

Structure-preserving image completion with multi-level dynamic patches

Bowen Liu¹ · Ping Li² · Bin Sheng³ · Yongwei Nie⁴ · Enhua Wu⁵

Published online: 24 October 2017
© Springer-Verlag GmbH Germany 2017

Abstract In this paper, we present a novel structure-preserving image completion approach equipped with dynamic patches. We formulate the image completion problem into an energy minimization framework that accounts for coherence within the hole and global coherence simultaneously. The completion of the hole is achieved through iterative optimizations combined with a multi-scale solution. In order to avoid abnormal structure and disordered texture, we utilize a dynamic patch system to achieve efficient structure restoration. Our dynamic patch system functions in both horizontal and vertical directions of the image pyramid. In the horizontal direction, we conduct a parallel search for multi-size patches in each pyramid level and design a competitive mechanism to select the most suitable patch. In the vertical

direction, we use large patches in higher pyramid level to maximize the structure restoration and use small patches in lower pyramid level to reduce computational workload. We test our approach on massive images with complex structure and texture. The results are visually pleasing and preserve nice structure. Apart from effective structure preservation, our approach outperforms previous state-of-the-art methods in time consumption.

Keywords Image completion · Patch-based approach · Dynamic patches · Structure preservation · Parallel search · Multi-scale solution

1 Introduction

Image completion is one of the challenging tasks in image editing. It plays an important role in impaired photograph reparation, object removal in photograph editing and disocclusion in image-based rendering. Given a deficit image, the basic goal of image completion is to fill the hole while keeping the overall visual effect realistic and harmonic. The main difficulties in the completion of task lie on structure reconstruction and texture representation, which are essential for achieving visually pleasing results.

One basic idea to accomplish image completion is to allow the information in the known region to propagate into the hole. For instance, the classical image inpainting technique by Bertalmio et al. [4] repairs small cracks on pixel level through propagating image Laplacians in the isophote direction. The pixel values in the known region diffuse into the hole according to partial differential equations (PDEs). Many diffusion-based methods work on the pixel level and perform well in local structure reservation in small gaps. However, in the task of completing large cavity, abnormal artifacts or strange structure may occur.

✉ Ping Li
pli@must.edu.mo
Bowen Liu
bowenliulbw@gmail.com
Bin Sheng
shengbin@sjtu.edu.cn
Yongwei Nie
nieyongwei@scut.edu.cn
Enhua Wu
ehwu@umac.mo

¹ Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong, Hong Kong
² Faculty of Information Technology, Macau University of Science and Technology, Macao, Macao
³ Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
⁴ School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
⁵ University of Macau and SKLCS-IOSCAS, Macao, Macao

Fig. 1 **a** The original image. Its resolution is 399×533 . **b** A hole is masked (magenta). **c** The result is generated with the state-of-the-art ImageMelding [11]. **d** Our approach

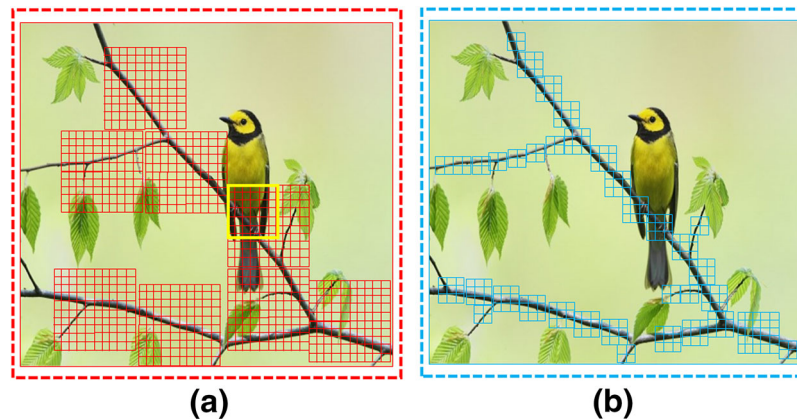
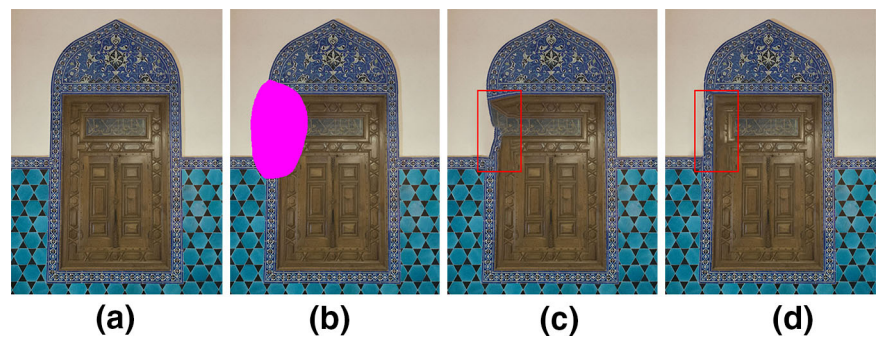


Fig. 2 Multi-size patches for structure preservation. Large patches are not always the best choice in structure preservation. **a** We try to capture the structure of the branches with 10×10 patches. Part of the bird's body and the leaves are absorbed into the patches. **b** We try to capture

the same structure with a series of patches in different sizes (2×2 , 3×3 and 5×5). The intake of irrelevant information is significantly reduced

When handling large holes, patch-based methods outperform diffusion-based techniques that work on pixel level. The seminal work of the patch-based completion by Efros and Leung [13] is under the topic of texture synthesis. In fact, texture synthesis can be regarded as an extreme situation which the cavity takes up most of the image. Due to the similarities between image completion and texture synthesis in some circumstances, many approaches such as [10, 12] successfully reproduce textures in the unknown region by amplifying texture synthesis automatically.

An important aspect of successful completion is to capture sufficient structural information and reconstruct structures smoothly within the cavity. Because the human visual system is sensitive to structural region [24]. In a lot of patch-based methods, structure restoration is conducted implicitly in the process of pursuing global coherence. These methods try to make the patches in the filled region similar to the known patches and hope to completely represent existing structure in the hole. However, disrupted structure may appear when the image contains complex structural information. As shown in Fig. 1c, even the completed image generated by the state-of-the-art technique [11] may present structural distortions in the hole. Figure 1d shows the completed result by the

approach we introduce in this paper. As can be clearly seen, the cavity is filled with nice textures while preserving the realistic structure.

In general, patch-based methods, which conduct structure restoration implicitly, tend to use large patches. The purpose of applying large patches is to capture structural information as much as possible. However, the optimization frameworks usually neglect the internal coherence of the hole. This may cause misarrangement of the patches and affect the overall completed effect. In addition, patches of large size may not be necessarily suitable for structure restoration. As observed in Fig. 2a, a large patch may absorb more irrelevant information and cause abnormal structure within the hole. Moreover, using large patch causes extra burden on hardware and higher time consumption.

In this paper, we introduce a new approach which balances the coherence between global and local to produce realistic completed image that preserves nice structure and texture. We formulate the image completion as an energy minimization problem that accounts for global and local coherence simultaneously. To achieve efficient structure restoration, we apply a dynamic patch system in our approach. Our dynamic patch system consists of two parts. The first part is the paral-

level search for multi-size patches. Different size patches enter the hole through a competitive mechanism. The second part is the multi-scale solution using multi-size patches. Large patches are applied in lower-resolution scale to maximize the structural information restoration. Small patches are used in higher-resolution scale to reduce the computational workload. Numerous results generated by our approach are shown. In sum, our approach has the following contributions:

- *Efficient structure restoration* The mixed use of different sizes of patches capture the structural information efficiently, avoiding the absorption of irrelevant information which causes abnormal structures.
- *Balanced computational workload* Multi-scale solution with dynamic patches adjusts the computational workload in the operation. It significantly reduces the computation in low pyramid level without sacrificing the visual effects and accelerates the completing process at the same time.
- *Parallel search and competitive mechanism* Parallel search for different size patches is conducted with GPU acceleration. A competitive mechanism is included to select the patch with minimum unit energy.

2 Related work

2.1 Image inpainting

Large quantities of methods and algorithms have been proposed for the task of image repairing. Bertalmio et al. [4] introduced an automatic technique that fills the user-defined region along the isophote lines without specification of the information sources. Their method has no requirements on the topology of the cavity. The information propagates according to a PDE which has roots in Navier–Stokes equation. Based on [4], researchers develop a lot of techniques through improving the mechanism of propagation and expanding the information sources. Chan and Shen [8] embedded Euler’s Elastica to deal with curves in the image and provided a numerical PDEs-based computational scheme. Levin et al. [21] conducted image inpainting with loopy belief propagation according to a specific image distribution. Ballester et al. [2] developed a variational framework for the inpainting problem. Their framework is based on joint interpolation on gray-level and gradient directions. Cai et al. [7] combined the image inpainting with the popular fully convolutional neural network. Ignácio and Jung [18] developed a block-based image inpainting technique in the wavelet domain. All the methods above are based on the pixel level and work effectively when the impaired region of the image is small and have limited structure. However, abnormal arti-

facts may occur when they are applied to large missing area or region with complex structure.

2.2 Texture synthesis and patch-based approaches

Patch-based methods which take advantages of texture synthesis techniques have also been proposed for image completion. Efros and Leung [13] introduced a nonparametric method for texture synthesis that based on the estimation of the sample image and similar neighbors. It fills the gap by pasting sample patches according to an estimation. Later research improved the search and sampling mechanism to achieve better structure preservation. Wei and Levoy [30] embedded a deterministic search process derived from Markov random field texture model and utilized tree-structured quantization for acceleration. Ashikhmin [1] developed a specific algorithm for natural images which include quasi-repeating patterns. This algorithm is fast and straightforward without changing the basic spatial frequencies. Liang et al. [22] designed a real-time sampling algorithm according to a nonparametric estimation of the local conditional MRF density function. Criminisi and Toyama [10] applied the propagation of information in inpainting into the exemplar-based method. Shen et al. [27] developed an image completion technique under the guidance of a gradient map. Hao et al. [16] proposed an image completion technique which enhanced with a mechanism that automatically captures perspective information from a single image. Hua and Wang [17] used pictures obtained from the Internet to restore structure in the image completion tasks. Iizuka et al. [19] designed a fully convolutional neural network for image completion which can maintain global and local consistency simultaneously. Liu et al. [23] used patches of different sizes in image completion for better structure preservation. Patch-based methods also have applications in other problems such as dense matching [26] and image fusion [25].

2.3 Image blending

The transition between the hole and the known region plays an important role in the overall effect of the completion. A sound completed result usually has an imperceptible edge which seamlessly connects the hole and the known region. To achieve consistent blending in image completion, researchers draw support from image blending. Burt and Adelson [6] introduced the seminal image stitching technique which combines images through pyramidal image decomposition. Shen et al. [28] designed a boosting Laplacian pyramids for multi-images exposure fusion. Sunkavalli et al. [29] applied similar multi-scale techniques with a preliminary treatment called image harmonization which transfers the appearance of an image. In 2003, Kwatra et al. [20] employed graph cut as a tool for seamless combination of images and textures. Gra-

cias et al. [15] developed a fast method to combine a set of registered images into a mosaic using watershed segmentation and graph cut optimization. Chen et al. [9] developed a system which is able to select suitable images from the Internet automatically and generate high-quality combined pictures with blending boundary optimization. Whitaker [32] proposed an image blending technique through level set comparison. A more recent technique is the ImageMelding by Darabi et al. [11]. It is designed for synthesizing transitional region between images with inconsistent color, texture and structural properties. Besides image blending, this method is versatile and successfully handle multiple tasks including image completion and object cloning. We regard ImageMelding as the state-of-the-art technique in this paper.

3 Image completion with dynamic patches

Image completion is not a simple copy–paste of the existing pixels, because images are more than collections of pixels. They all follow certain patterns. Textural and structural information is contained in these patterns which our visual system is sensitive to. One of the important aspects of image completion is to fill the hole with proper preservation of the structure and texture. To represent visually pleasing patterns in the hole, we formulated the problem into an energy minimization framework. Our goal is to fill the cavity with patches which are harmonious with both the global pictures and other patches within the hole. The challenge is to discover existing patterns in the known region and at the same time capture structural information efficiently and avoid interference of irrelevant information. Our approach is to directly optimize an objective function that respects coherence within the hole and global coherence. That is, the objective function consists of two terms: an external term and internal term. Then, the minimum of the objective function is achieved through an iterative optimization. The iterative optimization contains two phases: the search phase and the vote phase. In the search phase, we apply parallel search to retrieve different size patches. The most suitable patch is selected through a competitive mechanism. In the vote phase, all the patches are combined to fill the cavity. To make the patches seamlessly connected and avoid abnormal structure and texture, we calculate the minimum cost boundary between overlapped patches to achieve optimal seam. To accelerate the overall completing process, we apply a multi-scale solution in our approach. However, differently from previous methods, our approach uses large patches in lower-resolution scale in order to capture more structural information and provide a sound foundation for next scale.

3.1 Energy minimization

Given a color image I with a cavity C , we assume that the known region $S = I - C$ contains sufficient information for image completion. The completed image is obtained by minimizing the following objective function:

$$E(C|S) = \sum_{q \in C} \min_{p \in S} [w_1 E_1(Q, P) + w_2 E_2(Q)] \quad (1)$$

where E_1 is the energy term for external coherence and E_2 is the term for internal coherence. These two terms are described in detail in the following paragraphs. The w_1 and w_2 are the weighting factors that balance the preference between internal and external coherence. In all our experiments, we set $w_1 + w_2 = 1$. $Q = \mathcal{N}(q)$ is a $h \times h$ patch in the hole with target pixel q in its center. We set the target pixel q as the origin $Q_{0,0}$, and the other pixels in the patch are denoted as $Q_{i,j}$. Similarly, $P = f(\mathcal{N}(p))$ is a $h \times h$ patch in the known region after transformation f and p is its anchor pixel in the center. In our approach, the image is in $L * a * b$ space and all the pixels have five channels including three color channels L, a, b and two luminance gradient channels $\nabla_x L, \nabla_y L$.

Combined similarity measure The selection of metric is important in defining energy function. Most patch-based texture synthesis algorithms use the sum of squared differences (SSD) to measure the similarities between patches. However, the simple application of SSD may give preference toward uniform region [5] and cause disordered texture and abnormal structure in the completed image. Thus, we apply a modified metric based on the Bhattacharyya weighted distance function in [5]. Given any two patches, $A = \mathcal{N}(a)$ and $B = \mathcal{N}(b)$, we define the distance between them as:

$$D_m(A, B) = D_{\text{SSD}}(A, B) \cdot D_{BC}(A, B) \quad (2)$$

where

$$D_{\text{SSD}}(A, B) = \sum_1^h \|A_{i,j} - B_{i,j}\|^2 \quad (3)$$

$$D_{BC}(A, B) = \begin{cases} 1 & \rho_A = \rho_B, \\ \sqrt{1 - \sum_{i=1}^h \sqrt{\rho_A(i) \cdot \rho_B(i)}} & \text{else.} \end{cases} \quad (4)$$

The ρ_A and ρ_B are the histograms of the patches. The distance function is a multiplication of the simple SSD and Bhattacharyya metrics. However, different from [5], we embed

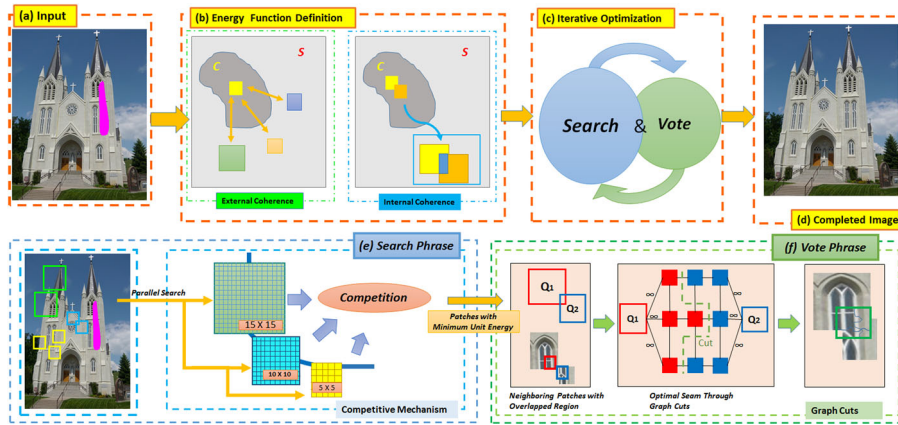


Fig. 3 An overview of our approach. **a** An image with a hole marked in magenta. **b** The definition of our energy function. Our energy function considers external and internal coherence simultaneously. **c** The optimum of our energy function is achieved by a “search and vote” iterative optimization. **e** The search phase. Parallel search is conducted in this

phase to retrieve multiple size patches. The retrieved patch candidates compete to enter the vote phase. **f** The vote phase. Patches are combined through calculating the optimal seam with graph cuts. **d** The result by our approach

the Bhattacharyya metric into a piecewise function. The purpose of this definition is to avoid the null measure when two patches have the exactly the same distribution. Our function will degrade into simple SSD when two patches follow the same distribution.

E_1 encodes the similarity between the patches within the hole and those in the known region. As shown in Fig. 3b, our algorithm searches suitable patches in the known region and calculates the distance between known patches and the target patch. The optimum patch minimizes the energy term below:

$$E_1(Q, P) = D_m(Q, P) \tag{5}$$

Note that the distance between two patches is calculated in both the color channels and the luminance channels.

E_2 constrains the coherence between the patches and its neighbor within the hole. Given a patch in the cavity Q , its adjoint patches Q' share common pixels with Q . As shown in Fig. 3b, we considered the overlapped region when calculating E_2 . The energy term that measures the differences between these overlapped regions is:

$$E_2(Q) = \sum_{Q' \in C} D_m(Q, Q'), \quad Q \cap Q' \neq \emptyset \tag{6}$$

The weighted factors w_1 and w_2 influence the balance between internal and external coherence. Users can adjust the weighted factors to decide completed effect. w_1 is usually set larger than w_2 to avoid trivial solution. When w_2 is set much larger than w_1 , patches may repeat too many times in the hole.

3.2 Iterative optimization with dynamic patches

To achieve global optimum completion is intractable in general, due to the massive solution space and time-consuming energy term evaluation. The energy function is non-convex and has lots of local minimum. Iterative optimization scheme is commonly applied to achieve approximate solutions. The idea that lies behind the scheme is to use the result of the previous iteration to initialize the next iteration. Then, the energy of the objective function is constrained not to increase in each iteration. Wexler et al. [31] proposed a “search and vote” iterative scheme to tackle the optimization. In our optimization, we also employ the “search and vote” scheme with advancing techniques. In the search phases, we introduce a parallel search strategy and a competitive mechanism to find the suitable patches. In the vote phases, we employ the graph cuts technique [20] to achieve seamless connection between patches. The pseudocode of our iterative optimization is shown in Algorithm 1.

3.2.1 Parallel search with competitive mechanism

Parallel search An effective method for patch searching is PatchMatch which is a randomized correspondence algorithm for structural image editing [3]. It first fills the hole by randomly picking up patches in the image and then looks for approximate nearest-neighbor matches between source patches rapidly. In the search process, good matches will propagate and random search is conducted to explore potential matches. In order to search for multiple size patches, we use a modified PatchMatch algorithm. We conducted a parallel search for different sizes of patches in the known region. The number of different size patches v is defined by the users.

Algorithm 1 Dynamic Patch-based Image Completion

Require: Image I , cavity C , source $S = I - C$, Number of different size patches v , Pyramid level L

Ensure: Final Image F

```

1: Initialize  $F$  through filling patches randomly
2: Compute image pyramid  $I_l, C_l, K(l_i), l_i = L, L - 1, \dots, 0$ 
3: for each pyramid level  $l_i$  do
4:   Define the patch sizes with Eq.11
5:   repeat
6:     for All  $q \in C$  do
7:       Parallel Search for  $v$  different size patches
8:       Retrieve the patch  $P$  that satisfies Eq.8
9:     end for
10:    Calculate the minimum cost boundary
11:    Combine all the patches
12:  until convergence
13:  Propagate solution to the next level
14: end for

```

Obviously, the more different patches are searched, the higher the computational cost. It requires high computing power and mass memory to store patches. Thus, we take advantages of the graphics processing unit (GPU) in our approach. GPU is a multi-thread, highly parallel processor that is suitable for massive data operations. The stream of GPU processing enables our multi-size patch search to be operated in the form of parallel acceleration. With the GPGPU ability emerged in CUDA, we can explore parallel programming model with shared memory and search for different size patches simultaneously. Different size patches found in this phase will be retrieved as candidates and enter the competition in the next stage.

Competitive mechanism Candidate patches have to compete to enter the hole and become part of the filling content. To simulate the process of competition, we calculate the unit energy term U_{h_i} of every $h_i \times h_i$ patch:

$$U_{h_i} = \frac{D_m(P_{h_i \times h_i}, Q_{h_i \times h_i})}{h_i^2} \quad (7)$$

The patch with the lowest U_{h_i} will be selected to enter the voting phases. Supposed that there are v candidates attending the competition, the winner satisfies:

$$P_{\text{win}} = \arg \min\{U_{h_i}\}, \quad i = 1, 2 \dots v \quad (8)$$

In our experiments, we defined $v = 3$ which means three different sizes of patches are searched simultaneously at each pyramid level in the search phases and they compete against each other to enter the hole. After all the winning patches are selected, they will be stored as material to form the contents of the cavity.

The parallel search and competitive mechanism allow us to use multi-sizes patches when completing a deficit image. The advantage of utilizing different sizes of patches is that it

enables us to carry out efficient structure and texture restoration. When capturing structure with large patches of a single size, irrelevant background information is easily absorbed into the patches. As shown in Fig. 2, when trying to capture the structure of the branches with a series large 10×10 patches, much irrelevant information (the background and the part of the bird) is absorbed. The result is significantly improved when replaced by the mixed use of three different sizes of patches. In addition, the mixed use of multi-size patches can lower the computational cost. Because small patches contain fewer pixels and require less calculation in energy evaluation.

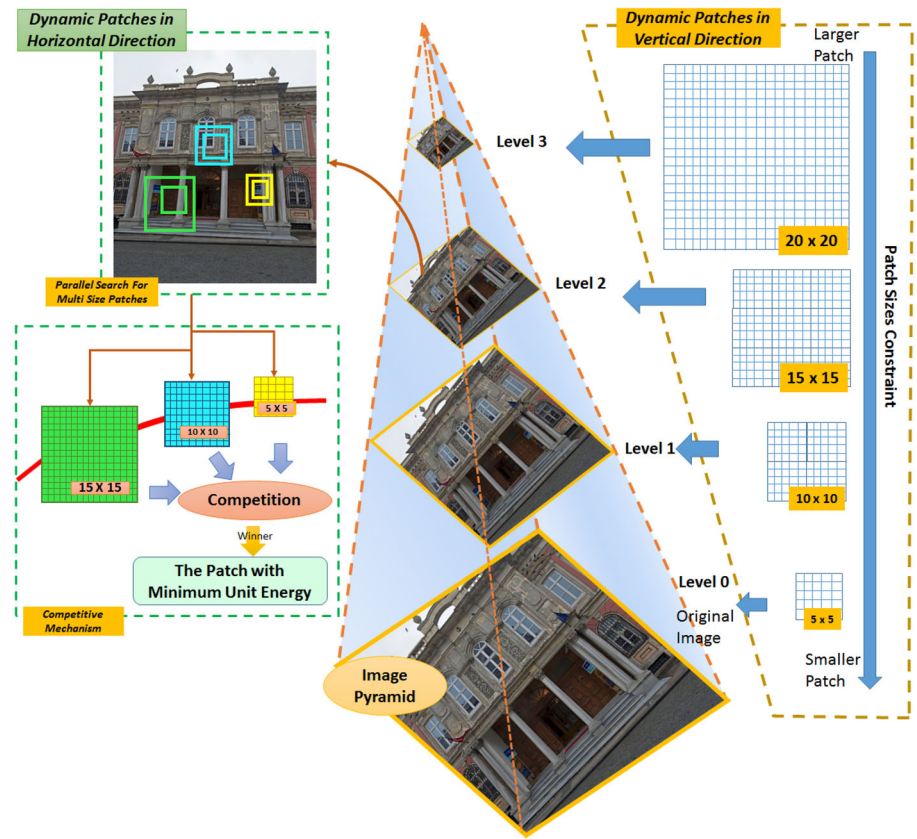
3.2.2 Vote phase with graph cuts

After retrieving all the suitable patches in the search phase, we fill the hole by combining the candidate patches into an entity. In [31], Wexler et al. proposed a weighted voting scheme that the color of the pixel is decided by the median of all votes. This scheme works well when using the fixed size patch. However, it cannot solve the problem of the discontinuity when combining different size patches. Mishandling the overlapping area of different size patches may lead to textural disorder or fracture of the structure. Thus, in order to seamlessly combine different size patches and enhance the coherence of the patches within the hole, we employ the techniques of graph cuts [20] for partly overlapped patches in the voting phase. Given two patches Q_1 and Q_2 that are overlapped along with their vertical edges (Fig. 3f), let s and t be the adjacent pixel positions in the overlap region and $Q_1(\cdot)$ and $Q_2(\cdot)$ be the corresponding pixel value at the specific position. Note that the two patches are only partially overlapped which means $Q_1 - Q_2 \neq \emptyset$. In order to cut the overlap region that can make the two patches match best, we have to lower the adaptive matching quality cost M proposed by [20]:

$$\begin{aligned}
M(s, t, Q_1, Q_2) &= \frac{\|Q_1(s) - Q_2(s)\| + \|Q_1(t) - Q_2(t)\|}{\|\nabla_{Q_1}^d(s)\| + \|\nabla_{Q_1}^d(t)\| + \|\nabla_{Q_2}^d(s)\| + \|\nabla_{Q_2}^d(t)\|} \\
& \quad (9)
\end{aligned}$$

Here d is the direction of the gradient and is the same as the edge direction between s and t . $\nabla_{Q_1}^d$ and $\nabla_{Q_2}^d$ are the gradients in the patches along the direction d . This adaptive matching quality cost penalizes more on seams crossing low-frequency regions than on those crossing high-frequency regions. The optimum seam that minimizes the cost M can be achieved through solving the path finding problem with graph cuts. Consider the graph shown in Fig. 3f, we want to find a minimum cost path through the 3×3 overlap region. The adjacent pixel nodes are connected with arcs which are labeled with the adaptive matching qual-

Fig. 4 An overview of the dynamic patch systems. Our dynamic patches system take effect in two directions of the image pyramid. In the horizontal direction, we conduct a parallel search for different size patches in each pyramid level. In the vertical direction, we constrain the algorithm to apply large patches in higher pyramid level and small patches in lower pyramid level



ity cost $M(s, t, Q_1, Q_2)$. Two additional nodes representing the patches Q_1 and Q_2 are added and are connected to the adjacent nodes with constraint arcs. The constraint arcs are weighted heavily to ensure that those connected nodes must come from specific patches. When we try to find a cut of minimum cost in the graph, we are separating node Q_1 and node Q_2 . This is a classical graph cut problem called min-cut [14] which has efficient solutions and is easy for implementation.

3.3 Multi-scale solution with dynamic patches

To further speed up convergence and strengthen global consistency, our approach follows a coarse-to-fine fashion in the implementation. We resized the image into L different resolutions to form an image pyramid. The “search and vote” iterations repeat in each pyramid level. The size of the patches changes continuously in the “search & vote” process. In order to further capture structural and textural information efficiently, we add constraints on the sizes of patches at each pyramid level. As shown in Fig. 4, patches in our approach vary vertically in each coarse level of the pyramid. The patch sizes adapt to the coarse level automatically to realize effective structure restoration and acceleration.

In our approach, we set constraints on the max/min size of patches allowed in each level. Let h_{\max} be the maximum patches allowed in all L pyramid levels. h_{\max} is user-defined and depends on hardware and application. Given a pyramid level l_i , we use a discrete function $K(\cdot)$ to map l_i to the size of the patches. In our approach, we define this function as:

$$K(l_i) = \left\lfloor \frac{h_{\max} - v}{1 + e^{-l_i \times \beta}} \right\rfloor \tag{10}$$

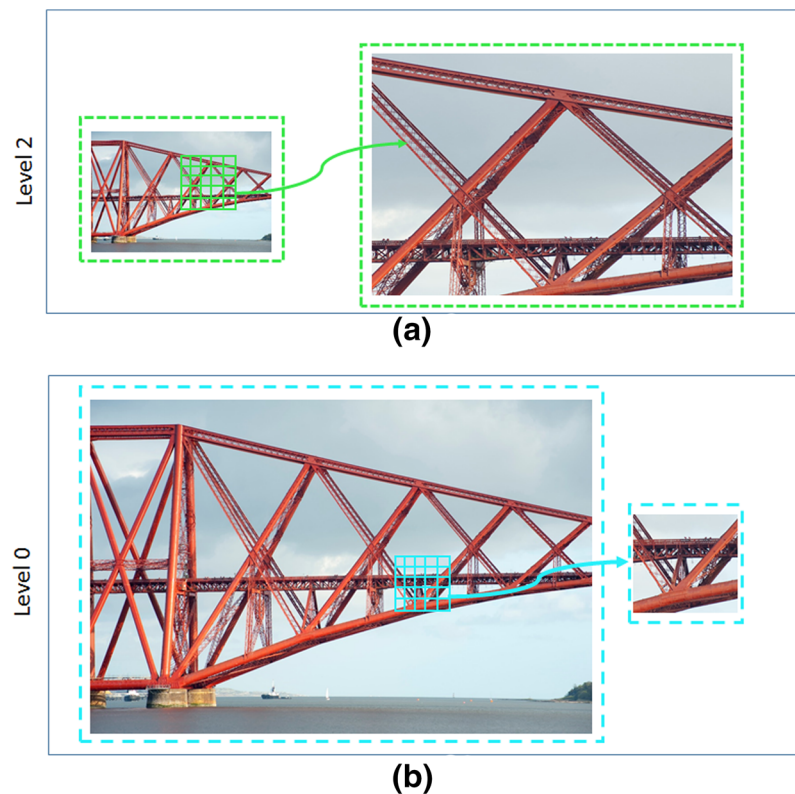
where β is a parameter which controls the intervals of different values and $\lfloor \cdot \rfloor$ is the floor operator. Let h_i be the size of the i th patch. Then, the size of the patches applied in pyramid level l_i satisfies :

$$\begin{cases} \text{Min}(h_1, h_2, \dots, h_v)_{l_i} \geq K(l_i), & \text{for } l_i > L/2 \\ \text{Max}(h_1, h_2, \dots, h_v)_{l_i} \leq K(l_i), & \text{for } l_i < L/2 \end{cases} \tag{11}$$

Equation 11 constrains the size of the patches that can be applied in coarse level l_i . It guarantees that large patches are applied in high pyramid level (which means in the image of low resolution) and small patches are used in low pyramid level (which means in the image of high resolution).

In fact, the actual form of $K(\cdot)$ can be different according to realistic applications. Equation 10 is one of the optional forms. The monotonicity of $K(\cdot)$ plays an important role in

Fig. 5 Using large patches in higher pyramid level can preserve more structure information than using those in lower pyramid level. **a** A 5×5 patch on a bridge picture of 660×440 resolution. **b** The same patch is applied on the same picture of 2200×1464 resolution



implementation. The $K(\cdot)$ used in our approach implies that the sizes of the patches shrink along with the falling of the pyramid level. In other words, an increasing $K(\cdot)$ ensures a series of shrinking patches along the vertical direction of the pyramid. Vice versa, a $K(\cdot)$ which is monotonically decreasing guarantees a series of enlarging patches. Large patches are employed in lower pyramid level.

In our approach, we utilize a monotonically increasing $K(\cdot)$ to set up limitation for the patches utilized in each pyramid level. The purpose of using an increasing function is that large patch is more effective in structure capturing in coarse images. Figure 5 compares the structural information captured in different pyramid levels with identical patches which are represented with the grid in the image. A 5×5 patch is applied to the same images in different scales. Their resolutions are 660×440 and 2200×1464 , respectively. It is shown that one 5×5 grid can capture the major structure of the bridge on the 660×440 images, but the same patch can only include minor local structures of the bridge in the image of higher resolution. Thus, applying large patches in higher level can grab more information than that in lower pyramid level. Also, in the aspect of computational workload, applying large patches in a higher pyramid level is much more economic. Because in lower pyramid level, the utilization of large patches has limited effect on structure reconstruction and consumes more computational resources than small patches.

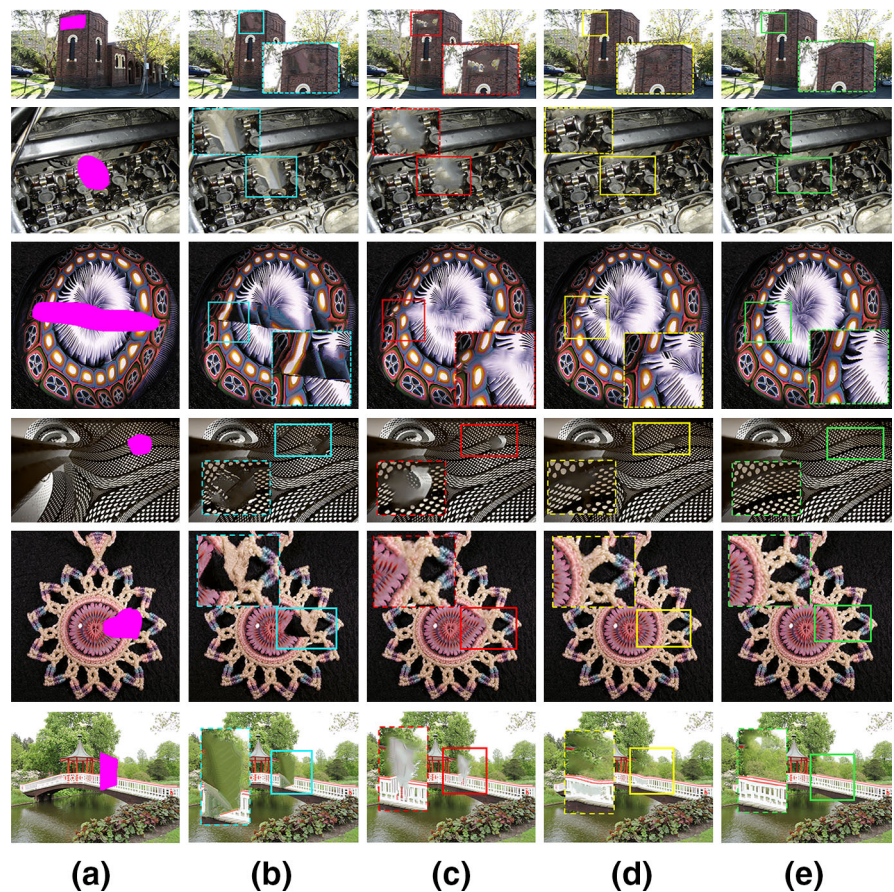
4 Results and discussion

To verify the performance of our approach, we apply it to images which contain different types of textures and structures. We compare our approach with some of the well-known methods, including the ImageMelding [11], the methods by Wexler et al. [31] and by Criminisi et al. [10]. The program used in our experiments is a collection of MATLAB and C++ functions. The program is run on an Intel Xeon E5-2470 V2 2.40 GHz computer with 8GRAM and a Nvidia Tesla graphic card. In the following subsections, we will discuss the visual effectiveness of our approach and its runtime performance. Also, a parameter analysis is presented. We analyze the internal and external coherence balanced by w_1 and w_2 and discuss how the size of patches in the vertical direction controlled by $K(\cdot)$ affects the results. Besides the subjective visual comparison, we also conduct a user study to evaluate the completed image generated by our approach.

4.1 Effectiveness

Figure 6 is the completed results by our approach tested on images which contain complex texture and structure. In Fig. 7, we apply our approach to object removal tasks. From the results, we can see that the restoration of the structures and texture plays an important role in the over-

Fig. 6 **a** The masked image to complete. **b** Results generated with the method by Criminisi et al. [10]. **c** Results generated with method by Wexler et al. [31]. **d** Results generated with ImageMelding [11]. **e** Our approach



all effect of the completion. As for the results generated by the compared methods, the presence of abnormal structures and texture causes unnatural distortion. When complex structures exist in the images, even the completed results by state-of-the-art ImageMelding [11] fail to represent reasonable structures in the hole. As shown in Fig. 6d, apparent abnormal fractures appear in the body of the architecture and lead to unpleasant incoherent. Unlike those results by previous methods, our results successfully preserve adequate structures with the aid of dynamic patches. Structural information is captured efficiently and well reproduced as seen in Fig. 6e.

The metric defined in our approach is a combination of the simple SSD and the Bhattacharyya distance. With the help of this combined distance, we appropriately maintain sufficient internal coherence within the hole. Combined with the dynamic patches we used, our results remarkably avoid the problem of interference information. Results shown in the first row of Fig. 6 (the church) are examples. The textured region of the image is embedded with some alphabets which are easily absorbed into the cavity causing artifact. Filling the hole with previous methods may easily be interfered by the irrelevant information as shown in Fig. 6d. The result generated by our approach (as illustrated in Fig. 6e) suc-

cessfully averts the disturbing information through allowing good matches to propagate within the cavity for the purpose of internal coherence.

4.2 Efficiency

Many vision problems are formulated into energy minimization problem. However, the energy minimization problems are seldom easy tasks. Finding global optimum for the objective function is often unpractical. We apply the “search and vote” scheme to guarantee the energy function not to increase in each iteration. In Fig. 8, we present the energy status throughout the iterations of optimization. Although initialization at each pyramid level causes a small energy inflation, we can see from the line chart that the energy continuously decreases in each iteration and becomes stable after several iterations. Although this approximate scheme cannot guarantee global optimum, our experimental results are visually pleasing.

In addition to quick convergence, our approach has advantages in time consumption compared to previous methods. Although seeking for different size of patches leads to higher computational cost, we utilize parallel programming techniques to overcome the difficulty. As illustrated in



Fig. 7 Our approach performs well on object removal tasks. Images in the left column are ground truth, while those in the right column are our results. Images in the middle column are masked images

Table 1, with the help of GPU acceleration, our approach saves approximately 68% of the time cost. We also list the statistics on time consumption of other methods; our approach is faster than the method by Criminisi et al. [10] and that by Wexler et al. [31]. When compared to the state-of-the-art ImageMelding [11], our approach still has advantages in time-saving. Our approach is just a prototype. The time cost of our approach can be further reduced if it is developed to maximize the capacity of the hardware.

4.3 Parameter analysis

4.3.1 External coherence versus internal coherence

The balance between internal and external coherence is controlled by w_1 and w_2 . They are user-defined coefficients and can be adjusted according to the pictures. Figure 9 illustrates the results generated with different configurations of our approach. A higher w_1 gives more preference to external coherence. However, when w_1 is weighted much larger than w_2 , the abnormal structure and disordered texture may appear again as can be seen in Fig. 9. An objective function with a larger w_2 emphasizes more on internal coherence. Nevertheless, the oversize w_2 may slightly blur the image since some patches are reused for too many times. Our experiments suggested that Twenty-Eighty Law may be a possible criterion for controlling the maximum ratio between w_1 and w_2 . A balanced configuration ($w_1 = w_2 = 0.5$) is still recommended.

As be seen in Fig. 9. The result generated by this configuration ($w_1 = w_2 = 0.5$) is realistic and has sharp structure.

4.3.2 Enlarging patches versus shrinking patches

We conduct an experiment by using enlarging patches in the vertical direction of the image pyramid and compare the results with those generated using shrinking patches. To apply enlarging patches in the vertical direction of the image pyramid, the $K(\cdot)$ should be a monotonically decreasing function. As illustrated in Fig. 10, the completed images using enlarging patches are blurred within the cavity. Some of the textual information is missed which causes obvious fracture of the structure. In the aspect of efficiency, the time consumption of image completion using enlarging patches is 22.82% higher than that of using shrinking patches (1452.207 vs. 1783.612 s), because patch searching and filling are much faster in low-resolution scales and using larger patches is more economical. In addition, the information captured in the coarse image provides a sound foundation for the completion in next pyramid level. As illustrated in Fig. 10, we fetch the intermediate outcomes in different pyramid level and notice that the visual effects have a significant difference. Using large patches in lower pyramid level cannot compensate for the poor foundation effectively, while costing extra computational resources. Thus, in general, applying shrinking patches along the vertical direction of the pyramid is more recommended.

4.4 User Study

In order to learn whether our approach generates better completed results than other methods from the user point of view, we conduct a user study. 30 different subjects were asked to rate a 9-point scale ([1–9] with 9 as the best) for the texture and structure of the completed image. Again, ImageMelding [11], the method by Criminisi et al. [10], the method by Wexler et al. [31] and our approach were compared. We presented each completed results to the subjects, and they were required to rate without knowing the exact technique was used. Four sets of test images (16 images in total) were tested, and we collected 480 data samples in total for analysis. Table 2 shows the descriptive statistics of the collected data.

From Table 2, the mean scores for our approach, ImageMelding, the method by Wexler et al. [31] and method by Criminisi et al. [10] are 7.98, 6.94, 5.50 and 5.23, respectively. We conduct an analysis of variance (ANOVA) to test whether the difference between the means is statistically significant. The F value is the test statistics which reflects the significance. The ANOVA result among the four groups is $F(3, 476) = 213.142$, $p < 0.001$ which suggests that there is a significant difference between the means of

Fig. 8 Energy evaluation. The energy variation of the objective function on a picture of a church is presented. Note that when the result generated by the previous pyramid level propagates to the next pyramid level, the energy inflates. This inflation is caused by the increase in the number of pixels when the resolution rises. The tendency of the energy also provides cues of convergence. When the energy stop falling, it suggests that image is completed and reaches to the optimum

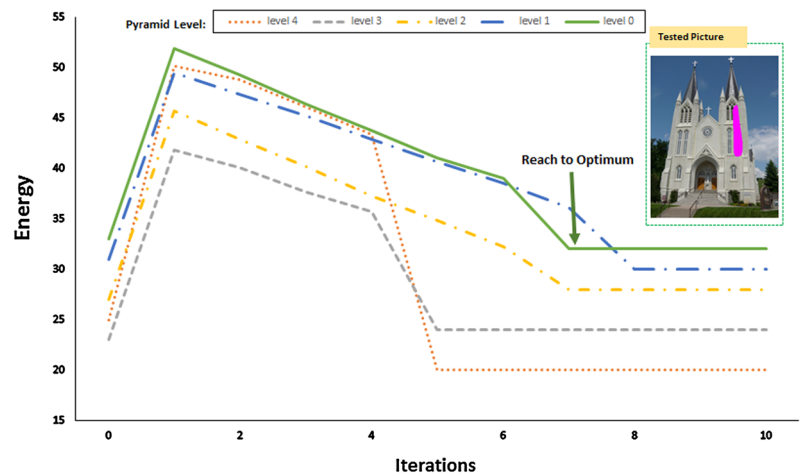


Table 1 Run-time performance

	Church	Engine	Crystal	Net	Heart	Bridge
Image size (pixels)	800 × 450	800 × 600	480 × 480	2000 × 1333	400 × 400	730 × 548
Time cost (s)						
Criminisi et al. [10]	2408.69	5001.221	3419.57	19931.94	1151.11	3505.27
Wexler et al. [31]	1852.84	3847.093	2630.44	15332.26	885.46	2696.36
ImageMelding [11]	822.663	1923.546	1122.996	7118.54	411.110	1251.882
Ours (without acceleration)	1245.112	2676.238	1699.669	9199.356	531.281	1617.816
Ours (with GPU acceleration)	741.138	1672.649	1011.708	5475.807	316.239	962.986

The time cost of completing the images in Fig. 6. With the GPU acceleration, our approach consumes less time than previous methods in general

Fig. 9 The results generated through different configurations of w_1 and w_2 . We use the Twenty-Eighty Law as a criterion to control the ratio between w_1 and w_2 . Notice that, when $w_1 = 0.8, w_2 = 0.2$, fracture occurs in the petal of the ice crystal. With the configuration of $w_1 = 0.2, w_2 = 0.8$, the structure is blurred and artifact occurs

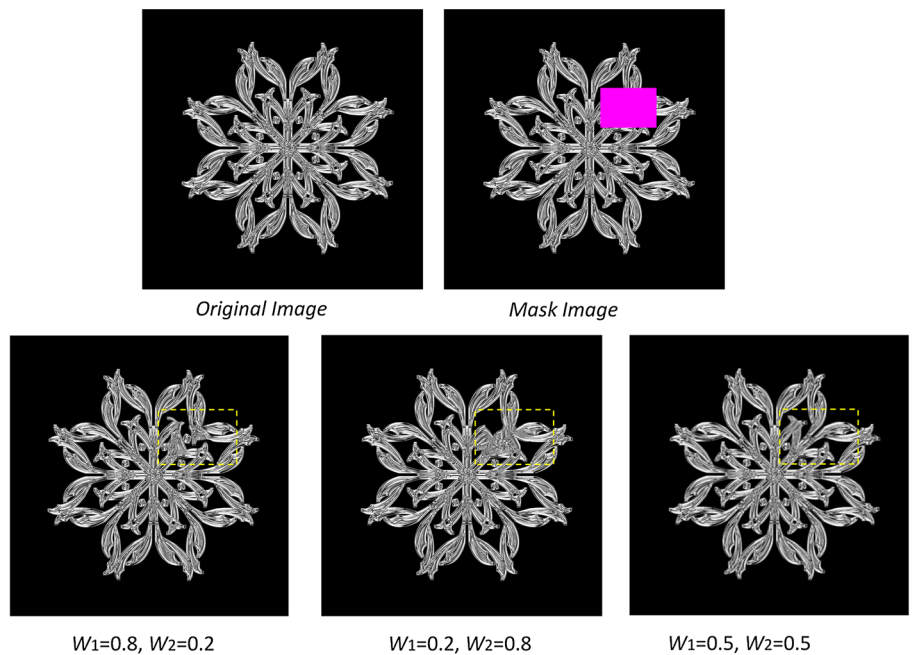


Fig. 10 Comparison of the effect of using enlarging patches and shrinking patches. Pictures on the left are results generated using a series of shrinking patches. Large patches are utilized in high level of the image pyramid. As it is shown, most structures are captured in level 5 and it provides a sound foundation for the iteration in lower level. The results on the right are generated using a series of enlarging patches. The size of patches increases along the vertical direction of the image pyramid. As can be seen, fracture of the structure occurs and the loss cannot be compensated by using large patches in lower pyramid level

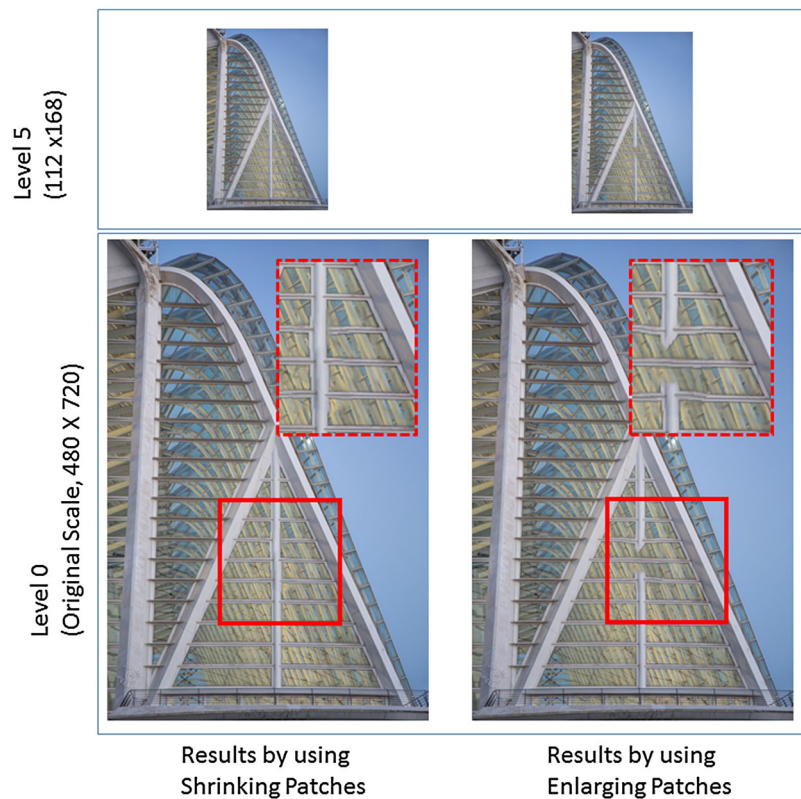


Table 2 User study statics

Methods	Mean	SD	Mean difference with our approach	95% confidence interval	
				Lower bound	Upper bound
Our approach	7.983	0.829	–	–	–
ImageMelding [11]	6.941	0.812	1.041	0.711	1.372
Wexler et al. [31]	5.500	0.502	2.483	2.153	2.813
Criminisi et al. [10]	5.230	1.459	2.750	2.419	3.080

The scores difference between our approach and other previous methods are presented in the third column of the table. Note that the lower and upper bounds of the estimated differences do not contain zero

four groups. Then, in post hoc test, we calculate the mean difference (MD) and compare our approach with ImageMelding (MD = 1.042, $p < 0.001$), method by Wexler et al. (MD = 2.48, $p < 0.001$) and method by Criminisi et al. (MD = 2.78, $p < 0.001$), all the results are still significant and suggest that the scores of our approach are higher. From the 95% confidence interval, it concludes that the results of our approach are clearly more perceptually pleasing than those by other methods.

5 Conclusion

In this paper, we present a dynamic patch-based method for image completion with efficient structure preservation. Compared to previous patch-based methods, our approach applies dynamic patches, which changes sizes in the horizontal and vertical direction of the image pyramid simultaneously.

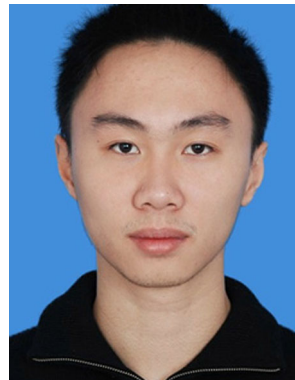
The introduction of the dynamic patches system allows us to capture structures and textures effectively and economically. Compared to the state-of-the-art ImageMelding, our approach does not suffer from abnormal structures and texture disorder. With the support of experiments and user study, our approach outperforms previous methods and generates visually appealing images. In the future, we are going to adapt our approach to other image editing tasks such as image cloning and satellite image processing. These tasks can be formulated into the similar framework of the image completion problem. Another possible direction is to optimize the patch search and assign mechanism simulating quantum behavior to the operators.

Acknowledgements The authors would like to thank all reviewers for their helpful suggestions and constructive comments. The work was supported by the National Natural Science Foundation of China (No.

61632003), and a Grant from the Research Grants Council of Hong Kong (No. 28200215).

References

- Ashikhmin, M.: Synthesizing natural textures. In: Proceedings of the 2001 Symposium on Interactive 3D graphics, pp. 217–226. ACM (2001)
- Ballester, C., Bertalmio, M., Caselles, V., Sapiro, G., Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Process.* **10**(8), 1200–1211 (2001)
- Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.: Patch-match: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* **28**(3), 24:1–24:11 (2009)
- Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 417–424 (2002)
- Bugeau, A., Bertalmio, M., Caselles, V., Sapiro, G.: A comprehensive framework for image inpainting. *IEEE Trans. Image Process.* **19**(10), 2634–2645 (2010)
- Burt, P.J., Adelson, E.H.: A multiresolution spline with application to image mosaics. *ACM Trans. Graph.* **2**(4), 217–236 (1983)
- Cai, N., Su, Z., Lin, Z., Wang, H., Yang, Z., Ling, B.W.K.: Blind inpainting using the fully convolutional neural network. In: The Visual Computer pp. 1–13 (2015)
- Chan, T.F., Kang, S.H., Shen, J.: Euler's elastica and curvature based inpainting. *SIAM J. Appl. Math.* **63**(2), 491–506 (2010)
- Chen, T., Cheng, M.M., Tan, P., Shamir, A., Hu, S.M.: Sketch2photo: internet image montage. *ACM Trans. Graph.* **28**(5), 124 (2009)
- Criminisi, A., Perez, P., Toyama, K.: Object removal by exemplar-based inpainting. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 721–728. IEEE (2003)
- Darabi, S., Shechtman, E., Barnes, C., Goldman, D.B., Sen, P.: Image melding: combining inconsistent images using patch-based synthesis. *ACM Trans. Graph.* **31**(4), 82:1–82:10 (2012)
- Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 341–346. ACM (2001)
- Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1033–1038. IEEE (1999)
- Ford Jr., L.R., Fulkerson, D.R.: *Flows in Networks*. Princeton University Press, Princeton (2015)
- Gracias, N., Mahoor, M., Negahdaripour, S., Gleason, A.: Fast image blending using watersheds and graph cuts. *Image Vis. Comput.* **27**(5), 597–607 (2009)
- Hao, C., Chen, Y., Wu, W., Wu, E.: Image completion with perspective constraint based on a single image. *Sci China Inf Sci* **58**(9), 1–12 (2015)
- Hua, M., Wang, W.: Effective structure restoration for image completion using internet resources. *Vis. Comput* **31**(6–8), 1113–1122 (2015)
- Ignácio, U.A., Jung, C.R.: Block-based image inpainting in the wavelet domain. *Vis. Comput.* **23**(9–11), 733–741 (2007)
- Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. *ACM Trans. Graph. (Proc. SIGGRAPH 2017)* **36**(4), 107:1–107:14 (2017)
- Kwatra, V., Schödl, A., Essa, I., Turk, G., Bobick, A.: Graphcut textures: image and video synthesis using graph cuts. *ACM Trans. Graph.* **22**(3), 277–286 (2003). doi:[10.1145/882262.882264](https://doi.org/10.1145/882262.882264)
- Levin, A., Zomet, A., Weiss, Y.: Learning how to inpaint from global image statistics. In: Proceedings of IEEE International Conference on Computer Vision, vol. 1, pp. 305–312 (2003)
- Liang, L., Liu, C., Xu, Y.Q., Guo, B., Shum, H.Y.: Real-time texture synthesis by patch-based sampling. *ACM Trans. Graph.* **20**(3), 127–150 (2001)
- Liu, B., Li, P., Sheng, B., Wu, E.: Image completion with dynamic patches. In: Proceedings of the Computer Graphics International Conference, p. 3. ACM (2017)
- Nill, N.B., Bouzas, B.: Objective image quality measure derived from digital image power spectra. *Opt. Eng.* **31**(4), 813–825 (1992)
- Qin, X., Shen, J., Mao, X., Li, X., Jia, Y.: Robust match fusion using optimization. *IEEE Trans. Cybern.* **45**(8), 1549–1560 (2015)
- Qin, X., Shen, J., Mao, X., Li, X., Jia, Y.: Structured-patch optimization for dense correspondence. *IEEE Trans. Multimed.* **17**(3), 295–306 (2015)
- Shen, J., Jin, X., Zhou, C., Wang, C.C.: Gradient based image completion by solving the Poisson equation. *Comput. Graph.* **31**(1), 119–126 (2007)
- Shen, J., Zhao, Y., Yan, S., Li, X., et al.: Exposure fusion using boosting Laplacian pyramid. *IEEE Trans. Cybern.* **44**(9), 1579–1590 (2014)
- Sunkavalli, K., Johnson, M.K., Matusik, W., Pfister, H.: Multi-scale image harmonization. *ACM Trans. Graph.* **29**(4), 125:1–125:10 (2010). doi:[10.1145/1778765.1778862](https://doi.org/10.1145/1778765.1778862)
- Wei, L.Y., Levoy, M.: Fast texture synthesis using tree-structured vector quantization. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 479–488. ACM Press/Addison-Wesley Publishing Co., New York (2000)
- Wexler, Y., Shechtman, E., Irani, M.: Space–time completion of video. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(3), 463–476 (2007)
- Whitaker, R.T.: A level-set approach to image blending. *IEEE Trans. Image Process.* **9**(11), 1849–1861 (2000)



Bowen Liu is currently a Ph.D. candidate at The Education University of Hong Kong. He finished his bachelor study in Jinan University and had an experience as an exchange student in Aalborg University Denmark. He has broad research interest in image processing, object detection, autonomous vehicle and smart transportation systems.



Ping Li is currently an assistant professor at Macau University of Science and Technology, who obtained his Ph.D. from The Chinese University of Hong Kong. His research interests include image/video stylization, big data visualization, GPU acceleration, and creative media. He has one image/video processing national invention patent and has excellent research project reported worldwide by ACM TechNews.



Enhua Wu completed his BSc in 1970 from Tsinghua University, Beijing, and received his Ph.D. degree from University of Manchester, UK, in 1984. He has been working at the State Key Lab. of Computer Science, Institute of Software, Chinese Academy of Sciences, since 1985, and at the same time invited as a full professor of University of Macau since 1997. Dr. Wu has been invited to chair or co-chair various conferences or program committees including Pacific Graphics 2005,

CASA2006, ACM VRST2010, 2013, 2015, WSCG2012 and as a keynote speaker for CyberWorlds2006, ACM VRST2010, ACM VRCAI2011, WSCG2012, etc. He is an Associate Editor-in-Chief of the Journal of Computer Science and Technology since 1995, and the editorial board member of TVC, CAVW, IJIG, IJVR. He is a member of IEEE & ACM, and a fellow of the China Computer Federation (CCF). His research interests include realistic image synthesis, physically based simulation and virtual reality.



Bin Sheng is an associate professor and director of Lab For Digital Media and Data Reconstruction at Shanghai Jiao Tong University. He obtained his Ph.D. from The Chinese University of Hong Kong. His research interests include image-based modeling and rendering, animation, image processing, natural phenomenon simulation, NPR, GPU processing, virtual reality, augmented reality, data visualization and biomedical imaging.



Yongwei Nie is now an assistant professor at the School of Computer Science and Engineering of South China University of Technology, Guangzhou, Guangdong, China, 510006. He received his doctor and bachelor degree from Computer School of Wuhan University in 2015 and in 2009. His research interests fall in Computer Graphics and Computer Vision.