Multiview Image Completion with Space Structure Propagation

Seung-Hwan Baek Inchang Choi Min H. Kim*
Korea Advanced Institute of Science and Technology (KAIST)
{shwbaek;inchangchoi;minhkim}@vclab.kaist.ac.kr

Abstract

We present a multiview image completion method that provides geometric consistency among different views by propagating space structures. Since a user specifies the region to be completed in one of multiview photographs casually taken in a scene, the proposed method enables us to complete the set of photographs with geometric consistency by creating or removing structures on the specified region. The proposed method incorporates photographs to estimate dense depth maps. We initially complete color as well as depth from a view, and then facilitate two stages of structure propagation and structure-guided completion. Structure propagation optimizes space topology in the scene across photographs, while structure-guide completion enhances, and completes local image structure of both depth and color in multiple photographs with structural coherence by searching nearest neighbor fields in relevant views. We demonstrate the effectiveness of the proposed method in completing multiview images.

1. Introduction

Digital photography has been stimulated from the perspective of collaboration, yielding casual multiview photographs of a scene. Even though multiview images are getting popular from the recent advances of digital imaging devices such as mobile cameras and light field cameras, consistent image editing of multiview photographs rarely has been discussed. For instance, most traditional image completion methods focus on an individual photograph [9, 31, 2]; hence, applying such single image completion methods is not able to retain the geometric consistency of space structures in images.

Traditional inpainting [17, 28] focuses on preserving coherency of *structure* in an image. In contrast, we propose a novel solution that allows for geometric coherency, so-called *space structure*, while completing multiview images. Multiview photographs allow us to estimate depth, utilized as geometric constraints. While completing multiview images, our method accounts for not only structure in an image, but also space structure across multiview images.

Our method first receives a target region from a user in one of the multiview photographs. The region is then completed by three steps: (1) the *preprocessing* step estimates dense depth maps and camera parameters using structure from motion, and completes the color and the depth of a reference view, (2) the *structure propagation* step conveys the space structure of the previous completions to the next image to be completed, and (3) the *structure-guided completion* enhances the quality of completion by incorporating the transferred space structure, searching for local similarity via the nearest neighbor field across multiview photographs. We perform the structure propagation and structure-guided completion steps for all images, resulting in multiple completed images for the first iteration. To enhance the image completion, our iterative approach repeats these two steps until we reach to the final completed images.

2. Related Work

Image completion synthesizes image structure to fill in larger missing areas using exemplar [9, 31, 2, 10, 17]. Our work is a branch of image completion, focusing on image completion of multiple photographs.

Single Image Completion Criminisi et al. [9] prioritized patches to be filled, and greedily propagated the patches from the known source region to the target region. Wexler et al. [31] solved the image completion problem with an expectation-maximization (EM) optimization using an image pyramid. Since the nearest neighbor field (NNF) search is the computational bottleneck of image completion, Barnes et al. accelerated the process by introducing an approximated search algorithm [2]. Darabi et al. [10] considered the gradients of images for patch-based synthesis and also introduced alpha-blending of multiple patches. Our method is based on the framework of Wexler et al. [31] with the main difference that our approach simultaneously completes the color and the depth of multiview photographs with structural coherence.

*Corresponding author

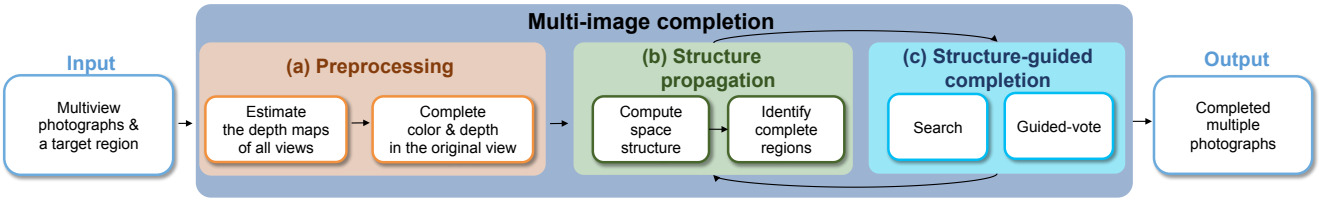


Figure 1. We take multiview photographs and a target region mask as input. (a) Preprocessing: per-view information is estimated such as camera parameters and depth maps. A color image and its depth map are completed as reference from a view (see Section 4). (b) Structure propagation: the space structure of the previously completed view is transferred to the next view to be completed. The target region is estimated with consideration of the space structure (Section 3.2). (c) Structure-guided completion: the local structure in each image is iteratively refined for color as well as depth (Section 3.3). We iterate *structure propagation* and *structure-guided completion*, resulting in multiple color images completed with geometric consistency.

Stereo Image Completion Since the disparity of similar colors depends on depth, completing the color and the depth on a target region in stereo images is a chicken-and-egg problem. Wang et al. [30] jointly synthesized the color and the depth using a greedy search [9]. They executed the image completion independently on left and right images and then evaluated them to constrain the perspective consistency. However, the consistency of image completion is not always guaranteed, particularly when processing a large holes. Morse et al. [23] initially inpainted the depth via a simple interpolation with a PDE solver, and exploited the inpainted depth to complete the color. Since PDE-based inpainting of the depth is even more challenging than completing the color due to lack of clues, the inaccurate depth completion in this method tends to reduce the quality of image inpainting significantly. Similar to traditional image-based rendering methods [32], the goal of these stereo inpainting methods is merely to remove target foreground objects by filling in the occlusion region, which is caused by the objects, using patches from the background. Note that large-scale structures in image completion cannot be accounted for the previous stereo completion methods; e.g., the synthesized structure in one image could be outside of the target region on other images due to perspective projection. Alternatively, Luo et al. [22] asked users to manually complete the target depth map and then performed image completion on the color images only. While this method can complete the target region with foreground objects, users need to manually correct the target region of both views considering the visibility of completed depth.

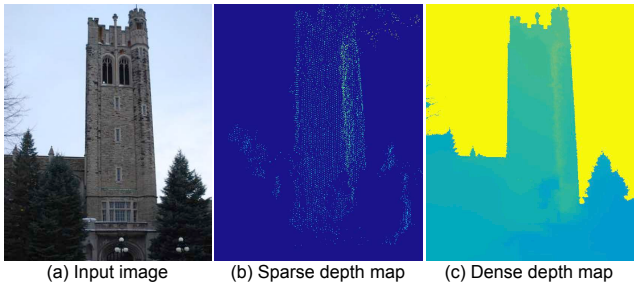
Multiple Image Completion Hays and Efros [16] embedded scene semantics into target regions by using millions of prior Internet photographs. Darabi et al. [10] considered a mixture of two patches from different images in applying a random correspondence search method [2]. Barnes et al. [3] proposed a 3D data structure for efficient NNF search from multiple source images. Wexler et al. [31] introduced a video completion method that propagates image completion with a smoothness constraint of consecutive frames. Even though these methods accounted for multiple images or patches, they presumed that the camera view point does not change significantly. Multiple images are merely employed to improve the quality of image completion. They all disregard the three-dimensional space structure of image completion. To the best of our knowledge, our method is the first work that accounts for the geometric consistency of space structure in image completion of multiple photographs.

Depth-Image-Based Rendering Image-based rendering (IBR) is a traditional method that synthesizes a novel view from multiple photographs [21, 15, 5]. IBR has been extended as depth-image-based rendering (DIBR) by adding depth information from multiview stereo [12, 32, 19, 6]. DIBR is used for synthesizing viewpoint with interpolation and extrapolation by means of direct and inverse projection on the segmented pixels from a novel viewpoint. The typical problems in DIBR viewpoint synthesis are completing the missing regions such as occlusions and cracks. In recent works of DIBR [24, 11, 14, 7], small or uniform regions can be restored by diffusion-based methods, such as [4]. Largely missing regions are restored by patch-based approaches [9, 29]. The ultimate goal of image completion in DIBR is to reconstruct missing regions using the background information, while keeping projected region intact. In contrast, we incorporate our patch-based approach to complete not only background but also foreground objects with geometric consistency in multiple photographs. Also, we complete the projected region as well as the missing region, in order to handle the artifacts of projected region arising from inaccurate depth values.

3. Multiple Image Completion

Our pipeline consists of three steps: *preprocessing*, *structure propagation*, and *structure-guided completion*. See Figure 1 for an overview.

Input We use multiview unstructured photographs I_i containing a static scene from different views $v_i \in V$ as input, where V is the set of every view. A user specifies an original image I_o, and give an original mask M_o indicating the



(a) Input image (b) Sparse depth map (c) Dense depth map

Figure 2. For given photographs (a), we estimate camera parameters and point clouds. A sparse depth map (b) is estimated by projecting the point clouds into each view. We propagate the sparse depth values using Matting Laplacian [20], resulting in a dense depth map per a view (c). Image courtesy of Olsson et al. [25].

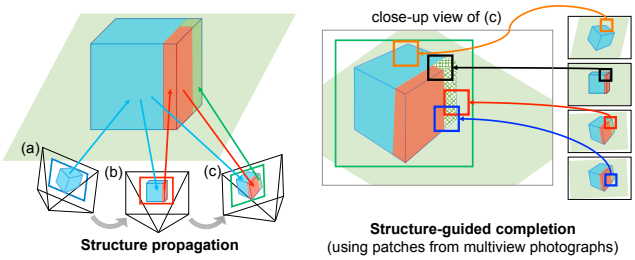


Figure 3. Schematic diagram of our multiview patch-based synthesis. *Structure propagation*: after completing the current view (a), we select a next nearest target view (b), and propagate the previous completion to (b). *Structure-guided completion*: we utilize multiple photographs (a), (b) and (c) to complete a target view.

completion region Ω_o to be completed on the original image I_o. Note that our method automatically completes the corresponding target region Ω_i on the i-th image, resulting in a consistent completed image $\hat{I}_i$.

3.1. Preprocessing

Since we target casually taken multiple photographs for image completion, *preprocessing* starts to estimate dense depth maps using structure from motion (SfM). See Figure 2 for example. We then conduct initial color and depth image completion on a reference view v_o. See Sections 3.3 and 4 for details on structure-guided image completion and SfM implementation.

3.2. Structure Propagation

Subsequent to estimating the camera parameters from SfM, we first select relevant views, where the initial completion is visible in multiple photographs. The completed color and depth in a target region of v_o is propagated to guide the completions of the selected views to preserve geometric consistency. See Figure 3 for overview.

Inverse & Direct Projection Let $\hat{V}$ be the set of completed multiviews, where $v_i \in \hat{V}$ is progressively completed from the original view $v_o \in \hat{V}$. In order to determine next view to be completed $v_j \notin \hat{V}$, we compare the lastly completed view v_i and other remaining views $v_j \notin \hat{V}$ in terms of the Euclidean distance of the camera positions and the orientation difference.

Once the most similar view v_j is selected, the completed color $\hat{I}_i$ and depth $\hat{D}_i$ of the homogeneous pixel coordinates $\dot{p}_i \in v_i$ are projected inversely to the homogenous global coordinates $\mathbf{p}$ via perspective projections of each view $v_i \in \hat{V}$:

$$\mathbf{p} = \hat{D}_i(p_i)\mathbf{E}_i^{-1}\mathbf{K}_i^{-1}\dot{p}_i, \tag{1}$$

where $\mathbf{K}$ and $\mathbf{E}$ are the intrinsic and extrinsic parameters of each view v_i in relation to $\dot{p}_i$ and $\mathbf{p}$.

Now we can accumulate each completion $\dot{p}_i$ in the global coordinates of a 3D point $\mathbf{p}$, thus project image completion of $\dot{p}_i$ to $\dot{p}_j$ in the next view v_j directly:

$$\dot{p}_j = \mathbf{K}_j\mathbf{E}_j\mathbf{p}. \tag{2}$$

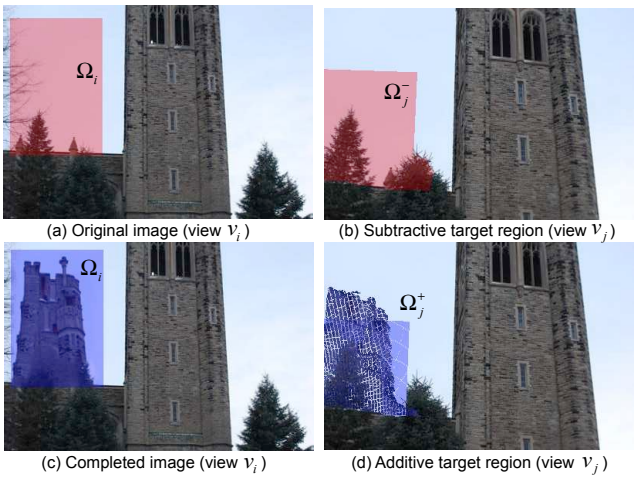
This mechanism of inverse and direct projection allows us to preserve geometric consistency in multiple photographs although this results in cracks due to perspective projection and occlusion. These cracks are handled in the stage of structure-guided completion (see Section 3.3).

Identifying Completion Regions Traditional stereo completion methods [30, 23] complete target region using background only; therefore, no additive structures are allowed in completion. In contrast, our structure propagation allows for introduction of additive structures in completion. To do so, we separate the target region Ω_j into two associated regions of *addition* and *subtraction*, Ω_j^+ and Ω_j^- in the new view v_j. We need two steps to identify the complete regions in the new view. First, we should identify the subtractive regions Ω_j^- to be removed in v_j. We inversely project each pixel $\dot{p}_j \in v_j$ to corresponding $\dot{p}_o \in v_o$ using Equations (1) and (2). We then determine if the corresponding pixel $\dot{p}_o$ belongs to the original region Ω_o. See Figures 4(a) and (b).

Second, we identify the additive regions Ω_j^+ where some additive structure to be introduced in completion. Note that this Ω_j^+ could be different from Ω_j^- depending on which structure is completed. Since the space structure exists only in the original view v_i, we project the completed pixel region Ω_i to v_j via inverse and direct projection. This Ω_j^+ therefore results in cracks and discontinuity. We expand the region to a convex hull that surrounds Ω_j^+ to solve these artifacts using patch-based synthesis (see Section 3.3). Finally, we define a union region of these two regions as follows: $\Omega_j = \mathrm{Conv}(\Omega_j^+) \bigcup \Omega_j^-$.

Note that this process does not account for a new structure to be completed in Ω_j^- in the next step. We solve this problem by iterating this completion workflow rounding the series of photographs.

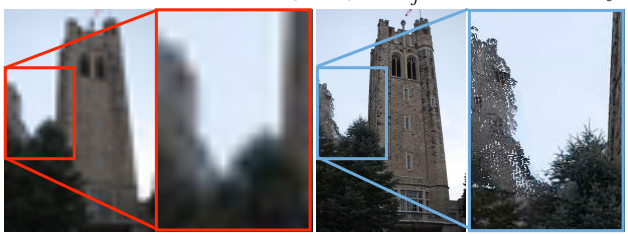
Compared to DIBR methods [24, 11, 14, 7], we synthesize missing regions as well as structure-guide regions in or-



(a) Original image (view v_i) (b) Subtractive target region (view v_j)

(c) Completed image (view v_i) (d) Additive target region (view v_j)

Figure 4. (a) We complete the target region Ω_i from the current view v_i. (b) Corresponding to Ω_i of the view v_i, we estimate the subtractive region Ω_j^- on a new view v_j. (c) & (d) Completion of view v_i introduces additive target region Ω_j^+ on the new view v_j.



(a) Color structure guide at low frequency (b) Color structure guide at high frequency

Figure 5. We use the low frequency shape (a) as structure guide. Since the projected guide includes severe corruption with cracks and missing details, we fix it with our multi-view patch synthesis.

der to to handle depth errors inevitably induced by completion. DIBR methods maintain I'_j and D'_j for Ω_j^+, and complete the remaining target region $\Omega_j^C = \{p \in \Omega_j^- | p \notin \Omega_j^+\}$. However, as the structure-guide region Ω_j^+ spreads over the image sparsely and contains artifacts due to the inaccurate depth values, this approach yields poor results as shown in Figure 6(c).

3.3. Structure-Guided Multiple Completion

3.3.1 Space Structures

Contrary to state-of-the-art methods [31, 2, 10], we have *additive* completion regions Ω_j^+ in addition to *subtractive* completion regions Ω_j^-. Prior to synthesizing the completion regions Ω_j, we first transfer the pre-completed color I_i and depth D_i on Ω_i to Ω_j^+ using Equations (1) and (2) as *structure guide*. We utilize this transferred information while synthesizing Ω_j as *guide*. We denote the structure guides of color and depth as I'_j and D'_j, respectively.

The traditional patch-based completion frameworks [31, 2] build an image pyramid that includes the target regions from coarse to fine levels. These methods search for a similar patch in the source region per the target *patch* centered at target pixel p_j, in order to substitute the color of p_j with that of the similar source pixel, so-called *search*. Now we have a source patch matched with target pixel p_j. As source patches corresponding to the neighbor pixels of p_j could have pixels that lie on the pixel position p_j, they interpolate the colors of those pixels to determine the final color of p_j, so-called *vote*. For each level, *search* and *vote* are performed iteratively, and this approach results in a completed image at the finest level.

Our proposed approach imposes the structure guides through this optimization problem. We synthesize the target region Ω_j while maintaining the projected structure guides of I'_j and D'_j on Ω_j^+. Our intuition is that the structure guide is roughly correct while the details are crisp (see Figure 5). Our approach harmonizes *patch-based synthesis with space-structure guide* in order to achieve geometric consistency in multi-image completion.

3.3.2 Coherence vs. Structure Guide

Our goal is to synthesize the target region Ω_j using the source regions of every view $\Psi_\forall = \left\{\Psi_k = A_k \setminus \Omega_k^{-1} | k \in V\right\}$, where A_k is the set of every pixel in view k. Our energy function consists of two terms, a *coherence* and a *guidance* term:

$$E = \sum_{p_j \in \Omega_j} \min_{q_\forall \in \Psi_\forall} \{E_{coherence}\left(p_j, q_\forall\right) + E_{guide}\left(p_j\right)\},$$

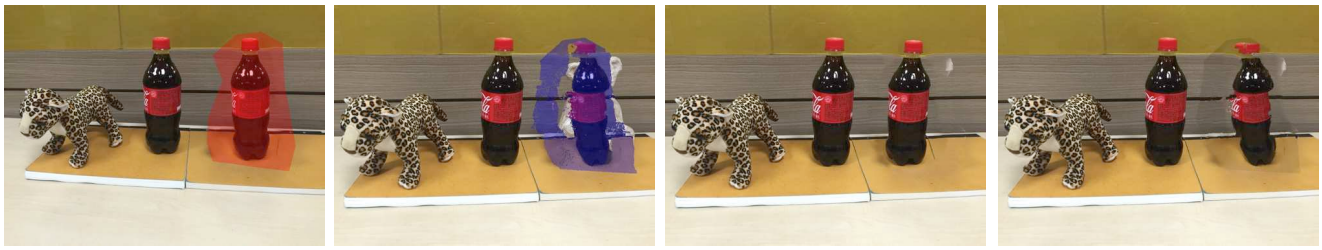
where p_j is a target pixel on the target region Ω_j, and $q_\forall$ is a source pixel on the source region $\Psi_\forall$.

The coherence term $E_{coherence}$ describes the similarity of the target and source patch in terms of color in CIELAB [8] and depth gradients, defined by:

$$\begin{aligned} E_{coherence}\left(p_j, q_\forall\right) =& C\left(\bar{I}_j\left(p_j\right), \bar{I}_\forall\left(q_\forall\right)\right) \\ &+ \alpha C\left(\nabla\bar{D}_j\left(p_j\right), \nabla\bar{D}_\forall\left(q_\forall\right)\right), \end{aligned} \tag{3}$$

where C is the L2 norm of the difference of two values, and α is the weighting constant. $\bar{I}_j(p_j)$ $\nabla\bar{D}_j(p_j)$ are the color patches and the depth gradient patches centered at pixel p_j.

Note that this function resembles that one from Darabi et al. [10] with the following difference. While Darabi et al. took both color and *color gradients* in their coherence term, we take both color and *depth gradients* for evaluating space-structure coherence. Our principal intuition for examining *depth gradients* is that, if we evaluate the difference of the depths directly, we cannot utilize many patches of the same *shape* but in a different distance to complete a depth map. Our light-weight shape representation of *depth gradients* allows us to evaluate the shape similarity of depth patches to synthesize depth information in missing regions, whereas traditional stereo completion methods [30] examined direct depth information. α varies depending on the level of quality of the reconstructed depth maps.



(a) Original input completion (b) Our intermediate projection (c) Our structure-guide completion (d) Traditional DIBR method

Figure 6. (a) shows an initial completion of the original view using our RGB-D completion method. (b) presents our intermediate projection of the completed structure to a new view. Note that the distorted space structure is caused by inaccurate depth. (c) shows the result of our completion. The distorted color and depth are updated iteratively through patch-based synthesis with structure guide. (d) compares our result with a traditional DIBR method, which complete the missing region with background and is incapable of handling depth inaccuracy.

The guidance term E_{guide} accounts for the similarity of the target patch and the structure guide for pixel $p_j \in \Omega_j^+$:

$$E_{guide}\left(p_j\right) = \beta C\left(I_j\left(p_j\right), I'_j\left(p_j\right)\right) + \gamma C\left(D_j\left(p_j\right), D'_j\left(p_j\right)\right), \quad (4)$$

where β and γ are the weighting factors for color and depth differences, respectively. The overall scales of β and γ balance the local image structure and the space structure guide. In order to solve this objective function, we complete the color image first by our patch-based completion considering structure guide I'_j. We then complete the depth using the completed color as well as the structure guide D'_j. This guide energy term is mainly handled by the *guided voting* stage. See Section 3.3.3 for detail on optimization.

3.3.3 Color and Depth Optimization

Our optimization is originated from Darabi et al. [31] with the main difference for structure-guided image completion. Our optimization consists of two steps: *search* and *guided vote*. Here we mainly describe the differences from the base algorithm.

Guide Initialization We first construct color and depth image pyramids of the original data and the structure-guide, $I_{j,l}, D_{j,l}$ and $I'_{j,l}, D'_{j,l}$, respectively, using a Lanczos kernel following Darabi et al. [10]. Here let $l \in [1, L]$ be the level of the pyramid and L be the number of levels ($I_{j,L} = I_j$). Accordingly, target completion region Ω_j and sub-region with structure guide Ω_j^+ are transformed into each level l, resulting in $\Omega_{j,l}$ and $\Omega_{j,l}^+$. Then, the structure guide $I'_{j,1}$ at the coarsest level is initially copied to the additive completion region $\Omega_{j,1}^+ \in I_{j,1}$ to impose the coarse image completion resemble the structure guide. The rest of completion region $\Omega_j \backslash \Omega_j^+$ is filled in by interpolating the gradients of the boundary pixels with a PDE solver. See Algorithm 1 for an overview.

Search For this search stage, we mainly use Equation (3) to find the NNF of the closest patch at pixel $q_\forall \in \Psi_\forall$ to match patch at pixel $p_j \in \Omega_j$. Here we are motivated to adopt a popular search method, proposed by Barnes et al. [2], which is based on random search and propagation. Since our method takes multiple images as input, we extend the search range from a single to multiple photographs, by concatenating them as one, similar to Barnes et al. [3].

Algorithm 1 COMPLETION()

Input: I_j, I'_j, D_j, D'_j
Output: $\hat{I}_j, \hat{D}_j$
1: Create image pyramids $I_{j,l}, I'_{j,l}, D_{j,l}, D'_{j,l}$
2: Initialize $\hat{I}_{j,1}, \hat{D}_{j,1}$
3: **for** $l = [1, ..., L]$ **do**
4: $\quad NNF_{j,l} \leftarrow \text{SEARCH}(\hat{I}_{j,l}, I_{\forall,l}, \nabla\hat{D}_{j,l}, \nabla D_{\forall,l})$
5: $\quad \hat{I}_{j,l}, \hat{D}_{j,l} \leftarrow \text{VOTE}(NNF_{j,l}, I_{\forall,l}, I'_{p_j,l}, \nabla D_{\forall,l}, D'_{p_j,l})$
6: **end for**

Guided Vote Our *guided vote* is the main difference from the state-of-the-art methods. From the search process, we collect colors and depth gradients of the patches from the NNF of neighbor pixels. We estimate *initial* color and depth gradients from the NNF, $I^*_{j,l}(p_j)$ and $\nabla D^*_{j,l}(p_j)$, centered at target pixel p_j using the weighted sum approach, proposed by Wexler et al. [31]. These coherent color and depth values are updated with a guidance of space structure, as described in Equations (3) and (4), as follows.

Since we obtain the initial colors $I^*_{j,l}$ from coherence search, to account for space structure, we update color at pixel p_j by linearly interpolating the coherent color $I^*_{j,l}(p_j)$ and structure guide $I'_{j,l}(p_j)$, where p_j is on the additive completion region Ω_j^+:

$$\hat{I}_{j,l}\left(p_j\right) \leftarrow \left(1 - \beta(l)\right) \times I^*_{j,l}\left(p_j\right) + \beta\left(l\right) \times I'_{j,l}\left(p_j\right), \quad (5)$$

where $\beta(l)$ is the frequency-dependent weight. We define $\beta(l)$ as a truncated function that interpolates the level of weighting linearly:

$$\beta\left(l\right) = \begin{cases} 1, & l \le l_s \\ (l_e - l)/(l_e - l_s), & l_s < l \le l_e \\ 0, & l_e < l \end{cases}, \quad (6)$$

where l_s and l_e are the user parameters that control the influence of structure guide I'_j.

Although we compare depth gradients in NNF search, our goal is to complete depth $\hat{D}_{j,l}$ rather than $\nabla D^*_{j,l}$ on the target region $\Omega_{j,l}$ and to achieve geometric consistency with structure guide $D'_{j,l}$. For the region $\Omega^C_{j,l} = \Omega_{j,l} \setminus \Omega^+_{j,l}$, we want to complete depth $\hat{D}_{j,l}$ from the voted depth gradients $\nabla D^*_{j,l}$. Also, we should preserve space structure $D'_{j,l}$ projected from the previous view, for the other region. Finally, the optimal $\hat{D}_{j,l}$ can be estimated by minimizing the following objective function:

$$\min_{\hat{D}_{j,l}} \sum_{\Omega^C_{j,l}} ||\nabla \hat{D}_{j,l} - \nabla D^*_{j,l}||^2, \tag{7}$$

with boundary conditions $\hat{D}_{j,l}|_{\delta\Omega^C_{j,l}} = D'_{j,l}|_{\delta\Omega^C_{j,l}}$. We optimize this objective equation with a constraint term of the boundary conditions: $\gamma \left(\hat{D}_{j,l}|_{\delta\Omega^C_{j,l}} - D'_{j,l}|_{\delta\Omega^C_{j,l}} \right)^2$, where a constant γ is a weighting constant that appears in Equation (4). By applying the Euler-Lagrange equation, the optimal $\hat{D}_{j,l}$ satisfies the Poisson equation: $\nabla^2 \hat{D}_{j,l} = \nabla \cdot D^*_{j,l}$. Algorithm 2 summarizes our completion method.

4. Implementation Details

Preprocessing Point clouds and camera parameters are estimated by using SfM methods [27, 13] from multiple photographs. We then project the point clouds into each view, resulting in a sparse depth map per a view using Equation (1) (see Figure 2b). We propagate sparse depth estimates to every pixel, yielding a dense depth map D_i per view (Figure 2c). Once we build depth information in photographs, for the target region Ω_o of input view v_o, we perform our completion method (Section 3.3) without any structure guide, resulting in a completed color image $\hat{I}_o$ and the completed depth map $\hat{D}_o$.

Algorithm 2 MULTICOMPLETION()

Input: $I_\forall, \Omega_o$
Output: $\hat{I}_\forall, \hat{D}_\forall$

```
Estimate per-view information K_∀, E_∀ and D_∀.
Î_o, D̂_o ← COMPLETION(I_o, ∅, D_o, ∅)
V̂ ← {v_o}
for iter = [1, ..., N] do
    while |V̂| < |V| do
        Select a view v_j to be completed
        Identify completion region Ω_j and Ω_j^+
        Generate structure guide I'_j and D'_j
        Î_j, D̂_j ← COMPLETION(I_j, I'_j, D_j, D'_j)
        V̂ ← V̂ ∪ {v_j}
    end while
    V̂ ← {v_o}
end for
```

Multiple Source Images We complete each view of multiple photographs with consideration of space structure, as described in Section 3. Different from traditional completion methods, we use multiple images as input, and therefore we extend the patch search range from a single to multiple photographs, by concatenating them as one following Barnes et al. [3]. When we complete view v_j, we select M source views uniformly on the sorted views w.r.t. the pose difference with view v_j, for computational efficiency. The concatenated image is used for our structure-guide completion (Section 3.3). whereas previous stereo completion methods [30, 22, 23] utilize one of the left and right images to complete the selected image.

Depth Denoising In general, a depth map reconstructed from the Poisson equation suffers from severe noise. We therefore conduct an additional process of depth denoising that smooths out the estimated depth in the target region. We first segment the initially-estimated depth map into superpixels [1] by using the color information in the completed color image. We then perform median filtering per each depth superpixel to clean noise. Matting Laplacian [20] is applied to remove quantization error over superpixel edges in the depth map.

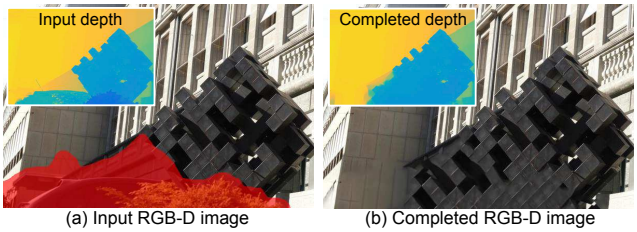
Parameters We use a 7×7 patch for color and depth. Parameter α in Equation (3) controls the balance between color and depth gradients in calculating individual coherence within each image. α is set to a lower value to weight the color similarity for local coherence of each image. Parameter β in Equation (4) is determined by l_s and l_e in Equation (6), which controls the balance between individual coherence and global geometric consistency. Parameter γ in Equation (4) controls the inaccurate depths in propagated structure guides. A higher value of γ suppresses the depth errors more, which is originated from Poisson reconstruction. In this experiment, we set the values of α, l_s, l_e and γ as 0.01, 0.3, 0.6, and 100. See Figure 9. In addition, we use ten levels of the image pyramid and the number of source views is set to three. We found that two iterations are enough for satisfying geometric consistency across multiple views. We complete three to six images selected from three to twenty images used for SfM, considering computational feasibility.

5. Results

Our Matlab and C++ implementation took approx. 300 seconds to complete color and depth of a 680×900 image with 2.5×10^4 target pixels. Computing time increases linearly proportional to the number of input images in general, taking 25 minutes for the multiple images with five photographs. To validate the performance of our method, we applied our method to three types of input data: stereoscopic images, a single RGB-D image and multiple images.

(a) Input (left) (b) Ours (left) (c) Ours (right) (d) [Wang et al.] (left) (e) [Wang et al.] (right)

Figure 7. Stereo image completion. Our method completes the stereo image (a) with complex structures, resulting in the completed images (b) and (c) with geometric consistency by creating or removing objects. Wang et al. [30] utilize background patches only in turn. They cannot preserve the structure of the scene, such as the pipe in the left-side of the scene (d) and (e). Also note that there are some notable artifacts in (d) and (e), which are originated from the individual completion of each view. Image courtesy of Scharstein et al. [26].



(a) Input RGB-D image (b) Completed RGB-D image

Figure 8. For a single RGB-D image data (a), our method completes the target region (red area), resulting in a consistent RGB-D image (b). Image courtesy of Kim et al. [18].

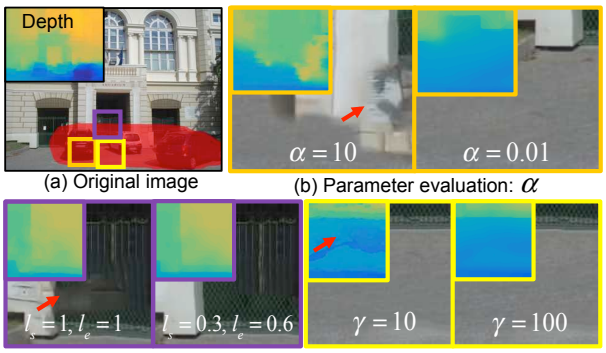
Figure 7 compares our method with Wang et al. [30] for a stereo image. Our method reconstructs the left and the right image with geometric consistency while completing the connection of left-most pipe naturally, while Wang et al. [30] fills the target region with background objects, inhibiting natural completion of the complex structure.

Figure 8 shows the result of completing color and depth in a single RGB-D image. Our method can achieve geometric consistency of not only color but also depth. Note that depth patches cannot be compared directly for shape similarity due to depth value differences. Our gradient-based depth synthesis allows us to compare shape similarity in terms of gradients through the NNF search.

To the best of our knowledge, multiple image completion has not been studied before. For evaluation, we therefore applied a stereo completion algorithm of Wang et al. [30] on a pair of stereo images consequently to process multiple images. Wang et al. [30] impose geometric consistency by comparing the colors of corresponding pixels in completed left and right images. Figure 10 demonstrates that our method can synthesize new structures consistently across multiple photographs, while maintaining coherence of each image. In contrast, Wang et al. [30] cannot account for large structures and fail to preserve both geometric consistency and coherence inside each image.

6. Discussion and Conclusions

Propagated depth maps could be unstable due to the errors originated from SfM (see Figure 6b). To solve this problem, we account for color and depth coherency in an image. The propagated space structure is simultaneously



(a) Original image (b) Parameter evaluation: α

(c) Parameter evaluation: $\beta(l_s, l_e)$ (d) Parameter evaluation: γ

Figure 9. (a) We investigate the effects of parameters: (b) α (orange box), (c) $\beta(l_s, l_e)$ (purple) and (d) γ (yellow). We fix other parameters while varying only one of them, with the following values: $\alpha = 0.01$, $l_s = 0.3$, $l_e = 0.6$ and $\gamma = 100$.

updated through iteration with consideration of spatial frequency. Nevertheless, the quality of depth maps could be degraded in case of dynamic or feature-less scenes, which potentially cause the failure of our method. Note that initial completion of the target image can be conducted by any other state-of-the-art single image completion methods [17, 10].

We have presented a novel multiple image completion method that preserves the geometric consistency among different views. The proposed method consists of structure propagation and structure-guided completion. Our structure-guided completion, which is designed as a single optimization framework, exhibits superior results in terms of coherency and consistency. Our versatile algorithm enables to complete not only multiple images, but also stereoscopic images. In addition it allows to fill the empty region with foreground as well as background objects, which has been challenging so far in the previous stereo completion methods [30, 23].

Acknowledgements

Min H. Kim gratefully acknowledges Korea NRF grants (2013R1A1A1010165 and 2013M3A6A6073718) and additional support by Samsung Electronics (G01140381) and an ICT R&D program of MSIP/IITP (10041313) in addition to Sung-Hyun Choi (Samsung) for helpful comments.

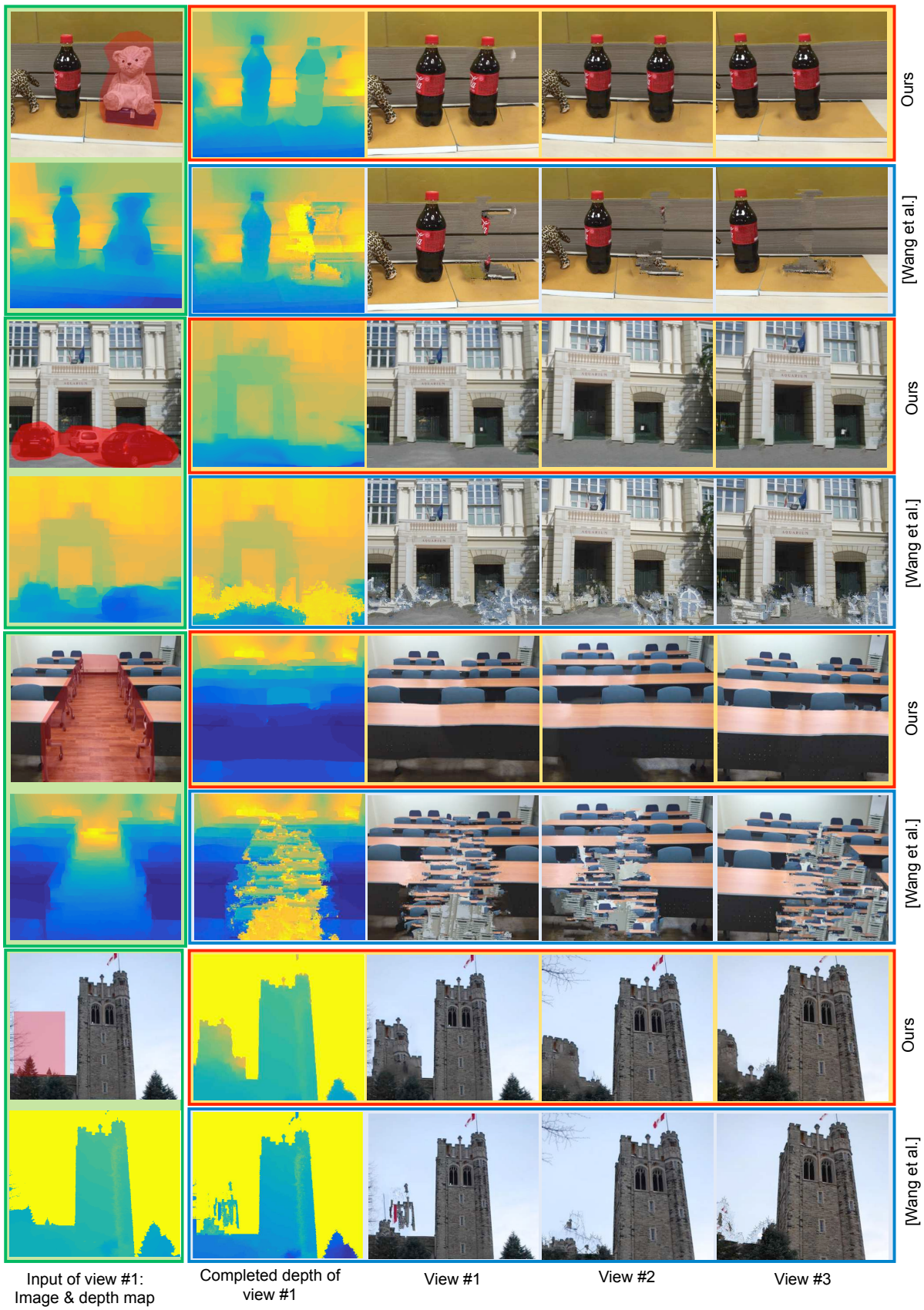


Figure 10. Multi-image completion. The first column shows input color and depth. Red color indicates completion regions. The second and the third column shows completed depth and color, respectively, compared with Wang et al. [30]. The fourth and the fifth column presents completions from different views. Refer to the supplemental materials for more results. Image courtesy of Olsson and Enqvist [25].

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 34(11):2274–2282, 2012.

[2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph. (TOG)*, 28(3):24:1–11, 2009.

[3] C. Barnes, F.-L. Zhang, L. Lou, X. Wu, and S.-M. Hu. PatchTable: Efficient patch queries for large datasets and applications. *ACM Trans. Graph. (TOG)*, 34(4):97:1–10, 2015.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. ACM SIGGRAPH '00*, pages 417–424, 2000.

[5] C. Buehler, M. Bosse, L. McMillan, S. J. Gortler, and M. F. Cohen. Unstructured lumigraph rendering. In *Proc. ACM SIGGRAPH '01*, pages 425–432, 2001.

[6] G. Chaurasia, S. Duchêne, O. Sorkine-Hornung, and G. Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. Graph. (TOG)*, 32:30:1–12, 2013.

[7] S. Choi, B. Ham, and K. Sohn. Space-time hole filling with random walks in view extrapolation for 3D video. *IEEE Trans. Image Processing (TIP)*, 22(6):2429–2441, 2013.

[8] CIE. Colorimetry. Technical report, 1986.

[9] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Processing (TIP)*, 13(9):1200–1212, 2004.

[10] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen. Image melding: combining inconsistent images using patch-based synthesis. *ACM Trans. Graph. (TOG)*, 31(4):82:1–10, 2012.

[11] I. Daribo and B. Pesquet-Popescu. Depth-aided image inpainting for novel view synthesis. In *Proc. IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, pages 167–170, 2010.

[12] C. Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Proc. SPIE*, volume 5291, pages 93–104, 2004.

[13] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 32(8):1362–1376, 2010.

[14] J. Gautier, O. L. Meur, and C. Guillemot. Depth-based image completion for view synthesis. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4, 2011.

[15] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. ACM SIGGRAPH '96*, pages 43–54, 1996.

[16] J. Hays and A. A. Efros. Scene completion using millions of photographs. *ACM Trans. Graph. (TOG)*, 26(3):4:1–8, 2007.

[17] J.-B. Huang, S. B. Kang, N. Ahuja, and J. Kopf. Image completion using planar structure guidance. *ACM Trans. Graph. (TOG)*, 33(4):129:1–10, 2014.

[18] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. H. Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph. (TOG)*, 32(4):73:1–12, 2013.

[19] J. Kopf, F. Langguth, D. Scharstein, R. Szeliski, and M. Goesele. Image-based rendering in the gradient domain. *ACM Trans. Graph. (TOG)*, 32(6):199:1–9, 2013.

[20] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. *ACM Trans. Graph. (TOG)*, 23(3):689–694, 2004.

[21] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. ACM SIGGRAPH '96*, pages 31–42, 1996.

[22] S. J. Luo, Y. T. Sun, I. C. Shen, B. Y. Chen, and Y. Y. Chuang. Geometrically consistent stereoscopic image editing using patch-based synthesis. *IEEE Transactions on Visualization and Computer Graphics. (TVCG)*, 21(1):56–67, 2015.

[23] B. Morse, J. Howard, S. Cohen, and B. Price. Patchmatch-based content completion of stereo image pairs. In *Proc. 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT) '12*, pages 555–562, 2012.

[24] K.-J. Oh, S. Yea, and Y.-S. Ho. Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-D video. In *Proc. Picture Coding Symposium (PCS) '09*, pages 1–4, 2009.

[25] C. Olsson and O. Enqvist. Stable structure from motion for unordered image collections. In *Proc. 17th Scandinavian Conference on Image Analysis*, number 12, pages 524–535. Springer, 2011.

[26] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. *Pattern Recognition: 36th German Conference, GCPR 2014*, pages 31–42, 2014.

[27] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graph. (TOG)*, 25(3):835–846, 2006.

[28] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum. Image completion with structure propagation. *ACM Trans. Graph. (TOG)*, 24(3):861–868, 2005.

[29] W. Sun, O. C. Au, L. Xu, Y. Li, and W. Hu. Seamless view synthesis through texture optimization. *IEEE Trans. on Image Processing (TIP)*, 23(1):342–355, 2014.

[30] L. Wang, H. Jin, R. Yang, and M. Gong. Stereoscopic inpainting: Joint color and depth completion from stereo images. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[31] Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 29(3):463–476, 2007.

[32] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. Graph. (TOG)*, 23(3):600–608, 2004.