STRUCTURE-GUIDED IMAGE COMPLETION VIA REGULARITY STATISTICS

Shuai Yang[1], *Jiaying Liu*[1*], *Sijie Song*[1], *Mading Li*[1] *and Zongming Guo*[1,2]

[1]Institute of Computer Science and Technology, Peking University, Beijing, China
[2]Cooperative Medianet Innovation Center, Shanghai, China

ABSTRACT

In this paper, we propose a novel hierarchical image completion approach using regularity statistics, considering structure features. Guided by dominant structures, the target image is used to generate reference images in a self-reproductive way by image data enhancement. The structure-guided image data enhancement allows us to expand the search space for samples. A Markov Random Field model is used to guide the enhanced image data combination to globally reconstruct the target image. For lower computational complexity and more accurate structure estimation, a hierarchical process is implemented. Experiments demonstrate the effectiveness of our method comparing to several state-of-the-art image completion techniques.

Index Terms— Image completion, structure detection, perspective transformation, image inpainting

1. INTRODUCTION

Image completion or image inpainting aims to fill the missing parts of an image and make the reconstructed image look natural. This important topic in image processing gains attentions with the popularity of digital life. And it is widely used in image editing applications such as watermark removal, panorama generation and cultural heritage restoration.

In the literature, image completion methods can be classified into two main categories. The first category is diffusion-based which propagates structures into the missing region. Bertalmio *et al.* [1] take use of the geometric and photometric information and propagate Laplacian descriptors along the isophote direction. The main defect of diffusion-based methods is the blurring artifacts when missing regions is large .

The second category concerns examplar-based methods. The main idea is to sample the pixels/patches in the known parts of the image and copy them to the missing region. According to inpainting strategies, exemplar-based methods can be categorized as greedy methods [2, 3] and global methods [4, 5, 6, 7, 8]. Greedy methods each time fill one pixel/patch by searching the best matches as samples and iteratively complete the missing region while global methods fill all missing pixels simultaneously by optimizing energy functions.

∗Corresponding author
This work was supported by National High-tech Technology R&D Program (863 Program) of China under Grant 2014AA015205, National Natural Science Foundation of China under contract No. 61472011 and Beijing Natural Science Foundation under contract No.4142021.

Among state-of-the-art global methods, Markov Random Field (MRF) model is widely used to build the energy function, for it is effective to realize global image consistency by defining the relationship of local adjacent pixels. In [5], candidate samples are searched throughout the whole image, leading to considerable processing time. To constrain the search space, Ruzic *et al.* [9] divide the image into several regions based on their context and search candidate samples in similar regions. Meanwhile, He and Sun [8] limit the search space to only 60 candidates using the statistics of patch offsets, obtaining gains in both speed and quality. In all of the works above, the basic operation is pixel/patch translations. However, broken structures cannot be restored by simply shifting the known pixels/patches into the unknown regions when image scenes contain complex transformations. Unfortunately, transformation operations other than translation are seldom available in MRF-based methods.

In fact, patch transformations such as rotation, scaling [7] and perspective transform [10] have been taken into account in coherence-based methods. The main problem for this kind of method is that more constraints need to be enforced when increasing the degrees of freedom or the result would fall into local optimum and suffer structure distortions.

In this paper, we use the MRF model to better realize global image consistency. At the same time, the search space is enriched by uniform structure-guided image data enhancement, without giving too much degrees of freedom, thus leading to fewer structure distortions. Specifically, dominant structure are extracted based on patch regularity and used as guidance to generate the enhanced images for reference. To combine the information of multiple reference images, we propose a hierarchical MRF-based image completion method using regularity statistics. Finally, we validate our method by comparing with state-of-the-art image completion algorithms on both man-made scenes and natural scenes.

The rest of this paper is organized as follows: Section 2 describes the proposed image completion approach. Experimental results are shown in Section 3 and concluding remarks are given in Section 4.

2. STRUCTURE-GUIDED IMAGE COMPLETION VIA REGULARITY STATISTICS

In this section, the proposed image completion method is presented. Given the target image I, its missing part is denoted

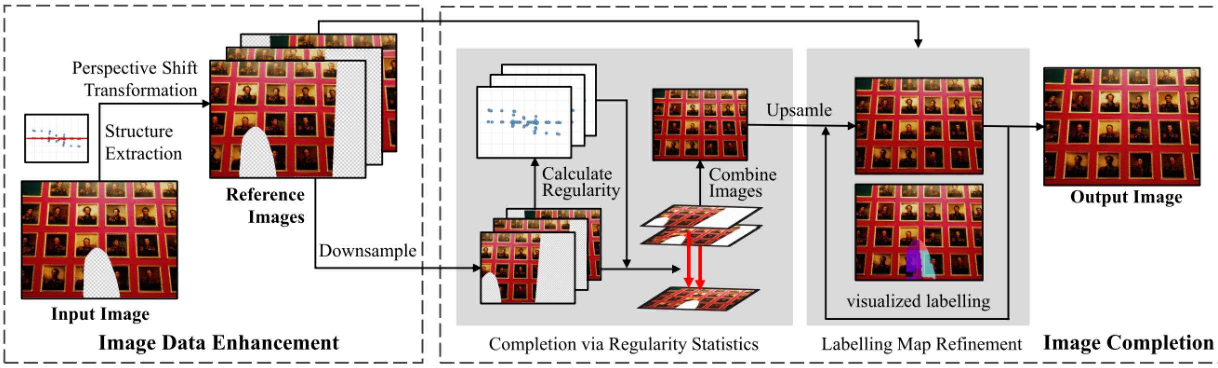


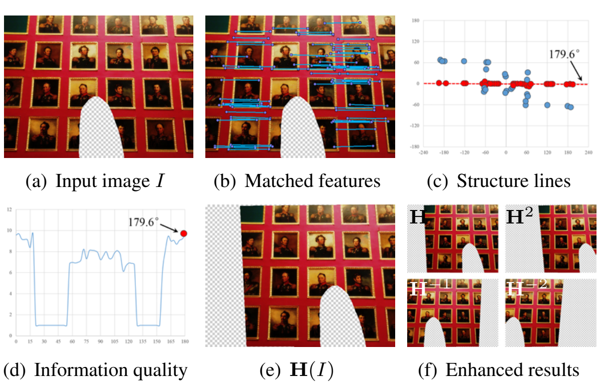
Fig. 1. Flow chart of the proposed hierarchical structure-guided image completion via regularity statistics.

as Ω, and its contour is indicated by $\delta\Omega$. The source regions, on the other hand, are denoted as $\Phi = I - \Omega$. Our goal is to fill Ω seamlessly using the information of Φ. Fig. 1 shows the main procedure of the proposed method.

2.1. Structure-Guided Image Data Enhancement

2.1.1. Dominant Structure Line Detection

Since human eyes are sensitive to structure consistency, by detecting and preserving linear structures, the completion quality can be well improved. Along structure lines, image patches demonstrate high regularity and this can be a significant cue on the determination of the lines. Inspired by [8], we detect the regularity using patch offsets. The frequency of the matched patches' relative spacial offsets is calculated. The most frequent ones form a set of dominant offsets and are allocated to unknown pixels for completion. The offsets extraction will be discussed in detail in Section 2.2. To take a step further, we analyse the dominant offsets which are likely distributed along dominant structure lines in the offset space, as shown in Fig. 2(c). We use a RANSAC-based voting approach [11] to detect the best fitting line as a dominant structure line. We repeat the RANSAC process over the outliers to search multiple dominant structure lines (the red line in Fig. 2(c)) until the number of inliers is less than a given threshold.



(a) Input image I (b) Matched features (c) Structure lines

(d) Information quality (e) $\mathbf{H}(I)$ (f) Enhanced results

Fig. 2. Structure-guided image data enhancement.

2.1.2. Perspective Shift Transformation

The ubiquitous foreshortening effects make the results of MRF-based image completion methods degrade severely, for they only perform translation operation. We put forward the concept of *Perspective Shift* in addition to traditional translation. Objects are shifted in a way that satisfies foreshortening effects. To accomplish this task, we estimate a homography matrix that performs an image registration transformation.

We begin with Speeded Up Robust Features (SURF, [12]) points detection and compute SURF descriptors for each feature point $\mathbf{k}$. Then, these feature points are matched (as shown in Fig. 2(b)) under two spacial constraints. Guided by the dominant structure line l, $\mathbf{k}_i$ and $\mathbf{k}_j$ are matched if their vector $\overrightarrow{\mathbf{k}_i\mathbf{k}_j}$ satisfies the *distance constraint* $\lambda_{min} < |\overrightarrow{\mathbf{k}_i\mathbf{k}_j}| < \lambda_{max}$ and the *angle constraint* $d_\pi(\overrightarrow{\mathbf{k}_i\mathbf{k}_j}, l) < \lambda_\theta$, where $d_\pi(\cdot,\cdot)$ is the included angle of two lines. λ_{min} ensures no feature point is matched with itself and λ_{max} is considered based on the idea of local similarity.

Then we perform a RANSAC-based voting algorithm over the matched feature points to find the best fitting transformation matrix $\mathbf{H}$. We repeat the RANSAC process over the outliers to obtain multiple perspective shift transformations. To find the optimal one, we define two measurements:

Information quantity. We use the percentage of the perspectively shifted known information in the missing regions to measure the information quantity:

$$R_{\text{quantity}}(\mathbf{H}) = |\mathbf{H}(\Phi) \cap \Omega| / |\Omega|, \tag{1}$$

where $\mathbf{H}(\cdot)$ is the perspective shift operation and $|\Omega|$ is the pixel number in the source region Ω.

Information quality. Textures after an ideal perspective shift operation should match those in the original image. We concentrate on the outer boundary of Ω with a width of λ_{max} pixels (denoted as $\Delta\Omega$) and define the information quality as:

$$R_{\text{quality}}(\mathbf{H}) = \sqrt{|\Delta\Omega| / \sum_{\mathbf{p}\in\Delta\Omega} (I(\mathbf{p}) - (\mathbf{H}(I))(\mathbf{p}))^2}. \tag{2}$$

Next, the optimal $\hat{\mathbf{H}}$ is obtained by solving:

$$\max_{\mathbf{H}} R_{\text{quality}}(\mathbf{H}) \text{ s.t. } R_{\text{quantity}}(\mathbf{H}) > \lambda_{\text{quantity}}. \quad (3)$$

To demonstrate how dominant structure lines guide the transformation estimation, we enumerate lines of different angles to impose the angle constraint and obtain corresponding $\hat{\mathbf{H}}$. The information quality for each $\hat{\mathbf{H}}$ is calculated and Fig. 2(d) shows how angle restriction affects the information quality, which validates that dominant structure lines are reliable to guide the algorithm to find $\mathbf{H}$ with relatively high $R_{\text{quality}}(\mathbf{H})$.

Moreover, if a certain image registration transformation exists, $\mathbf{H}^i$ could be possibly valid perspective shift transformation matrices as well (as shown in Fig. 2(f)). Intuitively, $\mathbf{H}^2$ represents a double perspective shift operation and $\mathbf{H}^{-1}$ represents an inverse perspective shift operation. We enrich the reference images with $\mathbf{H}^i(I),\ |i| \in \{1, 2, 3, 4\}$.

2.2. Global Completion via Regularity Statistics

In this section, we introduce our MRF-based image completion method. The MRF-based algorithms [5, 8] treat image completion as a labelling problem. In our work, we adopt the same idea of patch offset statistics as [8], which benefits from better texture and structure preservation. The target image I_0 and its enhanced results form a set of reference images denoted as $S_I = \{I_0, ..., I_W\}$. We denote patch offsets as $\mathbf{s} = (u, v, w)$, where (u, v) is the coordinates of the patch offsets and $w \in \{0, 1, ..., W\}$ indicates the w-th reference image. We match similar patches and calculate their offsets by:

$$\mathbf{s} = \arg\min_{\mathbf{s}} \|\Psi(\mathbf{p}+\mathbf{s}) - \Psi(\mathbf{p})\|_2^2, \quad (4)$$

where $\mathbf{p} = (x, y, w)$ is the position of a patch and $\Psi(\mathbf{p})$ is the patch centered at (x, y) in I_w. We argue that offsets near the dominant structure lines contribute more to the completion process, and optimize the offset histogram within I_0 by giving higher weights to those offsets. Given the statistics of all patch offsets from I_0 to I_i, we pick out the most K_i frequent ones and finally acquire a total number of $K = \sum_{i=0}^{W} K_i$ dominant offsets denoted as $S_{\mathbf{s}} = \{\mathbf{s}_i\}(i = 1, ..., K)$. Then image completion is realized by seeking the optimal labelling $L(\mathbf{p}) = i \in \{1, ..., K\}$ and copying the pixel value at $\mathbf{p} + \mathbf{s}_i$ to the pixel at $\mathbf{p}$.

2.2.1. Image Completion Model Based on MRF Prior

In this section, we describe our MRF energy function. Compared to the definition of [8], image gradient are taken into account to better preserve structure. Moreover, we reinforce the boundary treatment by using patch difference rather than pixel difference to accomplish better boundary consistency.

Given K dominant offsets, we define the energy function to evaluate the labelling:

$$E(L) = \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{N}_4} E(L(\mathbf{p}), L(\mathbf{q})) + \alpha \sum_{\mathbf{p}\in\Omega} E_d(L(\mathbf{q})), \quad (5)$$

where $L(\mathbf{x}) = i$ is the labelling that assigns the i-th dominant offset to the pixel at $\mathbf{p}$, $\mathcal{N}_4$ is the neighborhood system, $\alpha = 2$ is the weight to combine two energy terms:

Smoothness term: $E(L(\mathbf{p}), L(\mathbf{q}))$ penalizes the discontinuity within nearby pixels. It is defined as (for simplicity, we denote $i = L(\mathbf{p}), j = L(\mathbf{q})$):

$$\begin{aligned} E(i,j) = &\|I(\mathbf{p}+\mathbf{s}_i) - I(\mathbf{p}+\mathbf{s}_j)\|_1 \\ &+ \beta\|\nabla I(\mathbf{p}+\mathbf{s}_i) - \nabla I(\mathbf{p}+\mathbf{s}_j)\|_1, \end{aligned} \quad (6)$$

where ∇I is the magnitude of the image gradient, $\beta = 2$ is the weight to combine intensity and gradient terms.

Data term: $E_d(L(\mathbf{p}))$ is defined as:

$$E_d(i) = \begin{cases} +\infty, \text{if } \mathbf{p}+\mathbf{s}_i \notin \Phi \\ 0, \text{if } \mathbf{p}+\mathbf{s}_i \in \Phi \text{ and } \mathbf{x} \in \Omega\backslash\delta\Omega \\ d(\Psi(\mathbf{p}+\mathbf{s}_i), \Psi(\mathbf{x})), \text{other} \end{cases}, \quad (7)$$

where $d(\Psi(\mathbf{p}), \Psi(\mathbf{q}))$ is the patch difference measuring the consistency along the boundary between Ω and Φ:

$$\begin{aligned} d(\Psi(\mathbf{p}), \Psi(\mathbf{q})) = &\|G \otimes (\Psi(\mathbf{p}) - \Psi(\mathbf{q}))\|_1 \\ &+ \beta\|G \otimes (\nabla\Psi(\mathbf{p}) - \nabla\Psi(\mathbf{q}))\|_1, \end{aligned} \quad (8)$$

where G is the Gaussian weighting matrix and $\otimes$ is the pointwise product operator. Only the known pixels ($\Psi \cap \Phi$) are calculated.

Once the MRF graph is given, the energy optimization can be achieved using multi-label graph-cuts algorithm.

2.2.2. Hierarchical Implementation

We propose a hierarchical implementation for low computational complexity. Furthermore, the algorithm would be less sensitive to noises and local singularities, thus dominant offsets would be more reliable to demonstrate image regularity. To be specific, the target image is downsampled and completed using the proposed method to obtain a labelling map $\mathbf{s}(\mathbf{p}) = \mathbf{s}_i(\text{if } L(\mathbf{p}) = i)$ which describes the offset assignments. Then we multiply dominant offsets by two and upsample the labelling map using the nearest interpolation. To correct small misalignments, we propose an outside-in offset refinement algorithm based on pixel priority. The pixel priority is calculated according to [2]. The labelling map of the pixel with the highest priority is refined:

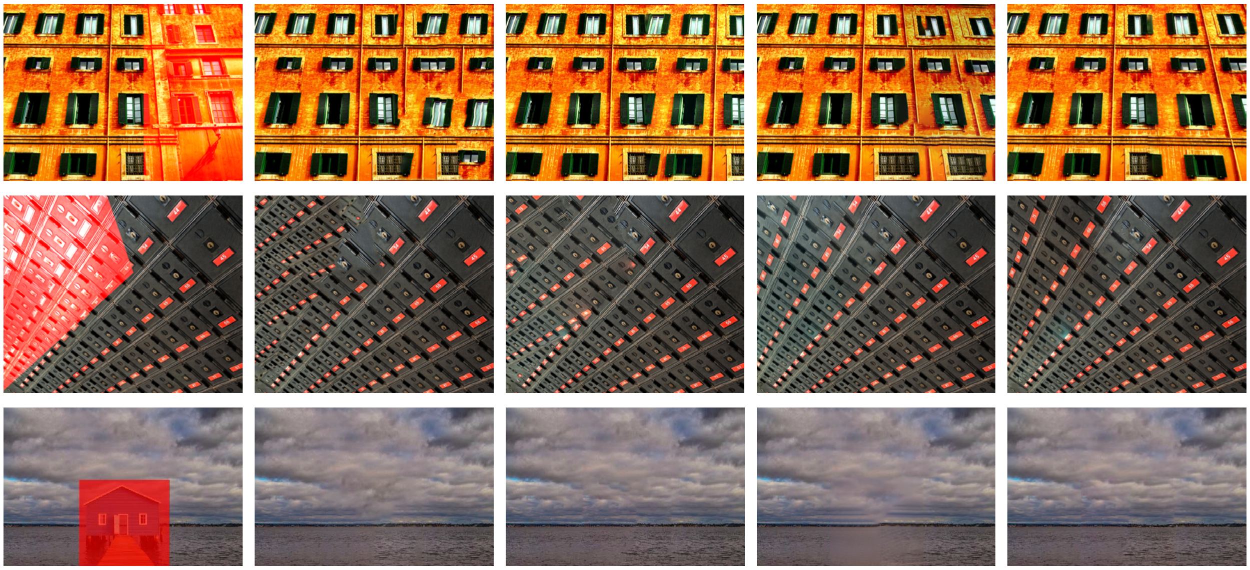
$$\widehat{\mathbf{s}}(\mathbf{p}) = \arg\min_{\mathbf{s}\in\mathcal{N}_5(\mathbf{s}(\mathbf{p}))} d(\Psi(\mathbf{p}+\mathbf{s}), \Psi(\mathbf{p})), \quad (9)$$

where $\mathcal{N}_5(\mathbf{s}) = \mathcal{N}_4(\mathbf{s}) \cup \{\mathbf{s}\}$. At each iteration of refining, the labelling map changes by at most one pixel but the final adjustment can be large thanks to the hierarchical process. Fig. 3 demonstrates that misalignments are corrected.

In the end, Poisson fusion [13] is used to hide seams.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method is implemented on Visual Studio 2013 platform. In the experiment, we set $\lambda_{min} = 10, \lambda_{max} = 200, \lambda_{\pi} = \pi/8, \lambda_{\text{quantity}} = 0.2$. The number of dominant offsets are $K_0 = 60, K_i = 10(i \in \{1, ..., W\})$. Our image completion approach is tested on varied images of man-made scenes[1] and natural/semi-natural scenes[2]. The results are



(a) Input image (b) Photoshop (c) He[8] (d) Huang[10] (e) Our result

Fig. 4. Comparison with state-of-the-art methods.

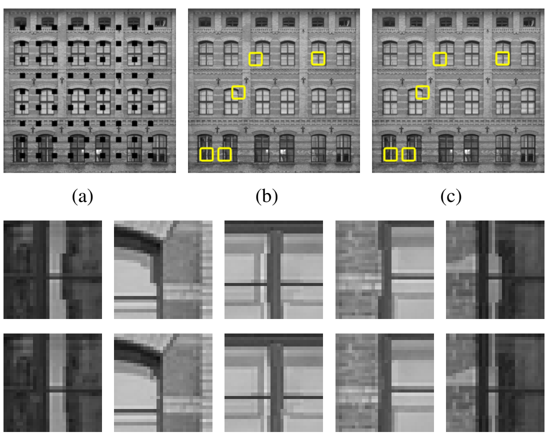
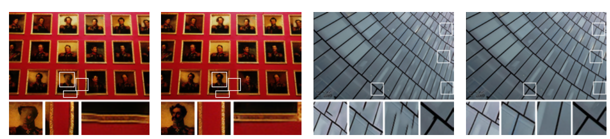


Fig. 3. The misalignments are corrected by offset map refinement. The first row: (a) original image, (b) completion result before refining and (c) after refining. The second row: patches with misalignments in (b). The third row: corresponding refined patches in (c).

compared to state-of-the-art image completion methods. The whole pictures and more experimental results can be found on our website[3].

We compare our approach with Photoshop Content Aware Fill [4, 6], Offset-Based method [8] and Planar Structure Guidance method [10]. Fig. 4 shows the results. In the man-made scenes, the structures crack in Photoshop's and He's results, for both methods search patches in only translation transformation space. Compared to Photoshop and He's method, our approach enhances the target image data to allow for a broader perspective transformation search space and suffers fewer artifacts. Meanwhile, Huang's method allows for a search space of more degrees of freedom and main structures are preserved. As shown in Fig. 5, compared to Huang's results, our approach suffers fewer distortions thanks to the perspective shift. In the semi-natural scenes, Huang's results suffer blurring artifacts. Our results owns better visual quality, which demonstrates the superiority of the proposed method.



(a) Huang[10] (b) Our result (c) Huang[10] (d) Our result

Fig. 5. Comparisons with Huang's work for the local images. Our approach suffers less structure line distortions

4. CONCLUSION

Given a target image with missing regions, the dominant structure lines of it is detected and used to guide the image data enhancement to obtain several transformed versions of the target image in a self-reproductive way. These enhanced images are combined to reconstruct the target image using the proposed regularity-statistics-based approach. The hierarchical implementation accelerates the algorithm and works for more robust structure feature detection. We validate the effectiveness of our method by comparisons with state-of-the-art image completion methods.

[1] https://sites.google.com/site/jbhuang0604/publications/struct_completion

[2] http://people.irisa.fr/Olivier.Le_Meur/publi/2013_TIP/index.html

[3] http://www.icst.pku.edu.cn/course/icb/Projects/HIC3D.html

5. REFERENCES

[1] M. Bertalio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Computer Graphics Proceedings*. ACM SIGGRAPH, 2000, pp. 417 – 424.

[2] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, pp. 1200 – 1212, September 2004.

[3] O. Le Meur, M. Ebdelli, and C. Guillemot, "Hierarchical super-resolution-based inpainting," *IEEE Transactions on Image Processing*, vol. 22, pp. 3779 – 3790, October 2013.

[4] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 463–476, March 2007.

[5] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2649 – 2661, November 2007.

[6] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 341–352, August 2009.

[7] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen, "Image melding: Combining inconsistent images using patch-based synthesis," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 82:1–82:10, July 2012.

[8] K. He and J. Sun, "Image completion approaches using the statistics of similar patches," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 2423 – 2435, December 2014.

[9] T. Ruzic and A. Pizurica, "Context-aware patch-based image inpainting using markov random field modeling," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 444 – 456, November 2014.

[10] J. B. Huang, S. B. Kang, N. Ahuja, and J. Kopf, "Image completion using planar structure guidance," *ACM Transactions on Graphics*, vol. 33, no. 4, pp. 129:1–129:10, July 2014.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus:a paradigm for model fitting with appfication to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[12] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *Proc. European Conf. Computer Vision*, vol. 110, no. 3, pp. 404–417, 2006.

[13] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Transactions on Graphics*, vol. 22, pp. 313 – 318, July 2003.