

# Virtual Inpainting for Dazu Rock Carvings Based on a Sample Dataset

HAIYAN WANG, Sichuan Fine Arts Institute, China

ZHONGSHI HE and DINGDING CHEN, Chongqing University, China

YONGWEN HUANG, Chongqing University of Science and Technology, China

YIMAN HE, Chongqing University, China

---

Numerous image inpainting algorithms are guided by a basic assumption that the known region in the original image itself can provide sufficient prior information for the guess recovery of the unknown part, which is not often the case in actual art-image inpainting. Sometimes, the art image that needs to be inpainted is so badly damaged that there is little prior information to serve as a good model to infer the appearance of the unknown fragment. Focusing on the lookup strategy for optimal patches, a novel semi-automatic exemplar-based inpainting framework based on a sample dataset is proposed in this article to solve such a problem with three steps: (1) reference images selection from the dataset using deep convolutional network, (2) sample image creation based on reference images with melding algorithm, and (3) exemplar-based inpainting according to the created sample image. Several comparative experiments over Dazu Rock Carvings with the state-of-the-art image completion approaches demonstrate the effectiveness of our contributions. First, the search space for candidate patches is extended from the known region to a sample image. It performs effectively for the inpainting case of little prior information existing in the original image itself. Furthermore, sample image creation is added to reduce the complexity of inpainting via multiple images and avoid the taboo of complete duplication in art restoration. Moreover, Poisson blending is used for post-procedure to improve the visual harmony between the reconstructed fragment and the known region in both color and illumination. Last but not least, our method is successfully applied in the virtual inpainting of Dazu Buddhist face images. The inpainted proposals can be a reference for the final actual artificial inpainting as well as a base for VR show.

CCS Concepts: • Applied computing → *Fine arts*;

Additional Key Words and Phrases: Exemplar-based inpainting, sample image creation, search space extension, Dazu Rock Carvings, image melding

**ACM Reference format:**

Haiyan Wang, Zhongshi He, Dingding Chen, Yongwen Huang, and Yiman He. 2019. Virtual Inpainting for Dazu Rock Carvings Based on a Sample Dataset. *ACM J. Comput. Cult. Herit.* 12, 3, Article 20 (June 2019), 17 pages.

<https://doi.org/10.1145/3303767>

---

This work was supported by Sichuan Fine Arts Institute Doctor Research Fund [Grant No. 18BS002] and Chongqing Graduate Research and Innovation Fund [Grant No. CYS17023].

Authors' addresses: H. Wang, Department of Art History, Sichuan Fine Arts Institute, 401331 Chongqing, China; email: why@scfai.edu.cn; Z. He and D. Chen, College of Computer Science, Chongqing University, 400044 Chongqing, China; emails: {zshe, dingding}@cqu.edu.cn; Y. Huang, School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, 401331 Chongqing, China; email: hyw@cqust.edu.cn; Y. He, Faculty of Architecture and Urban Planning, Chongqing University, 400044 Chongqing, China; email: 20144800@cqu.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2019 Association for Computing Machinery.

1556-4673/2019/06-ART20 \$15.00

<https://doi.org/10.1145/3303767>

## 1 INTRODUCTION AND MOTIVATION

Inspired by art restoration, Bertalmio et al. [1] first introduced the terminology of image inpainting to computer science in 2000. It refers to a guess recovery of the missing area in an image according to the prior information extracted from its known region. The aim is to make the recovery part look visually seamless and plausible, so an observer who is not familiar with the original image cannot detect the modifications. Various inpainting algorithms have been proposed since then. Guillemot and Le Meur [2] presented a comprehensive overview of the latest research, and the approaches can be roughly divided into two categories: diffusion-based inpainting and exemplar-based inpainting. Focusing on the structural components of an image, the former category gradually propagates local structures from the exterior to the interior of the losing area via partial differential equations or other parametric models relying on image smoothness priors. Total variation inpainting model has been applied in References [3–5]. The curvature-driven diffusion model [6] and the Euler elastic functional model [7] are introduced to consider the curve structures. Inherited by texture synthesis, the latter composites similar patches from the known part of the original image itself to reconstruct the unknown region according to image self-similarity priors. The known and unknown parts are assumed to share the same statistical distribution. Experimental results show that diffusion-based methods are smart in the recovery of the consistency of structure lines but undesirable in large-scale texture reconstruction, while exemplar-based approaches show a superior performance in the recovery of large-scale texture but less performance in the global structure reconstruction. Researchers paid more attention and effort to diffusion-based inpainting since 2000. After the seminal work of Criminisi et al. in 2003 [8] and the improvement in 2004 [9], exemplar-based inpainting has been actively explored in recent literature for its more potential practical use.

Criminisi's algorithm ensures consistency of strong structures as well as large texture with three steps: (1) computing the filling priority of each patch and bias towards the patches on the edge of the textures, (2) searching for the best similar patch, and (3) reconstructing the mask with data extracted from the matching patch. Buyssens et al. [10] presented a technical review of exemplar-based image inpainting. Various improvements have been proposed, including priority calculation [11–13], lookup strategy [14–18], and similarity metrics [19–20].

A minor change in the priority of filling order could lead to significant change in the final inpainted results. Instead of a gradient-based data term in the Criminisi algorithm, a tensor-based data term was proposed in Reference [11]. Xu et al. [12] used a sparsity-based data term and Martinez-Noriega et al. [13] amplified the data term in a nonlinear fashion. Furthermore, the existing lookup strategies mainly concentrate on the search speed. A simple but effective solution is to reduce the search space, such as using a search window with a defined size. Another solution is to compute an approximate nearest neighbor (ANN) instead of an exact nearest neighbor (NN) in a full search. References [14–16] exploited k-dimensional trees to compute such an ANN. PatchMatch [17] changed patch search to patch offsets search, and the ANN was initialized randomly, then refined iteratively. Its random search step was replaced with a hashing scheme in Reference [18] to improve the NN search. Moreover, similarity metrics plays an important role in searching for matching patches. The most widely used metric is the sum of square differences (SSD). However, a weighted Bhattacharya distance was proposed in Reference [19], and its variant was presented in Reference [20], leading the SSD bias towards uniform regions.

Though the terminology of image inpainting was derived from art restoration, the state-of-the-art research mainly involved test restoration (digital synthetic damage in an image), seldom involved in the practice of real art restoration. In fact, it is very necessary to introduce image inpainting techniques to virtual art restoration and heritage conservation, because traditional artificial inpainting not only wastes time and labor, but also likely puts precious cultural heritage at risk of secondary damage.

At the same time, “all methods are guided by the assumption that pixels in the known and unknown parts of the image share the same statistical properties or geometrical structures. This assumption translates into different local or global priors” [2]. For experimental images in most literature, the losing area is surrounded by a relatively large, simple or regular background or foreground, so the remainder of the original image itself can



(a) the red part is the area that needs to be inpainted. Courtesy of Reference [10].      (b) the whole face needs to be inpainted.

Fig. 1. Images that need to be inpainted.

Table 1. Sample Dataset

Number of the cave\niche	Total images
Beishan No.136, Cave of Prayer Wheel, A.D. 1142–1146.	258
Beishan No.180, Cave of Thirteen Incarnations of Avalokitesvara, A.D. 1116–1122.	418
Shimenshan No.6, Cave of Ten Avalokitesvaras, A.D. 1136–1141.	608
Baodingshan No.11, Niche of Sakyamuni Entering Nirvana, A.D. 1174–1252.	152
Baodingshan No.18, Sutra of Amitabha and His Pure Land, A.D. 1174–1252.	402
Baodingshan No.29, Cave of Full Enlightenment, A.D. 1174–1252.	374
	2,212

provide sufficient prior information for inpainting, as shown in Figure 1(a). However, it is not often the case in actual image restoration. An image that needs to be inpainted might be so badly damaged that there is little prior data in the original image itself as the necessary inferring base, as shown in Figure 1(b). Even the most advanced exemplar-based methods may fail the inpainting task in such cases. Figure 9 shows the unpleasant results.

Focusing on the above two issues, we propose a semi-automatic exemplar-based algorithm based on a sample dataset for Dazu Buddhist face inpainting. The Dazu Rock Carvings are well known as exquisite Buddhist art. Most of these carvings were carved in Southern Song Dynasty (1126–1279) and have been suffering from damage. The inpainting is becoming extremely urgent. Archaeological and art history scholars agree that Buddhist images made in a specific period and location share some universal characteristics. Accordingly, it is suitable for them to execute an inpainting method via a sample dataset.

The dataset in Table 1, set up by us, is composed of 2,212 Buddhist carvings' RGB head images of six different caves or niches at three different hills. The digital images are partly provided by the Dazu Institute and mainly photographed by the authors. The head (face and crown) part of each Buddhist statue image is captured in Photoshop, with a length-width ratio of 5:8.

The rest of this article is organized as follows: In Section 2, we describe the Criminisi algorithm and some inpainting methods via multiple images. Section 3 is our workflow of exemplar-based inpainting for Dazu rock carvings based on a sample dataset. The experimental analysis and result comparison are presented in Section 4. The conclusion, limitations, and our future work are discussed in Section 5.

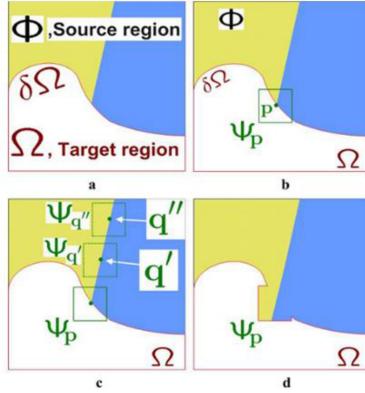


Fig. 2. Structure propagation by exemplar-based texture synthesis. Courtesy of Reference [9].

## 2 RELATED WORK

### 2.1 Criminisi Algorithm

As an outstanding exemplar-based inpainting algorithm proposed by Criminisi et al. [9], the Criminisi algorithm guides the filling process with a priority term based on edge strength. Illustrated in Figure 2, an image to be inpainted is considered as an original image  $I$  with the missing region (target region)  $\Omega$  and the known region (source region)  $\Phi$ .  $\delta\Omega$  is the contour of the target region.  $\Psi_p$  is a patch ( $9 \times 9$  pixels, the best empirical value) centered on the pixel  $p$ , which lies on the contour  $\delta\Omega$ . The optimal patch  $\Psi_q$  (with the minimum value of SSD) matching to  $\Psi_p$  is searched from candidate patches, e.g.,  $\Psi_{q'}$ ,  $\Psi_{q''}$  in  $\Phi$ . At last, the valid values from patch  $\Psi_q$  are pasted around  $p$  in  $\Omega$ . The above operation is iteratively repeated until  $|\Omega| = \phi$ .

The critical step and essential consideration in the Criminisi algorithm is the priority calculation of the filling order. Unlike the onion-peel filling order used in References [21–22], it gives higher priority to linear structures, ensuring the strong continuous structures or edges as well as the texture blocks with more known information to be inpainted first. The priority term  $P_p$  computes all pixels that  $p \in \delta\Omega$ , defined as:

$$P_p = C_p \cdot D_p, \quad (1)$$

where  $C_p$  is the confidence term and  $D_p$  is the data term.  $C_p$  has high values near the border of the initial filling region and decreases near the center of  $\Omega$ . The data term  $D_p$  favors patches in which the isophote is perpendicular to the front line at pixel  $p$  and it reflects the local image structure around  $\Omega$  and favors the continuation of structures that enter the missing region  $\Omega$ .  $C_p$  and  $D_p$  are respectively defined as:

$$\begin{cases} C_p = \frac{\sum_{q \in \Psi_p \cap (I - \Omega)} C_q}{|\Psi_p|} \\ D_p = \frac{|\nabla I_p^\perp \cdot n_p|}{\alpha} \end{cases}, \quad (2)$$

where  $|\Psi_p|$  is the size of  $\Psi_p$ , i.e., the number of pixels;  $\alpha$  is the normalization factor and can be ignored, as it is usually a constant;  $n_p$  is a unit vector orthogonal to the front  $\delta\Omega$  in the point  $p$ ;  $\nabla I_p^\perp$  is the isophote direction.

### 2.2 Search Space Extension

Some researchers noticed the limitation of search space in the exemplar-based method and challenged to extend the search space to yield some better guesses. Barnes et al. [23] extended the search space with rotations and scaling. Geometrical transformations were also added as candidates to account for the distance metrics in References [22, 24]. These methods are still limited in a single image.

Another solution went further and extended the search space to other similar images. The seminal work was Wilczkowiak et al. [25]. Based on the work of References [8, 22, 26–28], they filled holes by photomontage. Inherited from Reference [25], Hays et al. [29] first performed effective image completion via Internet-based images, but the completion results are just some acceptable proposals, because the inserted objects and textures are entirely novel and different from the original image. It is a kind of *creative inpainting* for what the original image *might be*. References [30–32] also extended the search space to Internet images, and their methods belong to *faithful inpainting* for what the original image *should be*, which is in accord with the initial aim of image inpainting. They faithfully reconstructed the mask due to the original image and the source images sharing the same physical content. Moreover, the selected source images and the original image need to be as close as possible to the same angle to ensure accuracy.

One challenge for completion via multiple images is the computing complexity. It seems appealing but is in fact difficult to apply in practice to searching matching patches directly from each image. Thus, the crucial step is the selection of source images from millions of images. The gist scene descriptor was introduced to select images depicting similar scenes in Reference [29]. The criteria of source images selection in Reference [30] was the total number of matched SIFT<sup>1</sup> key-points (at least 20) between the target image and the candidate image. Selection task in Reference [31] was completed via viewpoint invariant image search. Wang et al. [32] computed a set of reference points for the missing region and ignored the source images containing less than 8 reference feature points.

Another issue is the alignment or registration between the source images and target image. An infinity homography was introduced to register another image to the original image in References [25, 32]. Traditional template matching was applied in Reference [29] to cope with this case. Amirshahi et al. [30] employed the matched SIFT keypoints to estimate a homography matrix for global alignment and phase-only correlation for local alignment. While in Reference [31], geometric registration was achieved with multiple homographies and photometric registration was performed with a global affine transform on image intensity.

For the last problem of the consistency in color, exposure, and illumination, Poisson blending performed well to solve it in these algorithms.

### 3 THE PROPOSED ALGORITHM

Art restoration in practice is quite different from test image inpainting. It is not feasible for the traditional exemplar-based inpainting algorithms to perform the usual cases in which there is no matching patch in the rest of the original image in art restoration. In the existing completion methods via multiple images, the source image and the target image share the same image content, and the final result may tend toward direct duplication. However, it is not so lucky to have an image with the same content as the target image in art restoration. It is also a taboo to duplicate completely from a similar image, because there are no two identical statues among the Dazu Rock Carvings. Building upon the Criminisi algorithm and the effective search method based on deep convolutional network proposed in Reference [33], we propose a novel exemplar-based inpainting method via a sample dataset for Dazu Buddhist face inpainting following three stages:

*Stage 1: Similar Image Selection.* The top 10 similar images to the target image are selected as reference images from the sample dataset based on deep convolutional network;

*Stage 2: Sample Image Creation.* Sample images based on reference images are created using a melding algorithm;

*Stage 3: Image Inpainting.* Exemplar-based inpainting via the best sample image.

---

<sup>1</sup>SIFT is short for scale-invariant feature transform. It is an algorithm of extracting distinctive image features from scale-invariant key-points, proposed by D. G. David in 2004.

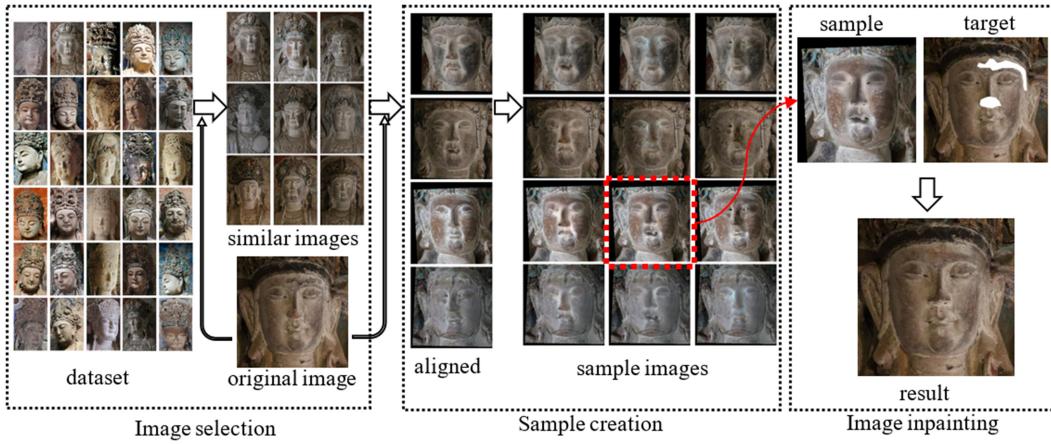


Fig. 3. Framework of exemplar-based inpainting via sample dataset.

The computing cost in Stage 3 is largely reduced by the work of Stages 1 and 2, as candidate patches are searched only in a sample image rather than in a dataset. Similar image selection can be well done with our method proposed in Reference [33]. Sample image creation is the critical work in this article, and it is also our core modification for image completion via multiple images. Image inpainting is our final task. Figure 3 shows the framework of our inpainting algorithm. It is worth mentioning that Stage 1 is automatic, whereas there is manual intervention in both Stage 2 and Stage 3.

### 3.1 Similar Image Selection

The feasibility of exemplar-based inpainting via a sample dataset is based on two facts. First, most Dazu rock carvings were engraved during the Southern Song Dynasty and thus had the unique esoteric Buddhism style. We also found that there was much similar information between the statues in the same cave or on the same subject in Reference [33]. Furthermore, Buddha statues cannot be built at will, and there are detailed descriptions for Buddhist imagery in Buddhist scriptures. The strict metrics and measurements of Buddhist image-making in China were concluded in Buddha's Theory of Imagery in Qing Dynasty. According to Buddhist scriptures, each kind of Buddhist statue has a fixed pattern, built not at liberty but with strict standards and criteria, as illustrated in Figure 4. Accordingly, there are many similarities between Buddhist images in Dazu. As shown in Figures 5(a) and 5(b), they look similar. If 5(a) is badly damaged like Figure 5(c), then the patch search space would be the similar image 5(b) rather than the remainder part of image 5(c) itself.

In this stage, we first select the top 10 similar images of the target image from the sample dataset using the effective two-step recognition method for Dazu Bodhisattva head images: feature extraction utilizing VGGNet and clustering with the K-means algorithm proposed in Reference [33]. Then manually mark labels on eyes, nose, and mouth to extract the face part of each selected image to reduce unrelated information.

### 3.2 Sample Image Creation

For the case that the damaged image itself can provide insufficient prior information, experienced art-image restorers can complete the inpainting task depending on similar artworks of the same artist or other artists who share the same ideology and art style. Hence, there might be a sample image in their mind before they start the actual work. Imitating this core idea, our work performs the inpainting task for Dazu Rock Carvings via sample images. Thanks to the stage of sample image creation, it largely reduces the computing complexity in inpainting via multiple images, because it avoids a naive patch searching from every similar image in order.

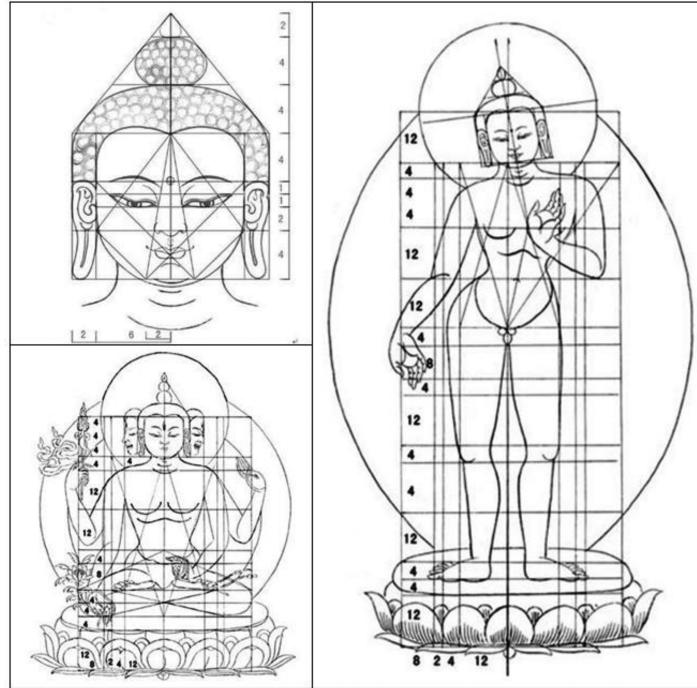


Fig. 4. Strict standards and criteria for built Buddhist statues; the figures refer to the metrics, 1 means the width of the index finger of the specific statue. Courtesy of Reference [34].



Fig. 5. Feasibility analysis.

This modification makes our algorithm distinctive and quite different from the existing completion methods via multiple images. Directly applying the most similar image as the sample image may lead the inpainting into the trap of complete duplication, which is a taboo in actual art-image inpainting. Sample image creation proposed in this article can guide the inpainted result inheriting from the reference images as well as avoiding simple duplication.

Sample images are created with three steps. First, the 10 most similar images are searched as reference images using the VGG-based method [33]. To ensure a coherent completion result, 14 feature points on eyes, nose, and mouth of each selected reference image are manually labelled and aligned with the target image. The alignment

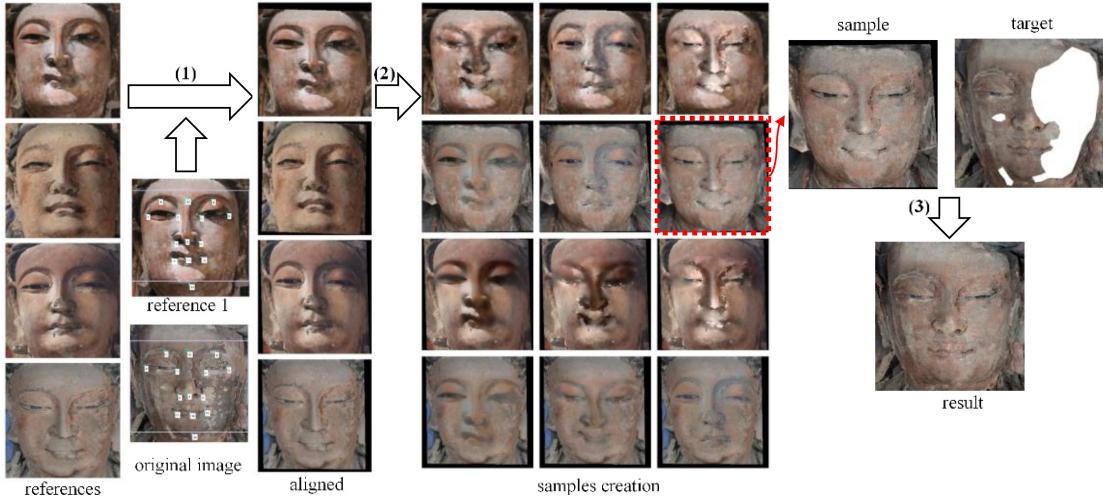


Fig. 6. Sample image creation.

between the original image and reference images is illustrated in Figure 6(1). The original image is usually the reference substance for the alignment. At last,  $C_{10}^2 = 45$  different reference images are melded in pairs to create sample images, and the best in vision is manually selected as a source image for exemplar-based inpainting, as shown in Figures 6(2) and 6(3). Usually, the sample with the least transformation and the most harmony in color, texture, and luminance is selected by art students.

For the melding step, we apply the work of Sunkavalli et al. [35] due to its impressive results. To composite two images taken from different sources or shot under different conditions, they presented a framework that explicitly matches the visual appearance of images through a process called image harmonization before blending:

*Pyramids Building.* Build pyramids from the reference image, the target image, and a uniform random noise image, respectively;

*Image Harmonization.* Modify the reference and noise pyramids to match the target pyramid;

*Image Composition.* Reconstruct the composition from the harmonized reference and noise pyramids.

Sample image creation might be particularly hard and badly constrained, e.g., the strict request of angle consistent (pairs of standard seven-eighths or three-quarter faces for harmonization experiments in Reference [35]); the algorithm does not always produce the expected result, and some created sample image may be worse (e.g., warping, blurry) than the references, as illustrated in Figure 6(2). Manual intervention is therefore necessary to choose a more plausible sample image among the 45 alternatives. Nevertheless, it is worthy for the effectiveness of art-image restoration.

### 3.3 Image Inpainting

The Criminisi algorithm is the basic part of our inpainting method. The target image  $I$  (the damaged original image) and the source image  $I_s$  (the created sample image) are loaded first. While the size of the damaged area  $|\Omega| \neq \phi$ , we compute the filling priority of the patches and select the target patch  $\Psi_p$  centered at  $p \in \delta\Omega$  with the highest  $P_p$ . Then the optimal patch that best corresponds to the unknown region is searched from  $I_s$ . Repeat the above procedures until the damaged part is full. At last, the completed region is blended with the known part of the original image to ensure the completed image  $I_c$  looking visually harmonious in color, illumination, and noise. The pseudo code of the exemplar-based inpainting via a sample image is shown below.

**ALGORITHM 1:** Exemplar-based inpainting via a sample image

---

**Input:** target/original image  $I$ , source/sample image  $I_s$   
**Output:** completed image  $I_c$

- 1 **load** images  $I$  and  $I_s$
- 2 **define** the damaged region of  $I$ :  $\Omega$
- 3 **while**  $|\Omega| \neq \phi$  **do**
- 4   compute priority  $P_p = C_p \cdot D_p, p \in \delta\Omega$  // Eq. (1)–(2)
- 5   select the target patch  $\Psi_p$  centered at  $p$  with the highest  $P_p$
- 6   search the matching patch  $\Psi_q$  from the source image  $I_s$  // Eq. (3)–(5)
- 7   copy and paste data in matching patch  $\Psi_q$  into corresponding unknown region in target patch  $\Psi_p$
- 8   update confidence  $C_p$ , frontline  $\delta\Omega$ , filling region  $\Omega$
- 9   repeat step 3–8 until  $|\Omega| = \phi$
- 10 **end**
- 11 blend the inpainted fragment with the target image using Poisson blending
- 12 **end**

---

The above algorithm is different from the Criminisi algorithm in two modifications: one is the similarity measure between patches in different sources of image, the other is Poisson blending processing for the restored image.

*Similarity metric.* The key problem in this algorithm is the similarity metric between patches in two images. A modified Bhattacharya distance proposed in Reference [19] and improved in Reference [20] is introduced to compute the similarity. It is defined as follows:

$$d_{(\text{SSD}, \text{BC})}(\Psi_p, \Psi_q) = d_{\text{SSD}}(\Psi_p, \Psi_q) \times (1 + d_{\text{BC}}(\Psi_p, \Psi_q)), \quad (3)$$

where

$$d_{\text{BC}}(\Psi_p, \Psi_q) = \sqrt{1 - \sum_{i=1}^k \sqrt{h_1(i)h_2(i)}}, \quad (4)$$

where  $h_1$  and  $h_2$  are the histograms of colors ( $L^*a^*b^*$ ) of  $\Psi_p$  and  $\Psi_q$ , respectively, and  $k$  is the number of bins of the histograms.

Meanwhile, pixel color value is incomplete for the image data in a patch, because the reference images and target image may be captured under various camera parameters. Therefore, image gradients including vertical and horizontal gradients ( $\nabla_x, \nabla_y$ ) of a patch are introduced into the patch representation due to its excellent performance in image melding in Reference [36]. Accordingly, there are five channels for every patch at each pixel, denoted as  $c = \{L, a, b, \nabla_x \text{ and } \nabla_y\}$ , which can improve the accuracy of the patch search, as it provides more complete information of a patch. Thus, the similarity metric, the sum of squared differences (SSD), is improved as:

$$d_{\text{SSD}}(\Psi_p, \Psi_q) = \sum_{n=1}^5 \sum_{v \in (N_p \cap (I - \Omega))} \left\| \Psi_p^{(n)}(v) - \Psi_q^{(n)}(v + p - q) \right\|^2, \quad (5)$$

where  $n$  represents the five channels, respectively.

*Results blending.* Because the matching patches are chosen from another image instead of the original image itself, there is inevitable visual disharmony in texture, color, illumination, contrast, noise, or blur between the known region and the restored fragments, demonstrated in our results before blending in Figure 10. We therefore present a post-process to match the aforementioned elements and avoid block artifacts.

Poisson blending introduced in Reference [37] is employed for our blending task. As an excellent gradient domain image processing method, Poisson blending can seamlessly blend an object or texture from a source image into a target image avoiding noticeable difference by guided interpolation with a guidance vector field  $\mathbf{v}$ . The guidance vector field  $\mathbf{v}$  for  $f$  (the interpolation in the damaged area  $\Omega$  of  $f^*$ , the known region of the original image) over  $\Omega$  is defined as a minimization problem:

$$\min_f \iint_{\Omega} |\nabla f - \mathbf{v}|^2, \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega}. \quad (6)$$

It is the unique solution of Poisson equation with Dirichlet boundary conditions:

$$\Delta f = \operatorname{div} \mathbf{v} \text{ over } \Omega, \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega}, \quad (7)$$

where  $\operatorname{div} \mathbf{v} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$  is the divergence of  $\mathbf{v} = (u, v)$ . Three Poisson equations are solved independently in three color channels for color images.

We execute Poisson blending after the matching patches are composited into the target image. The blended results in Figure 10 illustrate the great improvement in visual quality, as the known region and the reconstructed region look more harmonious.

## 4 EXPERIMENTAL RESULTS

We validate the effectiveness of our algorithm on a variety of experiments over Dazu Buddhist carvings' face inpainting based on the sample dataset in Table 1. All the algorithms are implemented on Matlab7.8.0 (R2009a).

To compare the proposed method to existing inpainting algorithms, we have chosen methods for which either the source code or an executable file are available. The reference methods include the basics of our proposed framework, the Criminisi algorithm [9], image completion using planar structure guidance proposed by Huang et al. [38] (which performs perfectly in recovery of repeated existing structures), and the completion algorithm using occlusion-free images from Internet photo-sharing sites proposed by Hays and Efros [29], which is mostly related to ours.

### 4.1 Comparison

**4.1.1 Digital Synthetic Damage Inpainting and Comparison.** We first validate our sample-based inpainting for two digital synthetic damage experiments. Compared to References [9], [29], and [38], our method yields the best qualitative result for Test 1, as shown in Figure 7. We find that results of References [9] and [29] disable to reconstruct the nose. The structure consistence of the nose bridge performs well but the nosewing is lost in results of Reference [9]. Almost the whole nose part is unrepaired in results of Reference [29] due to the fact that the selected similar image is not aligned with the target image, as shown in Figure 8, Test 1. Both the forehead and nose are well completed in results of Reference [38], but there are still some blurs. Whereas our blended result looks most visually pleasing, with almost no disturbing artifacts. This test shows that our method can deal with small mask and achieve impressive results.

The results of Test 2 in Figure 8 also illustrate the effectiveness of our method. Results of References [9] and [38] fail in this experiment for the loss of the eye and eyebrow. There is still unpleasant color inconsistency in our blended result compared to results of Reference [29]. It is partly attributable to the fact that Reference [29] is ideally suited for this type of problem and partly due to the huge color difference between the created sample and the target. That means it is still a challenge for color harmony in our method. Nevertheless, the result illustrates that our method can deal with the composition of different sources and achieve acceptable results compared to the related methods.

Furthermore, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are introduced to make some quantitative analysis for both Test 1 and Test 2 in Figure 8. The higher the PSNR and SSIM value are, the better the results are. The best results are marked in red boldface font.

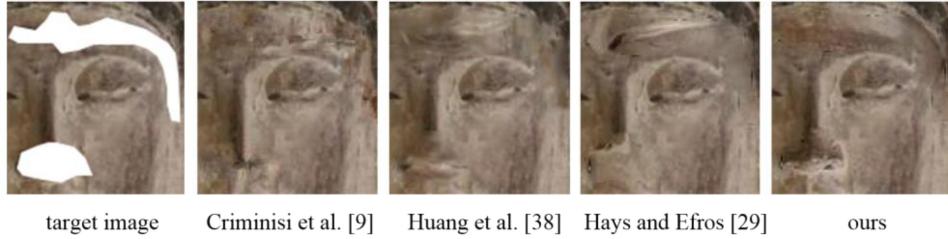


Fig. 7. Enlarged result sections of Test 1.

	original	target	Criminisi et al. [9]	Huang et al. [38]	Hays and Efros [29]	ours
<b>Test</b>						
<b>1</b>			PSNR=28.9905	PSNR=27.6512	PSNR=27.5051	<b>PSNR=29.8677</b>
			SSIM=0.9378	<b>SSIM=0.9432</b>	SSIM=0.9366	SSIM=0.9370
<b>Test</b>						
<b>2</b>			PSNR=30.4406	<b>PSNR=30.7763</b>	PSNR=29.0959	PSNR=30.1174
			SSIM=0.9266	<b>SSIM=0.9276</b>	SSIM=0.6251	SSIM=0.9003

Fig. 8. Comparison of the restored results for Test 1 and Test 2.

As shown in Test 1, compared to the other methods, our result gets the highest PSNR value, with 29.8677. The objective quantitative result is consistent with our qualitative comparison in vision. Whereas the SSIM values disaccord with the visual evaluation, for Reference [38] gets the highest SSIM value (0.9432), while ours (0.9370) is lower than References [38] and [9] (0.9378).

There is obvious inconsistency between the objective quantitative evaluation and subjective visual evaluation in Test 2. It is clearly illustrated that Huang's result gets the highest values of both PSNR and SSIM, with 30.7763 and 0.9276, respectively, even though it fails to repair the symmetry structure. However, for Hays' and ours,

which succeeds in restoring the structural part, not only the value of PSNR is a little lower than References [9] and [38], but also the SSIM values are lower. The SSIM value of Hays' result is particularly low, with only 0.6251, even though it is the best in vision.

The reason may be due to the fact that PSNR is based on the mean square error (MSE) between the corresponding pixels, and SSIM also involves luminance masking and contrast masking terms, although it emphasizes structural information error. Both References [9] and [38] search for matching patches from the known area of the original image, while Reference [29] and ours choose optimal patches from different source images instead. A similar image is directly as the searching space in Reference [29], and a created sample image is used as the source in our method, as illustrated in Figure 8. Accordingly, Reference [29] and ours produce more difference in texture, luminance, and contrast between the reconstructed image and the original image than the first two methods. The difference is not substantial when the mask is small. When the hole becomes larger, the difference is significant, such as the obvious difference of color and texture in Hays' and our results for Test 2. Therefore, PSNR and SSIM are not reliable indicators of the restored image quality.

Thus, we adopt visual evaluation and abandon quantitative comparison for the following complex or large-area damage inpainting tests in Section 4.1.2.

**4.1.2 Actual Damage Inpainting and Comparison.** Furthermore, we deal with several actual inpainting tests for large and complex damaged regions. The results of Test 3–Test 6 in Figure 9 demonstrate our method can produce the most visually pleasing result compared to the state-of-the-art methods. For example, References [9] and [38] fail to reconstruct symmetry structures for Test 3 and produce much artifacts or blurs for Test 4–Test 6. The result of Reference [29] for Test 5 is comparable to ours but there is dissymmetry completion for Test 3 and substantial blurs in the results of Test 4 and Test 6. While our method does a better job in symmetry structure inpainting (Test 3) and leads little block artifacts after blended (Tests 4 and 5). Nevertheless, there is still a little deformation in our result of Test 6.

The aforementioned experiments demonstrate that our method is more successful for the inpainting case in which image priors is insufficient in the original image. It performs better in continuing the structure and edge as well as deals with large damaged regions and achieves comparable—in many cases superior—results than the related methods. To some extent, the superiority of our method is obvious when the missing region becomes larger.

The Poisson blending processing is added following the completion of inpainting in our method. Its necessity is demonstrated in the comparison of the results of before and after blending in Figure 10. We can see that the blended results are much better than those before blending, as the known region and the repaired region look more harmonious in color, illumination, and texture.

## 4.2 Reconstruction of the Statue 5&6 in Beishan No.180

A more meaningful test for the reconstruction of Statues 5 and 6 in Beishan No.180 is performed in this subsection. They are illustrated to be constructed to substitute for the outside two damaged statues to keep the integrality of the cave [33]. The two statues seem quite different from the other eight statues and look like they are suffering from serious weathering even though they are deep in the cave. Therefore, we aim to reconstruct the two statues' faces to make them more harmonious with the other eight statues.

Figure 11 illustrates the overview of this work. Because the statues in the same cave or on the same subject are usually in the same modeling style, we select the reference images directly from the dataset of Beishan No.180 based on deep convolutional network, as shown in Figures 11(a) and 11(a'). Then, for each reference image, we make a face shot and align them with the original image, demonstrated as Figures 11(b) and 11(b'). Furthermore, Figures 11(c) and 11(c') show the sample image creation via paired melding. A sample is chosen for exemplar-based inpainting in Figures 11(d) and 11(d').

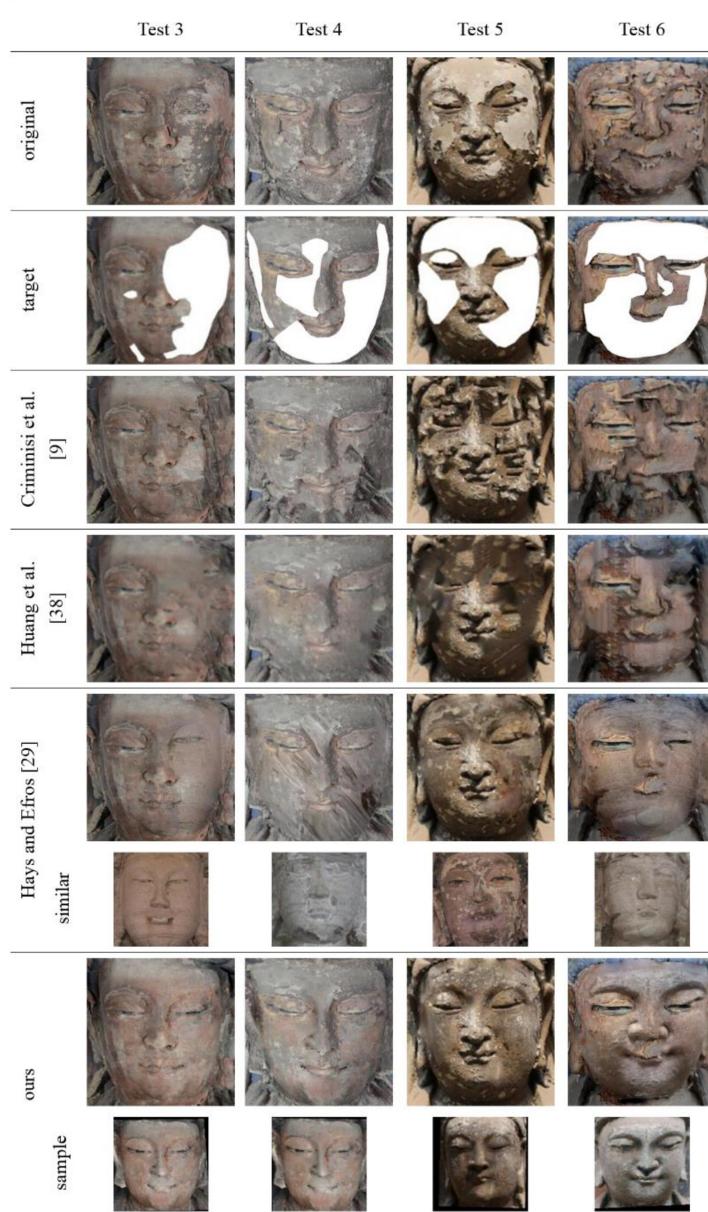


Fig. 9. Comparison of the inpainted results.

According to the comparison in Figure 12, we can see our method is absolutely more effective than the Criminisi algorithm and Huang's method in the face reconstruction for both Statues 5 and 6. Our results successfully reconstruct the eyes, nose, and mouth. While both of them fail to reconstruct these important elements. There is inconsistent completion of the wing of nose for Statue 5 in Hays and Efros' but it does better for the recovery of Statue 6. There is disharmony in color and texture with the original image in our result for Statue 6, comparable

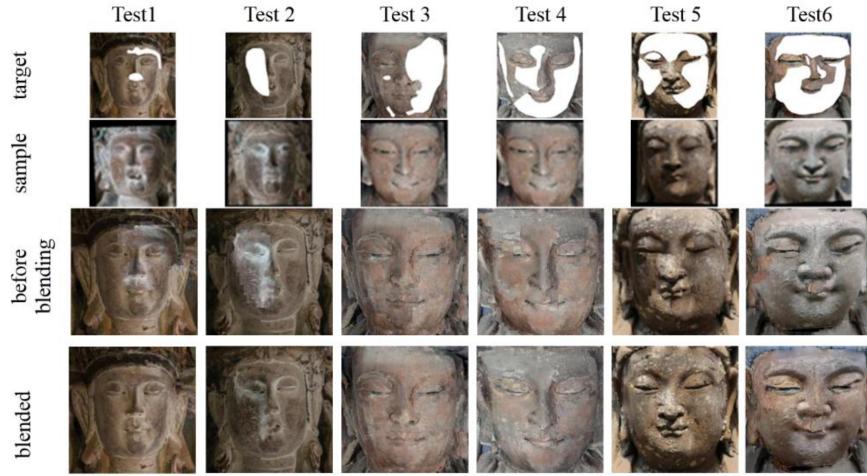


Fig. 10. Results comparison between before and after blending.

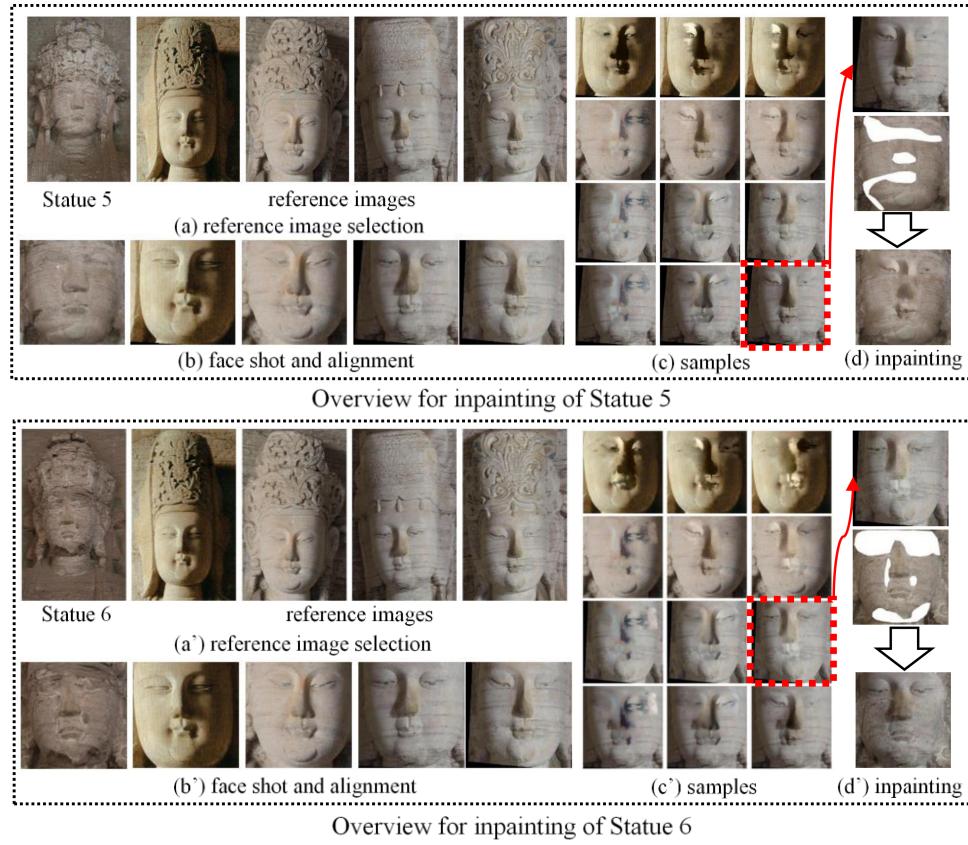


Fig. 11. Reconstruction of Statues 5 and 6 in Beishan No. 180.

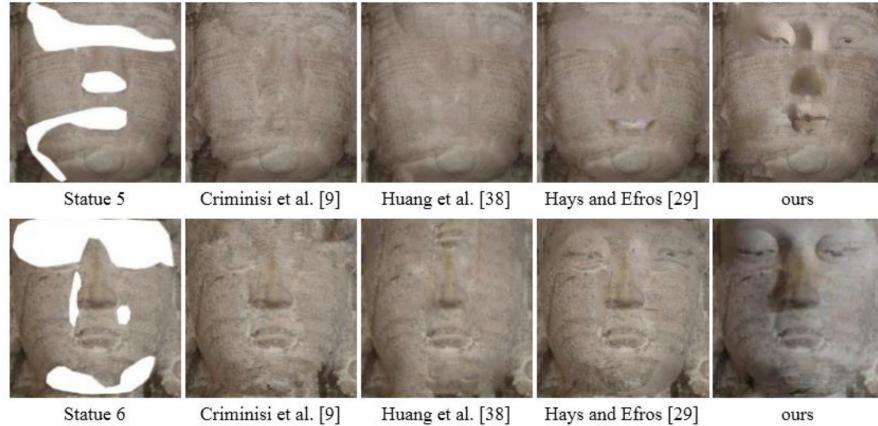


Fig. 12. Results comparison of Statues 5 and 6.

to Hays and Efros'. Nevertheless, the result is still acceptable. Although our result for Statue 5 is better than theirs, there is geometry inconsistency due to the quite different sources between the reconstructed part and the known region.

## 5 CONCLUSION

To cope with the inpainting case that the remainder of the damaged image itself provides little constraints for a plausible completion, we propose a novel semi-automatic exemplar-based inpainting framework based on a sample dataset. Various experiments on different inpainting cases of Dazu Buddhist face images demonstrate the relative merits of our algorithm compared to state-of-the-art completion techniques.

Our work has significant contributions. First, challenging the traditional patch-lookup strategy, we extend the search space from the known region of the original image itself to a sample image for an exemplar-based inpainting algorithm. Thanks to the guess-inferring based on *sample-similarity* in another image rather than the *self-similarity* in a single image, it performs effectively for the inpainting case when the original image is so badly damaged that there is insufficient prior information. Moreover, our method is successfully applied to the virtual exemplar-based inpainting for the Dazu Rock Carvings. It is very meaningful for the digital conservation of world cultural heritage. That is because our inpainting proposals can be used as reference for the actual artificial restoration to prevent the carvings from the inpainting risk of secondary damage. Meanwhile, the results can also be the basis for VR viewing of the Dazu Rock Carvings. Last but not least, the proposed inpainting method can also be exploited to solve similar problems with different art-image datasets.

Our exemplar-based inpainting via a sample dataset is doomed by high time complexity. Nevertheless, it is quite a dilemma to satisfy the constraints of both speed and quality in image restoration. For art-image restoration, it is worthy for quality assurance at the expense of time complexity. At the same time, a valid step of sample image creation is added in our searching procedure, which largely shrinks the search space from the whole dataset to one sample image. Therefore, the computing complexity is greatly reduced compared to a naive patch search all over the dataset. Above all, it avoids the inpainted results from directly duplicating the reference images.

There are still some limitations in our method for future improvement. First, sample image creation is not indeed creation, because the melding step in our work is more like a face change. Furthermore, too much manual intervention is the limitation of our method due to the strict request of angles consistent for image melding. In the

future, we would like to introduce other methods to create sample images, ensuring its validity and plausibility. We are also interested in research of the carvings' image inpainting at multiple scales and angles.

## REFERENCES

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. 2000. Image inpainting. In *Proceedings of the 27th Conference on Computer Graphics and Interactive Techniques*. 417–424.
- [2] C. Guillemot and O. Le Meur. 2014. Image inpainting: Overview and recent advances. *IEEE Signal Process. Mag.* 31, 1 (2014), 127–144.
- [3] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera. 2001. Filling in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Process.* 10, 8 (2001), 1200–1211.
- [4] T. Chan and J. Shen. 2001. Local inpainting models and TV inpainting. *SIAM J. Appl. Math.* 62, 3 (2001), 1019–1043.
- [5] A. Levin, A. Zomet, and Y. Weiss. 2003. Learning how to inpaint from global image statistics. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'03)*, Vol. 1. 305–313.
- [6] T. Chan and J. Shen. 2001. Non texture inpainting by curvature-driven diffusion. *J. Vis. Commun. Image Represent.* 12, 4 (2001), 436–449.
- [7] T. Chan, S. Kang, and J. Shen. 2003. Euler's elastica and curvature-based inpainting. *SIAM J. Appl. Math.* 63, 2 (2003), 564–592.
- [8] A. Criminisi, P. Pérez, and K. Toyama. 2003. Object removal by exemplar-based inpainting. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2. 18–20
- [9] A. Criminisi, P. Pérez, and K. Toyama. 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* 13, 9 (2004), 1200–1212.
- [10] P. Buyssens, M. Daisy, D. Tschumperlé, and O. Lezoray. 2015. Exemplar-based inpainting: Technical review and new heuristics for better geometric reconstructions. *IEEE Trans. Image Process.* 24, 6 (2015), 1809–1823.
- [11] O. Le Meur, J. Gautier, and C. Guillemot. 2011. Exemplar-based inpainting based on local geometry. In *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)*. 3401–3404.
- [12] Z. Xu and J. Sun. 2010. Image inpainting by patch propagation using patch sparsity. *IEEE Trans. Image Process.* 19, 5 (2010), 1153–1165.
- [13] R. Martinez-Noriega, A. Roumy, and G. Blanchard. 2012. Exemplar-based image inpainting: Fast priority and coherent nearest neighbor search. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP'12)*. 1–6.
- [14] Y. Wexler, E. Shechtman, and M. Irani. 2007. Space-time completion of video. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 3 (2007), 463–476.
- [15] K. He and J. Sun. 2012. Computing nearest-neighbor fields via propagation assisted KD-trees. In *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR'12)*. 111–118.
- [16] K. He and J. Sun. 2012. Statistics of patch offsets for image completion. In *Proceedings of the 12th European Conference on Computer Vision (ECCV'12)*. 16–29.
- [17] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. 2009. PatchMatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* 28, 3, Article 24 (2009), 1–11.
- [18] S. Korman and S. Avidan. 2011. Coherency sensitive hashing. In *Proceedings of the International Conference on Computer Vision (ICCV'11)*. 1607–1614.
- [19] A. Bugeau, M. Bertalmio, V. Caselles, and G. Sapiro. 2010. A comprehensive framework for image inpainting. *IEEE Trans. Image Process.* 19, 10 (2010), 2634–2645.
- [20] O. Le Meur and C. Guillemot. 2012. Super-resolution-based inpainting. In *Proceedings of the European Conference on Computer Vision (ECCV'12)*. 554–567.
- [21] R. Bornard, E. Lecan, L. Laborelli, and J.-H. Chenot. 2002. Missing data correction in still images and image sequences. In *Proceedings of the 10th ACM International Conference on Multimedia*. 355–361.
- [22] I. Drori, D. Cohen-Or, and H. Yeshurun. 2003. Fragment-based image completion. *ACM Trans. Graph.* 22, 3 (2003), 303–312.
- [23] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. 2010. The generalized PatchMatch correspondence algorithm. In *Proceedings of the European Conference on Computer Vision (ECCV'10)*, Vol. 6313. 29–43.
- [24] Y. Zhang, J. Xiao, and M. Shah. 2004. Region completion in a single image. *Eurographics*, Grenoble, France, Short Presentations.
- [25] M. Wilczkowiak, G. J. Brostow, B. Tordoff, and R. Cipolla. 2005. Hole filling through photomontage. In *Proceedings of the British Machine Vision Conference (BMVC'05)*.
- [26] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. 2004. Interactive digital photomontage. *ACM Trans. Graph.* 23, 3 (2004), 294–302.
- [27] Y. Wexler, E. Shechtman, and M. Irani. 2004. Space-time video completion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'04)*, Vol. 1. 120–127.
- [28] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick. 2003. Graphcut textures: Image and video synthesis using graph cuts. *ACM Trans. Graph.* 22, 3 (2003), 277–286.
- [29] J. Hays and A. A. Efros. 2008. Scene completion using millions of photographs. *Commun. ACM* 51, 10 (2008), 87–94.
- [30] H. Amirshahi, S. Kondo, K. Ito, and T. Aoki. 2008. An image completion algorithm using occlusion-free images from Internet photo sharing sites. *IEICE Trans. Fund. Electron. Commun. Comput. Sci.* E91-A, 10 (2008), 2918–2927.

- [31] O. Whyte, J. Sivic, and A. Zisserman. 2009. Get out of my picture! Internet-based inpainting. In *Proceedings of the British Machine Vision Conference (BMVC'09)*.
- [32] G. Wang, X. Chen, and S. Hu. 2014. Geometry-aware image completion via multiple examples. *Pacific Graphics Short Papers*. The Eurographics Association.
- [33] H. Wang, Z. He, Y. Huang, D. Chen, and Z. Zou. 2017. Bodhisattva head images modeling style recognition of Dazu Rock Carvings based on deep convolutional network. *J. Cult. Herit.* (2017).
- [34] [http://blog.sina.com.cn/s/blog\\_61b036130100wix3.html](http://blog.sina.com.cn/s/blog_61b036130100wix3.html).
- [35] K. Sunkavalli, M. K. Johnson, W. Matusik, and H. Pfister. 2010. Multi-scale image harmonization. *ACM Trans. Graph.* 29, 4 (2010), 125.
- [36] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen. 2012. Image melding: Combining inconsistent images using patch-based synthesis. *ACM Trans. Graph.* 31, 4 (2012), 82:1–82:10.
- [37] P. Pérez, M. Gangnet, and A. Blake. 2003. Poisson image editing. In *Proceedings of the ACM International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'03)*. 313–318.
- [38] J. B. Huang, S. B. Kang, N. Ahuja, and J. Kopf. 2014. Image completion using planar structure guidance. *ACM Trans. Graph.* 33, 4 (2014), 129.

Received June 2018; revised November 2018; accepted January 2019