Pacific Graphics (2014)
J. Keyser, Y. J. Kim, and P. Wonka (Editors)

Short Papers

Geometry-Aware Image Completion via Multiple Examples

Guihang Wang, Xuejin Chen† and Siyu Hu

CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System
University of Science and Technology of China

Abstract

When browsing through photographs taken during a trip, it can be a distressing discovery to find many other bystanders captured within the frame. A visually compelling snapshot preserves the desired subject in the foreground, and eliminates irrelevant persons or objects. In this paper, we present a new image completion algorithm that retains the geometric consistency of the scene using Internet photo collections. Our key insight allows us to successfully fill the missing regions with content derived from related images with distinct features, or from visible parts of the input image with repetitive textures. The final composition for the missing regions is guided by a confidence map combining the two types of example images. We validate our algorithm on a variety of scenes which are challenging for state-of-the-art completion techniques. The completion results successfully preserve the geometry consistency for complex structures in a wide range of depths.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—

1. Introduction

Photographs captured during a trip usually contain many unwanted additions, such as other tourists. While it is impossible to ask everyone to clear the area, we hope to attain a visually compelling image that is free of clutter. Image completion, which synthesizes a missing or destroyed region in an image, is the best option to achieve this goal.

Image completion is a popular subject of inquiry in the field of computer graphics. Example-based methods are effective to fill large holes in images. A popular method is to formulate the problem as a discrete labeling problem in order to find the optimal source for each pixel or patch in the unknown region from the visible parts [KT07, PKVP09]. A fast, randomized patch searching algorithm is proposed to accelerate a group of patch-based optimization approaches [BSFG09, BSGF10]. In example-based approaches, structure plays a critical role on the final completion. User-drawn sketches can be used to represent structures when synthesizing large regions [HZW*13]. For images containing a set of repetitive patterns, the statistics of the offsets [HS12] or mid-level planar structure constraints [HKAK14] can be used to implicitly preserve the structure in the unknown region. However, these methods using the input image itself fail to supplement regions with complex structures.

With the growth of photo-sharing websites, rich image resources bring new opportunities in many areas. A seminal work using millions of photographs is introduced in [HE07]. Though this approach demonstrates the great benefit of using images sourced from the Internet, only nature scenes composed of large textured areas can be completed well.

While there are usually hundreds of photographs for a landmark place, Amirshahi et al. present a system to complete the background while removing the other tourists using related photos on the Internet [AK08]. However, the full planar perspective motion model assumption used to align the reference image to the target image results in distinct misalignments in the composition for complex structures. Even with a local alignment, it fails to complete the low texture region due to the lack of feature correspondence. Multiple homography transformations are combined to complete missing regions in [WSZ09]. The algorithm inherits the limitations of homography transformation. Only nearly planar scenes with enough feature points can be completed. Reconstructed 3D point clouds are used to provide more accurate structure information for personal photo enhancement [ZGW*14]. The results become very blurry in regions without reconstructed geometry.

† Corresponding author

DOI: 10.2312/pgs.20141259

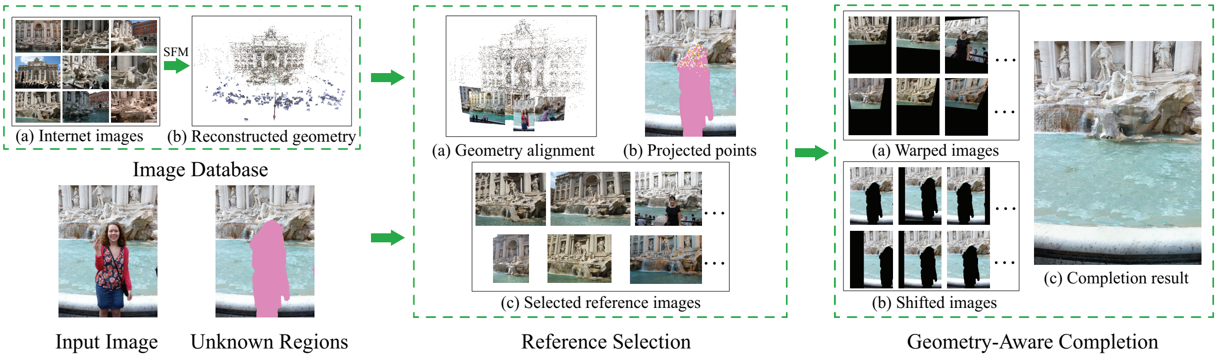

Figure 1: *Algorithm overview.*

For many scenes that contain a lot of freeform structures or unique objects, it is very challenging to generate geometrically correct content in the missing region. In this paper, we present a novel completion algorithm in a geometry-aware scheme using multiple examples retrieved from the Internet. Our approach takes the advantages of both the reconstructed geometry and the repetitiveness of flat regions to complete large regions for complex scenes.

2. Reference Image Selection

We complete a missing region Ω within an input image $\hat{I}$ by searching for the optimal content from the image itself, as well as from photographs of the same site collected from the Internet. Figure 1 shows the overview of our algorithm. We use Google to search for pictures that are similar to the input image. The images on the first two pages are selected to construct the image database. Using the SFM approach provided in [SSS06], a set of 3D points and camera parameters can be obtained offline.

We first register $\hat{I}$ into the reconstructed 3D model via SFM. Then the reconstructed 3D points are projected onto the input image plane and a set of reference points for the missing region are obtained. Each point is visible in several images included in the image database. The number of visible reference points for the missing region in a reference image I_i is denoted as n_i. Figure 2 shows a set of projected points in the missing region and the corresponding feature points from four reference images.

Reference images that contain less than eight reference feature points are ignored. We then select K_w of the best reference images according to the viewpoint deviation from the input image. For each candidate image I_i, we compute its score

$$S(I_i) = \frac{V_i \cdot \hat{V}}{\|V_i\|\|\hat{V}\|} + \frac{d_{max} - d_i}{d_{max}}, \quad (1)$$

where V_i and $\hat{V}$ are the view directions of the image I_i and

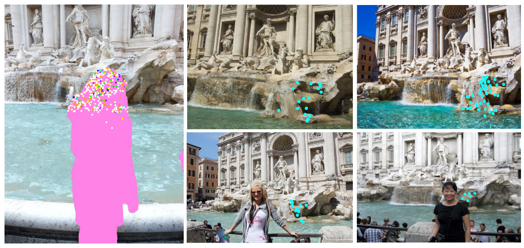

Figure 2: *The projected 3D points in the missing region can be found in other images. Note that most of the reconstructed points are located on objects with complex structures.*

the input image $\hat{I}$, respectively. d_i is the distance between the camera positions of I_i and $\hat{I}$. $d_{max} = \max d_i$.

3. Geometry-aware Image Completion

Our geometry-aware completion algorithm optimizes the composition on a global scale in order to select the optimal source for each pixel among multiple example images.

3.1. Warped Reference Images

We first warp each reference image I_i to the target image plane by a homography H_i that is estimated based on the correspondences of the projected points in the missing region. The K_w warped images $\{I_i^w\}_{i=1}^{K_w}$ are then added to the example set for completion. A group of warped reference images are shown in Figure 3(c). The warped images typically generate satisfactory examples for a region with rich structures, such as the rocks. However, there is usually a persistent large, flat area within the missing region. Even with very close views, the warped images lack the necessary for the bottom part of the missing region.

3.2. Shifted Input Images

Instead of using warped reference images, the visible parts of the original input image can often serve as superior examples due to texture repetition. We adopt the approach using shifted images based on offset statistics proposed in [HS12].

Given the missing region Ω, we first select the surrounding band region $\delta\Omega$ with a width of 50 pixels. For the $w \times w$ patch around each pixel $\mathbf{p} \in \delta\Omega$, we search for the most similar patch among all visible parts of the input image, excluding the area around the patch. Normalized cross-correlation is used to measure the patch similarity, and avoid the effects from different lighting. The offsets from all the pairs of similar patches are then accumulated in a 2D histogram. We select K_s of the highest peaks as the dominant offsets $\{\Delta_i\}_{i=1}^{K_s}$ for completion. The input image $\hat{I}$ is then shifted by the dominant offsets to generate K_s shifted images $I_i^s(\mathbf{p}) = \hat{I}(\mathbf{p}+\Delta_i)$. In our experiments, $K_w = 15$, $K_s = 20$, and $w = 18$.

3.3. Confidence Map

In order to preserve the geometric consistency in the missing region, the pixels where the reconstructed points are projected on $\hat{I}$ must have strong confidence in order to be chosen from the warped reference images. Given the positions of the visible points on the reference image I_i, we compute the confidence map of the region Ω in I_i^w as

$$\omega_i(\mathbf{p}) = \begin{cases} 0, & \text{if invalid } I_i^w(\mathbf{p}), \\ \frac{1}{n}\sum_{j=1}^{n_i} \mathcal{N}(\sigma, \mathbf{p}-\mathbf{p}_j), & \text{otherwise,} \end{cases} \tag{2}$$

where $\mathcal{N}$ is a Gaussian function, $\mathbf{p}_j$ is the position of a feature point on I_i^w, and $\sigma = 30$.When combining the K_w warped images, the confidence of each pixel $\mathbf{p}$ in Ω is

$$\varpi(\mathbf{p}) = \frac{1}{K_w}\sum_{i=1}^{K_w} \omega_i(\mathbf{p}). \tag{3}$$

Figure 3(a) shows the confidence map of a missing region.

3.4. Composition Optimization

Given the example set consisting of K_w warped images and K_s shifted images, we look for the optimal composition for the missing region Ω at the pixel level. In contrast to existing approaches, we not only complete the missing region Ω but also modify the boundary region $\delta\Omega$ with a width of 20 pixels to reduce the visual gaps.

The completion is formulated as a discrete labeling problem, with the cost function defined as

$$C(\mathbf{L}) = \sum_{\mathbf{p}\in\mathbf{R}} C_d\big(\mathbf{p}, L(\mathbf{p})\big) + \sum_{(\mathbf{p},\mathbf{q})\in\mathbf{N}} C_s\big(\mathbf{p},\mathbf{q},L(\mathbf{p}),L(\mathbf{q})\big). \tag{4}$$

The data term C_d for each pixel in Ω is defined as

$$C_d\big(\mathbf{p}, L(\mathbf{p})\big) = \varpi(\mathbf{p}) D(\bar{I}, I_{L(\mathbf{p})}, \mathbf{p}), \tag{5}$$

where $D(I_a, I_b, \mathbf{p})$ is the appearance difference composed of color difference and gradient difference, defined as

$$D(I_a, I_b, \mathbf{p}) = \|I_a(\mathbf{p}) - I_b(\mathbf{p})\| + \|\nabla I_a(\mathbf{p}) - \nabla I_b(\mathbf{p})\|. \tag{6}$$

The weighted average image $\bar{I}$ is employed as a reference for completion in order to reduce the effects of different lighting or different foregrounds in the example images.

$$\bar{I}(\mathbf{p}) = \sum_{i=1}^{K_w} \frac{\omega_i(\mathbf{p})}{K_w \varpi(\mathbf{p})} I_i^w(\mathbf{p}). \tag{7}$$

The data term C_d for each pixel $\mathbf{p} \in \delta\Omega$ is

$$C_d\big(\mathbf{p}, L(\mathbf{p})\big) = D(\hat{I}, I_{L(\mathbf{p})}, \mathbf{p}). \tag{8}$$

The smoothness term C_s is defined to penalize the difference between neighboring pixels. We define

$$C_s\big(\mathbf{p},\mathbf{q},L(\mathbf{p}),L(\mathbf{q})\big) = D(I_{L(\mathbf{p})}, I_{L(\mathbf{q})}, \mathbf{p}) + D(I_{L(\mathbf{p})}, I_{L(\mathbf{q})}, \mathbf{q}). \tag{9}$$

The composition is finally solved using the tool provided by Szeliski et al. [SZS*06]. Figure 3(d) shows the completion result for a complex scene. The structure is well-preserved, especially the rocks, however, gaps do exist as a result of different lighting effects. Finally, a Poisson blending is performed to reduce the lighting difference.

4. Results

In order to validate the effectiveness of our algorithm, we test it on a wide range of famous sites. We compare our results against four representative approaches, including Photoshop Content-Aware Fill [BSFG09, WSI07], Image Melding [DSB*12], Offset statistics [HS12] using the input image, and [WSZ09] using multiple images.

Figure 4 shows a series of comparisons on a wide range of challenging cases. In the first row, the missing region depicts the content at varying depths. Our method completes the actual structure for the rocks using reference images. The rim closest to the viewpoint is completed well by using examples from the input image itself. The approach [WSZ09] fails to generate an acceptable image completion due to the lack of information for the bottom region, which is mainly caused by homography transformations.

In the last row, our algorithm generates well-aligned structures for the building. The techniques using the input image generate quite messy structures due to lack of information. [WSZ09] produces a blurry composition. Furthermore, the structures of the columns and the door are not preserved due to the global homography transformation.

5. Conclusion

A novel image completion algorithm has been presented in this paper to preserve the geometric consistency in the pixel-level composition. By taking advantage of rich resources on the Internet, which are readily accessible to the public, our approach is able to complete large missing regions in a photograph. The geometric consistency is derived from the reconstructed 3D points. Our algorithm overcomes the drawbacks of low-level texture synthesis approaches. This is achieved by our geometry-aware completion with multiple examples from both reference images and the input image.

(a) Confidence map (b) Labels (c) Example images (d) Optimized Composition

Figure 3: *Optimized composition of multiple examples.*

Figure 4: *Comparisons with state-of-the-art algorithms.*

Acknowledgements

We appreciate the generous reviews from the anonymous reviewers. We thank Dr. Kaiming He, Dr. Jian Sun, and Dr. Oliver Whyte for providing their result images for comparison. We also thank Dr. Soheil Darabi, Prof. Noah Snavely, and Dr. Richard Szeliski for sharing their codes. This work was supported by the NSFC (No. 61103135), the Fundamental Research Funds for the Central Universities (No. WK2100060011), and Intel ICRI MNC.

References

[AK08] AMIRSHAHI H., KONDO S.: An image completion algorithm using occlusion-free images from internet photo sharing sites. *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences 91*, 10 (2008), 2918–2927. 1

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D.: Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. on Graph. 28*, 3 (2009), 24. 1, 3

[BSGF10] BARNES C., SHECHTMAN E., GOLDMAN D. B., FINKELSTEIN A.: The generalized patchmatch correspondence algorithm. In *ECCV*. 2010, pp. 29–43. 1

[DSB*12] DARABI S., SHECHTMAN E., BARNES C., GOLDMAN D. B., SEN P.: Image melding: combining inconsistent images using patch-based synthesis. *ACM Trans. on Graph. 31*, 4 (2012), 82. 3

[HE07] HAYS J., EFROS A. A.: Scene completion using millions of photographs. In *ACM Trans. on Graph.* (2007), vol. 26. 1

[HKAK14] HUANG J.-B., KANG S. B., AHUJA N., KOPF J.: Image completion using planar structure guidance. *ACM Trans. on Graph. (SIGGRAPH 2014) 33*, 4 (2014). 1

[HS12] HE K., SUN J.: Statistics of patch offsets for image completion. In *ECCV* (2012), pp. 16–29. 1, 3

[HZW*13] HU S.-M., ZHANG F.-L., WANG M., MARTIN R. R., WANG J.: Patchnet: a patch-based image representation for interactive library-driven image editing. *ACM Trans. on Graph. 32*, 6 (2013), 196. 1

[KT07] KOMODAKIS N., TZIRITAS G.: Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *Image Processing, IEEE Trans. 16*, 11 (2007). 1

[PKVP09] PRITCH Y., KAV-VENAKI E., PELEG S.: Shift-map image editing. In *ICCV* (2009), IEEE, pp. 151–158. 1

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3d. *ACM Trans. on graph. 25*, 3 (2006), 835–846. 2

[SZS*06] SZELISKI R., ZABIH R., SCHARSTEIN D., VEKSLER O., KOLMOGOROV V., AGARWALA A., TAPPEN M., ROTHER C.: A comparative study of energy minimization methods for markov random fields. In *ECCV*. 2006, pp. 16–29. 3

[WSI07] WEXLER Y., SHECHTMAN E., IRANI M.: Space-time completion of video. *PAMI 29*, 3 (2007), 463–476. 3

[WSZ09] WHYTE O., SIVIC J., ZISSERMAN A.: Get out of my picture! Internet-based inpainting. In *BMVC* (2009). 1, 3

[ZGW*14] ZHANG C., GAO J., WANG O., GEORGEL P., YANG R., DAVIS J., FRAHM J.-M., POLLEFEYS M.: Personal photograph enhancement using internet photo collections. *IEEE TVCG 20*, 2 (2014), 262–275. 1