Structure-Guided Image Inpainting Using Homography Transformation

Jiaying Liu, *Senior Member, IEEE*, Shuai Yang, Yuming Fang, *Senior Member, IEEE*, and Zongming Guo, *Member, IEEE*

***Abstract*—In this paper, we present a novel structure-guided framework for exemplar-based image inpainting to maintain the neighborhood consistence and structure coherence of an inpainted region. The proposed method consists of a data term for pixel validity and boundary continuity, a smoothness term to depict the compatibility of neighboring pixels for contextual continuity, and a coherence term to investigate image inherent regularities to ensure image self-similarity. To better reconstruct image structures, the method utilizes image regularity statistics to extract dominant linear structures of the target image. Guided by these structures, homography transformations are estimated and combined to globally repair the missing region using the Markov random field model. To reduce computational complexity, a hierarchical process is implemented to utilize the regularity effectively. The experimental results demonstrate that our method yields better results for various real-world scenes than existing state-of-the-art image inpainting techniques.**

***Index Terms*—Image inpainting, image self-similarity, homography transformation, linear structure, image completion.**

I. Introduction

IMAGE inpainting or image completion, an important topic in image processing, is carried out to reconstruct the missing parts of an image. In recent years, it has attracted much attention in the research community because of its intensive popularity in digital life. Image inpainting can be widely used in image editing applications, such as panorama generation, cultural heritage restoration, restoring images from scratches or text overlays, and loss concealment in impaired image transmission [1]. Generally, image inpainting can be classified into two categories: diffusion-based approaches and exemplar-based approaches [1].

Diffusion-based methods smoothly propagate information from known boundaries to missing regions. In the pioneering work by Bertalmio *et al.* [2], the authors made use of geometric and photometric information and propagated Laplacian descriptors along the isophote direction. Following that study [2], several improved mathematical models, including total variation (TV) [3], curvature driven diffusion (CDD) [4] and Mumford-Shah [5], were proposed. This class of techniques yields good results when inpainting long thin regions but is less effective in handling large holes, as it fails to consider global image structures and synthesize textures for image details.

To address the drawback of diffusion-based image inpainting techniques, exemplar-based methods sample pixels or patches in the known regions and fill the missing regions with textures synthesized from these samples, which effectively preserves image details. Therefore, in exemplar-based methods, many efforts are made in structure preservation to obtain promising inpainting results. Many structure preservation algorithms have been proposed, including isophote-based filling priority [6], searching along structure curves [7], [8], directional image inpainting [9] and guidance by statistics of patch offsets [10]. According to the inpainting strategies, exemplar-based methods can be classified as greedy methods [6], [9], [11] and global methods [12]–[15].

Greedy methods fill one target pixel/patch at a time by searching for the best matches as samples and iteratively complete the missing regions. Structure-based priority [6] was put forward to preserve structure continuity. It is realized by preferentially reconstructing the patches where the isophote direction and the filling direction are consistent. Based on filling priority, various studies have attempted to improve the greedy methods with respect to the priority definition [16], [17], matched patch searching [18], [19] and texture synthesis [20]. Compared with simply copying and pasting, texture synthesis methods using weighted average [19], [21], sparse representation [22], [23] and energy function optimization [24] have been proposed. Human-interactive methods have also been proposed to provide user-specified structure lines [7] for structure preservation. Nevertheless, greedy methods might run into a local optimum and introduce inconsistency when filling large missing regions with complicated structures.

Compared with greedy methods, global methods regard image inpainting as an optimization problem to assign the best candidate sample to each unknown pixel/patch. Recent studies focused on two kinds of global inpainting tools: a coherence energy function optimized using EM-like schemes [25], [26] and a Markov random field (MRF) energy function optimized using belief propagation [12], [24] or graph cuts [13], [27].

Manuscript received January 23, 2016; revised June 15, 2016 and July 26, 2017; accepted April 13, 2018. Date of publication April 30, 2018; date of current version November 15, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61772043 and in part by the CCF-Tencent Open Research Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Fatih Porikli.

The authors are with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China (e-mail: liujiaying@pku.edu.cn; williamyang@pku.edu.cn; FA0001NG@e.ntu.edu.sg; guozongming@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2018.2831636

Global methods based on the coherence energy function take image self-similarity into account and fill the unknown region such that the inpainted region shares the most similar patches with the known region. This kind of method gains its practicability using the fast patch search method of PatchMatch [14] and is employed in Adobe Photoshop Content Aware Fill. Based on PatchMatch, patch transformations such as rotation, scaling [28], [29], reflecting [30] and perspective transform [31] have been put forward to make the method robust with respect to complicated scenes.

Compared with coherence-based methods, global methods based on the MRF energy function naturally pose inpainting problems as labeling problems to assign each unknown pixel/patch a valid value (called a label). In the MRF model, overlapped patches or neighboring pixels are defined as adjacent nodes, and the relationship of each adjacent node pair is evaluated to ensure contextual continuity. In [12], [13], with the whole image available for sampling, the number of candidate labels could be considerably large, which limits MRF optimization efficiency. Priority belief propagation [12] and graph cuts [13] have been utilized to efficiently solve the optimization problem. To constrain the search space, Ruzic *et al.* [32] divided an image into several regions according to the context and searched candidate samples in similar regions. In addition, He and Sun [10] limited the search space for each unknown pixel to only 60 candidates using the statistics of the patch offsets, obtaining gains in both algorithm speed and inpainting quality. In addition, Liu and Caselles [15] utilized a hierarchical scheme to reduce computational complexity.

In this paper, we propose a novel model for exemplar-based single image inpainting. The main idea is to exploit the regularity of image self-similarity to guide and improve the inpainting process. The model is formulated using a data term, a smoothness term and a coherence term to evaluate the pixel validity, contextual continuity and image self-similarity, respectively. Moreover, we design an efficient EM-like optimization approach to solve the model. In the E step, guided by the linear structures of the target image, we heuristically estimate several homography transformations based on the repetitive regularity of the known region to form the search space. These homography transformations shift textures and structures into the unknown region. In the M step, the hole is filled by assigning each unknown pixel a transformation under the MRF energy constraint. Meanwhile, a hierarchical scheme is devised to obtain good structure preservation and low computational complexity.

Here, we intend to clarify the differences between the proposed method and other existing related methods. Although, according to our previous review, the optimization of the energy function is a common and effective approach to examplar-based image inpainting, there are some limitations that affect the inpainting qualities in the existing methods. In the aforementioned MRF-based methods [10], [12], [13], the basic operation is pixel/patch translation, which might yield broken structures in non-fronto-parallel scenes. Instead of pixel/patch translations, general transformations are used as the search space in the proposed method to enable non-fronto-parallel scene inpainting; thus, there is no broken structure in the inpainting results. Meanwhile, although some coherence-based methods [28], [29], [31] consider patch transformations, they suffer from structure distortions because these methods break images into overlapped patches and fail to search for and transform matched patches uniformly. On the contrary, the proposed method uses global transformations to uniformly shift valid information into the hole, preserving image structures well with less distortion. Another limitation of coherence-based methods is the blurring artifacts due to the patch compositing procedure. The proposed method has no such procedure and is free of this drawback; thus, it can preserve texture details. Hence, the proposed method is superior to the aforementioned MRF-based methods for non-fronto-parallel scene inpainting as well as to the aforementioned coherence-based methods because of its better structure and texture preservation.

In summary, the main contributions of our work include the following two aspects:

- In the proposed method, the exemplar-based image inpainting problem is formulated as a novel energy optimization problem for structure-guided transformation estimation and assignment. It contains a data term, a smoothness term and a coherence term, which take data validity, contextual continuity and self-similarity, respectively, into account.
- We propose an EM-like approach based on homography transformations to solve the proposed optimization problem. A heuristic transformation estimation strategy for search space establishment and a transformation assignment strategy for hierarchical exemplar-based image inpainting are presented for robust scene inpainting and better structure preservation.

The rest of this paper is organized as follows. In Section II, the proposed exemplar-based inpainting model is presented. The E step and M step of the optimization approach are described in Sections III and IV, respectively. More specifically, Section III introduces the homography transformation candidate estimation, and Section IV explains the hierarchical exemplar-based image inpainting. We validate our method by comparing it with state-of-the-art image inpainting algorithms using both artificial scenes and natural scenes in Section V. Concluding remarks are given in Section VI.

II. Proposed Image Inpainting Framework

In this section, we describe the main idea of the proposed model and the corresponding inpainting algorithm. First, we provide some notations of important concepts. Then, the general framework of the MRF for image processing is given. Finally, we introduce the details of the proposed image inpainting algorithm.

A. Notation

The following notations are used throughout this paper:

- I is the 2D target image. $p(x, y)$ is a pixel with (x, y) as its spatial coordinates, and its color is denoted as $I(p)$.
- Ω is the missing part of I, and its contour is indicated by $\delta\Omega$. $\Phi = I - \Omega$ is the source region.
- $\Psi(p)$ is the patch centered at pixel p.

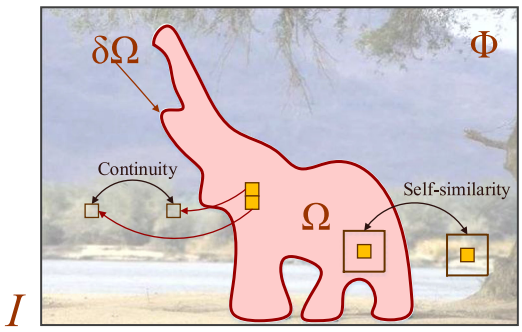


Fig. 1. Image priors for image inpainting. The inpainting problem involves filling Ω seamlessly using the information of Φ. The contextual continuity prior is measured based on the smoothness between adjacent pixels. The self-similarity prior is measured based on the patch difference.

- $\mathbf{H}$ denotes a homography transformation matrix. It projects pixel p onto $p' = \mathbf{H}(p)$, where $\mathbf{H}(\cdot)$ is the transform operation using $\mathbf{H}$.
- $S_x = \{x_i\}, i = 1, 2, \ldots, N$ represents a set of N elements x. The element x could be a pixel, line, or matrix.

B. The Proposed Inpainting Model

Formally, we define the inpainting problem as a labeling problem: Given the target image I, the goal is to fill Ω seamlessly using the information of Φ. Let us define a set of homography transformations $S_{\mathbf{H}} = \{\mathbf{H}_i\}, i = 1, \ldots, N_{\mathbf{H}}$ and a labeling function $L : \Omega \rightarrow S_{\mathbf{H}}$. The main idea is to assign transformation matrices (labels) to unknown pixels for inpainting. If $L(p) = \mathbf{H}_i$, then the unknown pixel $p \in \Omega$ is inpainted with the value of the known pixel $\mathbf{H}_i(p) \in \Phi$. To design an effective inpainting model to evaluate $S_{\mathbf{H}}$ and L, intrinsic image properties (as shown in Fig. 1) are utilized as priors:

1) The **contextual continuity prior** is a basic prior in image processing. In our method, we realize the global contextual continuity using a smoothness term E_s by measuring the local adjacent pixel smoothness. In addition, a data term E_d is designed to evaluate patch similarities along inpainting boundaries, which further emphasizes the boundary structural continuity.
2) The **self-similarity prior** is ubiquitous in both natural and artificial scenes. Essentially, textures are repeating two-dimensional patterns, and structures are repeating one-dimensional patterns [6]. The self-similarity property assumes that textures and structures are repetitive, and the patches in an image are considered to recur within the image. Therefore, in this paper, we use a coherence term E_c, defined as the sum of differences between matched patches, to evaluate this property.

By taking the above priors into consideration, we propose a novel inpainting model, which is formulated as follows:

$$E(L, S_{\mathbf{H}}) = \alpha E_d(L, S_{\mathbf{H}}) + E_s(L, S_{\mathbf{H}}) + E_c(L, S_{\mathbf{H}}), \quad (1)$$

where

$$E_d(L, S_{\mathbf{H}}) = \sum_{p \in \Omega} e_d(L(p), S_{\mathbf{H}}),$$

$$E_s(L, S_{\mathbf{H}}) = \sum_{(p,q) \in \mathcal{N}} e_s(L(p), L(q), S_{\mathbf{H}}),$$

$$E_c(L, S_{\mathbf{H}}) = \sum_{p \in I} e_c(L(\Psi(p)), S_{\mathbf{H}}).$$

E_d is the data term that considers the property of individual pixels, E_s is the smoothness term characterizing mutual influences among neighboring pixels, and α is the weight used to combine the data term and the smoothness term. By assuming the properties of positivity and Markovianity [33], we can define E_d and E_s using the well-studied MRF model. The mathematical formulation of these two terms will be exploited to realize the contextual continuity in Section IV. Meanwhile, E_c is the coherence term used to measure image self-similarity. It is defined as the sum of differences e_c between image patches $\Psi(p)$ in the inpainted image and their matched patches in the source region:

$$e_c(L(\Psi(p)), S_{\mathbf{H}}) = \min_{q \in \Phi, \mathbf{H} \in S_{\mathbf{H}}} \|\Psi_L(p) - \mathbf{H}(\Psi(q))\|_2^2, \quad (2)$$

where $\Psi_L(p)$ is the image patch in the inpainted image and $\mathbf{H}$ is utilized to make the metric robust to scaling, rotation and perspective transformations. For simplicity, e_c is set to 0 for patches that have unknown pixels, have pixels assigned with different labels, or go across the inpainting boundary.

C. Algorithm Overview

Given the model, we need to find the optimal $S_{\mathbf{H}}$ and L to minimize (1). Here, we propose an EM-like optimization approach. The main idea is to update one variable iteratively while keeping another variable fixed. In the E step, we set L as the fixed variable, and the optimization problem becomes estimating the transformation matrices that best depict the regularity statistics of the target image. In the M step, we set the estimated $S_{\mathbf{H}}$ as the fixed variable, and the problem can be reformulated as a standard MRF label assignment problem.

Transformation estimation step (E step): We attempt to estimate $S_{\mathbf{H}}$ while setting L as the fixed variable. Since $S_{\mathbf{H}}$ is the range of the labeling function L in the data term and smoothness term, with L fixed, $S_{\mathbf{H}}$ in these two terms should be fixed as well. Thus, in the i-th iteration, the optimal $S_{\mathbf{H}}^i$ satisfies:

$$\begin{aligned} S_{\mathbf{H}}^i &= \arg\min \alpha E_d(L^{i-1}, S_{\mathbf{H}}^{i-1}) + E_s(L^{i-1}, S_{\mathbf{H}}^{i-1}) \\ &\quad + E_c(L^{i-1}, S_{\mathbf{H}}^i), \\ &= \arg\min E_c(L^{i-1}, S_{\mathbf{H}}^i). \end{aligned} \quad (3)$$

Since the coherence of the patches in the unknown region is not calculated, we only focus on the known region:

$$S_{\mathbf{H}}^i = \arg\min \sum_{p \in \Phi} \min_{q \in \Phi, \mathbf{H} \in S_{\mathbf{H}}} \|\Psi_L(p) - \mathbf{H}(\Psi(q))\|_2^2. \quad (4)$$

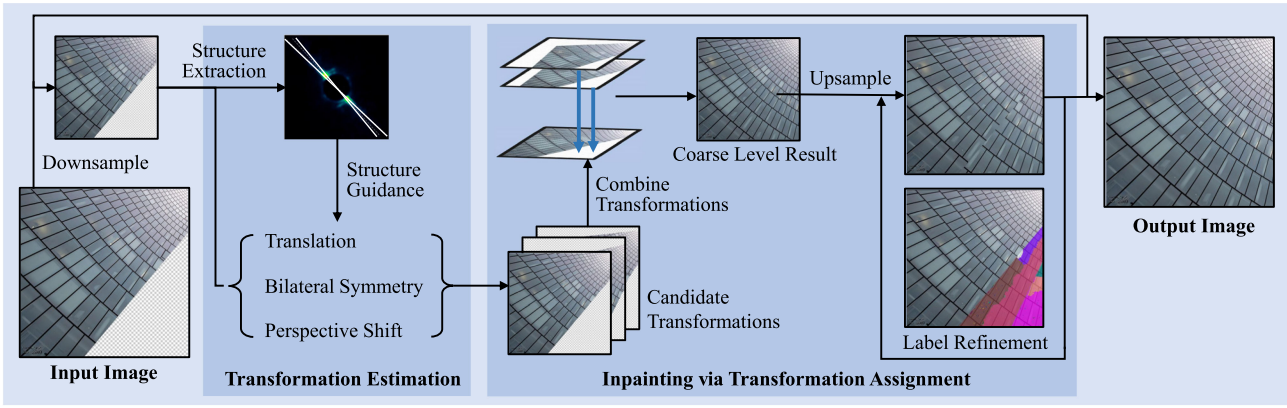


Fig. 2. Flow chart of the proposed structure-guided image inpainting approach using homography transformation.

We heuristically solve the above equation by seeking the regularities of the repetitive patterns in Φ and estimating the transformations that best agree with these regularities, which will be described in detail in Section III.

Transformation assignment step (M step): The coherence of the patches in Ω is zero, as these patches are filled by shifting the values from Φ to Ω using exact transformations in $S_{\mathbf{H}}$. Therefore, if $S^i_{\mathbf{H}}$ is fixed, for any L, $E_c = \sum_{p\in\Phi} \min \|\Psi(p) - \mathbf{H}(\Psi(q))\|_2^2$ is constant. Then, (1) can be rewritten as a standard MRF function:

$$\begin{aligned} L^i &= \arg\min \alpha E_d(L^i, S^i_{\mathbf{H}}) + E_s(L^i, S^i_{\mathbf{H}}) + E_c(L^i, S^i_{\mathbf{H}}), \\ &= \arg\min \alpha E_d(L^i, S^i_{\mathbf{H}}) + E_s(L^i, S^i_{\mathbf{H}}), \\ &= \arg\min \alpha \sum_p e_d(L^i(p)) + \sum_{(p,q)} e_s(L^i(p), L^i(q)), \end{aligned} \quad (5)$$

where $S^i_{\mathbf{H}}$ forms the range of the labeling function L^i. This problem can be solved using the graph cuts algorithm. To decrease computational complexity, a pyramid implementation is proposed. The hierarchical solution will be presented at length in Section IV.

The framework of the proposed method is illustrated in Fig. 2 with two main procedures: structure-guided homography transformation estimation and assignment. In the estimation step, dominant linear structures are extracted to guide the estimation of three kinds of transformations: translation, bilateral symmetry and perspective shift. These transformations depict the self-similarity property of an image and are assigned to each unknown pixel based on a global optimization of the MRF in the assignment step. After that, the target image is inpainted by transforming the known pixel values into its unknown region. Moreover, a hierarchical implementation is adopted. A low-resolution image is first built and restored using the proposed transformation-based inpainting algorithm. It is then upsampled to its full resolution, with the quality of its inpainted region improved using the proposed label refinement technique. We include the pseudocode of the transformation-based inpainting in Algorithm 1.

Algorithm 1: Transformation-Based Inpainting

Input: Input image I, mask Ω
Output: Inpainted image I
1: Initialize $i = 0$ and $L^i(\Omega) = 0$ and $S^i_{\mathbf{H}} = \varnothing$
2: **while** $\Omega \neq \varnothing$ **do**
3: $\quad i \leftarrow i + 1$
4: $\quad S^i_{\mathbf{H}} \leftarrow$ TransformationEstimation(I, Ω, L^{i-1})
5: $\quad L^i \leftarrow$ TransformationAssignment$(I, \Omega, S^i_{\mathbf{H}})$
6: $\quad I, \Omega \leftarrow$ Inpainting$(I, \Omega, L^i, S^i_{\mathbf{H}})$
7: **end while**

III. Structure-Guided Homography Transformation Estimation

In this section, we propose a heuristic algorithm for the transformation estimation problem in (4). The statistics of the matched image patches and feature points are calculated to obtain the regularity of the source region. The dominant linear structures are extracted using the regularity statistics. These statistics are then used to estimate a set of homography transformation matrices for sampling under the guidance of dominant linear structures.

First, we give the notation of the linear structure and the transformation. Since human eyes are sensitive to the structure continuity, by detecting and preserving dominant linear structures, the inpainting quality can be improved significantly. In this paper, we define dominant linear structures as a set of lines $S_l = \{l_i\}, i = 1, 2, \ldots, N_l$. $l \in S_l$ is called a dominant structure line, and its direction is denoted as $\phi(l) \in [0, \pi]$.

The homography transformation matrix, which is used to map an image from a two-dimensional view plane onto another view plane, takes the form of:

$$\mathbf{H} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix}, \quad (6)$$

where $a_1, ..., a_8$ are transformation parameters. Specifically, $\mathbf{H}$ projects $p(x, y)$ to its corresponding pixel $\mathbf{H}(p) = p'(x'/w', y'/w')$, where

$$[x', y', w']^T = \mathbf{H}[x, y, 1]^T. \quad (7)$$

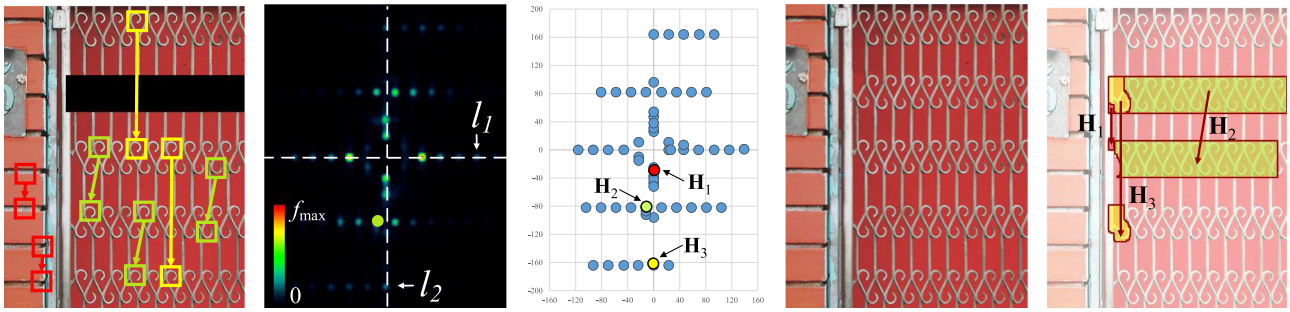


Fig. 3. Translation transformation estimation. (a) Target image and matched patches in Φ. (b) Visualized distribution of displacement vectors. The distribution shows high sparsity. The white dotted lines are the extracted dominant structure lines. Note that the textures of the fence demonstrate regularity in the horizontal direction. The proposed method successfully detects the horizontal structure line l_1 even if there are only a few horizontal lines in the image. (c) Estimated $S_{\mathbf{H}_T} = \{\mathbf{H}_{Ti}\}(i = 1, 2, \ldots, N_T)$ in the $(dx\text{-}dy)$ space. (d) Inpainting result and (e) visualized labeling. The missing region is mainly reconstructed using three translation transformations, which are represented as red, green and yellow circles in (c).

In this paper, we focus on three kinds of homography transformations:

- Translation: $S_{\mathbf{H}_T} = \{\mathbf{H}_{Ti}\}, i = 1, 2, \ldots, N_T$.
- Bilateral symmetry: $S_{\mathbf{H}_S} = \{\mathbf{H}_{Si}\}, i = 1, 2, \ldots, N_S$.
- Perspective shift: $S_{\mathbf{H}_P} = \{\mathbf{H}_{Pi}\}, i = 1, 2, \ldots, N_P$.

and $S_{\mathbf{H}} = S_{\mathbf{H}_T} \cup S_{\mathbf{H}_S} \cup S_{\mathbf{H}_P}$.

A. Dominant Structure Line and Translation Transformation

We first consider the simplest case: translation transformation, the matrix of which is defined as:

$$\mathbf{H}_T = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix}, \tag{8}$$

where (dx, dy) is the displacement vector. For simplicity, we use $\mathbf{v} = (dx, dy)$ to represent the whole $\mathbf{H}_T$, and the calculation of the transformed pixel in (7) can be simplified to $\mathbf{H}_T(p) := p + \mathbf{v} = (x + dx, y + dy)$. Inspired by [10], we use the statistics of the displacement vectors of the matched patches in the source region to estimate translation transformation matrices and the dominant structure lines.

The displacement vectors can be found using:

$$\mathbf{v}(p) = \arg\min_{\mathbf{v}_d} \|\Psi(p + \mathbf{v}) - \Psi(p)\|_2^2 \text{ s.t. } |\mathbf{v}| > \tau. \tag{9}$$

The constraint $|\mathbf{v}| > \tau$ is added to prevent insignificant small displacements. Given all these matched patches, we calculate the frequency of their displacement vectors:

$$f(\mathbf{v}) = \sum_{p \in \Phi} \omega(\mathbf{v}(p) = \mathbf{v})/|\Phi|. \tag{10}$$

where $\omega(\cdot)$ is 1 when the argument is true and 0 otherwise. $|\Phi|$ is the number of pixels in the source region Φ.

Since image patches demonstrate high similarity along linear structures, displacement vectors are likely distributed along the dominant structure line in the $(dx\text{-}dy)$ space. Then, we define the dominant structure line l as the line with the most displacement vectors lying on it:

$$\hat{l} = \arg\max_{l} f(l) = \arg\max_{l} \sum_{\mathbf{v} \in l} f(\mathbf{v}), \tag{11}$$

where $f(l)$ is the sum of the frequencies of the displacement vectors lying on l. The best fitting line $\hat{l}$ is extracted via RANSAC-based voting [34]. We repeat the voting process for all outliers to search for multiple dominant structure lines $S_l = \{l_i\}, i = 1, 2, \ldots, N_l$ until $f(l_{N_l+1}) < \lambda_{\text{line}} f(l_1)$.

Since the displacement vectors near the dominant structure lines better depict the principal structures of the target image, they contribute more to the inpainting process compared with other displacement vectors. Thus, we use S_l to refine the frequency of the displacement vectors as follows:

$$\hat{f}(\mathbf{v}) = \begin{cases} 2f(\mathbf{v}), \text{ if } \exists l \in S_l \rightarrow \mathbf{v} \in l \\ f(\mathbf{v}), \text{ other} \end{cases}. \tag{12}$$

Finally, we choose a number of N_T displacement vectors with the greatest $\hat{f}(\mathbf{v})$, and their corresponding transformation matrices form $S_{\mathbf{H}_T} = \{\mathbf{H}_{Ti}\}, i = 1, 2, \ldots, N_T$. As shown in Fig. 3, the distribution of these vectors demonstrates sparsity, and most of the displacement vectors assemble around the estimated $S_{\mathbf{H}_T}$, which can well describe the self-similarity of the target image and thus be used to predict the desired texture and structure patterns.

B. Bilateral Symmetry Transformation

Objects in natural images often possess bilateral symmetry and can be utilized for inpainting. We estimate bilateral symmetry transformation matrices with the form of:

$$\begin{aligned} \mathbf{H}_S &= \mathbf{H}_{St}\mathbf{H}_{Sr} \\ &= \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos 2\theta & \sin 2\theta & 0 \\ \sin 2\theta & -\cos 2\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned} \tag{13}$$

where θ is the direction of the axis of symmetry. A bilateral symmetry transform operation can be decomposed into a

reflection operation $\mathbf{H}_{Sr}$ using the line passing through the origin as the axis and a translation operation $\mathbf{H}_{St}$.

Given a dominant structure line $l \in S_l$, $\mathbf{H}_{Sr}$ is determined by

$$\theta = \phi(l) + \pi/2. \tag{14}$$

This means that the axis is perpendicular to the dominant structure line l, which retains the direction of l after reflection.

Then, we estimate $\mathbf{H}_{St}$ using the method proposed in the previous section. The main difference is that in (9), the matched patches $\Psi(p+\mathbf{v})$ are searched for in the reflected image $\mathbf{H}_{Sr}(I)$ rather than in I itself.

C. Perspective Shift Transformation

The ubiquitous foreshortening effects severely degrade the results of MRF-based image inpainting methods [10], [13], as they only perform the translation operation. We put forward the concept of *perspective shift* in addition to the traditional translation operation. Objects are shifted in a way that satisfies the foreshortening effects. To accomplish this task, we estimate perspective shift transformation matrices $\mathbf{H}_P$ of the following form

$$\mathbf{H}_P = \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & 1 \end{bmatrix}. \tag{15}$$

To find non-fronto-parallel self-similarities for $\mathbf{H}_P$ estimation, feature points are used. We begin with speeded up robust features (SURF, [35]) point detection and compute SURF descriptors for each feature point. Then, these feature points are matched under the guidance of the dominant structure line l. Specifically, two feature points (assuming that they are located at pixels p and q) are matched if their vector $\overrightarrow{pq}$ satisfies the angle constraint:

$$d_\pi(\phi(\overrightarrow{pq}), \phi(l)) < \lambda_\theta, \tag{16}$$

where $d_\pi(\phi_1, \phi_2) = \min(|\phi_1 - \phi_2|, \pi - |\phi_1 - \phi_2|)$. The angle constraint facilitates the estimation of orientational consistent transformations.

Then, we apply a RANSAC-based voting algorithm to the matched feature point pairs to find the best fitting homography matrix $\mathbf{H}$ as a candidate perspective shift transformation. We repeat the RANSAC-based voting process for all outliers to obtain a set of candidate perspective shift transformations.

To determine the optimal candidate, we define two measurements:

- *Information magnitude:* Let $\Lambda(\mathbf{H}) = \mathbf{H}(\Phi) \cap \Omega$ denote the region perspectively shifted from the source region to the missing region using $\mathbf{H}$. The information magnitude is defined as the percentage of $\Lambda(\mathbf{H})$ in Ω:

$$R_{\text{mag}}(\mathbf{H}) = |\Lambda(\mathbf{H})|/|\Omega|. \tag{17}$$

  $\mathbf{H}$ with low $R_{\text{mag}}(\mathbf{H})$ is not desirable, as it can only fill a small portion of the missing region.
- *Information quality:* The boundary consistency is taken into account. As shown in Fig. 4(b) and (c), we concentrate on the outer boundary of Ω, which has a width of $\lambda_{\Delta\Omega}$ pixels (denoted as $\Delta\Omega$). The information quality is defined as the correlation coefficient of the pixel values in $\Delta\Omega$:

$$R_{\text{quality}}(\mathbf{H}) = \text{Cov}(\mathbf{X}, \mathbf{Y})/\sigma(\mathbf{X})\sigma(\mathbf{Y}). \tag{18}$$

  In the above equation, $\mathbf{X}$ is a vector whose elements are raster-scanned pixel values of I in $\Delta\Omega$; $\mathbf{Y}$ is similarly defined. Higher $R_{\text{quality}}(\mathbf{H})$ indicates a more seamless inpainting result.

Then, $\mathbf{H}_P$ is obtained by solving:

$$\max_{\mathbf{H}} R_{\text{quality}}(\mathbf{H}) \text{ s.t. } R_{\text{mag}}(\mathbf{H}) > \lambda_{\text{mag}}. \tag{19}$$

Moreover, if a certain perspective shift transformation $\mathbf{H}_P$ exists, $\mathbf{H}_P^i$ could possibly be valid perspective shift transformation matrices as well. Intuitively, $\mathbf{H}_P^2$ represents a double perspective shift operation, and $\mathbf{H}_P^{-1}$ represents an inverse perspective shift operation. Thus, we update $S_{\mathbf{H}_P} = S_{\mathbf{H}_P} \cup \{\mathbf{H}_P^i\}, |i| \in \{1, \ldots, \lambda_{\mathbb{Z}}\}$, if $R_{\text{mag}}(\mathbf{H}_P^i) > \lambda_{\text{mag}}$. Fig. 4(d) gives an example of the estimated $S_{\mathbf{H}_P}$, with $\lambda_{\mathbb{Z}} = 4$.

Finally, $S_{\mathbf{H}} = S_{\mathbf{H}_T} \cup S_{\mathbf{H}_S} \cup S_{\mathbf{H}_P}$ is obtained. The pseudocode of the transformation estimation algorithm is summarized in Algorithm 2.

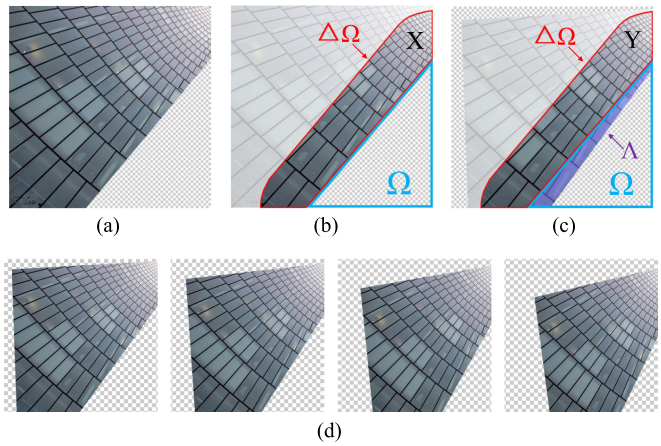


Fig. 4. Perspective shift transformation estimation. (a) The target image. (b) (c) The outer boundary of I and $\mathbf{H}_P(I)$, respectively. The violet region Λ is generated by $\mathbf{H}_P$ to fill the blue triangular region Ω. $R_{\text{mag}}(\mathbf{H}_P)$ is the percentage of the violet region in the blue triangle. Meanwhile, $R_{\text{quality}}(\mathbf{H}_P)$ is the correlation between $\mathbf{X}$ and $\mathbf{Y}$, which evaluates the consistency of the boundary between Λ and $\mathbf{X}$. (d) Estimated perspective shift transformations. From left to right: $\mathbf{H}_P(I), \mathbf{H}_P^2(I), \mathbf{H}_P^3(I)$ and $\mathbf{H}_P^4(I)$.

IV. Hierarchical Exemplar-Based Image Inpainting

In this section, we give detailed definitions of the data term and the smoothness term in (5), which could be solved using graph cuts. In addition, a hierarchical implementation is proposed to achieve lower computational complexity and better structure estimation.

A. MRF Energy Function Definition

After homography transformation estimation, we obtain a set of transformation matrices $S_{\mathbf{H}} = \{\mathbf{H}_i\}, i = 1, \ldots, N_{\mathbf{H}}$. To accomplish contextual-continuous inpainting, we seek the

Algorithm 2: Transformation Estimation

Input: Input image I, mask Ω
Output: Transformations $S_{\mathbf{H}}$

Initialize $S_l = S_{\mathbf{H}_T} = S_{\mathbf{H}_S} = S_{\mathbf{H}_P} = \varnothing$
$\triangle$ *Dominant structure line estimation:*
match patches to compute $\mathbf{v}$ (9) and $f(\mathbf{v})$ (10)
estimate l_1 (11) and $i \leftarrow 1$
while $f(l_i) \geq \lambda_{line} f(l_1)$ **do**
    $S_l \leftarrow S_l \cup \{l_i\}$ and $i \leftarrow i + 1$
    estimate l_i (11)
end while
$\triangle$ *Translation estimation:*
refine $\hat{f}(\mathbf{v})$ by S_l (12)
for $i = 1 \rightarrow N_T$ **do**
    seek $\mathbf{H}_T$ with the i-th highest $\hat{f}(\mathbf{v})$
    $S_{\mathbf{H}_T} \leftarrow S_{\mathbf{H}_T} \cup \{\mathbf{H}_T\}$
end for
$\triangle$ *Bilateral symmetry estimation:*
for all $l \in S_l$ **do**
    compute $\mathbf{H}_{Sr}$ by l (14)
    estimate $\mathbf{H}_{St}$ using the statistics of matched patches between I and $\mathbf{H}_{Sr}(I)$ as in *Translation estimation*
    $\mathbf{H}_S = \mathbf{H}_{St}\mathbf{H}_{Sr}$ and $S_{\mathbf{H}_S} \leftarrow S_{\mathbf{H}_S} \cup \{\mathbf{H}_S\}$
end for
$\triangle$ *Perspective shift estimation:*
match feature points under angle constraint (16)
estimate candidate transformations $\{\mathbf{H}\}$ using RANSAC-voting over matched feature points
seek $\mathbf{H}_P$ from $\{\mathbf{H}\}$ (19)
for all i such that $|i| \in \{1, \ldots, \lambda_{\mathbb{Z}}\}$ **do**
    if $R_{\text{mag}}(\mathbf{H}_P^i) > \lambda_{\text{mag}}$ **then**
        $S_{\mathbf{H}_P} \leftarrow S_{\mathbf{H}_P} \cup \{\mathbf{H}_P^i\}$
    end if
end for
$S_{\mathbf{H}} = S_{\mathbf{H}_T} \cup S_{\mathbf{H}_S} \cup S_{\mathbf{H}_P}$

optimal labeling function that minimizes the following MRF energy function:

$$\alpha E_d(L) + E_s(L) = \alpha \sum_{p \in \Omega} e_d(L(p)) + \sum_{(p,q) \in \mathcal{N}} e_s(L(p), L(q)), \tag{20}$$

where α is the weight used to combine two energy terms, and

- *Smoothness term:* $e_s(L(p), L(q))$ penalizes the discontinuity of nearby pixels. It is defined as follows (for simplicity, we assume that $L(p) = \mathbf{H}_i$, $L(q) = \mathbf{H}_j$):

$$e_s(L(p), L(q)) = d_p(\mathbf{H}_i(p), \mathbf{H}_j(p)) + d_p(\mathbf{H}_i(q), \mathbf{H}_j(q)), \tag{21}$$

where $d_p(\cdot, \cdot)$ measures the similarity between two pixels:

$$d_p(p, q) = \|I(p) - I(q)\|_1 + \beta \|\nabla I(p) - \nabla I(q)\|_1, \tag{22}$$

where ∇I is the magnitude of the image gradient and β is the weight used to combine the intensity and gradient terms. Fig. 5(a) presents the definition of the smoothness term. Since $(\mathbf{H}_j(p))$ and $(\mathbf{H}_j(q))$ are two adjacent pixels in the source region, their values satisfy contextual continuity. Therefore, the continuity between p (with a value of $I(\mathbf{H}_i(p))$) and q (with a value of $I(\mathbf{H}_j(q))$) can be measured based on the similarity between $(\mathbf{H}_i(p))$ and $(\mathbf{H}_j(p))$. The same is true for $(\mathbf{H}_i(q))$ and $(\mathbf{H}_j(q))$.

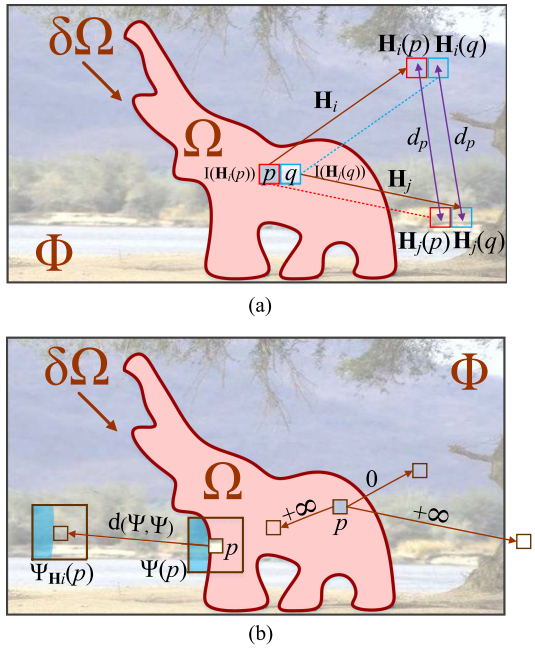


Fig. 5. Illustration of the MRF energy function. We use translation transformations as examples to achieve a more intuitive presentation. The small boxes represent pixels, and the large boxes represent image patches. The transformation $\mathbf{H}$ is represented by an arrow. (a) The smoothness term measures the similarity between pixel $(\mathbf{H}_i(p))$ and $(\mathbf{H}_j(p))$ as well as between pixel $(\mathbf{H}_i(q))$ and $(\mathbf{H}_j(q))$ (see the small boxes to which the purple arrows point). (b) The data energies are written next to their corresponding arrows. The energy of the pixel on $\delta\Omega$ is measured based on patch differences in the blue area $(\Psi \cap \Phi)$. The energy of the inner pixel depends on whether the pixel to which the considered arrow points is valid. (a) Smoothness term. (b) Data term.

- **Data term:** $e_d(L(p))$ is defined as:

$$e_d(L(p)) = \begin{cases} +\infty, \text{if } \mathbf{H}_i(p) \notin \Phi \\ 0, \text{if } \mathbf{H}_i(p) \in \Phi \wedge p \in \Omega \backslash \delta\Omega \\ d_\Psi(\Psi_{\mathbf{H}_i}(p), \Psi(p)), \text{other} \end{cases}, \tag{23}$$

where $\Psi_{\mathbf{H}_i}(p)$ is the patch centered at p in the inversely transformed image $\mathbf{H}_i^{-1}(I)$ and $d_\Psi(\cdot, \cdot)$ is the patch difference that measures the consistency along the boundary between Ω and Φ. $d_\Psi(\cdot, \cdot)$ is calculated as follows:

$$d_\Psi(\Psi_1, \Psi_2) = \|G \otimes (\Psi_1 - \Psi_2)\|_1 + \beta \|G \otimes (\nabla\Psi_1 - \nabla\Psi_2)\|_1, \tag{24}$$

where G is the Gaussian weight matrix and $\otimes$ is the pointwise product operator. Only known pixels in the patch are computed, as shown in Fig. 5(b) in blue.

Once the MRF graph is given, the energy optimization is achieved using multi-label graph cuts.[1]

[1] http://vision.csd.uwo.ca/code/

Algorithm 3: Hierarchical Inpainting

Input: Input image I, mask Ω, level K
Output: Inpainted image I

```
Initialize I^(0) = I
{I^(k)}, {Ω^(k)} ←Downsample(I, K), k = 1, ..., K − 1
L^(K−1) ←TransformationBasedInpainting (I^(K−1), Ω^(K−1))
for k = K − 2, ..., 1, 0 do
    L^(k) ←Upsample(L^(k+1)) (25)
    while Ω^(k) ≠ ∅ do
        select p ∈ δΩ^(k) with highest priority
        L^(k)(p) ←LabelRefinement(L^(k)(p)) (26)
        update Ω^(k)
    end while
end for
I ← Inpainting(I, Ω, L^(0))
```

B. Hierarchical Implementation

At the coarse level, image inpainting benefits from low computational complexity. Furthermore, it becomes less sensitive to noise and local singularities. Thus, estimated transformations are more reliable for demonstrating image regularity. Based on this consideration, we propose a hierarchical implementation of our inpainting algorithm. The pseudocode of the proposed hierarchical inpainting is given in Algorithm 3.

The target image is first downsampled by a factor of 2 to form a K-level image pyramid $\{I^{(k)}\}, k = 0, \ldots, K-1$, where the superscript k denotes the image pyramid level, with $k = 0$ and $k = K-1$ corresponding to the highest and lowest resolution levels, respectively. $L^{(K-1)}$ is obtained using the proposed EM-like optimization approach to inpaint $I^{(K-1)}$. Instead of upsampling the inpainting results directly, which leads to blurring artifacts, we upsample the labeling function using the nearest neighbor interpolation:

$$L^{(k)}(p) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot L^{(k+1)}(\lfloor p/2 \rfloor), \tag{25}$$

where $\lfloor \cdot \rfloor$ is the rounding down operation. However, the nearest neighbor interpolation may cause small misalignments along the outer boundaries $\delta\Omega$ and inner boundaries between different label assignments. To solve this issue, we propose an outside-in labeling refinement method based on pixel priority. The pixel priority is calculated according to the formula given in [6]. Regarding the pixel p with the highest priority on $\delta\Omega$, its label is refined as follows:

$$\widehat{L}(p) = \mathbf{H}' = \arg \min_{\mathbf{H}' \in \mathcal{N}(L(p))} d_\Psi(\Psi_{\mathbf{H}'}(p), \Psi(p)), \tag{26}$$

where $\mathcal{N}(\mathbf{H})$ contains transformation matrices that project p to the neighborhood of $\mathbf{H}(p)$, as shown in Fig. 6. Specifically, we define $S_\mathcal{N}$ as a set of 5 translation transformation matrices with $(dx, dy) \in \{(-1,0), (1,0), (0,-1), (0,1), (0,0)\}$. Then, $\mathcal{N}(\mathbf{H}) = \{\mathbf{H}_T \mathbf{H} | \mathbf{H}_T \in S_\mathcal{N}\}$. Compared with [6], our search range for the matched patches is only five pixels rather than the whole Φ. In each iteration of refinement, the labeling function changes by at most one pixel, but the final adjustment can be large thanks to the hierarchical process. Fig. 7 demonstrates that image structures are preserved by the proposed labeling refinement method.

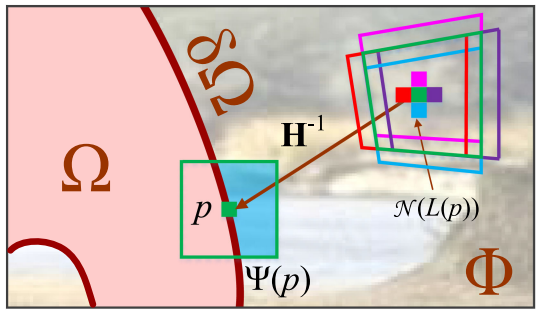


Fig. 6. The proposed labeling refinement method. Before refinement, $L(p) = \mathbf{H}$ is assigned to the p with the highest priority. The red, magenta, purple and blue pixels are neighbors of $\mathbf{H}(p)$. Their corresponding patches are inversely transformed using $\mathbf{H}'^{-1} \in \mathcal{N}(\mathbf{H})$ to measure the similarity with $\Psi(p)$ in the blue area. The central pixel of the most matched patch is used to fill p, and the label of p is updated accordingly.

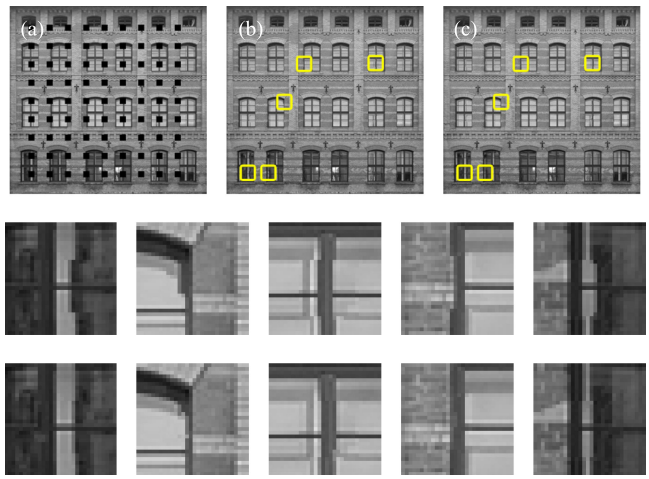


Fig. 7. Correction of the misalignments via labeling refinement. The first row: (a) original image, (b) inpainting result before refinement and (c) inpainting result after refinement. The second row: patches with misalignments in (b). The third row: corresponding refined patches in (c).

In the end, we obtain the optimal labeling for the target image at the original resolution to fill the missing region. In addition, a Poisson fusion [36] is used to further hide seams.

V. Experimental Results and Analysis

The proposed method is implemented using C++ on the Visual Studio 2013 platform and tested using various images of natural/semi-natural scenes[2] and artificial scenes.[3] We use the state-of-the-art image inpainting methods [10], [14], [24], [25], [31] to conduct a comparison experiment. In the experiment,

[2]http://people.irisa.fr/Olivier.Le_Meur/publi/2013_TIP/index.html
[3]https://sites.google.com/site/jbhuang0604/publications/struct_completion

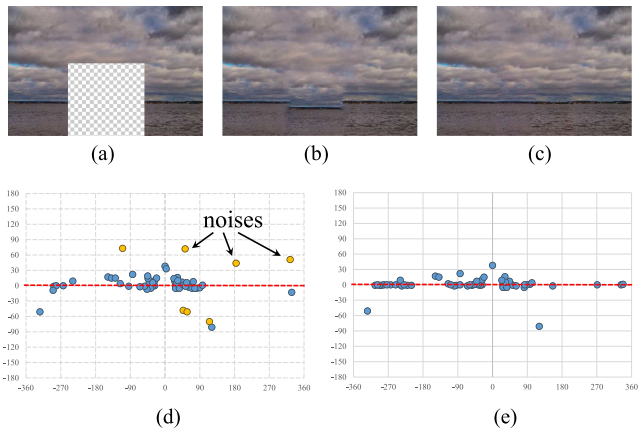


Fig. 8. Refinement of the frequency of displacement vectors using dominant structure lines. (a) The target image. (b) (c) Inpainting results obtained using different translation transformations. (d) Translation transformations for (a) and the dominant structure line (red dotted lines) detected. (e) Translation transformations after refinement. Noises (in yellow) are effectively suppressed.

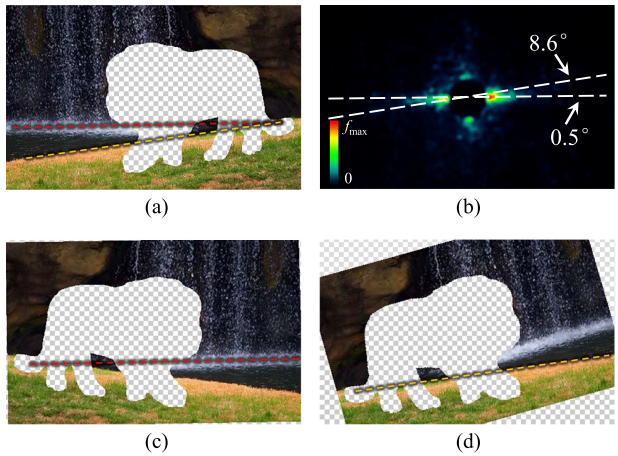


Fig. 9. Bilateral symmetry transformation. (a) Target image. (b) Dominant structure lines (dotted lines). The dominant structure lines detected exactly match the ground plane (yellow line) and water plane (red line) in (a). (c) (d) Results obtained via reflection transformations $\mathbf{H}_{Sr}$. The samples available for the inpainting of the ground and water are dramatically enriched.

we set the following thresholds: $\lambda_{\text{line}} = 0.6, \lambda_{\Delta\Omega} = 200, \lambda_\theta = \pi/8, \lambda_{\mathbb{Z}} = 4$, and $\lambda_{\text{mag}} = 0.2$. The minimum allowed length of displacement vectors is $\tau = \max(W, H)/15$, where W and H are the width and height of the target image, respectively. The weights α and β used to balance different energy terms are set to 2. The number of transformations is $N_{\mathbf{H}} = 60$.

A. Guidance of the Dominant Structure Lines

Our method extracts the dominant structure line to guide the homography transformation estimation. We investigate how structures are well preserved using this guidance.

Translation Transformation. In Section III-A, the frequency of the displacement vectors is refined using the dominant structure lines. Fig. 8 illustrates how the refinement affects the inpainting results. The original translation transformation matrices suffer from noises that lead to poor local minima. After frequency refinement based on the dominant structure line (sea level), the noise offsets are suppressed, and the sea level is better preserved.

Bilateral Symmetry Transformation. Dominant structure lines determine the reflection operation $\mathbf{H}_{Sr}$ of the bilateral symmetry transformation. Fig. 9 illustrates the reflected images $\mathbf{H}_{Sr}(I)$. The dominant structure lines detected exactly match the ground plane and water plane. Thus, the reflected results effectively enrich the samples available for the inpainting of the ground and water.

Perspective Shift Transformation. For perspective shift transformation estimation, dominant structure lines take the role of the angle constraint in (16). Fig. 10(c) shows the point pairs that were matched under the direction guidance of the dominant structure line l, with $\phi(l) = 137.1°$. To demonstrate how dominant structure lines guide the transformation estimation results, we enumerate different angles to guide the direction and obtain the corresponding perspective transformation matrix $\mathbf{H}_P$. The corresponding R_{quality} is calculated. Fig. 10(d) shows how angle

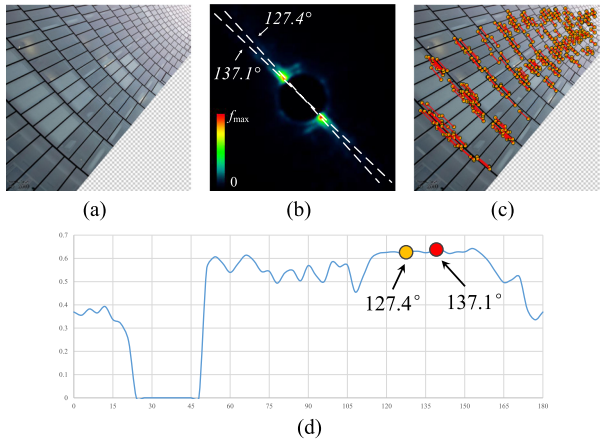


Fig. 10. Use of dominant structure lines to guide the main direction of the perspective shift transformation. (a) The input image. (b) Dominant structure lines (dotted lines). (c) Matched feature points. (d) Illustration of how the angle constraint affects $R_{\text{quality}}(\mathbf{H}_P)$ and how dominant structure lines lead to relatively high $R_{\text{quality}}(\mathbf{H}_P)$.

restriction affects R_{quality}, proving that the dominant structure lines can reliably guide the algorithm to find $\mathbf{H}_P$ with relatively high information quality.

B. Comparisons Using Natural/Semi-Natural Scenes

We compare our approach with Photoshop's content-aware fill tool [14], [25], He's MRF-based method [10], Le Meur's hierarchical SR-based method [24] and Huang's planar structure guidance method [31] using natural/semi-natural scenes. The software corresponding to Le Meur's method and Huang's method are available on their respective project websites[23]. The results of He's method are from the project websites.[4]

[4]http://research.microsoft.com/en-us/um/people/kahe/eccv12/

Fig. 11. Comparison with state-of-the-art methods using natural and semi-natural scenes: *Tiger* (top), *House* (middle) and *Trashcan* (bottom). (a) Original pictures with unknown regions, (b) Photoshop results, (c) Le Meur's results [24], (d) Huang's results [31], and (e) results of the proposed method.

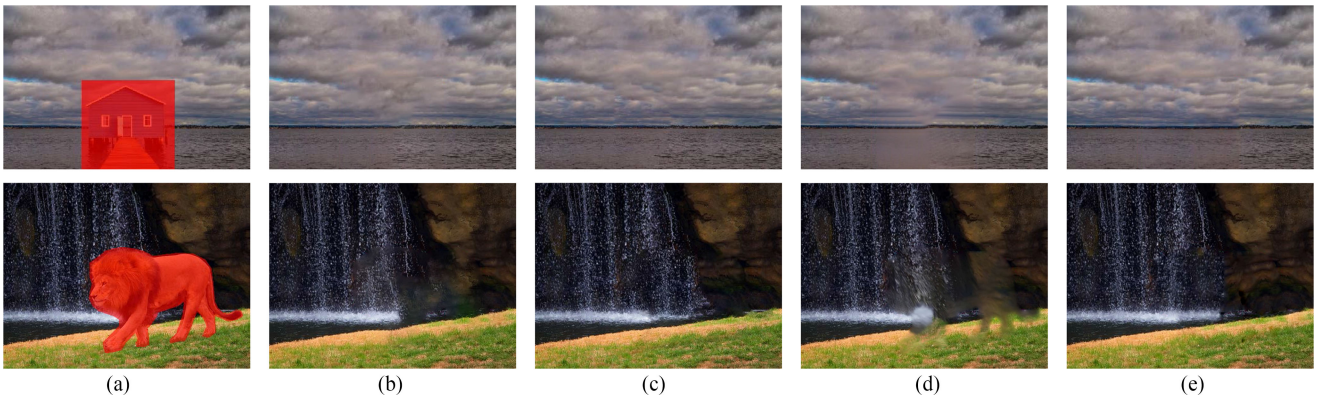


Fig. 12. Comparison with state-of-the-art methods using natural and semi-natural scenes: *Sea* (top) and *Lion* (bottom). (a) Original pictures with unknown regions, (b) Photoshop results, (c) He's results [10], (d) Huang's results [31], and (e) results of the proposed method.

Figs. 11 and 12 show the inpainting results. Please enlarge these figures on the screen to achieve a better comparison. The coherence-based methods (Photoshop and Huang's work) tend to produce blurring artifacts, leaving textures that are not well preserved. For example, in the *Tiger* image, the wood grain of the trunk is not well reconstructed. In Huang's result, the structure of the trunk is even incomplete. Although Le Meur's method preserves the wood grain, it overpropagates the trunk structure on the left side. By comparison, our method uses the strong structure guidance and successfully reconstruct a richly textured and structure-preserved trunk. Meanwhile, as shown in the *House* image, the bilateral symmetry transformations enable the proposed method to recover the left part of the house. In addition, for images without strong symmetry, such as *Lion* in the last row of Fig. 12, the proposed method has advantages over He's work in filling the ground plane and water plane thanks to the bilateral symmetry transformations (see Fig. 9). For natural/semi-natural scenes, the proposed method is better than or comparable to the state-of-the-art methods with respect to both texture and structure preservation.

C. Comparisons Using Regular Artificial Scenes

We compare our approach with Photoshop's content-aware fill tool [14], [25], He's MRF-based method [10] and Huang's planar structure guidance method [31]. The test images and results are taken from Huang's project website[3]. Fig. 13 shows the results for challenging artificial scenes. Please enlarge these figures on the screen to achieve a better comparison. In Photoshop's and He's results, the structures appear severely cracked, as both methods search for patches in only the translation transformation space and fail to reconstruct the structures with

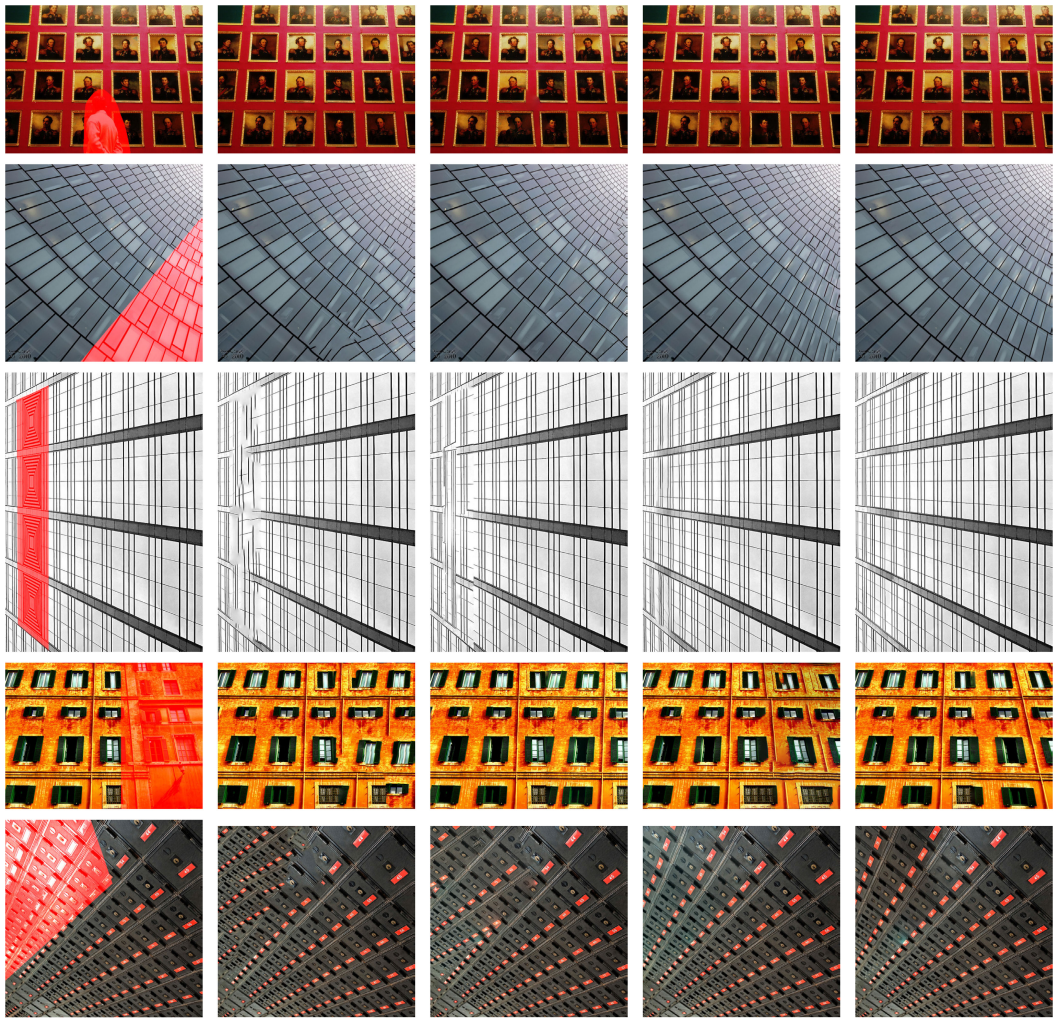


Fig. 13. Comparison with state-of-the-art methods using regular artificial scenes. From top to bottom: *Exhibition*, *Pane*, *Glass*, *Windows* and *Locker*. (a) Original pictures with unknown regions, (b) Photoshop results, (c) He's results [10], (d) Huang's results [31], and (e) results of the proposed method.

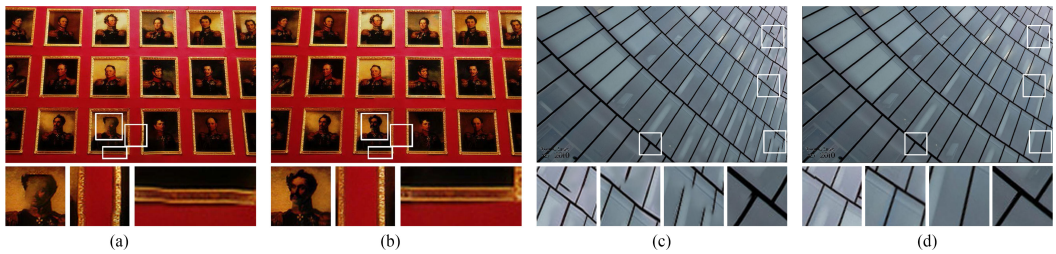


Fig. 14. Comparisons with Huang's work using local images: *Exhibition* (left) and *Pane* (right). Our approach suffers fewer structure line distortions. (a) Huang [31]. (b) Our result. (c) Huang [31]. (d) Our result.

foreshortening effects. Compared with Photoshop and He's method, the proposed method and Huang's method allow for a broader perspective transformation search space and suffer fewer artifacts.

Huang's method has the same base algorithm as that of Photoshop in regard to objective function optimization [25] and PatchMatch [14]. By guiding the patch searching and propagation using mid-level structure cues, Huang's method allows for a search space with more degrees of freedom without the problem of falling into poor local minima. Therefore, structures are well preserved. Compared with Huang's results, our approach suffers less distortion thanks to the perspective shift. As shown in Fig. 14, although mid-level constraints are utilized, the structure lines are more or less distorted in Huang's results because the search for each patch is relatively separated. In the *Exhibition* image (left), the picture frames are distorted, and the head portrait is blurred. By comparison, the proposed method perspectively shifts all known pixel values to-

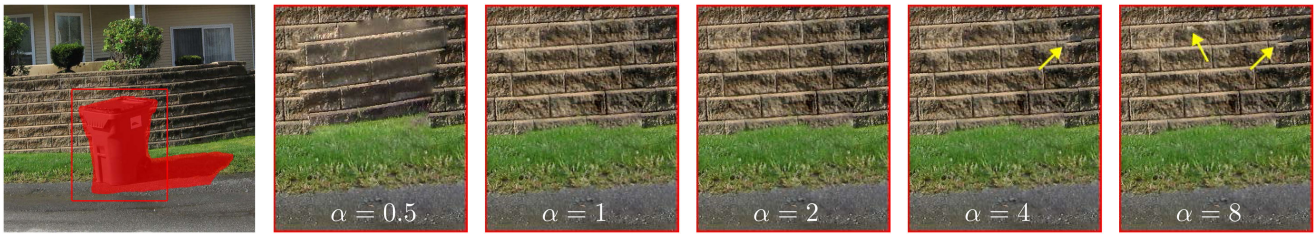


Fig. 15. Effect of the date term weight α on the inpainting result. From left to right: input image, inpainting results of the proposed method with $\alpha = 0.5$, $\alpha = 1$, $\alpha = 2$, $\alpha = 4$ and $\alpha = 8$. The area in the red rectangle in the original image is enlarged in the results.

TABLE I
THE AVERAGE RUNNING TIMES (SECONDS) OF DIFFERENT METHODS

Image	Photoshop	He[6] [10]	Le Meur [24]	Huang [31]	Ours
Natural	0.55	11.54	446.42	84.67	6.56
Artificial	0.64	11.66	1135.91	126.84	16.18
Textural	0.18	1.07	239.97	73.65	3.42

wards the unknown regions, and adjacent pixels are dealt with uniformly, leading to better inpainting quality. In the *Pane* image (right), the window frames are distorted and even appear as cracked in Huang's results, while the proposed method synthesizes physically plausible inpainting results.

D. Running Time

We compare the running times of different methods using the *Natural* image set (five images from Figs. 11 and 12), *Artificial* image set (five images from Fig. 13) and *Textural* image set (thirty-two images from the texture dataset[5]). The average image sizes of these three sets are 345 $\times$ 466, 532 $\times$ 640 and 200 $\times$ 200, respectively. Their average missing rates are 19.02%, 23.43% and 59.04%, respectively. Table I shows the average running times for these images using an Intel Xeon 3.00 GHz CPU E5-1607 and 16 GB RAM. It can be observed that the time costs of the two MRF-based approaches (He's method[6] and the proposed method) are similar in magnitude. The Photoshop commercial software is much faster than the proposed method because it is well tuned and fully parallelized. Meanwhile, the time-consuming super-resolution process in Le Meur's method becomes its major computational burden. Moreover, the efficiency of Huang's method is limited because its released software is implemented using MATLAB. Because the proposed method is not multi-threaded, it uses only one core. Our method can be further sped up by matching patches and feature points in parallel.

E. Effect of the Parameters

A crucial aspect of our approach is the transformation assignment performed to accomplish contextual continuity. In (20), the data term and smoothness term are combined using the weight α. Since the data term and smoothness term emphasize the boundary continuity and inner continuity, respectively, we can control the boundary structure using different values of α, as shown in Fig. 15. The use of a higher α value can better preserve the boundary structures. However, an extremely high α value will overemphasize the local continuity and may yield some artifacts around the boundary (see yellow arrows in Fig. 15). The best inpainting result is obtained with an intermediate weight of 2.

The other parameter, i.e., β, denotes the weight used to combine the intensity term and gradient term in (22) and (24). The gradient term makes the proposed algorithm robust to illuminance and color variations. It can, however, also affect the contextual continuity. We found via an experiment that $\beta = 2$ is a good choice for the pixel/patch similarity measurement.

Fig. 16. Inpainting of Markov-process-disenabled scenes with/without user guidance. (a) Input image. (b) Our inpainting result without user guidance. (c) User guidance, in which user specifies the horizontal symmetry transformation. (d) Inpainting result with user guidance.

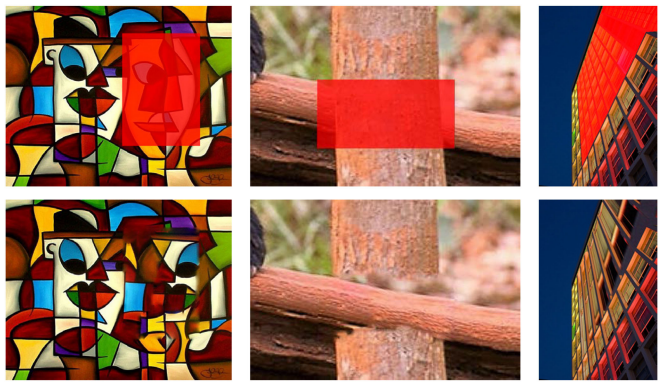


Fig. 17. Failure cases. Top: Three kinds of images without well-defined structures. *Artwork* with complex structure fragments (left), *Tree* with structure ambiguity (middle) and *Building* with high missing rate (right). Bottom: The corresponding inpainting results.

F. Limitations

Since our method is based on the MRF model, the reconstruction of Markov-process-disenabled scenes will be challenging.

[5]http://graphics.stanford.edu/projects/texture/demo/synthesis_misc.html

[6]We use the C++ re-implementation of He's method in the OpenCV 3.0 xphoto module to test the running time: https://github.com/Itseez/opencv_contrib/tree/master/modules/xphoto

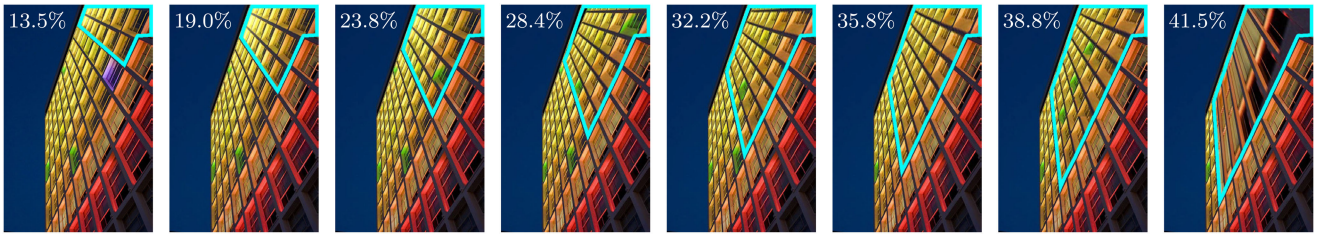


Fig. 18. Effect of the missing rate on the inpainting result for the *Building* image. The cyan frames in the images indicate the inpainted regions. The missing rates are shown in the upper-left corner. Since the sky in the image provides no valid information for inpainting, it is not considered when determining the missing rate in this experiment; hence, the missing rate is $|\Omega|/(|I| - |\text{sky region}|)$.

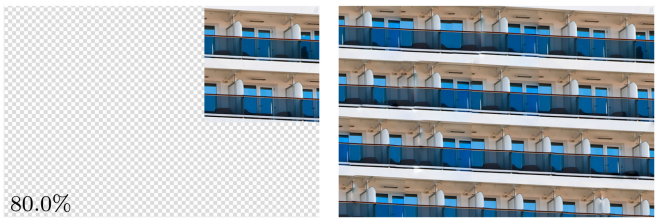


Fig. 19. Plausible inpainting result (right) obtained from the well-structured *Balcony* image (left) using the proposed method with 80.0% of the valid information missing.

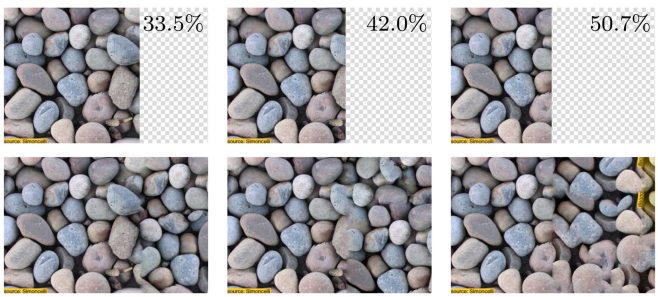


Fig. 20. Effect of the missing rate on the inpainting result for the *Pebbles* image. Top: Input images, with missing rates given in the upper-right corner. Bottom: The corresponding inpainting results.

For example, our method fails to detect the symmetry of a face when half of the face is covered, as shown in Fig. 16(a), thus yielding an odd result [Fig. 16(b)]. One possible solution is to introduce user guidance, as is done in many human-interactive methods [7], [30]. As shown in Fig. 16(c), (d), by artificially specifying a horizontal symmetry transformation, our inpainting result becomes much more reasonable.

Our method may fail for images without well-defined structures. Fig. 17 shows the failure cases under three challenging situations. In the first case, the artwork is composed of complex structure fragments without dominant structures. In the second situation, the trunk is reconstructed in the wrong direction due to the structure ambiguity between the source region and missing region. The last case concerns the missing rate. With too much information lost, it is hard to estimate valid dominant structure lines to guide the inpainting process.

We further investigate the critical missing rate at which our method is no longer effective. We find via experiments on multiple images that the critical missing rate is image-dependent. As shown in Fig. 18, our method demonstrates robustness with respect to information loss for well-structured images; an extreme example is given in Fig. 19, in which the missing rate reaches 80.0%. For natural scenes, the critical missing rate is a little lower. The results in Fig. 20 show that with increased missing rate, our method fails to preserve the boundary structures (42.0% missing rate) and even the inner textures (50.7% missing rate). Empirically, it seems that the performance of our method will decrease when the missing rate of the valid information becomes greater than $40\%-60\%$.

VI. Concluding Remarks

In this paper, we introduce a novel inpainting model that takes both contextual continuity and self-similarity into account. An efficient EM-like optimization approach is proposed to solve the inpainting problem. Given a target image with missing regions, our approach can detect its dominant structure lines and use these features to automatically guide the homography transformation estimation. These estimated transformations are combined to reconstruct the target image using the proposed hierarchical inpainting approach. The hierarchical implementation accelerates the algorithm and offers robust structure feature detection. We validate the effectiveness of our method via comparisons with state-of-the-art image inpainting algorithms using both natural/semi-natural and artificial scenes. The experimental results demonstrate that the image inpainting results can be greatly improved using the proposed inpainting model.

References

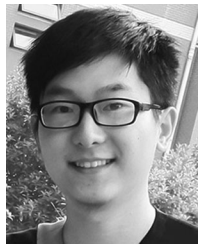
[1] C. Guillemot and O. Le Meur, "Image inpainting: Overview and recent advances," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 127–144, Jan. 2014.
[2] M. Bertalio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. Annu. Conf. Comput. Graph. Interactive Techn.*, 2000, pp. 417–424.
[3] T. F. Chan and J. Shen, "Mathematical models for local nontexture inpaintings," *SIAM J. Appl. Math.*, vol. 62, no. 3, pp. 1019–1043, 2001.
[4] T. F. Chan and J. Shen, "Nontexture inpainting by curvature-driven diffusions (CDD)," *J. Vis. Commun. Image Represent.*, vol. 12, no. 4, pp. 436–449, 2001.
[5] A. Tsai, A. Yezzi, and A. S. Willsky, "Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1169–1186, Aug. 2001.
[6] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
[7] J. Sun, L. Yuan, J. Jia, and H.-Y. Shums, "Image completion with structure propagation," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 861–868, Jul. 2005.
[8] H. Huang *et al.*, "Mind the gap: Tele-registration for structure-driven image completion," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 174:1–174:10, Nov. 2013.

[9] C.-W. Fang and J.-J. Lien, "Rapid image completion system using multiresolution patch-based directional and nondirectional approaches," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2769–2779, Jul. 2009.
[10] K. He and J. Sun, "Image completion approaches using the statistics of similar patches," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2423–2435, Dec. 2014.
[11] A. Bugeau, M. Bertalmio, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2634–2645, Apr. 2010.
[12] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2649–2661, Nov. 2007.
[13] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 151–158.
[14] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 341–352, Aug. 2009.
[15] Y. Liu and V. Caselles, "Exemplar-based image inpainting using multiscale graph cuts," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1699–1711, May 2013.
[16] N. C. Tang, C.-T. Hsu, C.-W. Su, T. K. Shih, and H. Y. M. Liao, "Video inpainting on digitized vintage films via maintaining spatiotemporal continuity," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 602–614, Feb. 2011.
[17] A. Mosleh, N. Bouguila, and A. B. Hamza, "Video completion using bandlet transform," *IEEE Trans. Multimedia*, vol. 14, no. 6, pp. 1591–1601, May 2012.
[18] P. Pérez, M. Gangnet, and A. Blake, "Patchworks: Example-based region tiling for image editing," Microsoft Research, Redmond, WA, USA, Tech. Rep. MSR-TR-2004-04, Jan. 2004.
[19] A. Wong and J. Orchard, "A nonlocal-means approach to exemplar-based inpainting." in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 2600–2603.
[20] R. Martinez-Noriega, A. Roumy, and G. Blanchard, "Exemplar-based image inpainting: Fast priority and coherent nearest neighbor search," in *Proc. Mach. Learn. Signal Process.*, Sep. 2012, pp. 1–6.
[21] O. Le Meur, J. Gautier, and C. Guillemot, "Examplar-based inpainting based on local geometry," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3401–3404.
[22] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1153–1165, May 2010.
[23] Z. Li, H. He, H. Tai, Z. Yin, and F. Chen, "Color-direction patch-sparsity-based image inpainting using multidirection features," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1138–1152, Mar. 2015.
[24] O. Le Meur, M. Ebdelli, and C. Guillemot, "Hierarchical super-resolution-based inpainting," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3779–3790, Oct. 2013.
[25] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 463–476, Mar. 2007.
[26] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[27] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: Image and video synthesis using graph cuts," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 277–286, 2003.
[28] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The generalized patchmatch correspondence algorithm," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 29–43.
[29] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen, "Image melding: Combining inconsistent images using patch-based synthesis," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 82:1–82:10, Jul. 2012.
[30] J. Huang, J. Kopf, N. Ahuja, and S. B. Kang, "Transformation guided image completion," in *Proc. IEEE Int. Conf. Comput. Photography*, Apr. 2013, pp. 1–9.
[31] J. B. Huang, S. B. Kang, N. Ahuja, and J. Kopf, "Image completion using planar structure guidance," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 129:1–129:10, Jul. 2014.
[32] T. Ruzic and A. Pizurica, "Context-aware patch-based image inpainting using Markov random field modeling," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 444–456, Nov. 2014.
[33] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Secaucus, NJ, USA: Springer-Verlag, 2001.
[34] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[35] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, vol. 110, no. 3, pp. 404–417.
[36] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, pp. 313–318, Jul. 2003.

Jiaying Liu (S'08–M'10–SM'17) received the B.E. degree in computer science from the Northwestern Polytechnic University, Xian, China, in 2005, and the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2010.

She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored more than 100 technical articles in refereed journals and proceedings and holds 24 granted patents. Her current research interests include image/video processing, compression, and computer vision. She was a Visiting Scholar with the University of Southern California, Los Angeles, CA, USA, from 2007 to 2008. In 2015, she was a Visiting Researcher with Microsoft Research Asia, supported by Star Track for Young Faculties.

Dr. Liu served as a TC member in the IEEE CAS MSA and APSIPA IVM, and a APSIPA Distinguished Lecturer from 2016 to 2017. She is a Senior Member of CCF.



Shuai Yang received the B.S. degree in computer science from Peking University, Beijing, China, in 2015. He is currently working toward the Ph.D. degree at the Institute of Computer Science and Technology, Peking University, Beijing, China.

His current research interests include image inpainting, depth map enhancement, and image stylization.

Yuming Fang (M'13–SM'17) received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from the Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore. Currently, he is a Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, 3-D image/video processing, etc. He serves as an Associate Editor of the IEEE ACCESS and is on the editorial board of *Signal Processing: Image Communication*.

Zongming Guo (M'09) received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively.

He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication.

Dr. Guo is an Executive Member of the China Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, and the Wang Xuan News Technology Award and the Chia Tai Teaching Award in 2008. He received the Government Allowance granted by the State Council in 2009. He received the Distinguished Doctoral Dissertation Advisor Award from Peking University in 2012 and 2013.