Vis Comput (2014) 30:833–843
DOI 10.1007/s00371-014-0961-2

ORIGINAL ARTICLE

Stereoscopic image completion and depth recovery

Tai-Jiang Mu · Ju-Hong Wang · Song-Pei Du · Shi-Min Hu

Published online: 8 May 2014

Abstract In this paper, we have proposed a novel patch-based method for automatic completion of stereoscopic images and the corresponding depth/disparity maps simultaneously. The missing depths are estimated in local feature space and a patch distance metric is designed to take the appearance, depth gradients and depth inconsistency into account. To ensure the proper stereopsis, we first search for the proper stereoscopic patch in both left and right images according to the distance metric, and then iteratively refine the images. Our method is capable of dealing with general scenes including both frontal-parallel and non-frontal-parallel objects. Experimental results show that our method is superior to previous ones with better stereoscopically consistent content and more plausible completion.

Keywords Stereoscopic · Completion · Depth inconsistency · Distance metric

1 Introduction

3D techniques have gained great success in recent years. The success in turn inspires the development of stereoscopic 3D capture and display techniques, which brings in masses of stereoscopic 3D content, making 3D editing techniques and applications on the agenda, some of which have been explored, such as stereo cloning [27,28,38], stereoscopic warping [10,30], disparity mapping [23], stereoscopic retargeting [2,6,24], stereo painting [20], just to name a few.

In this paper, we focus on image completion for stereoscopic image pairs. We define "holes" by the missing or removed regions, which are filled by proper synthesized content based on other parts of the image. Massive researches have been conducted for single image completion, some of which have been proved to be of good quality [1,13,21,37]. Completing the left and right images separately may produce good results for monocular viewing, while this may probably break the stereo correspondence and affect the 3D viewing experience. Especially for the content which is missing in one view but present in the other (refer to Figure 1(c) in [29]), known as *half occlusion*. Additionally, depth/disparity is available for producing more visual plausible results [8,14] when it comes to stereoscopic 3D. More importantly, depth constraint should be carefully taken into consideration to guarantee a comfortable and consistent 3D viewing experience [17,22,33].

This paper has proposed a patch-based refine method to complete the missing regions in a stereoscopic image pair such that the filled color image content is stereoscopic as well as consistent, simultaneously recovering the underlying depth/disparity maps. Our method takes the *depth inconsistency* (see Sect. 3.3) into account and achieves better stereoscopically consistent completion than previous method [29]. As the depth maps usually contain some deviations, especially in those smooth regions, we do not seek to pre-complete full disparity maps as performed in [29]. Instead, the depth maps and color images are simultaneously completed in a *mutual promotion* way using an iterative fashion. Current best matched patches, according to a depth gradient domain patch distance metric, will be found out for the unknown regions in each iteration and the missing depths will be inferred using *depth shift* (see Sect. 3.4.4) from its

T.-J. Mu (✉) · S.-P. Du · S.-M. Hu
TNList, Department of Computer Science and Technology,
Tsinghua University, Beijing 100084, China
e-mail: mmmutj@gmail.com

J.-H. Wang
Tsinghua-Tencent Joint Laboratory for Internet Innovation
Technology, Beijing 100084, China

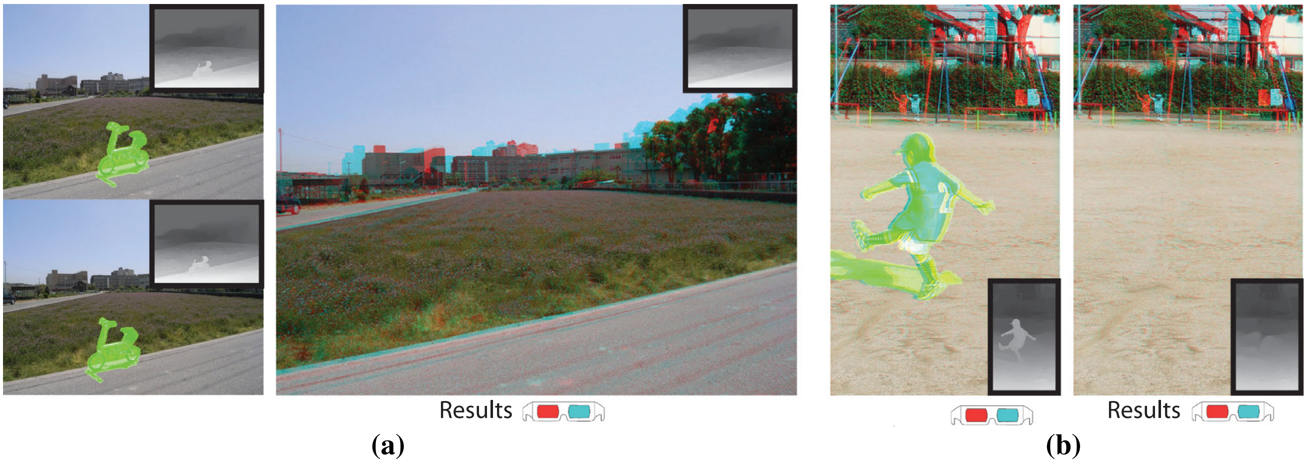


Fig. 1 We present a novel technique for stereoscopic image completion and depth recovery. Given a stereoscopic image pair with undesired regions (*green*) and its disparity maps, we complete the non-frontal-parallel road, plants in **(a)** and ground in **(b)** in a stereoscopically consistent manner, resulting in reasonable disparity maps (*inset*). The result anaglyphs can be better viewed with red-cyan glasses

local feature space (*pixel color* and *pixel location*). The estimated depth map, in turn, helps to find out better patch candidates. The designed gradient domain metric of patch distance also helps to make our method capable of handling both frontal-parallel and non-frontal-parallel scenes (demonstrated in Figs. 1, 3) to recover structures spanning a large depth range, relieving the assumption, adopted in [29,39], that the "holes" should be filled with content from further regions than the original ones. This assumption is not suitable for all cases and can even be wrong in some extreme cases.

Specifically, this paper makes the following contributions:

- a patch-based refine scheme which produces stereoscopically consistent results and can be used to handle the depth inconsistency problem, one that can bring in inconsistent completion in previous methods;
- a depth gradient domain patch distance metric, which is suitable for completing both frontal-parallel and non-frontal-parallel scenes;
- a disparity estimation method, which estimates disparity using depth shift in local feature space, facilitating simultaneously images and depth map completion.

This paper is organized as follows: Section 2 will review the related work; our method will be presented in detail in Sect. 3; then we will show our results in Sect. 4; finally, Sect. 5 concludes our work and discusses the future work.

2 Related work

Single Image Completion Single image completion concerns about two things. One is to find out the optimal content for each unknown (target) pixel/patch, and the other is to synthesize the content in a visually plausible manner. The filling content can be propagated or directly chosen from known (source) pixels/patches. In the former case, *diffusion-based* methods, like [3], usually involve solving partial differential equations (PDEs), propagating the source surroundings into the target regions along salient structures with the assumption of smoothness in color. However, these methods are less capable of handling large target areas where highly textured contents or semantic structures are desired. The latter case is often referred to as *exemplar-based* methods. In these methods, image completion problem is regarded as assigning each pixel/patch in the target regions a source one. The problem can be further formed as a *discrete Markov Random Fields (MRFs)* optimization [21,37] or just a patch searching problem. The MRFs can be optimized via belief propagation(BP) [21,37]. Though the *MRF-based* methods can achieve satisfying results via global optimization, they are indeed time consuming, usually quadratical in the number of image pixels. In patch searching, source patches are matched to the target regions to give the best candidates. Criminisi et al. [7] fill the blank left behind when the foreground objects are removed and the filling order is incorporated to prefer the one with high structure information and more available surroundings. Methods like [7], performed in greedy fashions, are easily fallen into local optima as well as sensitive to initialization and the optimization strategy.

Searching for the best match for some pixels/patches, typically known as the approximate nearest-neighbor fields (ANNF) problem, can be computationally expensive with the increase of search space and scale of problem. Barnes et al. [1] assume that there is a good chance that at least one good match can be found in a randomly assigned ANNF and iteratively refine the ANNF by propagating the good one to its neighbors. He et al. [12] make significant improvement in ANNF by combining candidate distributions and query dependency, which is further used to find out the main patch offsets to combine shifted images to complete the target regions [13].

With additional depth information available, He et al. [14] extended *PatchMatch* [1] to RGB-D(depth) to simultaneously inpainting color image and depth map.

Stereoscopic image/video editing With the rapid growth of consumption in 3D production, such as 3D movies, 3D cameras and so on, developing individual stereoscopic 3D media editing tools becomes more and more popular in the research group. To avoid visual discomfort [17,22] when watching stereoscopic content, disparities of presented content should be bounded in the comfort zone [17,22] as suggested in [5,31,33]. Stereo motion also affects the perceived level of visual comfort [11] and should be carefully manipulated [19].

As to specific editing tools, Lo et al. [27] and Luo et al. [28] both focus on stereoscopic image cloning, copying the stereo content of interest somewhere else and then compositing it into a new 3D scene. Tong et al. [38] make it different by copying intended content from 2D images. Niu et al. [30] extend 2D image warping to 3D. Du et al. [10] make it possible to switch among different perspectives for stereoscopic images using feature correspondences between the image pair. Lang et al. [23] exploit a nonlinear and locally adaptive fashion to remap the disparity range of stereoscopic images and videos. Lee et al. [25] also develop a nonlinear disparity mapping technique using disparity histogram. Yan et al. [41] propose a linear depth mapping method to adjust the depth range of stereo videos according to the viewing conditions. Didyk et al. [9] provide a near real-time solution to display stereoscopic videos on multi-view autosteresocopic screens. Zhang et al. [42] and Liao et al. [26] convert conventional videos into stereoscopic ones. Stereoscopic image retargeting is also addressed in [2,6,24].

Wang et al. [39] develop a stereoscopic inpainting system for simultaneous color and depth recovery. They complete the half occlusion regions using 3D warping of the counterpart in the other image and then greedily search for the optimal patch for unknown pixels using depth-assist texture synthesis [7]. Morse et al. [29] accomplish completing the target regions in a stereo pair by exploiting *PatchMatch* to search for candidate patches within both left and right images guided by pre-completed depth maps.

Hervieu et al. [15] present a two-step algorithm for inpainting stereo images by first diffusing disparities into holes and then synthesizing textures using a variation of [7]. Later, in [16], they propose to inpaint one of the stereoscopic images in a texture-by-number manner using depth layers and fill-in the corresponding pixels of the other using disparity map.

The methods proposed by Wang et al. and Morse et al. are most related to ours. The carefully designed depth gradient patch distance metric makes our method produce more consistent results and relieves the depth assumption, i.e., the "holes" should be filled with content from further regions than the original ones. Our method, thus, is more practical to general scenes including both frontal-parallel and non-frontal-parallel objects.

3 Stereoscopic image completion

This paper addresses the image completion problem for stereoscopic image pairs as shown in Fig. 1. In particular, given a source image pair, (I^l, I^r), which is assumed rectified, and its disparity maps, (D^l, D^r), users are asked to draw loosely corresponding contours, $(\partial\Omega^l, \partial\Omega^r)$, to identify the undesired regions, (Ω^l, Ω^r), as described in Sect. 3.1. The stereoscopic image completion is then formulated as an optimization to maximize the coherence and stereo consistency of the completion results in Sect. 3.2. The final completed color images $(\hat{I}^l, \hat{I}^r)$ and disparity maps $(\hat{D}^l, \hat{D}^r)$ are iteratively synthesized using a stereoscopic patch refine scheme and local disparity estimation (Sect. 3.4).

3.1 Preprocessing

Disparity estimation Estimating a full dense disparity map from a stereo image pair is still a challenge in computer vision. All disparity maps used in this paper are estimated using the method proposed by Smith et al. [35] for its good performance and sub-pixel disparities.

Region selection We simply ask users to specify loosely corresponding contours $(\partial\Omega^l, \partial\Omega^r)$ on both views to identify the undesired regions (Ω^l, Ω^r), though more sophisticated scheme can be used, such as the contour transfer proposed by Luo et al. [28].

3.2 Completion via optimization

In research community, single image completion is usually formed as a global optimization problem, seeking to minimize the following objective function which measures the coherence [1,34,40] between the target and source regions of the image:

$$d_{coh}(T, S) = \sum_{t \in T} \min_{s \in S} d(t, s), \tag{1}$$

where T is the missing region (target), S is the rest of the image (source), t and s are patches traversing target and source regions, respectively, and $d(t, s)$ is a distance measure between t and s.

When it comes to stereoscopic media, additional information is available, i.e., another view and depth about the scene. Thanks to the human visual system, objects can be perceived even if they appear just in one view, which implies that some missing content in one view can be inferred from the other. Also, depth information should be taken into account when measuring difference between patches. More importantly, to guarantee the stereo consistency within missing regions (Ω^l, Ω^r), the stereo correspondences of the filled content in the missing region should be kept in mind.

In particular, let $(\Psi^l, \Psi^r) \triangleq (I^l \setminus \Omega^l, I^r \setminus \Omega^r)$, denoting the source regions. Also let $M(\cdot)$ define a mapping for stereo correspondences between $\hat{I}^l$ and $\hat{I}^r$ with disparity maps $(\hat{D}^l, \hat{D}^r)$, i.e., $M(t)$ represents the corresponding *right(left)* view patch centered at $(x + \hat{D}^{l(r)}(x, y), y)$ when t is a *left(right)* view patch centered at (x, y). We then extend the objective function for single image completion in Eq. 1 to stereoscopic image pair to take both appearance coherence and stereoscopic consistency into consideration. Finally, we define the stereoscopic image completion as a minimization of the following objective function [29]:

$$SCC(\hat{I}^l, \hat{I}^r, \hat{D}^l, \hat{D}^r | \Psi^l, \Psi^r, \Omega^l, \Omega^r, D^l, D^r) = \sum_{t \in \Omega^l \cup \Omega^r} \min_{s \in \Psi^l \cup \Psi^r} \tilde{d}(t, s) + \lambda_{sc} \cdot \sum_{t \in \Omega^l \cup \Omega^r} \tilde{d}(t, M(t)), \quad (2)$$

where $\tilde{d}(\cdot, \cdot)$ is a measure of difference between patches concerning about both appearance and depth (see Eq. 3), and λ_{sc} controls the importance of stereoscopic consistency. Similar to Eq. 1, the first term tries to fill the target coherently with source information in either of the views; meanwhile, the second term ensures that the filled content in target regions looks stereoscopically consistent.

This optimization can be approximately achieved using an E–M (*expectation–maximization*) style approach, iteratively alternating between patch matching and value update, just like the completion presented in [29], and we will explain the details in Sect. 3.4.

3.3 The metric for patch distance

As we can see from Eq. 2, the distance measurement between patches plays a key role in the whole algorithm. Depth has been directly adopted in previous methods [14,29,39] to measure difference between patches, encouraging to choose patches from similar depth layer for the target regions. This performs well for frontal-parallel scenes, for the scenes can be simplified as different depth layers. In non-frontal-parallel scenes, depths at different pixels, even close ones, can change a lot. Therefore, we adapt the measurement to evaluate difference in depth gradients. In particular, let $\mathbf{G}(p)$ denote the depth gradient field of patch p, and then we define the distance metric between patches p_1 and p_2 in Eq. 2 as:

$$\tilde{d}(p_1, p_2) = (1 - \lambda_g) d_c(p_1, p_2) + \lambda_g \cdot d_g(\mathbf{G}(p_1), \mathbf{G}(p_2)), \quad (3)$$

where $d_c(\cdot, \cdot)$ is the sum of square difference in color space, $d_g(\cdot, \cdot)$ measures the difference in depth gradient field, and λ_g controls the weight. Further, $d_g(\cdot, \cdot)$ is defined as sum of *L-2* distance in gradient field, namely,

$$d_g(\mathbf{g_1}, \mathbf{g_2}) = \sum_{\mathbf{x}} \|\mathbf{g_1}(\mathbf{x}) - \mathbf{g_2}(\mathbf{x})\|, \quad (4)$$

where $\mathbf{g_1}(\mathbf{x})$ and $\mathbf{g_2}(\mathbf{x})$ denote the corresponding depth gradient vectors for each location $\mathbf{x}$ within the patches.

A significant difference between methods proposed in [15,29,39] and ours is that we do not exploit the assumption that the target should be filled with contents farther than the original. This assumption is unsuitable for non-frontal-parallel scenes and will produce inconsistencies when structured content is to be inpainted, such as a straight line crossing the missing region, where the missing depths of the straight line should be extended from both the available farther and nearer parts of the line. This assumption can even be wrong in some extreme cases, where the "holes" are the farthest parts of the image, for all the available depths are nearer than the missing depths. Our measurement evaluates the depth similarity in gradient domain and is free of comparing depth values, thus suitable for both frontal-parallel and non-frontal-parallel scenes.

Morse et al. [29] try to maintain the stereoscopic consistency through stereo propagation, namely, including offsets from stereo-corresponding patch and its neighbors in the candidate set. This method could handle scenes without depth inconsistency. However, when depth inconsistency occurs (e.g., occlusions) in the underlying disparity maps, stereoscopic inconsistency may exist between completed color images.

To better understand this point, we refer to Fig. 2 for illustration. In the top row, two views are presented with red dashed contours marking the target regions. The horizontal object (far) is occluded by the vertical object (near) in the right view. According to the pre-completed disparity map for the left view in [29], the stereo-corresponding point for target point A (yellow) is expected to be point $A^{'} = M(A)$ (green, in source region) in the right view. Then there is good chance that the patches around $A^{'}$ are chosen to be candidates for target point A, resulting in a stereoscopically inconsistent completion in the bottom row.

Looking into Fig. 2 further, we observe that depth inconsistency occurs at point A, i.e. A mismatches $A^{''} = M(A^{'})$ (green, let us name it as A's *stereo reflection*), which is the

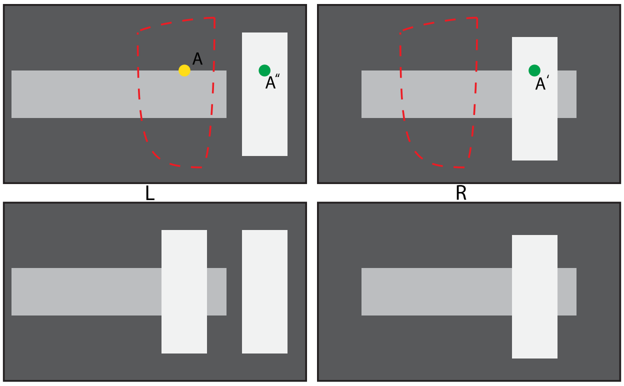


Fig. 2 Depth inconsistency occurs at target point A (*yellow*) for the Left view. Stereo propagation may result in inconsistent completion (*bottom row*) when the depth inconsistency exists in underlying disparity maps

stereo-corresponding point of $A^{'}$. This implies that depth inconsistent patch pair can differ a lot, especially when an occlusion occurs.

So, when a candidate source patch is selected, we should account for the similarity between the stereo-corresponding pair of patches, i.e., the severer the depth inconsistency is, the less similarity is required. In particular, let $DI(p)$ denote the depth inconsistency at location p, defined as the shift between p and its *stereo reflection*. Then the cost of choosing source patch s for target location t is calculated as:

$$F(s,t) = \mathcal{L}_{\lambda_{ic}}\Big(\mathcal{L}_{\lambda(s,t)}\big(\tilde{d}(s,t), \tilde{d}(t, M(t))\big), DI_{\epsilon}(t)\Big), \qquad (5)$$

where $\mathcal{L}_{\lambda}(A,B) \triangleq (1-\lambda)\cdot A + \lambda \cdot B$ is the linear combination between A and B with correlation λ, and $\lambda(s,t)$ controls the weight of similarity between the stereo-corresponding patch pair when source patch s is considered for target patch t. t is mapped to $M(t)$ based on the estimated disparity maps as described in Sect. 3.4.4. λ_{ic} is introduced to trade off the depth inconsistency. $DI_{\epsilon}(t)$ is a cut of $DI(t)$ designed to tolerate depth inconsistency to a certain degree with a threshold ϵ, i.e., $DI_{\epsilon}(t)$ is set to zero only when $DI(t) < \epsilon$, otherwise remains the same as $DI(t)$. According to previous analyses, $\lambda(s,t)$ decreases with the increase in depth inconsistency, so we take the following definition,

$$\lambda(s,t) = \lambda_m \cdot e^{-DI_{\epsilon}(t)}, \qquad (6)$$

where λ_m is the maximal weight for stereo consistency.

$F(s,t)$ is the final metric of patch distance when a source patch s is considered for a target patch t. The effectiveness of taking depth inconsistency into account is demonstrated in Fig. 6.

Algorithm 1 Pseudo-code for stereoscopic image completion

```
1: Pre-sort scan order;
2: Random initialization;
3: for each iteration do
4:     Calculate disparity gradient;
5:     for all $p \in \Omega^l \cup \Omega^r$ do
6:         Refine the optimized source patch;
7:         Update color and disparity values;
8:     end for
9: end for
```

3.4 Iterative color and depth synthesis

3.4.1 Overview of the algorithm

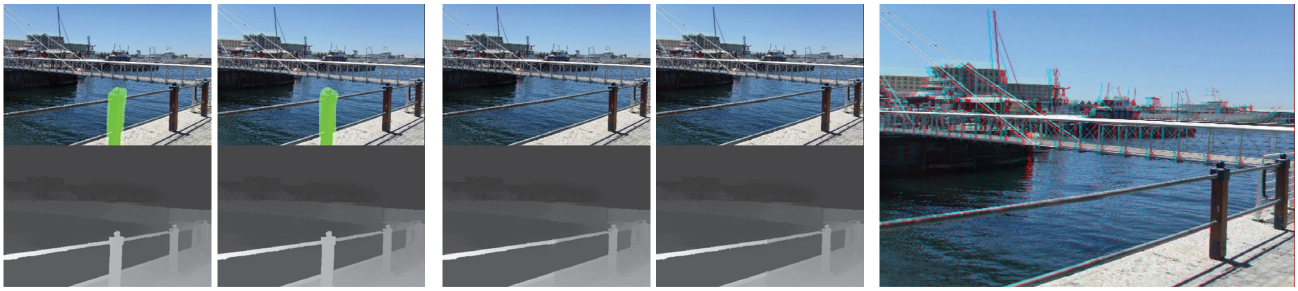
The inputs to the optimization in Eq. 2 include the source image pair (I^l, I^r), the corresponding disparity maps (D^l, D^r), and "holes" (Ω^l, Ω^r), marking the missing regions to be completed. The algorithm seeks an assignment, $(\hat{I}^l, \hat{I}^r)$ and $(\hat{D}^l, \hat{D}^r)$, for the pixels in "holes" to minimize the objective function in Eq. 2. As discussed in [40], an iterative algorithm can be applied to optimize the objective function when the following two local conditions are satisfied at each point p in $\Omega^l \cup \Omega^r$:

(i) All patches containing p appear exactly somewhere in $\Psi^l \cup \Psi^r$;
(ii) All patches containing p agree on the values at p.

So, the iterative algorithm should try to find patches meeting the above two conditions. Instead of searching the optimized candidate source patch for each unknown pixel in a greedy way as Wang et al. [39], an E-M style approach [29] with *PatchMatch* is adopted to satisfy the two conditions by propagating the optimized candidate from neighbors. We present our pseudo-code for stereoscopic image completion in *Algorithm* 1. Firstly, the filling order is prepared in *Step* 1 and *Step* 2 initializes the source candidate patch for each unknown pixel in $\Omega^l \cup \Omega^r$ with available patches in $\Psi^l \cup \Psi^r$; then in each iteration, disparity gradient is estimated to facilitate patch distance calculation; later, we refine the current best candidate patch in a stereoscopically consistent manner for unknown pixels with the scan order obtained in *Step* 1, followed by updating color images and disparity maps in *Step* 7. Compared to [29], our algorithm does not take pre-completed disparity maps as inputs; instead, the disparity maps and color images are completed simultaneously. We explain the details of the algorithm in the following sections.

3.4.2 Scan order and initialization

The order of the unknown pixels to be processed is responsible for completion results of high quality. We hope that



(a) Input images with mask(green) and disparity maps **(b)** Left and right completion results **(c)** Result anaglyph

Fig. 3 Completion for different depth layers in non-frontal-parallel scene. The seawater, embankment and guardrail, occupying different depth ranges, are stereoscopically completed with reasonable disparity maps

the patch with more already known information will be handled first. So we use the confidence term when computing patch priorities as in [7], differently extended to stereoscopic image pair, i.e., the order is sorted for unknown pixels in both views together, and thus the two views are processed together.

In standard *PatchMatch* algorithm, all patch offsets for the unknown pixels are randomly distributed. In stereo case, this random distribution can be improved in the following manner: if pixel $p(x, y) \in \Psi^l$ and $(x + D^l(x, y), y) \in \Omega^r$, then the offset for location $(x + D^l(x, y), y)$ in right view is set to $(-D^l(x, y), 0)$, with initial disparity $-D^l(x, y)$; Similarly, if pixel $p(x, y) \in \Psi^r$ and $(x + D^r(x, y), y) \in \Omega^l$, then the offset for location $(x + D^r(x, y), y)$ in left view is set to $(-D^r(x, y), 0)$, with initial disparity $-D^r(x, y)$; the offsets and disparities for remaining unknown pixels are still randomly initialized in $\Psi^l \cup \Psi^r$.

Similar to the mutual completion in [39], this improvement is based on the fact that some occluded parts in one view can be visible in the other one due to the viewpoint difference in the stereoscopic image pair. This is also the reason why we do not need a pair of accurately corresponding contours to indicate the missing regions. However, due to the greedy optimization strategy adopted in [39], the offset for a missing pixel is fixed ever since it is visited and will no longer change. This is obviously sensitive to the accuracy of underlying disparity maps. In our initialization, it only performs as a "good" guess and the imperfect ones will be refined later.

The effectiveness of taking scan order initialization into account is demonstrated in Figs. 5 and 6, respectively.

3.4.3 Stereoscopic patch refinement

Like *PatchMatch*, for each pixel $p \in \Omega^l \cup \Omega^r$, the offset of p is refined by finding the best matched one within a set of candidate patches, $\mathscr{C}(p)$, followed by a random search around the best one to jump out of local minima. The candidate set $\mathscr{C}(p)$ contains two parts:

- the one currently found in p and the ones propagated from its 4-connected neighbors with higher priorities or in source regions;
- the one found in p's stereo-corresponding point in the other view.

The first part is adopted in the original *PatchMatch* algorithm, and the second one is referred to as cross-image propagation, which has also been exploited in [4,29].

3.4.4 Color and depth update

The new color for a pixel p in the missing regions is computed as the average of color values at p for all patches containing p. The disparity value is updated in a similar way, except that the disparity values at p for all pathes containing p will be "shifted" before averaging, since the depth similarity is measured in gradient domain. This makes sure that the second condition in 3.4.1 is satisfied and thus guarantees the optimization.

The disparity maps used in [39] contain many flat blocks (constant values) due to the limitation of the adopted stereo matching algorithm [36]. Consequently, their segmentation-based plane fitting disparity estimation will also recover flat blocks in the final disparity maps, which makes it unsuitable to complete non-frontal-parallel scenes. The diffusion-based inpainting methods adopted to pre-complete the disparity maps in [15,29] try to propagate disparity values near the hole boundaries inside. This can produce apparent seams in the center of target regions when the underlying disparities seem to be changing gradually, which is common in non-frontal-parallel scenes.

When we choose a source patch as a candidate for a target location, the depth of the target location can be regarded as a shift to the source one. Based on this rule, we propose to estimate the mean value of the *depth shift* from source to target. We also notice that pixels of similar color and close position are likely to have the same depth shift. So, the depth shift is weighted in the local feature space, pixel color $\boldsymbol{c}$ and pixel location $\boldsymbol{x}$. Specifically, we consider a target location t with its neighbors $\mathcal{N}_t$ and its candidate source location s. In our implementation, $\mathcal{N}_t$ contains offsets to t referring to pixels with higher priorities than t or in source region. Finally, the mean *depth shift* from s to t is modeled as:

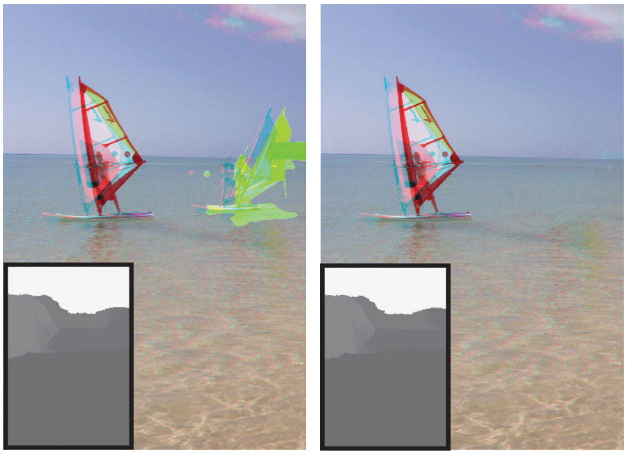
$$\overrightarrow{d}(s,t) = \sum_{v \in \mathcal{N}_t} w_{s,t}(v) \cdot (d_{s+v} - d_{t+v}), \tag{7}$$

where

$$w_{s,t}(v) = \frac{g_x(\|v\|/\sigma_x) g_c(\|\boldsymbol{c}_{t+v} - \boldsymbol{c}_s\|/\sigma_c)}{\sum_{u \in \mathcal{N}_t} g_x(\|u\|/\sigma_x) g_c(\|\boldsymbol{c}_{t+u} - \boldsymbol{c}_s\|/\sigma_c)}. \tag{8}$$

In the above equation, g_x and g_c are the Gaussian kernel functions for pixel location $\boldsymbol{x}$ and pixel color $\boldsymbol{c}$, respectively, and σ_x, σ_c are the corresponding bandwidths. Finally, the disparity for target location t is estimated as:

$$\hat{d}_t = d_s - \overrightarrow{d}(s,t). \tag{9}$$



(a) Input anaglyph and masks(green)

(b) Result anaglyph and disparity map

Fig. 4 Serious errors exist in the blue sky, sail boat (the *left one*), water (*whiter values* in disparity maps mean closer distance to the viewer). The water and sky are well completed after the sail boat (the *right one*) and the ball are removed

4 Implementations and results

Parameters In our experiments, all color images are handled in *CIE L*a*b** color space due to its close relationship to human vision. To facilitate calculation, both color images and disparity maps are normalized into [0, 1]. In our implementation, the size of patch is 15 × 15. Although the depth information can provide some structure cues about the scenes, compared to color information, it is not so supportive as expected when searching for the better patches, for its huge simplification about the scenes, e.g., all the background can be in the same depth layer. So, we take small values, 0.1 $\sim$ 0.2, for λ_g when calculating difference between patches in Eq. 3. The depth inconsistency toleration threshold ε is set to 3.0 for all examples in this paper. The maximal stereo-consistent weight λ_m in Eq. 6 is tuned to be 0.35. The depth inconsistency trade-off weight, λ_{ic}, is set to 0.2. The bandwidths, σ_x and σ_c, are set to 0.2 and 0.1 respectively to account for the influence from neighboring pixels when estimating disparities in Eq. 9.

Figures 1 and 3 exhibit our ability to complete non-frontal-parallel scenes when the foreground bicycle and boy are removed, respectively, noting those well-completed straight lines. In Fig. 4, our method can still achieve satisfying result though the underlying disparity maps suffer from serious errors.

Next, we compare our method with Wang et al.'s [39] and Morse et al.'s [29] (the depth maps are not available) using two frontal-parallel examples (Figs. 5, 6) both provided in their papers. The comparison shows that our patch refine method is comparable to Wang et al.'s when applied to frontal-parallel scenes and results in more stereoscopically consistent completion than Morse et al. do.

Figure 5 shows a task to complete a frontal-parallel background when the foreground bonsai (sub-figure (a), masked in green) is removed. As expected, our method can also faithfully recover the occluded windows in one view from the available windows in the other view, and the completion is superior to what have been done in [39] (only left view is published) and [29]. While [29] [sub-figure (d)] has difficulty in completing continuous roof edges that are occluded in both views for they employ a scan line order, our method [sub-figure (c)] can extend the roof in a reasonable way using the scan order sorted for the stereoscopic image pair together mentioned in Sect. 3.4.2 to assist the roof edges to be completed earlier.

Figure 6 demonstrates the depth-inconsistent problem discussed in Sect. 3.4.3. In depth-consistent areas, such as the occluded chimney in the top of the right view, missing objects can be well recovered using our method, the stereo propagation in [29] or the mutual completion in [39]. However, when completing depth-inconsistent missing regions, e.g., the air-condition occluded by the basketball stands in the left view, the stereo propagation in [29] prefers to the inconsistent source content, e.g., the fence, resulting in inconsistent completion. Our stereoscopic patch refine scheme can produce stereo-consistent content [sub-figure (b)] as well as avoid the depth-inconsistent problem. The imperfect completion of air-condition in sub-figure (d) also proves that the mutual completion in [39] relies heavily on accurate disparity maps for they adopt a greedy optimization strategy when searching patches, i.e., the candidate patch for a missing pixel will no

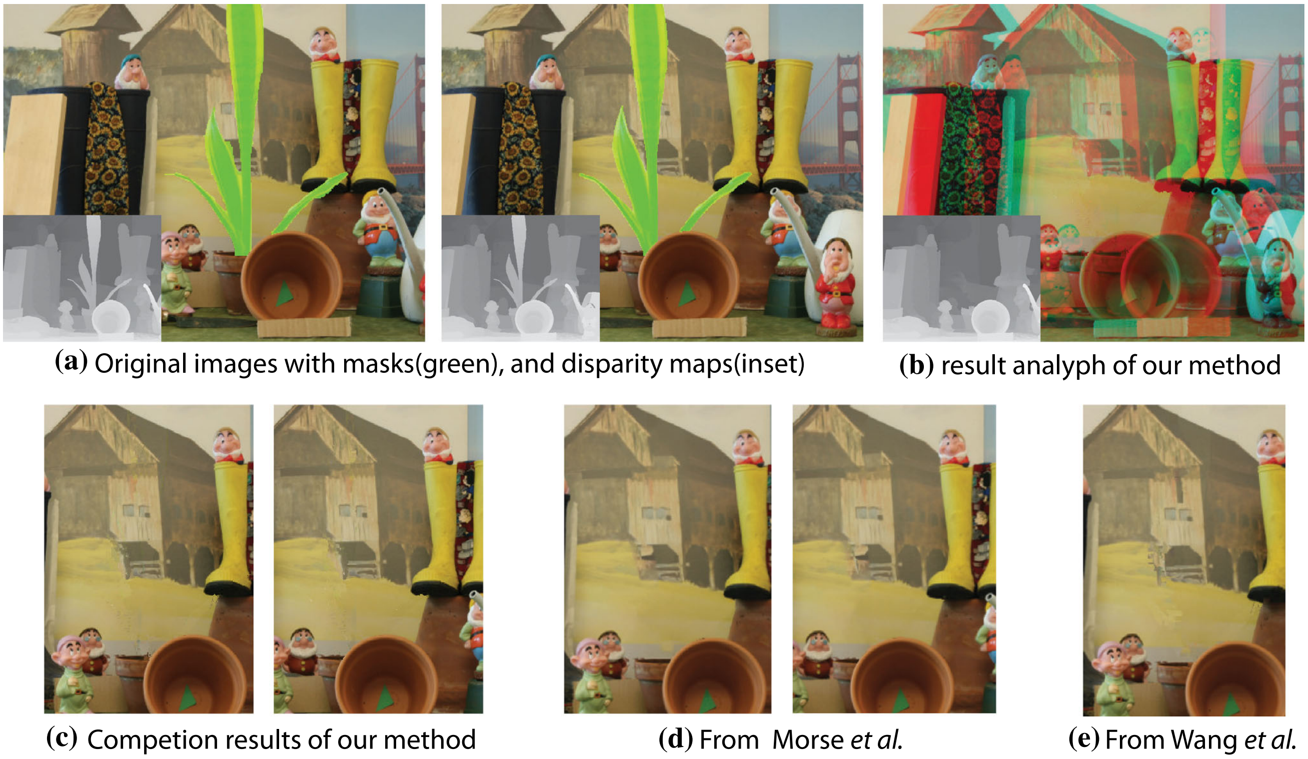

(a) Original images with masks(green), and disparity maps(inset)
(b) result analyph of our method
(c) Competion results of our method
(d) From Morse et al.
(e) From Wang et al.

Fig. 5 Completion for half-occluded background in frontal-parallel scenes. Half-occluded windows in original images (**a**) are faithfully recovered using our method (**b**, **c**) with underlying disparity maps(*inset*). Roof edges are better recovered using our method compared to the results (**d**) reported in [29]. **e** is the left view result in [39]

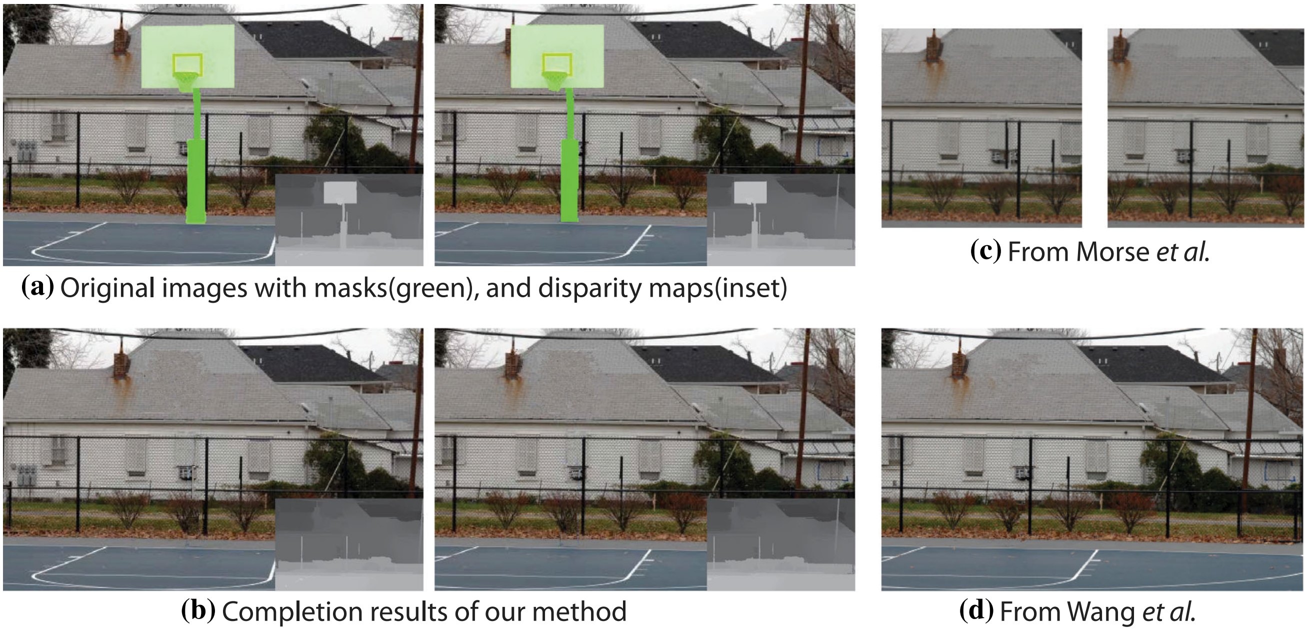
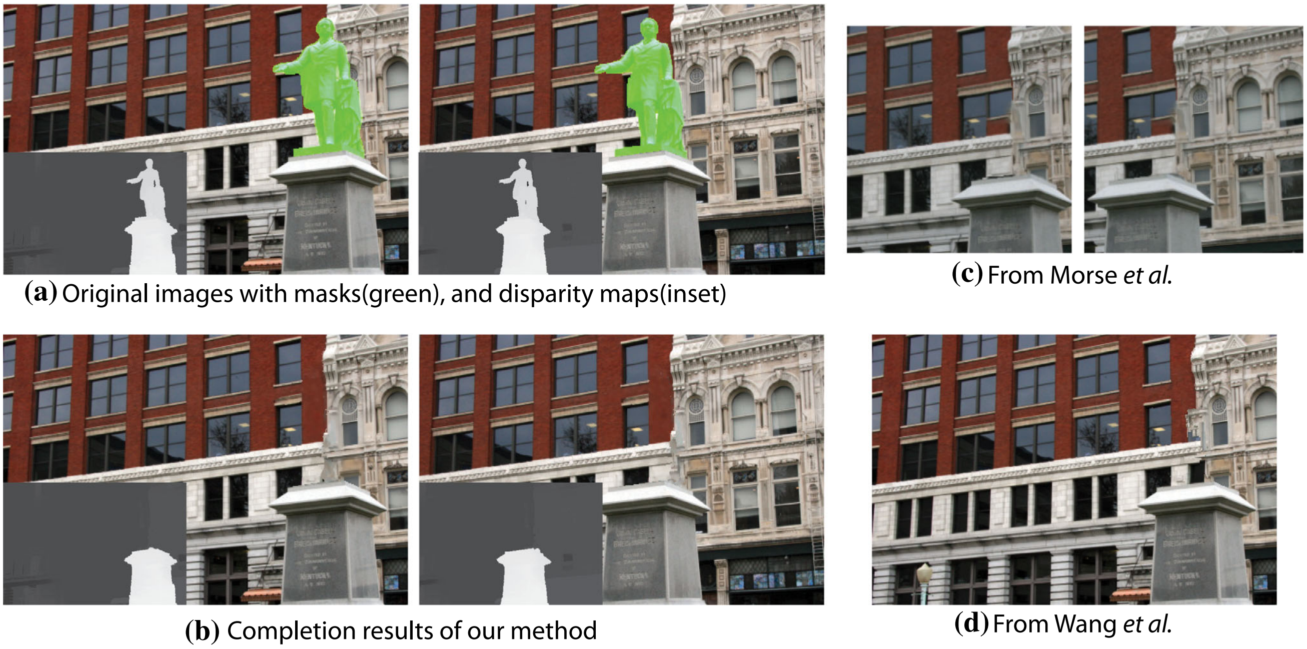

(a) Original images with masks(green), and disparity maps(inset)
(c) From Morse et al.
(b) Completion results of our method
(d) From Wang et al.

Fig. 6 Completion for depth inconsistent regions. Depth inconsistency appears at the occluded air-condition in the left view (**a**). Our stereoscopic patch refine scheme can produce stereo-consistent completion (**b**), while [29] completes the air-condition with inconsistent content, e.g., the fence, in source region in **c**. **d** is the right view result from [39]



(a) Original images with masks(green), and disparity maps(inset)

(b) Completion results of our method

(c) From Morse *et al.*

(d) From Wang *et al.*

Fig. 7 Comparison with [29] and [39]. Our method **b** produces more faithful and reasonable results than both **c** [29] and **d** [39]. Note the windows and the walls

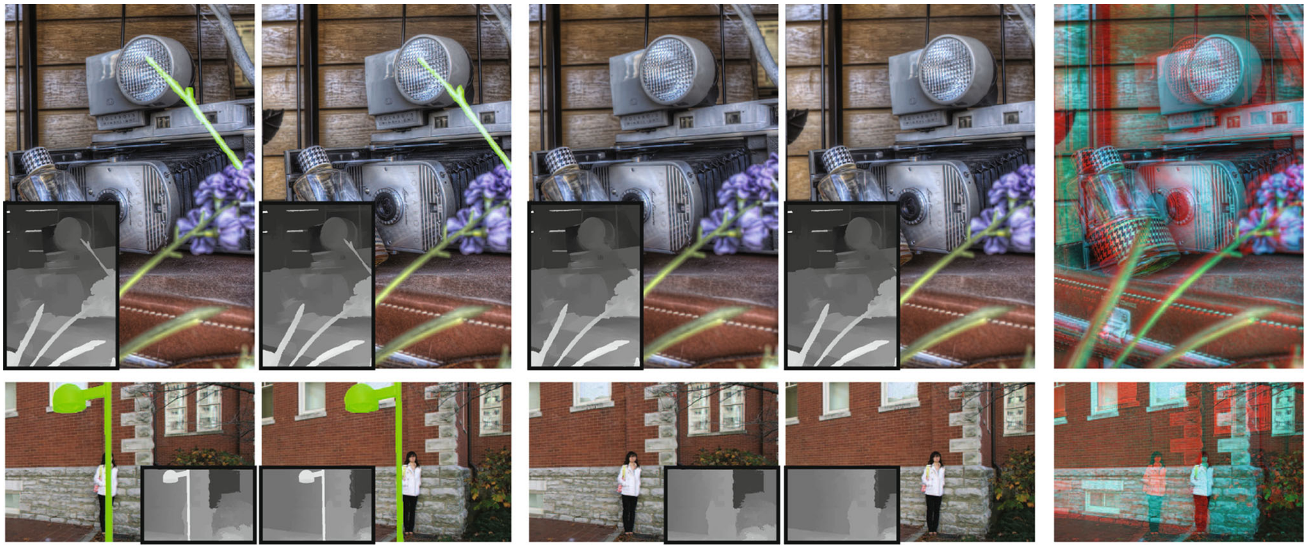


Fig. 8 More results. From left to right: original left and right images with masks (*green*) and disparity maps (*inset*), completion results of color images and disparity maps (*inset*) for the both views, and the result anaglyphs (🟥🟦)

longer change once the pixel is visited. Our method, on the other hand, will iteratively refine the candidate patch.

Figure 7 shows another comparison with [29] and [39]. Note the windows and the walls. The patch searching strategy adopted by our method produces more faithful and reasonable results than [29] (scan line searching) and [39] (greedy searching). More results are presented in Fig. 8.

Limitations Currently, our method is less capable of supporting interactive editing. The time complexity is linear with the production of iterative times, the size of "holes" and the size of patch. The method proposed by He et al. [12] can be used to speed up the patch matching process. The parameters are selected empirically and should be better justified in future.

5 Conclusions and future work

In this paper, we have presented a method that fills the missing "holes" in both color images and disparity maps simultaneously when undesired regions are removed in a stereoscopic image pair. Stereo-consistent completion results are ensured through a novel patch matching scheme which takes the depth inconsistency, one that can cause inconsistent completion in previous methods, into account. The carefully designed patch distance metric makes our method capable of handling non-frontal-parallel scenes, as well as frontal-parallel ones, and free of employing the unnecessary depth constraint, i.e., missing regions should be filled with contents further than the original. Experimental results show that the proposed algorithm is capable of producing more stereoscopically consistent content or more reasonable completion.

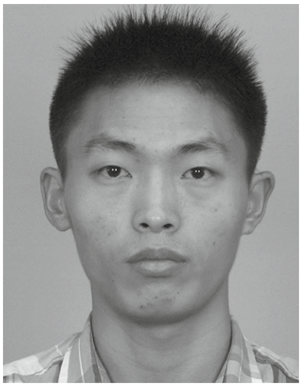
As future works, an intuitive extension is to investigate the problem of stereoscopic video completion [32]. In stereoscopic video completion, the number of available source patches can be repetitive as well as huge. To reduce the searching time, using the stereo correspondences between views to study the distribution of candidate patches [12] will be critical. Furthermore, we could also extend some interesting applications of visual media [18], such as interactive editing and composition, to stereoscopic cases.

Acknowledgments We would like to thank the anonymous reviewers for their helpful comments. This work was supported by the National Basic Research Project of China (Project Number 2011CB302205), the Natural Science Foundation of China (Project Number 61272226/ 61120106007), the National High Technology Research and Development Program of China (Project Number 2013AA013903) and Tsinghua University Initiative Scientific Research Program.

References

1. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. ACM Trans. Graph. **28**(3), 24:1–24:11 (2009)
2. Basha, T., Moses, Y., Avidan, S.: Geometrically consistent stereo seam carving. In: ICCV, pp. 1816–1823 (2011).
3. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: SIGGRAPH, pp. 417–424 (2000).
4. Bleyer, M., Rhemann, C., Rother, C.: Patchmatch stereo - stereo matching with slanted support windows. In: BMVC, pp. 1–11 (2011).
5. Celikcan, U., Cimen, G., Kevinc, E.B., Capin, T.: Attention-aware disparity control in interactive environments. Vis. Comput. **29**(6–8), 685–694 (2013)
6. Chang, C.H., Liang, C.K., Chuang, Y.Y.: Content-aware display adaptation and interactive editing for stereoscopic images. IEEE Trans. Multimedia **13**(4), 589–601 (2011)
7. Criminisi, A., Pérez, P., Toyama, K.: Object removal by exemplar-based inpainting. In: IEEE CVPR, pp. 721–728 (2003).
8. Dahan, M.J., Chen, N., Shamir, A., Cohen-Or, D.: Combining color and depth for enhanced image segmentation and retargeting. Vis. Comput. **28**(12), 1181–1193 (2012)
9. Didyk, P., Sitthi-Amorn, P., Freeman, W.T., Durand, F., Matusik, W.: Joint view expansion and filtering for automultiscopic 3d displays. ACM Trans. Graph. **32**(6), 221:1–221:8 (2013)
10. Du, S.P., Hu, S.M., Martin, R.R.: Changing perspective in stereoscopic images. IEEE Trans. Vis. Comput. Graph. **19**(8), 1288–1297 (2013)
11. Du, S.P., Masia, B., Hu, S.M., Gutierrez, D.: A metric of visual comfort for stereoscopic motion. ACM Trans. Graph. **32**(6), 222:1–222:9 (2013)
12. He, K., Sun, J.: Computing nearest-neighbor fields via propagation-assisted kd-trees. In: IEEE CVPR, pp. 111–118 (2012).
13. He, K., Sun, J.: Statistics of patch offsets for image completion. In: ECCV, pp. 16–29 (2012).
14. He, L., Bleyer, M., Gelautz, M.: Object removal by depth-guided inpainting. ÖAGM / AAPR Workshop **2011**, 1–8 (2011)
15. Hervieu, A., Papadakis, N., Bugeau, A., Gargallo, P., Caselles, V.: Stereoscopic image inpainting: distinct depth maps and images inpainting. In: ICPR, pp. 4101–4104 (2010).
16. Hervieux, A., Papadakis, N., Bugeau, A., Gargallo, P., Caselles, V.: Stereoscopic image inpainting using scene geometry. In: ICME, pp. 1–6 (2011).
17. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. J. Vis. **8**(3), 33:1–33:30 (2008)
18. Hu, S.M., Chen, T., Xu, K., Cheng, M.M., Martin, R.R.: Internet visual media processing: a survey with graphics and vision applications. Vis. Comput. **29**(5), 393–405 (2013)
19. Kellnhofer, P., Ritschel, T., Myszkowski, K., Seidel, H.P.: Optimizing disparity for motion in depth. Comput. Graph. Forum **32**(4), 143–152 (2013)
20. Kim, Y., Winnemoller, H., Lee, S.: Wysiwyg stereo painting with usability enhancements. IEEE Trans. Vis. Comput. Graph. PrePrint(99), 1–1 (2014).
21. Komodakis, N., Tziritas, G.: Image completion using efficient belief propagation via priority scheduling and dynamic pruning. Trans. Image Proc. **16**(11), 2649–2661 (2007)
22. Lambooij, M.T.M., IJsselsteijn, W.A., Heynderickx, I.: Visual discomfort and visual fatigue of stereoscopic displays: a review. J. Imaging Sci. Technol. 53(3), 030,201–030,201–14 (2009).
23. Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., Gross, M.: Nonlinear disparity mapping for stereoscopic 3d. ACM Trans. Graph. **29**(4), 75:1–75:10 (2010)
24. Lee, K.Y., Chung, C.D., Chuang, Y.Y.: Scene warping: layer-based stereoscopic image resizing. In: IEEE CVPR, pp. 49–56 (2012).
25. Lee, S., Kim, Y., Lee, J., Kim, K., Lee, K., Noh, J.: Depth manipulation using disparity histogram analysis for stereoscopic 3d. Vis. Comput. **30**(4), 455–465 (2014)
26. Liao, M., Gao, J., Yang, R., Gong, M.: Video stereolization: combining motion analysis with user interaction. IEEE Trans. Vis. Comput. Graph. **18**(7), 1079–1088 (2012)
27. Lo, W.Y., van Baar, J., Knaus, C., Zwicker, M., Gross, M.: Stereoscopic 3d copy and paste. ACM Trans. Graph. **29**(6), 147:1–147:10 (2010)
28. Luo, S.J., Shen, I.C., Chen, B.Y., Cheng, W.H., Chuang, Y.Y.: Perspective-aware warping for seamless stereoscopic image cloning. ACM Trans. Graph. **31**(6), 182:1–182:8 (2012)
29. Morse, B., Howard, J., Cohen, S., Price, B.: Patchmatch-based content completion of stereo image pairs. In: 3DIMPVT, pp. 555–562 (2012).
30. Niu, Y., Feng, W.C., Liu, F.: Enabling warping on stereoscopic images. ACM Trans. Graph. **31**(6), 183:1–183:7 (2012)
31. Pollock, B., Burton, M., Kelly, J., Gilbert, S., Winer, E.: The right view from the wrong location: depth perception in stereoscopic multi-user virtual environments. IEEE Trans. Vis. Comput. Graph. **18**(4), 581–588 (2012)
32. Raimbault, F., Kokaram, A.: Stereo-video inpainting. J. Electron. Imaging **21**(1), 1–10 (2012)

33. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: predicting visual discomfort with stereo displays. J. Vis. **11**(8), 11:1–11:29 (2011)
34. Simakov, D., Caspi, Y., Shechtman, E., Irani, M.: Summarizing visual data using bidirectional similarity. In: IEEE CVPR, pp. 1–8 (2008).
35. Smith, B., Zhang, L., Jin, H.: Stereo matching with nonparametric smoothness priors in feature space. In: IEEE CVPR, pp. 485–492 (2009).
36. Sun, J., Li, Y., Kang, S.B., Shum, H.Y.: Symmetric stereo matching for occlusion handling. In: IEEE CVPR, pp. 399–406 (2005).
37. Sun, J., Yuan, L., Jia, J., Shum, H.Y.: Image completion with structure propagation. ACM Trans. Graph. **24**(3), 861–868 (2005)
38. Tong, R.F., Zhang, Y., Cheng, K.L.: Stereopasting: interactive composition in stereoscopic images. IEEE Trans. Vis. Comput. Graph. **19**(8), 1375–1385 (2013)
39. Wang, L., Jin, H., Yang, R., Gong, M.: Stereoscopic inpainting: joint color and depth completion from stereo images. In: IEEE CVPR, pp. 1–8 (2008).
40. Wexler, Y., Shechtman, E., Irani, M.: Space-time completion of video. IEEE Trans. Pattern Anal. Mach. Intell. **29**(3), 463–476 (2007)
41. Yan, T., Lau, R., Xu, Y., Huang, L.: Depth mapping for stereoscopic videos. Int. J. Comput. Vis. **102**(1–3), 293–307 (2013)
42. Zhang, G., Hua, W., Qin, X., Wong, T.T., Bao, H.: Stereoscopic video synthesis from a monocular video. IEEE Trans. Vis. Comput. Graph. **13**(4), 686–696 (2007)



Tai-Jiang Mu received his BS degree in computer science from Tsinghua University in 2011. He is currently a Ph.D. candidate in the Department of Computer Science and Technology, Tsinghua University, Beijing. His research interests include computer graphics, stereoscopic image/video processing and stereoscopic perception.

Ju-Hong Wang is deputy chair of technical committee of Tencent Technology Company Limited. She received a master degree from Beijing University of Post and Telcommunication. Her research area is computer graphics and multimedia.

Song-Pei Du received his BS degree in computer science from Tsinghua University in 2009. He is currently working toward the Ph.D. degree in the Department of Computer Science and Technology, Tsinghua University, Beijing. His research interests include computer graphics, geometric modeling, texture synthesis, image processing and stereoscopy.

Shi-Min Hu is currently a professor in the Department of Computer Science and Technology at Tsinghua University, Beijing. He received the Ph.D. degree from Zhejiang University in 1996. His research interests include digital geometry processing, video processing, rendering, computer animation, and computer aided geometric design. He is associate Editor-in-Chief of The Visual Computer, and on the editorial boards of IEEE TVCG, Computer-Aided Design and Computer and Graphics.