Robust Algorithm for Exemplar-based Image Inpainting

Wen-Huang Cheng[1], Chun-Wei Hsieh[2], Sheng-Kai Lin[2], Chia-Wei Wang[2], and Ja-Ling Wu[1,2]
[1]Graduate Institute of Networking and Multimedia
[2]Department of Computer Science and Information Engineering
National Taiwan University
{wisley, mikemage, seka, nacci, wjl}@cmlab.csie.ntu.edu.tw

Abstract

This paper presents a robust algorithm for exemplar-based image inpainting. We focus on the development of a generic priority function to provide robust performance. With user intervention, the algorithm is easy to be applied to any image contents with different characteristics. We conduct experiments on several kinds of still images and demonstrate the effectiveness of the proposed algorithm. The performance is evaluated in terms of the visual quality and the perceptual reasonability.

1. Introduction

In multimedia signal processing, image inpainting [1] is the technology generally applied to the problem of automatic filling-in the missing regions of an image in a visually plausible way. In the last few years, it has been extensively studied in research fields and benefited in most electronic imaging applications [10], such as digital photo editing [3] (e.g., foreground removal), image restoration [11] (e.g., artwork repairing), and multimedia transmission [12] (e.g., lost data recovering).

In the literature, conventional schemes that proposed for image inpainting can be divided into two categories: *texture-oriented* [2] and *structure-oriented* [3]. The texture-oriented scheme (usually known as the texture synthesis scheme) generates the target region with available sample textures from its surroundings [2]. This approach is specifically useful for the images with large texture areas. On the other hand, the structure-oriented scheme obtains the missing regions via data fusion techniques, such as the bilinear interpolation. As shown in [3], the linear structures (i.e., edges or object boundaries) can be preserved and propagated to the target regions. However, since most ordinary images are not composed of pure texture or pure structure, better performance is expected for those taking advantages of both schemes.

Approaches taking both the texture and the structure into account are then presented. Bertalmio *et al* [4] propose to decompose each original image into two separate component images of different characteristics. One of them is processed by texture-oriented scheme and the other by structure-oriented one. The two processed components are added back to reconstruct the missing regions. Similar idea is shown in [5], instead of generating two new component images, the original one is segmented into two kinds of sub-regions. These works first attempt to provide a unified processing mechanism for image inpainting, but they suffer serious blurring artifacts and are shown to be merely suitable for small target regions.

Recently, the category of exemplar-based inpainting scheme is proposed, in which the visible parts of the image serve as an source set of examples to infer the target regions. That is, the missing data is "replicated" rather than "synthesized" from available information. Compared with other kinds of approaches, exemplar-based approach is very effective in reducing the undesired blurring artifacts and applicable to both the small and large image gaps. Harrison [8] first presents an exemplar-based algorithm focusing on the object removal, but in this work the linear structures are often overruled by nearby noises. Next, Drori *et al* [6] iteratively approximates the missing regions from coarse to fine levels using the principle of self-similarity. They show impressive experimental results but the reported computation time is extremely long (between 83 and 158 minutes for a 384×223 sized image). Although some fast algorithms [7] are developed, the visual quality is accordingly degraded. Criminisi *et al* [9] exploit a patch-based algorithm, in which the filling order is decided by a predefined priority function to ensure that the linear structures will propagate before texture filling to preserve the connectivity of object boundaries. The performance is compatible to previous techniques and better speed efficiency is obtained. Nevertheless, we will show that the priority function adopted in [9] is not well-defined and becomes very unreliable after numbers of iterations. Further, for different ordinary images, the variations of relative importance between the texture and the structure characteristics have still been ignored.

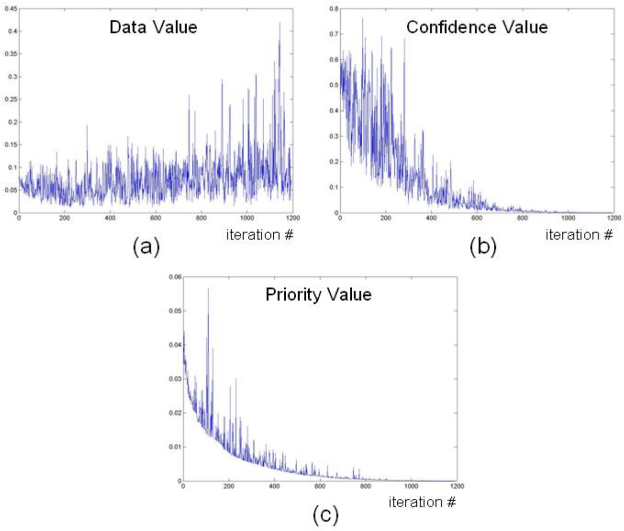

Figure 1 Behavior Illustrations of (a) data term, (b) confidence term, and (c) priority function for the filling-in process of an image inpainting task.

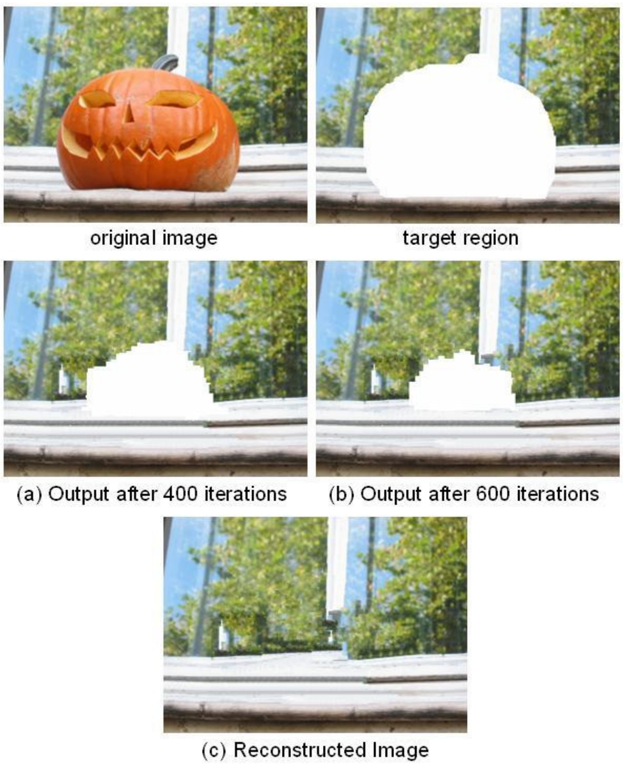

Figure 2 Dropping effect on reconstruction with large target region.

In this paper, instead of creating a new inpainting scheme, we try to generalize the priority function for the family of algorithms given in [9] to provide more robust performance. And the generalized algorithms can be applied to any image contents with different characteristics. In the experiments, we show that our approach is superior to what was early proposed and would be more reliable and flexible for practical applications.

The rest of this paper is organized as follows. Section 2 presents the key observations and the major ideas for conducting robust image inpainting. Experimental results and conclusion are given in Sections 3 and 4, respectively.

2. Robust image inpainting

First, for the convenience of explanation, similar notations that used in the inpainting literature [3][9] are adopted. The entire image Λ is composed of two disjoint regions: the source region Φ and the target region Ω. The source region is defined to be the visible part and the target region is the missing one. Additionally, $\delta\Omega$ represents the pixel set of the target region boundary.

Generally, an exemplar-based inpainting algorithm includes the following four main steps:

1) *Initializing the Target Region*, in which the initial missing areas are extracted and represented with appropriate data structures.
2) *Computing Filling Priorities*, in which a predefined priority function is used to compute the filling order for all unfilled pixels $p \in \delta\Omega$ in the beginning of each filling iteration.
3) *Searching Example and Compositing,* in which the most similar example is searched from the source region Φ to compose the given patch Ψ_p (of size $N \times N$ pixels) that centered on the given pixel p.
4) *Updating Image Information,* in which the boundary $\delta\Omega$ of the target region Ω and the required information for computing filling priorities are updated.

As shown in [9], the reconstructed visual quality and the reasonability of the filled image are mainly influenced by the filling order. Therefore, it's our belief that better performance is reachable via developing a robust priority function. Besides, in this section, we will show that it's also important to take the content characteristics of different images into account for promoting robustness.

2.1. Reforming the priority function

In Criminisi's original algorithm [9], the priority function P(p) for each unfilled pixel $p \in \delta\Omega$ is defined as the product of two terms as follows.

$$P(p) = C(p)D(p), \tag{1}$$

where $C(p)$ is called the *confidence term* and $D(p)$ is called the *data term*. The confidence term is used to capture the texture property and the data term represents the structure characteristic. They are respectively defined as follows [9].

$$C(p) = \frac{\sum_{q \in \Psi_p \cap \Phi} C(q)}{|\Psi_p|}, \quad 0 \le C(p) \le 1, \tag{2}$$

and

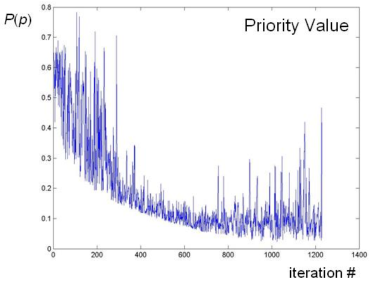

Figure 3 Priority values of adding the confidence terms and data terms of Fig. 1.

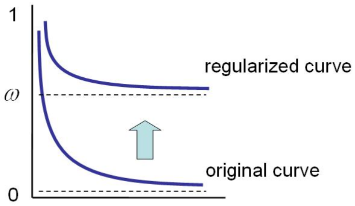

Figure 4 Illustration of the curve regularization.

$$D(p) = \frac{|\nabla_p^{\perp} \cdot n_p|}{\alpha},\ 0 \le D(p) \le 1, \tag{3}$$

where α, $\nabla_p^{\perp}$, and n_p are respectively the normalization factor (e.g., α = 255), the isophote vector, and the unit vector orthogonal to the front $\delta\Omega$ in the point p. The detailed definition and mathematical derivation are referred to [9, Sec. 3].

In our experiments, we find that the confidence value drops rapidly to zero as the filling process proceeds, which makes the computed priority values undistinguishable, and in turn, results in incorrect filling orders. For short, we name this phenomenon as "dropping effect" in the rest of this paper. As shown in Fig. 1, even if the data values increases, the shape of the priority curve is completely dominated by the exponentially dropping confidence values. The drawback becomes more serious and noticeable when filling large target regions. Fig. 2 gives a real example. In the beginning, structures of the salient window frames are correctly propagated in both the horizontal and the vertical directions (cf. Fig. 2(a)). However, when the dropping effect occurs, the uncertainty of the priority values makes the target region starting to be randomly filled and breaks the vertical frame structures sharply, as shown in Fig. 2(b). Errors continuously propagate to the central part of the reconstructed image and noticeable visual artifact results (cf. Fig. 2(c)).

It is well known that numerical multiplication is essentially sensitive to extreme values and their influences are usually over amplified. The original priority function $P(p)$, defined in multiplicative form, needs to be reformulated to avoid the prescribed deficiency. From mathematical viewpoint, the response of an additive function is linearly proportional to its input and is more stable to unexpected noise and variations.

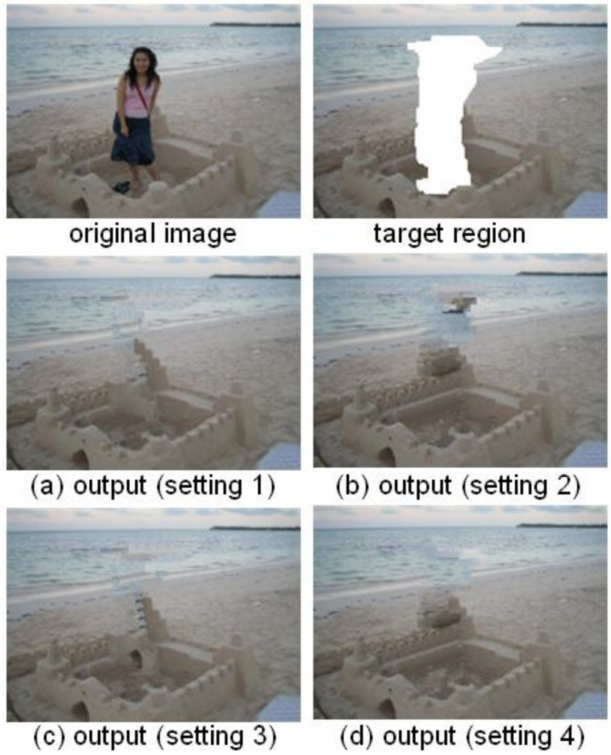

Figure 5 Importance of component weights. (a) to (d) are reconstruction examples with different settings of component weights.

Therefore, first, we modify the priority function to an additive rather than a multiplicative form as follows:

$$P(p) = C(p) + D(p). \tag{4}$$

Obviously, from Fig. 3, the resulting treatment of the confidence and the data terms is more balanced. However, direct combination of $C(p)$ and $D(p)$ is still unreasonable due to the significant differences in their curve behavior (cf. Fig. 1). For example, as shown in Fig. 3, the priority values are dominated by the confidence and the data terms in the left and the right halves, respectively. This implies that where should be filled next depends not only on the image characteristics but is also time dependent. Basically, the values of $D(p)$ are consistent except there places where strong structures exist, i.e., the sharp peaks in Fig. 1. The $C(p)$ values are purposely designated to decay but the decreasing rate is too fast such that the value variations of $C(p)$ are significant. Therefore, in our work, we use a regularizing function $R_C(p)$ to smooth the curve of the confidence term $C(p)$ to match that of the data term as follows.

$$R_C(p) = (1-\omega) \times C(p) + \omega,\ 0 \le \omega \le 1, \tag{5}$$

where ω is the regularizing factor for controlling the curve smoothness. Without loss of generality, in our work, ω is empirically set to 0.7. Note that the curve becomes smoother but its shape is fully preserved. For example, the value range of $C(p)$ is regularized to [ω, 1] (cf. Fig. 4), but the interrelationship of $C(p)$ is unchanged.

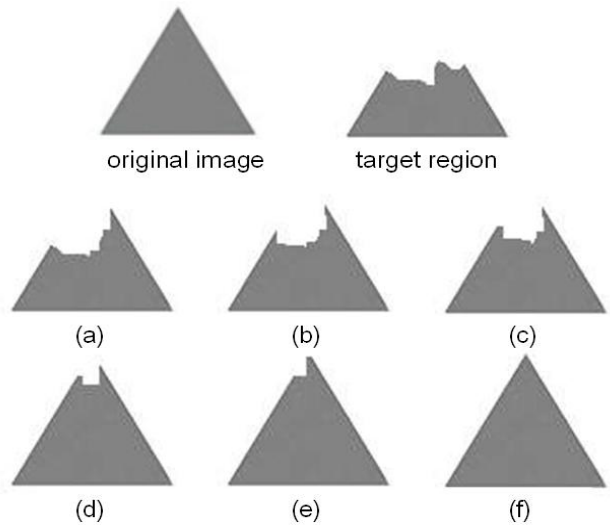

Figure 6 Filling-in process of the proposed algorithm on a synthesis image. (The filling order is from (a) to (f).)

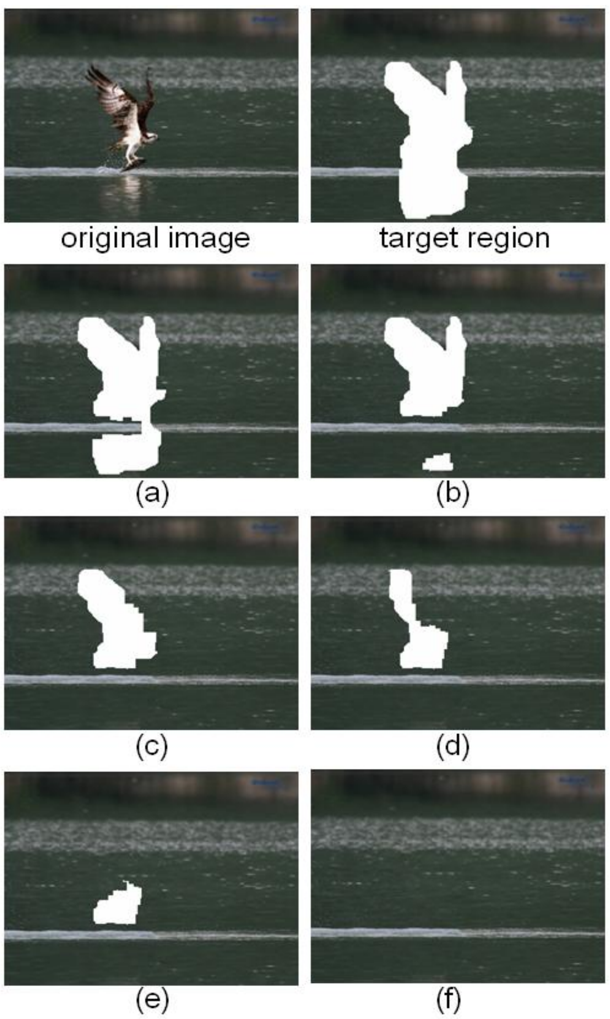

Figure 7 Filling-in process of the proposed algorithm on a natural image. (The filling order is from (a) to (f).)

In this way, the new priority function is able to resist undesired noises and robust to the aforementioned over amplified phenomenon. More importantly, we assure that the decided filling order is more trustful and depends only on the image itself by parallelizing the curve behaviors of the function component terms.

2.2. Selecting the component weights

For different images, the relative importance of texture and structure characteristics should be carefully considered. For example, in Fig. 5, the target region is filled with four different weight settings of confidence and data terms. Higher weights for the data and the confidence terms are respectively given in the settings 1, 3 and the settings 2, 4. Correspondingly, the generated sand towers have stronger structures with higher data weight (cf. Fig. 5(a) and (c)) and the others are mainly filled with the sand textures (cf. Fig. 5(b) and (d)). For this image, the outputs with the settings 1, 3 are obviously preferred by end users because the reconstructed tower contours are more reasonable. It is also worthy to notice that, even with similar settings, the results may be visually different such as the presence and absence of the central tower door in Fig. 5(a) and (c), respectively. There are strong connections between the component weights and the resulting image. The combination of the priority function needs more flexibility to achieve better visual quality according to the image contents. Therefore, we define the robust priority function $RP(p)$ as follows.

$$RP(p) = \alpha \cdot R_C(p) + \beta \cdot D(p)\,,\ 0 \le \alpha, \beta \le 1\,, \qquad (6)$$

where α and β are respectively the component weights of the confidence and the data terms. Note that $\alpha + \beta = 1$.

Based on our observations, the bottleneck of exemplar-based inpainting algorithm lies in the search for best matching example from source regions. Even if we need more computations in the proposed priority function (two multiplications and one addition versus one multiplication), the overall computation is fast enough and very close to that of [9]. Our experiments confirm this claim. Therefore, it may not be reasonable but currently acceptable to decide an optimal setting of the component weights by exhausted computer search. Of course, finding more efficient searching algorithms will be one of future works.

3. Experimental results

To verify the effectiveness of the proposed priority function $RP(p)$, we conduct some experiments and compare the so-obtained results with the conventional approach [9], i.e., $P(p)$. The testing data are carefully selected from both natural and synthetic images. For fair comparison, except the priority functions, all the other processing algorithms (e.g., exemplar searching) are the same for both approaches.

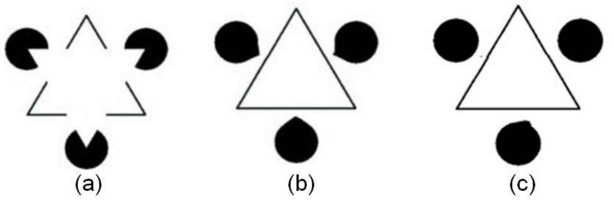

Figure 8 Comparison of reconstructions on (a) "Kanizsa triangle" between (b) the conventional approach and (c) the proposed algorithm.

Figure 9 Comparison of reconstructions on the image in Fig. 5 between (a) the conventional approach and (b) the proposed algorithm.

The most important factor that influences the reconstructed quality is the pixel filling order. Figs. 6 and 7 illustrate two examples of the reconstructed steps with the proposed priority function. In Fig. 6, the inpainting results (cf. Fig. 6(f)) of a partially removed triangle are obtained, as shown in Fig. 6(a) to (e). Clearly, the structures (i.e., the triangle boundary) are filled far before the remaining internal textures to preserve the shape. With the reliable priority values, the two boundaries are correctly propagated until reaching their convergence. A perfect triangle is then obtained. In contrast, the same experiment is performed with the conventional approach and an incomplete triangle is generated as shown in Fig. 8 of [9]. Due to the aforementioned dropping effect, the pixels near the upper triangle vertex are randomly filled and the shape continuity is seriously destroyed. Besides, according to the filling progress, we noticed that the boundary structure and the internal texture are almost propagated simultaneously (cf. Fig. 8 of [9]). That's due to the ignorance of the image characteristics (i.e., stronger structure) and the fixed assignment of the component weights. Fig. 7 gives another example of object removal in a natural image, in which two horizontal structures are clearly observed. Similar to Fig. 6, the structures are correctly propagated until all surface edges are completed (cf. Fig. 7(a) to (e)) without the dropping effect. Even with complex water textures, the reconstructed image is visually plausible as given in Fig. 7(f). It demonstrates that the proposed priority function effectively avoids the danger of structure distortion and is well-defined for both of the synthetic and the natural images.

Next, an example of the "Kanizsa triangle" [9] is illustrated in Fig. 8(a). The results of the conventional

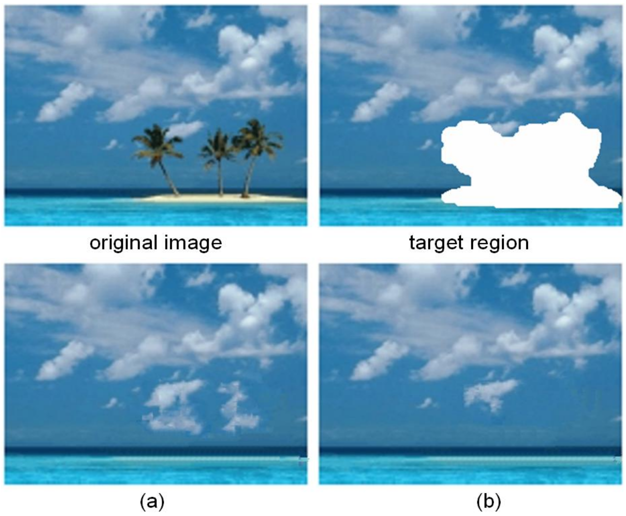

Figure 10 Comparison of reconstructions on a natural image between (a) the conventional approach and (b) the proposed algorithm.

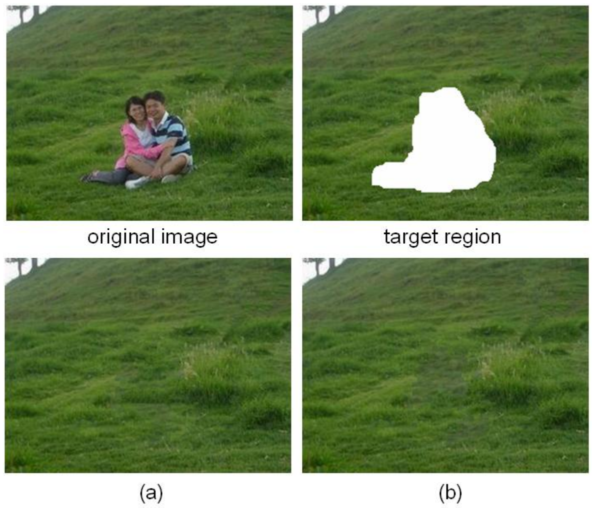

Figure 11 Reconstruction examples of the proposed algorithm on a natural image with different component weights in (a) and (b).

approach and the proposed algorithm are given in Figs. 8(b) and (c), respectively. Although the imperfections of circle distortions are reported in both algorithms, the causes are different. In Fig. 8(b), it's not reasonable for the structures to outward propagate against the circle center. The result should be more likely to be two intersected linear lines according to the exemplar-based inpainting algorithm. To analyze further, we find that the imperfections arose from the filling disorder such that redundant pixels are added to change the curvature. On the contrary, in the proposed approach, the filling orders are correctly computed and most of the reconstructed circles are next to perfect. Based on our observations, the only failure (i.e., the circle below the triangle) is caused by the false search of the best matching example. At the same time, the decided priority values and the filling order are still reasonable in the proposed priority function $RP(p)$.

Figs. 9(a) and (b) respectively show the filled results of the image in Fig. 5. By assigning suitable component weights, the performance of the proposed algorithm is visually superior to that of the conventional approach. Note that both the coastline and the castle contours are smoothly connected. Moreover, the water and sand textures are also seamlessly propagated. In Fig. 9(a), the whole image may be visually acceptable, but some artifacts that caused by the dropping effect are noticeable. First, compared with Fig. 9(b), the sand tower is not complete. The left side is sharply cut out. Fixed component weights make the sand textures be filled before the castle structures. Similar problem occurs around the coastline above the tower spire. The areas of beach and sea are falsely mixed.

Finally, flexibility of the component weights is demonstrated to improve the perceptual reality and shown be able to take user's preference into account. Fig. 10 gives one example. Both the results of the conventional approach (Fig. 10(a)) and the proposed algorithm (Fig. 10(b)) are visually plausible, but our results are more similar to the scene of the original image. As shown in Fig. 10(a), the generated cloud is too close to the horizontal line and becomes visually strange. By fixed assignment of the component weights, there's no opportunity to adjust the resultant output. In contrast, with the help of dynamic setting selection, the finest result can be subjectively picked up by end users from the set of possible candidates. By this way, it's our belief that the result of Fig. 10(b) is more realistic and matches our perception well. On the other hand, even if the content characteristics of an image are understood in advance, the flexibility of weights selection helps to satisfy the preferences for different users. For example, Fig. 11 illustrates an image with grass-texture. Two outputs (Figs. 11(a) and (b)) are obtained from the proposed algorithm by setting different component weights. Both the target regions are reasonably filled. However, unlike the previous example (i.e., Fig. 10), it's difficult to tell which one would be superior in reality or visual perception. The chosen result depends on the subject of different users. Therefore, we involve end users in the recovering procedure and provide them all possible candidates. In this way, end users are not only passive content receivers but also active content creators.

Conclusions

This paper presents a robust algorithm for example-based image inpainting, which can be adapted to any image contents of different characteristics. The main contribution of this work is the development of a generic priority function for fairly integrating the structure and the texture information to facilitate the image reconstruction. Experimental results show that the proposed algorithm is effective in both the visual quality improvement and the user preference consideration. The computational complexity of the proposed algorithm is dominated by two tasks: exemplar search and component weight selection. Therefore, our future work will focus on the investigation of efficient searching scheme and on the automatic discovery of component weights for different kinds of images.

Acknowledgements

This work was partially supported by the CIET-NTU(MOE) and National Science Council of Taiwan, China under the contract No. NSC93-2622-E-002-033, NSC93-2752-E-002-006-PAE and NSC93-2213-002-006.

References

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *ACM Comput. Graph. (SIGGRAPH 2002)*, July 2000, pp. 417-424.

[2] A. Efros and T. Leung, "Texture synthesis by nonparametric sampling," in *Proc. Int. Conf. Computer Vision*., Kerkyra, Greece, Sept. 1999, pp. 1033-1038.

[3] R. Bornard, E. Lecan, L. Laborelli, and J.-H. Chenot, "Missing data correction in still images and image sequences," in *ACM Multimedia Conf.*, 2002, pp. 355-361.

[4] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. Image Processing*, vol. 12, no. 8, pp. 882-889, August 2003.

[5] H. Grossauer, "A combined PDE and texture synthesis approach to inpainting," *Lecture Notes in Computer Science*, vol. 3022, pp. 214-224, 2004.

[6] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," in *ACM Comput. Graph. (SIGGRAPH 2003)*, 2003, pp. 303-312.

[7] S. Borikar, K. K. Biswas, and S. Pattanaik, "Fast algorithm for completion of images with natural scenes," *CS Technical Report*, University of Central Florida, April 2004.

[8] P. Harrison, "A nonhierarchical procedure for re-synthesis of complex texture," in *Proc. Int. Conf. Central Europe Computer Graphics, Visualization, and Computer Vision*, Plzen, Czech Republic, Feb. 2001.

[9] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Processing*, vol. 13, no. 9, pp. 1200-1212, Sept. 2004.

[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. New Jersey: Prentice-Hall, 2002.

[11] S.-C. Pei, Y.-C. Zeng, and C.-H. Chang, "Virtual restoration of ancient Chinese paintings using color contrast enhancement and lacuna texture synthesis," *IEEE Trans. Image Processing*, vol. 13, no. 3, pp. 416-429, March 2004.

[12] J. Park, D.-C. Park, R. J. Marks, and M. A. El-Sharkawi, "Content-based adaptive spatio-temporal methods for MPEG repair," *IEEE Trans. Image Processing*, vol. 13, no. 8, pp. 1066-1077, August 2004.