V. 2006. Surf: Speeded up robust features. In *ECCV*, 404–417.

BLACK, M. J., AND ANANDAN, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding 63*, 1, 75–104.

BROX, T., AND MALIK, J. 2011. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell. 33*, 3, 500–513.

BROX, T., BRUHN, A., PAPENBERG, N., AND WEICKERT, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, 25–36.

BRUHN, A., AND WEICKERT, J. 2005. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *ICCV*, 749–755.

CHEN, T., CHENG, M., TAN, P., SHAMIR, A., AND HU, S. 2009. Sketch2photo: internet image montage. *ACM Trans. Graph. 28*, 5.

CHIA, A. Y. S., ZHUO, S., GUPTA, R. K., TAI, Y., CHO, S., TAN, P., AND LIN, S. 2011. Semantic colorization with internet images. *ACM Trans. Graph. 30*, 6, 156.

CHO, M., SHIN, Y. M., AND LEE, K. M. 2008. Co-recognition of image pairs by data-driven monte carlo image exploration. In *ECCV*, 144–157.

DALE, K., JOHNSON, M. K., SUNKAVALLI, K., MATUSIK, W., AND PFISTER, H. 2009. Image restoration using online photo collections. In *ICCV*, 2217–2224.

DOSOVITSKIY, A., SPRINGENBERG, J. T., RIEDMILLER, M. A., AND BROX, T. 2014. Discriminative unsupervised feature learning with convolutional neural networks. In *NIPS*, 766–774.

FURUKAWA, Y., CURLESS, B., SEITZ, S. M., AND SZELISKI, R. 2010. Towards internet-scale multi-view stereo. In *CVPR*, 1434–1441.

HACOHEN, Y., SHECHTMAN, E., GOLDMAN, D. B., AND LISCHINSKI, D. 2011. Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. Graph. 30*, 4, 70.

HACOHEN, Y., SHECHTMAN, E., AND LISCHINSKI, D. 2013. Deblurring by example using dense correspondence. In *ICCV*, 2384–2391.

HAYS, J., AND EFROS, A. A. 2007. Scene completion using millions of photographs. *ACM Trans. Graph. 26*, 3, 4.

HORN, B. K. P., AND SCHUNCK, B. G. 1981. Determining optical flow. *Artif. Intell. 17*, 1-3, 185–203.

JOHNSON, M. K., DALE, K., AVIDAN, S., PFISTER, H., FREEMAN, W. T., AND MATUSIK, W. 2011. Cg2real: Improving the realism of computer generated images using a large collection of photographs. *IEEE Transactions on Visualization and Computer Graphics 17*, 9, 1273–1285.

JOSHI, N., MATUSIK, W., ADELSON, E. H., AND KRIEGMAN, D. J. 2010. Personal photo enhancement using example images. *ACM Trans. Graph. 29*, 2.

KIM, J., LIU, C., SHA, F., AND GRAUMAN, K. 2013. Deformable spatial pyramid matching for fast dense correspondences. In *CVPR*, 2307–2314.

KRÄHENBÜHL, P., AND KOLTUN, V. 2011. Efficient inference in fully connected crfs with gaussian edge potentials. In *NIPS*, 109–117.

LIAO, J., LIMA, R. S., NEHAB, D., HOPPE, H., SANDER, P. V., AND YU, J. 2014. Automating image morphing using structural similarity on a halfway domain. *ACM Trans. Graph. 33*, 5, 168:1–168:12.

LIPMAN, Y., YAGEV, S., PORANNE, R., JACOBS, D. W., AND BASRI, R. 2014. Feature matching with bounded distortion. *ACM Trans. Graph. 33*, 3, 26.

LIU, C., YUEN, J., AND TORRALBA, A. 2011. Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell. 33*, 5, 978–994.

LIU, F., GLEICHER, M., WANG, J., JIN, H., AND AGARWALA, A. 2011. Subspace video stabilization. *ACM Trans. Graph. 30*, 1, 4.

LIU, S., YUAN, L., TAN, P., AND SUN, J. 2013. Bundled camera paths for video stabilization. *ACM Trans. Graph. 32*, 4.

LIU, Z., YUAN, L., TANG, X., UYTTENDAELE, M., AND SUN, J. 2014. Fast burst images denoising. *ACM Trans. Graph. 33*, 6, 232.

LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision 60*, 2, 91–110.

MARTIN-BRUALLA, R., GALLUP, D., AND SEIRZ, S. M. 2015. Time-lapse mining from internet photos. In *ACM SIGGRAPH*, vol. 34, 62.

MARTIN-BRUALLA, R., GALLUP, D., AND SEITZ, S. M. 2015. 3d time-lapse reconstruction from internet photos. In *ICCV*, 1332–1340.

NOAH, S. 2011. Scene reconstruction and visualization from internet photo collections: A survey. *Transactions on Computer Vision and Applications 3*, 44–66.

REVAUD, J., WEINZAEPFEL, P., HARCHAOUI, Z., AND SCHMID, C. 2015. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *CVPR*, 1164–1172.

REVAUD, J., WEINZAEPFEL, P., HARCHAOUI, Z., AND SCHMID, C. 2016. Deepmatching: Hierarchical deformable dense matching. *CoRR abs/1506.07656*.

RHEMANN, C., HOSNI, A., BLEYER, M., ROTHER, C., AND GELAUTZ, M. 2011. Fast cost-volume filtering for visual correspondence and beyond. In *CVPR*, 3017–3024.

SEN, P., KALANTARI, N. K., YAESOUBI, M., DARABI, S., GOLDMAN, D. B., AND SHECHTMAN, E. 2012. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph. 31*, 6, 203.

SHEN, L., AND TAN, P. 2009. Photometric stereo and weather estimation using internet images. In *CVPR*, 1850–1857.

SHEPARD, D. 1968. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 23rd ACM National Conference*, ACM '68, 517–524.

SHI, B., INOSE, K., MATSUSHITA, Y., TAN, P., YEUNG, S., AND IKEUCHI, K. 2014. Photometric stereo using internet images. In *International Conference on 3D Vision*, 361–368.

SHIH, Y., PARIS, S., BARNES, C., FREEMAN, W. T., AND DURAND, F. 2014. Style transfer for headshot portraits. *ACM Trans. Graph. 33*, 4, 148:1–148:14.

SIMO-SERRA, E., TRULLS, E., FERRAZ, L., KOKKINOS, I., FUA, P., AND MORENO-NOGUER, F. 2015. Discriminative learning of deep convolutional feature point descriptors. In *ICCV*, 118–126.

SIMONYAN, K., VEDALDI, A., AND ZISSERMAN, A. 2014. Learning local feature descriptors using convex optimisation. *IEEE Trans. Pattern Anal. Mach. Intell. 36*, 8, 1573–1585.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3d. *ACM Trans. Graph. 25*, 3, 835–846.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2008. Modeling the world from internet photo collections. *International Journal of Computer Vision 80*, 2, 189–210.

TAU, M., AND HASSNER, T. 2016. Dense correspondences across scenes and scales. *IEEE Trans. Pattern Anal. Mach. Intell. 38*, 5, 875–888.

TOLA, E., V.LEPETIT, AND FUA, P. 2008. A fast local descriptor for dense matching. In *CVPR*.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *ICCV*, 839–846.

WASSERMAN, L. 2013. *All of statistics: a concise course in statistical inference*. Springer.

WEDEL, A., RABE, C., VAUDREY, T., BROX, T., FRANKE, U., AND CREMERS, D. 2008. Efficient dense scene flow from sparse or dense stereo data. In *ECCV*, 739–751.

WEINZAEPFEL, P., REVAUD, J., HARCHAOUI, Z., AND SCHMID, C. 2013. DeepFlow: Large displacement optical flow with deep matching. In *ICCV*, 1385–1392.

XU, L., JIA, J., AND MATSUSHITA, Y. 2012. Motion detail preserving optical flow estimation. *IEEE Trans. Pattern Anal. Mach. Intell. 34*, 9, 1744–1757.

YAN, Z., ZHANG, H., WANG, B., PARIS, S., AND YU, Y. 2016. Automatic photo adjustment using deep neural networks. *ACM Trans. Graph. 35*, 2, 11.

YANG, H., LIN, W., AND LU, J. 2014. DAISY filter flow: A generalized discrete approach to dense correspondences. In *CVPR*, 3406–3413.

YÜCER, K., JACOBSON, A., HORNUNG, A., AND SORKINE, O. 2012. Transfusive image manipulation. *ACM Trans. Graph. 31*, 6, 176.

ZACH, C., POCK, T., AND BISCHOF, H. 2007. A duality based approach for realtime TV-L1 optical flow. *Pattern Recognition*, 214–223.

ZARAGOZA, J., CHIN, T., TRAN, Q., BROWN, M. S., AND SUTER, D. 2014. As-projective-as-possible image stitching with moving DLT. *IEEE Trans. Pattern Anal. Mach. Intell. 36*, 7, 1285–1298.

ZHANG, C., GAO, J., WANG, O., GEORGEL, P., YANG, R., DAVIS, J., FRAHM, J., AND POLLEFEYS, M. 2014. Personal photograph enhancement using internet photo collections. *IEEE Transactions on Visualization and Computer Graphics 20*, 2, 262–275.