Image editing by object-aware optimal boundary searching and mixed-domain composition

Shiming Ge¹ (✉), Xin Jin², Qiting Ye^{1,3}, Zhao Luo^{1,3}, and Qiang Li⁴

© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract When combining very different images which often contain complex objects and backgrounds, producing consistent compositions is a challenging problem requiring seamless image editing. In this paper, we propose a general approach, called *object-aware image editing*, to obtain consistency in structure, color, and texture in a unified way. Our approach improves upon previous gradient-domain composition in three ways. Firstly, we introduce an iterative optimization algorithm to minimize mismatches on the boundaries when the target region contains multiple objects of interest. Secondly, we propose a mixed-domain consistency metric for measuring gradients and colors, and formulate composition as a unified minimization problem that can be solved with a sparse linear system. In particular, we encode texture consistency using a patch-based approach without searching and matching. Thirdly, we adopt an object-aware approach to separately manipulate the guidance gradient fields for objects of interest and backgrounds of interest, which facilitates a variety of seamless image editing applications. Our unified method outperforms previous state-of-the-art methods in preserving global texture consistency in addition to local structure continuity.

Keywords seamless image editing; patch-based synthesis; image composition; mixed-

domain; gradient-domain composition

1 Introduction

Seamless image editing has been an active research field in recent years. It is widely applied in panorama mosaicing [1], photo composition [2, 3], manipulating large collections of photos [4], and so on. Seamless image editing involves combining source regions with target images in a visually natural way. Since input images often contain multiple objects with differently textured backgrounds, a natural-looking editing result should meet the expectations of human visual perception [5] and preserve not only local structure continuity within boundaries but also consistent color and texture transitions between source and target images.

In general, when combining images with very different structures and textures, a successful image editing algorithm should preserve the following properties, to produce results in agreement with our visual perception:

1) **Local structure continuity.** When the input images include multiple objects or different structures, the composition should not break the local salient structures in the overlap region, to avoid structure collision or discontinuity.

2) **Smooth color transitions.** The composition should blend the colors between input images to avoid blurring due to differences along the boundary.

3) **Global texture consistency.** When the input images have very different textures, the process should take into account not only gradients but also colors to preserve texture consistency.

Much research has been conducted to solve these issues.

1 Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100095, China. E-mail: geshiming@iie.ac.cn (✉).

2 Beijing Electronic Science and Technology Institute, Beijing 100070, China.

3 School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China.

4 School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China.

Manuscript received: 2017-02-24; accepted: 2017-12-01

In this paper, we propose a general approach which we call *object-aware image editing* (OAIE), based on optimal editing region selection and mixed-domain composition. Our main contributions include:

- 1) an iterative optimization algorithm to compute optimal boundaries for regions containing multiple objects, which can be used to preserve local structure continuity;
- 2) a mixed-domain measure to address smooth color transitions and global texture consistency, allowing us to formulate editing as a unified minimization problem;
- 3) a patch-based approach to encoding texture consistency which requires neither search nor voting.

Based on these contributions, OAIE can simultaneously eliminate local structure artifacts, provide smooth color transitions, and preserve global texture consistency.

2 Related work

In the image editing literature, many methods have been proposed to address inconsistency in structures, colors, and textures.

2.1 Optimal seam methods

Optimal seam methods are often applied for seamlessly compositing images and textures; they achieve good local structure continuity. They seek a partition curve (an optimal seam) in the overlap region to minimize the difference between the two input images, in order to make the seam as invisible as possible. Generally, objects are visually salient in images [6], so an optimal seam should pass around objects to avoid structure collisions. Dynamic programming [7] or graph cut [8] methods are usually applied to find the optimal seam. Optimal seam methods can handle image composition well when there are only small structure and color differences between input images in the overlap region. However, if the differences are large, it is hard to find a pleasing optimal seam. To combine inconsistent images, Darabi et al. [9] proposed a novel method, *image melding* (IM), which uses a Poisson equation solver to find suitable colors by minimizing an energy function based on mixed L_2/L_0 norms for colors and gradients. This method can produce a gradual transition between source images without sacrificing texture sharpness. However, when there

are too many edges or complex textures around the boundary, the editing result can look unnatural. Tao et al. [10] proposed *error-tolerant image composition* (ETIC), which minimizes the curl of the target gradients on the foreground-background boundary. However, the editing result may have a very sharp boundary and color leakage may occur.

2.2 Gradient-domain methods

Gradient-domain methods are usually adopted to obtain *smooth color transitions*. They make use of known gradient information to produce the final composition by interpolation. The basic idea is to reconstruct the image from gradient fields with specified boundary conditions. Pérez et al. [11] proposed an effective image blending approach, *Poisson image editing* (PIE). By solving a Poisson equation with user-specified Dirichlet boundary conditions, this approach can blend colors seamlessly between input images. However, its effectiveness heavily depends on careful alignment of the structures of the input images along the user-drawn boundary. Agarwala et al. [2] combined graph cut optimization and gradient-domain composition to create photomontages. Zomet et al. [12] proposed an L_1 -based gradient-domain stitching method to eliminate visible seams in the overlap region by use of gradient fields. Jia et al. [13] proposed an easy blending method, *drag-and-drop pasting* (DDP), by optimizing boundary conditions for gradient-domain composition. Farbman et al. [14] introduced an alternative, mean-value coordinate based approach, to carry out seamless cloning via a weighted combination of values along the boundary. Bhat et al. [15] proposed a unified variational model, *GradientShop*, to perform a number of image and video editing tasks. It contains many filters and uniformly uses quadratic optimization, lowering the computation time. Li et al. [16] performed multi-scale editing by applying a nonlinear filter bank to adjacent pixels at each level of a Gaussian pyramid, to eliminate visual artifacts. However, visual artifacts still occur when the trade-off parameter is too large or too small. In Ref. [17], the authors presented a gradient-based variational model for video editing, which addresses the problem of propagating gradient-domain information along the optical flow of the video. With this method,

a user can edit a frame by modifying the texture of an object's surface and then propagate this edit throughout the video. Bie et al. [18] incorporated the users' intent in outlining the source patch, to tackle structure conflicts between the source image patch and the target image. Hua et al. [19] added an extra edge-aware constraint term in a general gradient-domain optimization framework, enforcing similar image filtering effects while preserving edges. In Ref. [20], Zhang et al. performed image copy-and-paste with optimized gradients, where they created a gradient transition map in the cloning area and then used an interpolation-based method to calculate the composition results from the reconstructed gradient map.

2.3 Patch-based synthesis

Patch-based texture synthesis is usually used to address *global texture consistency*, and has been successfully applied to various editing tasks on still images, video, and stereo pairs. Darabi et al. [9] proposed a novel method, *image melding* (IM), which adopts patch-based synthesis to find suitable colors, and minimizes an energy function based on mixed L_2/L_0 norms for colors and gradients. This method uses an iterative search-and-voting blending scheme, and can produce a gradual transition between inconsistent source images without sacrificing texture sharpness. However, when there are too many edges or complex textures around the boundary, or limited texture sources for synthesis, the editing result can be unnatural. Ma and Xu [21] proposed an efficient manifold preserving edit propagation method which searches feature space using an adaptive neighborhood size, which reduces time and memory costs without reducing visual fidelity. Barnes et al. [4] proposed a fast patch-based optimization method, *PatchTable*, for efficient computational photography. In Ref. [22], Luo et al. extended the patch-based synthesis framework from 2D to 3D for stereoscopic image editing. They introduced a depth-dependent patch-pair similarity measure and a joint patch-pair search. Chen et al. [23] presented an interactive system called *sketch2photo* to compose a realistic picture from a simple freehand sketch annotated with text labels. Their system found several photographs in agreement with the sketch and text

labels by searching the Internet and automatically selected suitable photographs to generate a high quality composition. In Ref. [24], Zhang et al. extended patch-based synthesis to plenoptic images captured by consumer-level lenslet-based devices for interactive light field editing. They represented the light field as a set of images captured from different viewpoints and performed patch-based image synthesis on all affected layers of the central view, and then propagated the edits to all other views. To address the heavy computational burden of gradient-domain operators, Ref. [25] proposed a patch-based synthesis method using a Laplacian pyramid to improve searching for correspondences with enhanced awareness of edge structures.

3 Optimal boundary search

3.1 Definition

As illustrated in Fig. 1, seamless image editing requires combining a region of interest Ω_r (with external region boundary $\partial\Omega_r$) from a source image f_s into the target image f_t . Generally, Ω_r may include one or more objects of interest $O_i, i = 1, \dots, n$ (e.g., $n = 3$). We define a target of interest Ω_i (with external target boundary $\partial\Omega_i$) as a subregion surrounding the corresponding O_i (with external object boundary ∂O_i), and note that $O_i \subseteq \Omega_i \subseteq \Omega_r$. Then the optimal editing region selection problem requires finding the optimal

Fig. 1 Optimal editing region selection. A region of interest from the source image may include one or more objects of interest. Optimal editing region selection requires finding optimal targets of interest which contain objects of interest, to minimize local structure mismatches. As optimal target boundaries should not intersect each other, we break the band connectivity with several shortest boundary cuts or connections, giving an effective algorithm for finding the shortest closed path.

targets of interest $\{\Omega_i\}$, $i = 1, \dots, n$, which minimize local structure mismatches. The task can be turned into one of optimizing each target boundary $\partial\Omega_i$ between ∂O_i and $\partial\Omega_r$ for each object of interest O_i . Let $\Omega = \cup\Omega_i$, $O = \cup O_i$, $B_i = \Omega_i \setminus O_i$, and $B = \cup B_i$. Then the optimal editing region Ω is separated into two parts including objects of interest O and backgrounds of interest B . The extra region $\Omega_r \setminus \Omega$ is neglected in image editing. The cut C_i connects the boundaries between object of interest O_i and Ω_r , and the cut C_{ij} connects the boundaries between O_i and O_j .

We further define various operators, including the gradient operator $\nabla = [\partial/\partial x, \partial/\partial y]$, the Laplacian operator $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$, the divergence operator $\nabla \cdot = \partial/\partial x + \partial/\partial y$, the L_2 norm $\|\cdot\|$, and the pixel count $|R|$ in a region R .

3.2 Optimal boundary for single object

Consider the minimization problem for gradient-domain composition as in Ref. [11], which seeks an image f to approximate the target gradient field $\mathbf{v} = \nabla f_s$ (from the source image) in a least-squares sense with the given user-specified Dirichlet boundary condition:

$$\min_f \int_{\Omega} \|\nabla f - \mathbf{v}\|^2 dx dy, \quad f|_{\partial\Omega} = f_t|_{\partial\Omega} \quad (1)$$

Letting $f' = f - f_s$, Eq. (1) can be written as

$$\min_f \int_{\Omega} \|\nabla f'\|^2 dx dy, \quad f'|_{\partial\Omega} = (f_t - f_s)|_{\partial\Omega} \quad (2)$$

Note that the variational energy $\int_{\Omega} \|\nabla f'\|^2 dx dy$ in Eq. (2) will approach zero if and only if the boundary condition $f'|_{\partial\Omega} = k$ (where k is a constant). This observation has been utilized in DDP [13] to define the optimal boundary condition when including a single object of interest.

3.3 Optimal boundaries for multiple objects

In this paper, we extend the case to placing multiple objects within Ω_r . Then, the resulting optimal boundaries can be computed by minimizing the following boundary energy function:

$$E_B(\{\partial\Omega_i\}, \{k_i\}) = \sum_{i=1}^n \sum_{p \in \partial\Omega_i} \|(f_t(p) - f_s(p)) - k_i\|^2 \quad (3)$$

where the constant k_i is a vector for a color image or a scalar for a grayscale image. We may consider each target boundary $\partial\Omega_i$ independently due to their

independence of each other. Since $\partial\Omega_i$ may pass through all pixels in the region $\Omega_r \setminus O$, it is intractable to simultaneously estimate $\{\partial\Omega_i\}$ and $\{k_i\}$. To solve this problem, an iterative optimization algorithm is proposed, which operates in an alternate manner.

In Step 3, computing the optimal boundary for O_i is equivalent to finding a shortest closed path in a graph $\mathcal{G}_i$. The nodes in $\mathcal{G}_i$ are pixels within the band $\Omega_r \setminus O$ while the edges represent 4-connectivity relationships between neighboring pixels. For $\partial\Omega_i$, the cost $((f_t(p) - f_s(p)) - k_i)^2$ is defined on each node as the color difference with respect to k_i . Following Ref. [13], we break the band connectivity with the shortest boundary cut C_i connecting ∂O_i and $\partial\Omega_r$, as shown by black solid lines in Fig. 1, and remove all edges crossing the cut from the corresponding graph $\mathcal{G}_i$. In addition, to ensure each optimal boundary encloses only one object, we construct the boundary connection set $\{C_{ij}\}$ between objects of interest $\{O_i\}$, shown as black dotted lines in Fig. 1, and then apply 2D dynamic programming [26] to find the shortest closed path which connects the two sides of C_i as well as passing through the associated boundary connections.

3.4 Obtaining objects of interest

To compute the optimal boundaries, we need to firstly obtain the objects of interest $\{O_i\}$. Level sets [27] can be used to perform automatic object segmentation. However, as the editing of objects of interest usually needs user interaction (e.g., to only select certain objects), we rely on obtaining objects of interest by interactive segmentation techniques.

Algorithm 1 Computing optimal boundaries

Step 1: Initialise the boundaries $\partial\Omega_i = \partial O_i$, $i = 1, \dots, n$.

Step 2: Given each current $\partial\Omega_i$, compute the optimal k_i by taking the derivative of Eq. (3) and setting to zero:

$$\frac{\partial E_B(\{\partial\Omega_i\}, \{k_i\})}{\partial k_i} = 0 \Leftrightarrow k_i = \frac{1}{|\partial\Omega_i|} \sum_{p \in \partial\Omega_i} (f_t(p) - f_s(p)), i = 1, 2, \dots, n \quad (4)$$

where $|\partial\Omega_i|$ is the length of the boundary $\partial\Omega_i$. So k_i is the average color difference across the boundary $\partial\Omega_i$.

Step 3: Using the current k_i , the boundary $\partial\Omega_i$ is optimized using the shortest path algorithm.

Step 4: Repeat Steps 2 and 3 until the energy in Eq. (3) does not decrease in two successive iterations.

To do so, a user simply draws a box surrounding the object of interest, and then GrabCut [28] is applied to extract the object of interest. Note that even if the objects of interest are not obtained very accurately, our computation of optimal boundaries still can avoid structure inconsistency since the optimal boundaries usually pass through smooth regions rather than object edges.

4 Object-aware editing

4.1 Mixed-domain composition

Optimal editing region selection can preserve *local structural continuity* well. However, if the backgrounds of the input images differ greatly in color and texture (e.g., pasting a source of interest from a smooth region into a target image with coarse textures), color bleeding or texture artifacts still may occur when reconstructing from a gradient field, even with well-optimized boundary conditions. To address this issue, we must enhance color and texture consistency between images. Therefore, we perform mixed-domain composition by solving the following minimization problem:

$$\min_f E_C = \min_f \left(\sum_{p \in \Omega} D_g(p) + \lambda \sum_{p \in B} D_c(p) \right) \quad (5)$$

where the consistency energy E_C provides a unified consistency metric for measuring gradients and colors. The gradient-domain energy is defined on all targets of interest to obtain *smooth color transitions*, using $D_g(p) = \|\nabla f(p) - \mathbf{v}(p)\|^2$, while the color-domain energy is defined on backgrounds of interest to preserve *global texture consistency*. We adopt a patch-based metric for encoding texture consistency and define $D_c(p) = \|\mathbf{f}(p) - \mathbf{f}_t(p)\|^2 = \sum_{q \in \mathcal{N}_w^B(p)} (f(q) - f_t(q))^2$, where $\mathbf{f}(p)$ and $\mathbf{f}_t(p)$ are the background pixel sets in $w \times w$ image patches $\mathcal{N}_w(p)$ centered at p , and $\mathcal{N}_w^B(p) = \mathcal{N}_w(p) \cap B$. Unlike the IM method [9], this patch-based approach neither needs search nor voting. The trade-off parameter λ is used to balance the influence of the two terms.

Defining a binary mask M , the minimization problem may be written as

$$\min_f \sum_{p \in \Omega} (D_g(p) + \lambda M(p) D_c(p)) \quad (6)$$

where $M(p)$ is 1 if $p \in B$, and 0 otherwise. Based on the Euler–Lagrange equation, we obtain:

$$\nabla^2 f + \lambda M |\mathcal{N}_w^B| f = \nabla \cdot \mathbf{v} + \lambda M |\mathcal{N}_w^B| f_t, \quad \forall p \in \Omega \quad (7)$$

We make use of backward differences to discretely approximate Eq. (7), leading to a large sparse linear system:

$$(|\mathcal{A}(p)| - \lambda M(p) |\mathcal{N}_w^B(p)|) f(p) - \sum_{q \in \mathcal{A}(p) \cap \Omega} f(q) = \sum_{q \in \mathcal{A}(p) \cap \partial \Omega} f_t(q) + \sum_{q \in \mathcal{A}(p)} v_{pq} - \lambda M(p) |\mathcal{N}_w^B(p)| f_t(p) \quad (8)$$

where $|\mathcal{A}(p)|$ represents the number of available pixels in a 4-neighborhood $\mathcal{A}(p)$ centered at p ; it satisfies $|\mathcal{A}(p)| \leq 4$. v_{pq} is the projection of $\mathbf{v}((p+q)/2)$ on the edge $[p, q]$ in the direction of $\vec{pq}$, given by $v_{pq} = \mathbf{v}((p+q)/2) \cdot \vec{pq}$. The computational complexity for constructing the sparse linear system is $O(|\Omega|)$; the space complexity for storing the coefficient matrix is $O(|\Omega|)$. Many efficient technologies such as Ref. [29] can be adopted for solving the sparse linear system. We set $w = 9$ and $\lambda = 1$ in this paper.

4.2 Object-aware gradient manipulation

Unlike traditional composition methods [11, 15] which manipulate the guidance gradient field for the target region using a unified operator, OAIE manipulates the guidance gradient fields of O and B separately with independent operators, in an *object-aware* way.

4.2.1 Transformation operators

For objects of interest O , the gradient field can be manipulated by modifying ∇f_s using linear or nonlinear transformations:

$$\mathbf{v}_o = \tau(\nabla f_s), \quad \forall p \in O \text{ and } \tau \in \Lambda \quad (9)$$

where the transformation operator τ is used to change the appearance of targets of interest (e.g., their texture, illumination, contrast, color, etc.) and can be chosen from the transformation set $\Lambda = \{\text{clone}, \text{illuminate}, \text{smooth}, \text{recolor}, \dots\}$. Here, *clone* provides seamless cloning, *illuminate* provides local illumination change, *smooth* provides texture smoothing (e.g., cartoonization), *recolor* provides object recoloring, and so on. Note that different operators can be used for each object.

4.2.2 Mixture operators

For backgrounds of interest B , the gradient field is

manipulated by modifying or mixing source and target gradient fields:

$$\mathbf{v}_b = \phi(\nabla f_s, \nabla f_t), \quad \forall p \in B \text{ and } \phi \in \Theta \quad (10)$$

where the mixture operator ϕ is chosen from the set $\Theta = \{src, max, min, avg, \dots\}$ where *src* stands for using the source gradient, *max/min/avg* stand for picking the larger / smaller / average gradient from ∇f_s and ∇f_t , etc.

4.2.3 Combination

Finally, the gradient fields from Eqs. (9) and (10) are combined to form the new guidance gradient field in editing region Ω . This new guidance gradient field is substituted into Eqs. (5)–(8).

4.3 Algorithm comparison

OAIE performs composition with gradient fields in an optimized target region Ω with boundary conditions on $\partial\Omega$, as well as colors in B . By contrast, PIE reconstructs images from gradient fields in Ω_r with non-optimized boundary conditions on $\partial\Omega_r$ while DDP performs this process with gradient fields in Ω and optimized boundary conditions on $\{\partial\Omega_i\}$. ETIC optimizes the boundary conditions by minimizing the curl of the gradients on $\partial\Omega_r$. IM carries out image composition with patch-based synthesis using the textures in $\Omega_r \setminus \Omega$.

5 Experimental results

To demonstrate the effectiveness of our OAIE approach, we have implemented several seamless image editing applications and compare the results with those from four state-of-the-art methods: PIE [11], DDP [13], ETIC [10], and IM [9]. Various such results are given below.

5.1 Seamless cloning

Seamless cloning aims to copy a region of interest from a source image and seamlessly insert it into a target image. In this process, the guidance gradient fields of the editing region Ω and backgrounds of interest B are manipulated with Eqs. (9) and (10) respectively.

Figure 2 shows a seamless cloning example which combines a region of interest containing multiple objects with a target image. The result of direct cloning is shown in Fig. 2(c), with object boundaries and optimal boundaries marked in blue and red

respectively. Results generated by the four methods PIE, DDP, ETIC, and IM all have artifacts in the background region, e.g., around the object boundaries. However, OAIE obtains a more natural result since it considers a unified constraint on both object appearance and background texture simultaneously.

Figure 3 provides a further seamless cloning example which inserts a source region with a smooth background into a target image with rich textures. None of the other four methods preserve texture consistency well (leading to unnatural colors or textures), while OAIE obtains consistency of both local structures and global textures.

5.2 Selective modification

Selective modification aims to adjust local or global appearance (e.g., texture, color, illumination, etc.), so image cloning methods such as DDP, ETIC, and IM are not applicable. We compared our results from OAIE with those from the PIE method. During selective modification, optimal editing region selection is neglected, so $f_s = f_t$, while the gradient fields of O and B can be obtained using the same or different operators.

Figure 4 demonstrates examples of global and local selective modification. Figure 4(a) demonstrates global recoloring of Fig. 3(h) with the *recolor* operator. This operator multiplies the RGB channels of the original image by different values respectively to form the source image, and then performs seamless composition. Figure 4(b) shows the result of using operator *illuminate* to reduce local specular reflections (shown in blue in Fig. 4(a)). The operator *illuminate* modifies the gradient field using $\mathbf{v} = \alpha^\beta |\nabla f|^{-\beta} \nabla f$ where $\alpha = \beta = 0.2$, following Ref. [11]. We see that when either global or local operators are applied to the cloned composition, the resulting textures still remain natural, verifying the robustness of OAIE.

Figure 5 presents another example of selective modification which changes the local color of a target of interest. In Fig. 5(a), the object and target boundaries are marked in blue and red respectively. The PIE result exhibits color bleeding artifacts around the object, while OAIE avoids this problem effectively. To explain why, let us consider the difference between the minimization problems

Fig. 2 Seamless cloning for combining a region containing multiple objects with a target image. Results generated by the four other methods have artifacts in the background while OAIE provides a more natural result since it considers a unified constraint on both object appearance and background texture simultaneously.

solved by PIE and OAIE. Compared to PIE, the minimization problem for OAIE includes an extra texture consistency term defined using a patch-based approach. It has been demonstrated in Ref. [9] that patch-based synthesis can produce a gradual transition between images without sacrificing texture sharpness. This allows OAIE to produce a smooth transition across image edges without color bleeding and texture blurring.

5.3 Seamless stitching

Seamless stitching aims to combine multiple source images into a panorama mosaic. Figure 6 illustrates an example of stitching two remote sensing images with a very large tone difference. Optimal seam methods will lead to obvious artifacts due to the color difference between the source images in the overlap region (shown in blue in Fig. 6(b)). Since consistency constraints are imposed for both color and texture, OAIE achieves a more globally natural result than

PIE, as shown in Figs. 6(c) and 6(d).

In addition, OAIE can stitch multiple target parts, as shown in Fig. 7, where several face parts from different face images are stitched into an integrated face portrait. The upper row shows the target image and source parts marked in red in the source images, while the bottom row gives the results generated by cloning, PIE, DDP, and OAIE respectively. Various structure or color artifacts can be seen in the results generated from PIE and DDP. In contrast, the result from OAIE is more natural. A potential application of this editing tool is portrait synthesis for police work.

6 Discussion

6.1 Limitations

Like other methods, our method has some limitations when the background structures between source and target images in the editing region collide. As

Fig. 3 Seamless cloning when combining images with large differences in texture. OAIE obtains consistent local structures and global textures while none of the other methods preserve texture consistency well (leading to unnatural colors or textures).

Fig. 4 Selective modification to provide global recoloring and local illumination change. Even if global or local operators are applied to the cloned composition, the resulting texture still remains natural.

shown in Fig. 8, the result generated by our method suffers from color bleeding around the background structures, while the results from PIE and DDP have serious structure artifacts.

6.2 Computational efficiency

Our method is simple and efficient. Firstly, the user selects and drags the region of interest from the source image to the target image. Then, simple interactive segmentation allows the user to indicate the objects of interest. This usually takes about 1–2 seconds of computation depending on the sizes of the objects. After specifying the type of editing required, the composition is performed automatically by using dynamic programming to find optimal boundaries and solving a large sparse linear system for mixed-domain composition. The computational complexity of dynamic programming is $O(n)$, while solving a large sparse linear system takes time $O(n)$ [29] where n is the number of variables.

6.3 Patch-level similarity

In our mixed-domain composition, we adopt patch-

based synthesis to preserve texture consistency, so texture similarity is measured at patch-level rather than at pixel-level. Generally, patch-based synthesis can preserve more texture cues, giving less smooth results than pixel-based synthesis. Figure 9 shows an example, where the results provide a close up of part of Fig. 2, using different synthesis methods. The results demonstrate that patch-based synthesis provides more rich texture information.

6.4 Editing quality estimation

Generally, subjective measures are used to evaluate editing quality. This is usually performed by using the human eyes to check the naturalness of the editing results in terms of structure, color, and texture. To verify the advantages of our method, we used blind image quality evaluation as proposed in Ref. [30], considering the image naturalness in the local editing regions. The experimental results showed that our method can give better quality.

7 Conclusions

In this paper, we have proposed a general approach called OAIE for seamless image editing. It jointly performs optimal editing region selection and mixed-domain composition, allowing it to cope with visual inconsistency in local structures and global textures simultaneously. In particular, OAIE provides a unified mixed-domain consistency measure for gradients and colors, in which texture consistency is encoded with a patch-based approach which does

Fig. 5 Selective modification for changing colors of a target of interest. OAIE can provide a smooth transition across image edges without color bleeding and texture blurring, while the PIE result exhibits color bleeding artifacts around the object.

Fig. 6 Seamless stitching two remote sensing images with very large tone difference. Optimal seam methods lead to obvious artifacts due to the color difference between source images. The PIE result contains inconsistent textures while OAIE achieves a more globally natural result.

not require search. Compared to four state-of-the-art methods, our unified approach is more powerful in preventing global color and texture inconsistencies, while preserving local structure continuity.

Acknowledgements

This work was supported in part by the National Key Research and Development Plan (Grant No. 2016YFC0801005), the National Natural Science

Foundation of China (Grant Nos. 61772513 and 61402463), and the Open Foundation Project of Robot Technology Used for Special Environment Key Laboratory of Sichuan Province in China (Grant No. 16kftk01).

References

- [1] Philip, S.; Summa, B.; Tierny, J.; Bremer, P. T.; Pascucci, V. Distributed seams for gigapixel panoramas. *IEEE Transactions on Visualization and Computer Graphics* Vol. 21, No. 3, 350–362, 2015.
- [2] Agarwala, A.; Dontcheva, M.; Agrawala, M.; Drucker, S.; Colburn, A.; Curless, B.; Salesin, D.; Cohen, M. Interactive digital photomontage. *ACM Transactions on Graphics* Vol. 23, No. 3, 294–302, 2004.
- [3] Cheng, M.-M.; Zhang, F.-L.; Mitra, N. J.; Huang, X.; Hu, S.-M. RepFinder: Finding approximately repeated scene elements for image editing. *ACM Transactions on Graphics* Vol. 29, No. 4, Article No. 83, 2010.
- [4] Barnes, C.; Zhang, F.-L.; Lou, L.; Wu, X.; Hu, S.-M. PatchTable: Efficient patch queries for large datasets and applications. *ACM Transactions on Graphics* Vol. 34, No. 4, Article No. 97, 2015.
- [5] Li, J.; Tian, Y.; Huang, T. Visual saliency with statistical priors. *International Journal of Computer Vision* Vol. 107, No. 3, 239–253, 2014.

Fig. 7 Seamless stitching multiple target parts. Various structure or color artifacts (e.g., around the glasses and mouth) can be found in the results generated by the PIE and DDP methods, while the OAIE result is more natural.

Fig. 8 Limitations of our method and other methods. Color bleeding exists in our result, while the results from PIE and DDP have serious structure artifacts.

Fig. 9 Composition using patch-based synthesis (above) and pixel-based synthesis (below). Patch-based synthesis provides richer texture cues.

- [6] Li, J.; Duan, L. Y.; Chen, X.; Huang, T.; Tian, Y. Finding the secret of image saliency in the frequency domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 37, No. 12, 2428–2440, 2015.
- [7] 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*, 341–346, 2001.
- [8] Kwatra, V.; Schödl, A.; Essa, I.; Turk, G.; Bobick, A. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics* Vol. 22, No. 3, 277–286, 2003.
- [9] Darabi, S.; Shechtman, E.; Barnes, C.; Goldman, D. B.; Sen, P. Image melding: Combining

- inconsistent images using patch-based synthesis. *ACM Transactions on Graphics* Vol. 31, No. 4, Article No. 82, 2012.
- [10] Tao, M. W.; Johnson, M. K.; Paris, S. Error-tolerant image compositing. *International Journal of Computer Vision* Vol. 103, No. 2, 178–189, 2013.
- [11] Pérez, P.; Gangnet, M.; Blake, A. Poisson image editing. *ACM Transactions on Graphics* Vol. 22, No. 3, 313–318, 2003.
- [12] Zomet, A.; Levin, A.; Peleg, S.; Weiss, Y. Seamless image stitching by minimizing false edges. *IEEE Transactions on Image Processing* Vol. 15, No. 4, 969–977, 2006.
- [13] Jia, J.; Sun, J.; Tang, C.-K.; Shum, H.-Y. Drag-and-

- drop pasting. *ACM Transactions on Graphics* Vol. 25, No. 3, 631–637, 2006.
- [14] Farbman, Z.; Hoffer, G.; Lipman, Y.; Cohen-Or, D.; Lischinski, D. Coordinates for instant image cloning. *ACM Transactions on Graphics* Vol. 28, No. 3, Article No. 67, 2009.
- [15] Bhat, P.; Zitnick, C. L.; Cohen, M.; Curless, B. GradientShop: A gradient-domain optimization framework for image and video filtering. *ACM Transactions on Graphics* Vol. 29, No. 2, Article No. 10, 2010.
- [16] Li, X. Y.; Gu, Y.; Hu, S.-M.; Martin, R. R. Mixed-domain edge-aware image manipulation. *IEEE Transactions on Image Processing* Vol. 22, No. 5, 1915–1925, 2013.
- [17] Sadek, R.; Facciolo, G.; Arias, P.; Caselles, V. A variational model for gradient-based video editing. *International Journal of Computer Vision* Vol. 103, No. 1, 127–162, 2013.
- [18] Bie, X.; Wang, W.; Sun, H.; Huang, H.; Zhang, M. Intent-aware image cloning. *The Visual Computer* Vol. 29, Nos. 6–8, 599–608, 2013.
- [19] Hua, M.; Bie, X.; Zhang, M.; Wang, W. Edge-aware gradient domain optimization framework for image filtering by local propagation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2838–2845, 2014.
- [20] Zhang, Y.; Ling, J.; Zhang, X.; Xie, H. Image copy-and-paste with optimized gradient. *The Visual Computer* Vol. 30, No. 10, 1169–1178, 2014.
- [21] Ma, L.-Q.; Xu, K. Efficient manifold preserving edit propagation with adaptive neighborhood size. *Computers & Graphics* Vol. 38, 167–173, 2014.
- [22] Luo, S. J.; Sun, Y. T.; Shen, I. C.; Chen, B. Y.; Chuang, Y. Y. Geometrically consistent stereoscopic image editing using patch-based synthesis. *IEEE Transactions on Visualization and Computer Graphics* Vol. 21, No. 1, 56–67, 2015.
- [23] Chen, T.; Cheng, M.-M.; Tan, P.; Shamir, A.; Hu, S.-M. Sketch2Photo: Internet image montage. *ACM Transactions on Graphics* Vol. 28, No. 5, Article No. 124, 2009.
- [24] Zhang, F. L.; Wang, J.; Shechtman, E.; Zhou, Z. Y.; Shi, J. X.; Hu, S. M. PlenoPatch: Patch-based plenoptic image manipulation. *IEEE Transactions on Visualization and Computer Graphics* Vol. 23, No. 5, 1561–1573, 2016.
- [25] Lee, J. H.; Choi, I.; Kim, M. H. Laplacian patch-based image synthesis. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2727–2735, 2016.
- [26] Mortensen, E. N.; Barrett, W. A. Intelligent scissors for image composition. In: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, 191–198, 1995.
- [27] Sethian, J. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Sciences*. Cambridge University Press, 1999.
- [28] Rother, C.; Kolmogorov, V.; Blake, A. “GrabCut”: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics* Vol. 23, No. 3, 309–314, 2004.
- [29] Krishnan, D.; Szeliski, R. Multigrid and multilevel preconditioners for computational photography. *ACM Transactions on Graphics* Vol. 30, No. 6, Article No. 177, 2011.
- [30] Mittal, A.; Soundararajan, R.; Bovik, A. C. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters* Vol. 20, No. 3, 209–212, 2013.

Shiming Ge is an associate professor in the Institute of Information Engineering, Chinese Academy of Sciences. He received his B.S. and Ph.D. degrees from the University of Science and Technology of China. His research mainly focuses on computer vision, deep learning, and AI security.

He is a senior member of the IEEE and a member of the ACM.

Xin Jin received his Ph.D. degree in computer science from Beihang University. Currently, he is an assistant professor in Beijing Electronic Science and Technology Institute, China. His research mainly focuses on visual computing and visual media security.

Qiting Ye received his B.S. degree in computer science from Peking University in 2015. He is now a master student in the Institute of Information Engineering, Chinese Academy of Sciences. His major research interests lie in computer vision and deep learning.

Zhao Luo is currently a master candidate in the Institute of Information Engineering, Chinese Academy of Sciences. He received his B.S. degree from the University of Electronic Science and Technology of China. His research interests are object tracking and deep learning.

Qiang Li is currently a professor in Southwest University of Science and Technology. He received his Ph.D. degree from the University of Science and Technology of China. His research mainly focuses on the Internet of things and intelligent information processing.

Open Access The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits

unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.