Journal of Information Processing Vol.24 No.6 897–907 (Nov. 2016)

[DOI: 10.2197/ipsjjip.24.897]

Regular Paper

Texture Transfer Based on Energy Minimization for Painterly Rendering

Yuta Tsunematsu[1] Norihiko Kawai[1,a)] Tomokazu Sato[1,b)] Naokazu Yokoya[1,c)]

Received: April 13, 2016, Accepted: August 10, 2016

Abstract: Non-photorealistic rendering (NPR) creates images with artistic styles of paintings. In this field, a number of methods of converting photographed images into non-photorealistic ones have been developed, and can be categorized into filter-based and exemplar-based approaches. In this paper, we focus on the exemplar-based approach and propose a novel method which transfers a style of a reference pictorial image to a photographed image. Specifically, we first input a pair of target and reference images. The target image is converted by minimizing an energy function which is defined based on the difference in intensities between an output image and a target image, and the pattern dissimilarity between an output image and a reference image. The proposed method transfers structures and colors of textures in the reference image and generates continuous textures by minimizing the energy function. In experiments, we demonstrate the effectiveness of the proposed method using a variety of images and examine the influence of parameter changes and intensity adjustment for pre-processing on resultant images.

Keywords: texture transfer, non-photorealistic rendering, painterly rendering, exemplar-based rendering, energy minimization

1. Introduction

Non-photorealistic rendering (NPR), which creates images with artistic styles of paintings such as oil paintings and pencil drawing, has been investigated in the field of computer graphics. Such rendering approaches enable various expressions according to the preference and sensibility of creators even if the target scene is the same. In this field, a number of methods for converting photographed images to non-photorealistic styles have been developed [16], and can be categorized into filter-based and exemplar-based approaches.

Many of the filter-based methods focus on strokes and convert sets of pixel values to various strokes by imitating actual paintings composed of multiple strokes. A number of filter-based methods have been proposed so far [10], [11], [12], [15], [22]. For example, Haeberli [10] proposed a method which selectively converts a part in a target image into a painterly style until the entire image is converted. In this method, users manually select a direction and a size of a brush, and the color of the brush is automatically determined according to the intensities of the target image. In [10], a method that automatically determines the positions and directions of strokes using intensity gradients is also proposed. Hertzmann [11] proposed a method that draws strokes based on an actual drawing manner that first draws a rough sketch with a large brush and then draws details with smaller brushes. Among relatively recent works, Zeng et al. [22] proposed a semantics-driven method. This method first parses a target image using methods in Refs. [9], [19], [20] and represents the image by multiple layers. The method then generates strokes in each layer considering an orientation field obtained by anisotropic diffusion and object categories. Such filter-based methods can generate various expressions of paintings by adjusting parameters for strokes. However, available painting styles are limited according to methods. In addition, much knowledge and many experiences of users are required to obtain their desired results.

There have also been a number of exemplar-based methods that transfer painterly styles in reference images to target images [2], [8], [13], [17], [18], [21]. These methods just require users to input reference images with their favorite styles and can convert target images more easily than filter-based methods. However, most of the conventional methods evaluate the local similarity and copy pixel values of the most similar patches in the reference image to the target image, and the process is iterated from the region with already determined pixel values to that with pixel values to be determined by raster scan. Such successive copy often does not preserve the spatial coherence for the entire image. For example, in the case where the reference image does not contain appropriate patches for some local area in the target image, once an implausible patch is copied there in the target image in a step during the raster scan, implausible textures tend to be continuously copied in subsequent steps. Although the methods iterate the raster scan to try to obtain more natural images, the number of iterations is very heuristic because there is no criterion to represent the plausibility for the entire image. In addition, while some methods only transfer the structures of textures, users often desire the transfer of both structures and colors of textures.

1 Nara Institute of Science and Technology, Ikoma, Nara 630–0192, Japan
a) norihi-k@is.naist.jp
b) tomoka-s@is.naist.jp
c) yokoya@is.naist.jp

© 2016 Information Processing Society of Japan

For these issues, this paper focuses on an exemplar-based approach, and proposes a method to transfer structures and colors of textures in an energy minimization framework. The contribution of this paper is optimally determining all pixel values by minimizing an energy function that evaluates the plausibility for the entire image based on both the relationship between output and target images and the relationship between output and reference images, rather than evaluating the similarity only for local regions and successively determining pixel values. The proposed approach transfers both structures and colors in the reference image to the target image with fewer texture disconnections than the successively copy approach while preserving the characteristics of the target image.

The rest of this paper is organized as follows. The next section reviews related work. Section 3 describes the proposed method. Section 4 shows three kinds of experiments. Finally we conclude this paper in Section 5.

2. Related Work

This section reviews exemplar-based texture transfer methods. Hetzmann et al. [13] proposed a texture transfer method that refers a pair of a photo and its converted image with a painterly style. Tang et al. [18] improved the method with respect to the computational cost by employing CUDA. These methods convert a target image by learning the relationship between the photo and its converted image. These methods also work with color transfer methods such as the method in Ref. [5]. In Ref. [5], a target image is converted so that the image has colors that are used in a reference image. In the method, colors in target and reference images are classified into 11 basic ones according to the Basic Color Terms (BCTs) [4]. Within each category, colors in the target image are transformed to ones used in the reference image. Although the combination of Refs. [13], [18], and [5] transfers both textures and colors, users have to prepare a pair of a photo and its converted image according to their preference.

For this problem, texture transfer methods that need only a single reference image in addition to a target image have been proposed [2], [8], [17], [21]. Efros et al. [8] proposed a texture synthesis method and applied it to texture transfer. The method firstly copies similar texture patches in a reference image to a target image successively while overlapping the partial regions of the patches. In the overlapping regions, the optimal boundaries between the patches are calculated using the method in Ref. [7] to achieve the natural connection of textures. Ashikhmin [2] proposed a texture transfer method by extending his texture synthesis method [1], which successively determines pixel values by raster scan. During the raster scan, the correspondences of pixels that were already made in the raster scan are used for determining the pixel in a reference image corresponding to a target pixel in a target image for texture coherence and speeding up of synthesis.

Wang et al. [21] and Lee et al. [17] have focused on directions of strokes, which are important characteristics in artistic paintings and are not considered in the methods in Refs. [2], [8], [13]. The method in Ref. [21] segments a target image and creates a direction field based on a medial axis of each region. Users then manually cut out patches from a reference image and they are used for converting the target image. Lee et al. [17] proposed an automatic transfer method, which is a method extended from Ref. [2], considering directions of strokes. The method creates a smooth direction field using the method in Ref. [14], which preserves edge directions smoothly around important features. The transferred image is then generated in a similar way to Ref. [2]. The methods in Refs. [17], [21] mainly transfer structures of textures and do not consider color changes of target images.

In summary, it should be noted that existing exemplar-based texture transfer methods basically successively transfer partial textures. Therefore, the order of transfer largely affects results. In addition, the method often does not preserve the spatial coherence for the entire image. On the other hand, the proposed method optimally determines the intensity of each pixel independently in an energy minimization framework, resulting in generating continuous textures in any part in the output image.

3. Texture Transfer by Energy Minimization

Figure 1 shows the overview of the proposed method. The proposed method converts target image T into output image O considering both the relationship between output and target images for preserving the structure of textures in target image T and the relationship between output and reference images for reflecting the characteristics of textures and colors in reference image R to output image O. Note that the resolution of output image O is the same as that of target image T but is not necessarily the same as that of reference image R.

In the following, Section 3.1 defines an energy function, Section 3.2 explains a method for minimizing the energy, and Section 3.3 describes a coarse-to-fine framework for energy minimization.

3.1 Definition of Energy Function

We define an energy function E based on the weighted sum of SSD_{OT} (Sum of squared differences between intensities in output image O and target image T) for preserving the structure of textures in target image T and $SSSD_{OR}$ (Sum of SSD between patches in output image O and reference image R) for reflecting the characteristics of textures and colors in reference image R as follows:

$$E = SSD_{OT} + \alpha SSSD_{OR}, \tag{1}$$

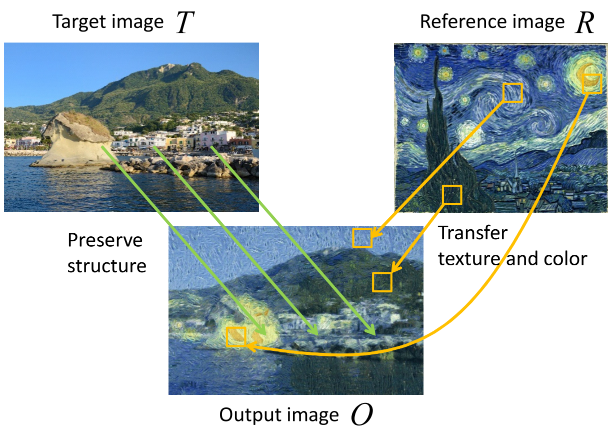


Fig. 1 Overview of the proposed method.

© 2016 Information Processing Society of Japan

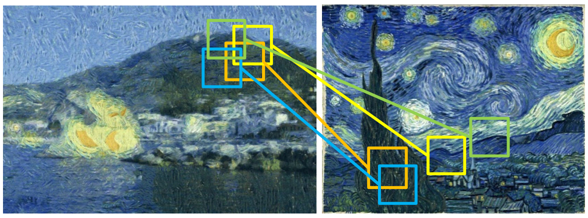


Fig. 2 Illustration of pairs of similar patches.

where α is a weight for controlling how much textures in the reference image are reflected to the output image. SSD_{OT} is defined based on the difference in values of all pixels in output image O and target image T as follows:

$$SSD_{OT} = \sum_{\mathbf{x}_i \in O} \|\mathbf{I}_O(\mathbf{x}_i) - \mathbf{I}_T(\mathbf{x}_i)\|^2, \quad (2)$$

where $\mathbf{I}_O(\mathbf{x}_i)$ and $\mathbf{I}_T(\mathbf{x}_i)$ are vector representations of RGB values at pixel position $\mathbf{x}_i$ of pixel i in the output image O and target image T, respectively. This term preserves the structure of the target image from largely changing.

$SSSD_{OR}$ totalizes SSD values for all the pairs of patches centered at all the pixels in output image O and some patches in reference image R and is defined as follows:

$$SSSD_{OR} = \sum_{\mathbf{x}_i \in O} \sum_{\mathbf{p} \in W} \|\mathbf{I}_O(\mathbf{x}_i + \mathbf{p}) - \mathbf{I}_R(\mathbf{x}_j + \mathbf{p})\|^2, \quad (3)$$

where $\mathbf{I}_R(\mathbf{x}_j + \mathbf{p})$ is a vector representation of RGB values at pixel position $\mathbf{x}_j + \mathbf{p}$ in a reference image R. Pixel j is an arbitrary pixel in reference image R, but automatically determined when the energy function is minimized. $\mathbf{p}$ is a shift vector to indicate a pixel in a square patch W. Here, if $\mathbf{x}_i + \mathbf{p}$ is out of the image region, the calculation is skipped. If $\mathbf{x}_j$ satisfies the condition that $\mathbf{x}_j + \mathbf{p}$ $(\exists \mathbf{p} \in W)$ is out of the image region, such pixel j is excluded from the candidate list for corresponding pixels in R in advance.

It should be noted here that since the neighboring patches in O overlap each other and the energy minimization described in Section 3.2 makes a patch centered at an arbitrary pixel in output image O similar to some patch in reference image R as shown in **Fig. 2**, continuous textures are usually generated at an arbitrary part in output image O in spite of not explicitly considering the spatial coherence of textures between adjacent patches.

3.2 Energy Minimization

Energy function E in Eq. (1) is iteratively minimized by a framework of greedy algorithm to find a good local minimum. Here, after initializing pixel values in output image O and pixels j corresponding to $\forall i (\in O)$ in Eq. (1) (how to initialize them is described in detail in Section 3.3), energy function E is minimized by iterating the following two processes until a condition for energy convergence is satisfied: (i) searching the reference image for similar texture patches to determine j in Eq. (3) and (ii) updating pixel values I_O in the output image.

The process (i) basically searches the reference image for such central position of the patch that satisfies the following equation with respect to each pixel $\mathbf{x}_i$ in output image O while all the pixel values in output image O fixed.

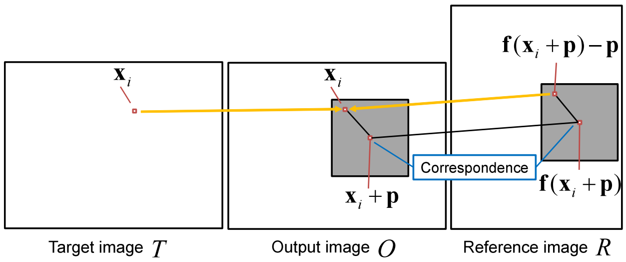


Fig. 3 Pixel relationship to calculate pixel value $\mathbf{I}_O(\mathbf{x}_i)$.

$$\mathbf{f}(\mathbf{x}_i) = \arg\min_{\mathbf{x}_j \in R} \sum_{\mathbf{p} \in W} \|\mathbf{I}_O(\mathbf{x}_i + \mathbf{p}) - \mathbf{I}_R(\mathbf{x}_j + \mathbf{p})\|^2. \quad (4)$$

Here, since it takes much time if we exhaustively search for the most similar patch, we employ PatchMatch [3], which limits the range of search for similar patches, to speed up the searching process by following the state-of-the-art methods in the field of texture synthesis such as Ref. [6].

The process (ii) updates each pixel value vector $\mathbf{I}_O(\mathbf{x}_i)$ of output image O independently so that energy E can be minimized while keeping all the pairs of similar patches between output image O and reference image R determined by process (i) fixed. In the following, we describe a method to determine pixel value $\mathbf{I}_O(\mathbf{x}_i)$.

Here, E can be resolved into each partial energy $E_{\mathbf{x}_i}$, which contains only pixel value vector $\mathbf{I}_O(\mathbf{x}_i)$ for pixel i as a parameter. As illustrated in **Fig. 3**, pixel positions related to $\mathbf{x}_i$ in the output image are $\mathbf{x}_i$ in the target image and $\mathbf{f}(\mathbf{x}_i + \mathbf{p}) - \mathbf{p}$ in the reference image, and the partial energy can be represented as follows:

$$E_{\mathbf{x}_i} = \|\mathbf{I}_O(\mathbf{x}_i) - \mathbf{I}_T(\mathbf{x}_i)\|^2 + \alpha \sum_{\mathbf{p} \in W} \|\mathbf{I}_O(\mathbf{x}_i) - \mathbf{I}_R(\mathbf{f}(\mathbf{x}_i + \mathbf{p}) - \mathbf{p})\|^2, \quad (5)$$

where the relationship between the total energy E and each partial energy $E_{\mathbf{x}_i}$ is represented as follow:

$$E = \sum_{\mathbf{x}_i \in O} E_{\mathbf{x}_i}. \quad (6)$$

Since each partial energy $E_{\mathbf{x}_i}$ is independent with respect to each pixel value in output image O, we can minimize energy E by minimizing each partial energy $E_{\mathbf{x}_i}$ independently. Pixel value $\mathbf{I}_O(\mathbf{x}_i)$ that minimizes partial energy $E_{\mathbf{x}_i}$ can be calculated by differentiating $E_{\mathbf{x}_i}$ and finding an extremal value as follows.

$$\mathbf{I}_O(\mathbf{x}_i) = \frac{\mathbf{I}_T(\mathbf{x}_i) + \alpha \sum_{\mathbf{p} \in W} \mathbf{I}_R(\mathbf{f}(\mathbf{x}_i + \mathbf{p}) - \mathbf{p})}{1 + \alpha N_W}, \quad (7)$$

where N_W is the number of pixels in patch W.

3.3 Coarse-to-fine Approach

Resultant images generated by the proposed method largely depend on the patch size. If the patch size is fixed through the entire process, only textures with a specific scale are reflected on resultant images. For this problem, we employ a coarse-to-fine approach, which generates image pyramids and iterates energy minimization while changing the resolution of the target, reference, and output images. This approach can also avoid bad local minima and reduce computational cost as well as reflect textures with different sizes on resultant images.

Specifically, we first generate image pyramids with M levels,

© 2016 Information Processing Society of Japan

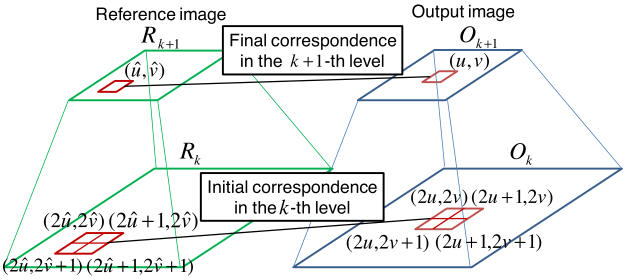


Fig. 4 Initialization of pixel correspondences.

Algorithm 1 Proposed method

Input: T, R, O
Generate image pyramids with M levels consisting of T_k, R_k and O_k $(k = 0, 1, \ldots, M-1)$
for $k = M-1$ to 0 **do**
  if $k = M-1$ **then**
    Copy T_k to O_k
    Initialize $\mathbf{f}(\mathbf{x}_i)$ for $\forall i$ in a random manner
  else
    Initialize $\mathbf{f}(\mathbf{x}_i)$ for $\forall i$ using correspondences for the $k+1$-th level
    O_k is initialized by Eq. (7)
  end if
  while a condition for energy convergence is not satisfied **do**
    for $\forall i \in O_k$ **do**
      Find $\mathbf{f}(\mathbf{x}_i)$ that satisfies Eq. (4) using PatchMatch
    end for
    for $\forall i \in O_k$ **do**
      Update pixel value $\mathbf{I}_{O_k}(\mathbf{x}_i)$ by Eq. (7)
    end for
  end while
end for
Output: $O(= O_0)$

which contain images whose width and height are recursively reduced by half. Here, we represent the k-th level images of the target, reference, and output images as T_k, R_k, and O_k, respectively. The width and height of k-the level image are $1/2^k$ of those of the original image. We then iterate energy minimization from $M-1$-th level to 0-th level layers successively using a certain size of patch. In the $M-1$-th level, pixel values in output image O_{M-1} are initialized by copying target image T_{M-1} to O_{M-1}, and initial pixel positions $\mathbf{f}(\mathbf{x}_i) = \mathbf{x}_j$ corresponding $\mathbf{x}_i$ are determined in a random manner by following PatchMatch algorithm [3]. In the other levels k, initial values of O_k are determined by Eq. (7) after initializing pixel positions $\mathbf{f}(\mathbf{x}_i) = \mathbf{x}_j$ corresponding $\mathbf{x}_i$. Here, as shown in **Fig. 4**, since pixel (u, v) in O_{k+1} is the combination of four pixels $(2u, 2v)$, $(2u+1, 2v)$, $(2u, 2v+1)$, and $(2u+1, 2v+1)$ in O_k, given that $(\hat{u}, \hat{v})$ in R_{k+1} is corresponding to (u, v) in O_{k+1}, we set initial pixels corresponding to the four pixels as $(2\hat{u}, 2\hat{v})$, $(2\hat{u}+1, 2\hat{v})$, $(2\hat{u}, 2\hat{v}+1)$, and $(2\hat{u}+1, 2\hat{v}+1)$, respectively.

Finally, we summarize the specific procedure of the proposed method in Algorithm 1.

4. Experiments

We conducted experiments using four photographs as target images and four paintings with various characteristics as reference images (**Fig. 5**) with a standard PC (Core i7-3930K 3.2 GHz CPU and 16 GB memory). The resolutions of target and reference images were resized so that the short sides become 300 pixels for experiments to unify the scales of textures in target and reference images. The characteristics of the target images (T1 to T4) and reference images (R1 to R4) are as follows:

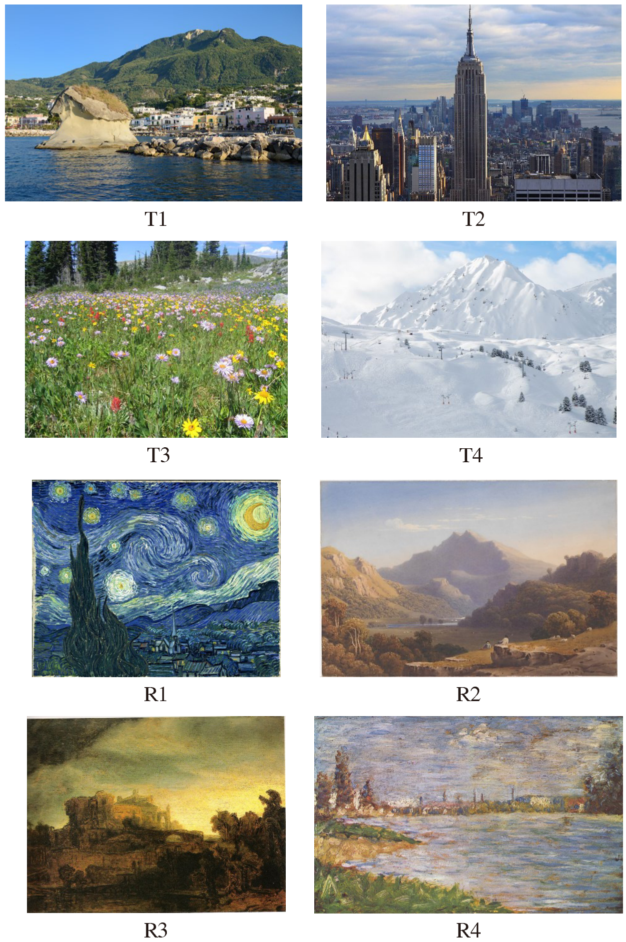


Fig. 5 Target (T1 to T4) and reference (R1 to R4) images.

T1 Landscape with detailed textures of buildings in one area and nature.
T2 City landscape with many detailed textures of buildings and cloud texture.
T3 Flower meadow.
T4 Winter mountain with little texture and low contrast.
R1 Landscape with small buildings and nature with long strokes and a small number of colors.
R2 Photorealistic countryside landscape.
R3 Landscape with some buildings and nature with dark colors.
R4 Landscape with nature with clear brush strokes with natural colors.

For the coarse-to-fine approach, we generated four-layer image pyramids ($M = 4$). We considered the energy to be converged when the ratio of energy decrease is less than 0.3% per layer.

In the experiments, we first examine the effectiveness of the proposed method and the influence of parameters on resultant images. We then examine transfered results with and without intensity adjustments in pre-processing for energy minimization because intensity adjustments may be effective in the proposed framework especially when the distribution of intensities between target and reference images are largely different.

© 2016 Information Processing Society of Japan

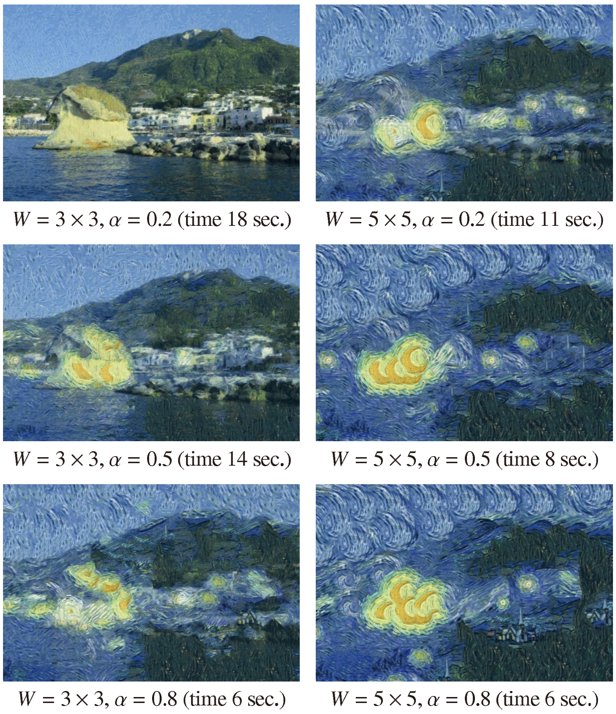


Fig. 6 Results with different parameters for target image T1 and reference image R1.

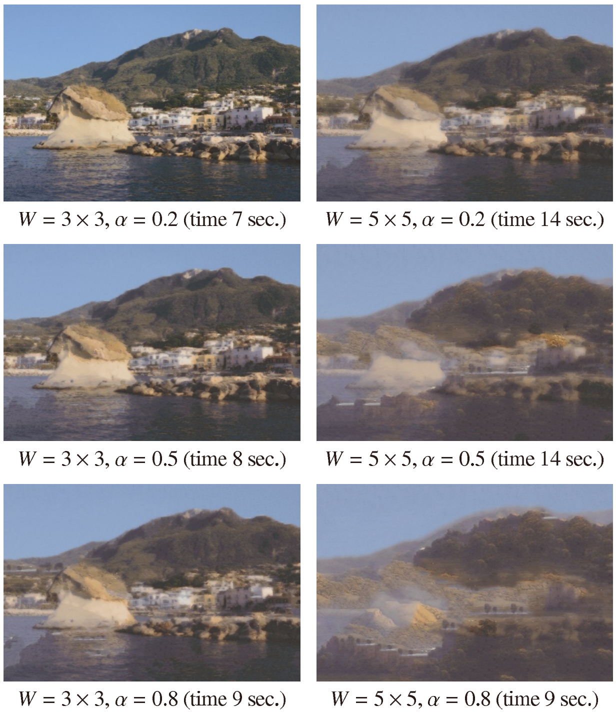


Fig. 7 Results with different parameters for target image T1 and reference image R2.

In the following, Section 4.1 describes experiments with different patch sizes and weights without intensity adjustment and Section 4.2 describes experiments with intensity adjustment. Section 4.3 compares experimental results obtained by the proposed and conventional methods using the target and reference images in paper [17].

4.1 Experiments with Different Parameters

This section discusses experimental results obtained while changing patch sizes and weights for balancing two cost terms. Specifically, we employed 3×3 and 5×5 pixels as patch size and 0.2, 0.5, and 0.8 as weight α. In this paper, we show converted images and processing times for some combinations of target and reference images as shown in Figs. 6 to 16. In the following, we first discuss individual results and then describe the overall tendency with different parameters and various combinations of target and reference images.

Figures 6 to **9** show results for target image T1, which has detailed texture in one part and a high color contrast for the entire image. From the results, we confirmed that the target image was surely converted to painterly style, reflecting the characteristics of textures in the reference images, but the amount of reflectance largely depended on the parameter values. For example, when $W = 3 \times 3$ and $\alpha = 0.2$, the conversion was slight but the color of the reference images were surely reflected. When $W = 3 \times 3$ and $\alpha = 0.5$, the strokes in the reference images were also reflected. However, as for reference images R1 and R4, which have clear strokes, the reflectance of strokes is insufficient as shown in Figs. 6 and 9. When $W = 5 \times 5$, the long strokes and relatively large textures were reflected to the results; however, as a side effect, some detailed textures such as buildings got lost as weight α became larger. For the combination of T1 and R1 with $W = 3 \times 3$

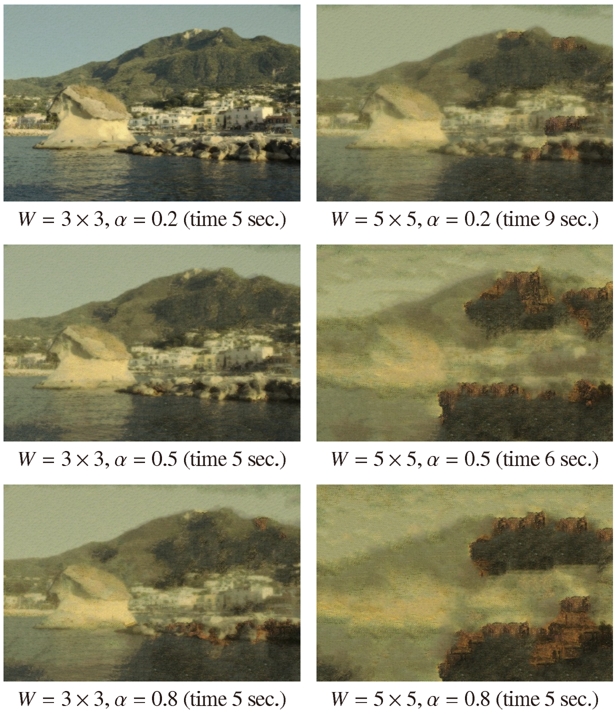
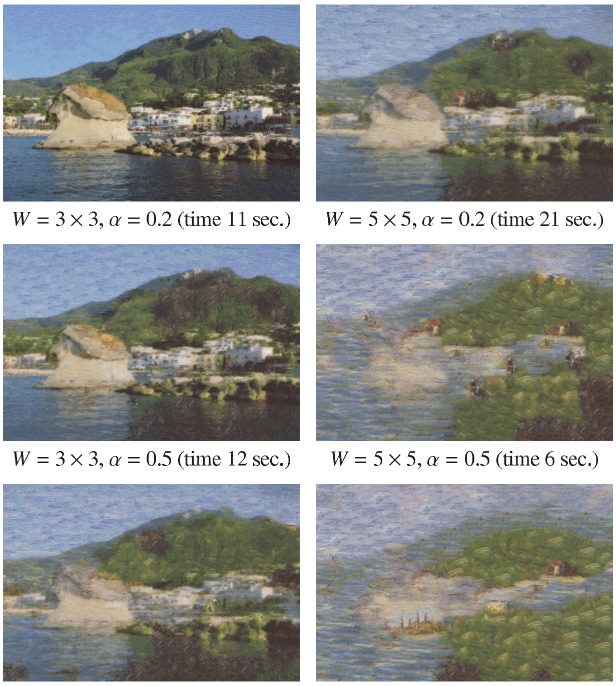


Fig. 8 Results with different parameters for target image T1 and reference image R3.

and $\alpha = 0.5$, we show the relationship between the normalized energy and the number of iterations in **Fig. 10**. From the graph, we confirmed that the energy is steadily reduced by each iteration.

Figures 11 and **12** show results for target image T2, which has detailed textures of buildings and cloud textures. Since the reference images do not contain textures corresponding to buildings in the target image, the detailed textures of buildings became destroyed as the patch size and the weight got larger. However, when $W = 3 \times 3$ and $\alpha = 0.5$, the reflectance of the strokes in the

© 2016 Information Processing Society of Japan



$W = 3 \times 3$, $\alpha = 0.2$ (time 11 sec.) $W = 5 \times 5$, $\alpha = 0.2$ (time 21 sec.)
$W = 3 \times 3$, $\alpha = 0.5$ (time 12 sec.) $W = 5 \times 5$, $\alpha = 0.5$ (time 6 sec.)
$W = 3 \times 3$, $\alpha = 0.8$, (time 10 sec.) $W = 5 \times 5$, $\alpha = 0.8$ (time 6 sec.)

Fig. 9 Results with different parameters for target image T1 and reference image R4.

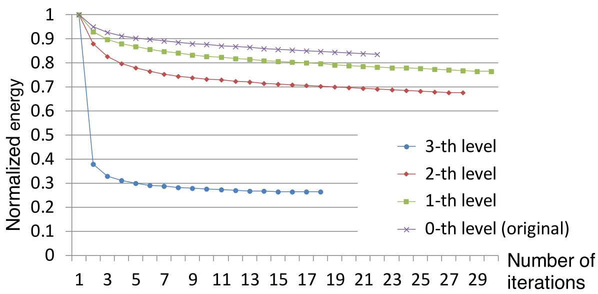


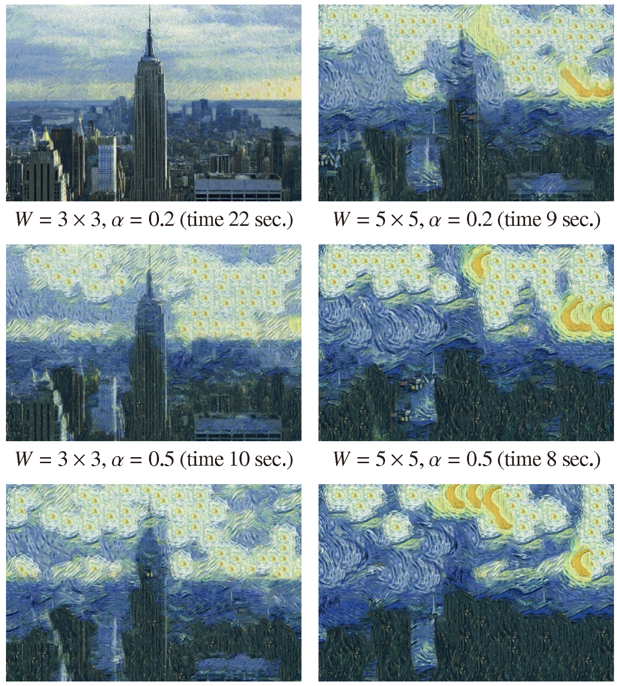
Fig. 10 Relationship between energy and number of iterations.

reference images is slight but actually noticeable while keeping the rough structure of textures in the target image.

Figures 13 and **14** show results for target image T3, which has detailed textures of flowers and grass. From the results, we can see that the color tones of reference images are well reflected to the target image, respectively. When $W = 3 \times 3$ and $\alpha = 0.8$ in Fig. 14, the strokes of the reference image are also surely reflected. However, we also confirmed that the detailed texture of grass became lost when $W = 5 \times 5$.

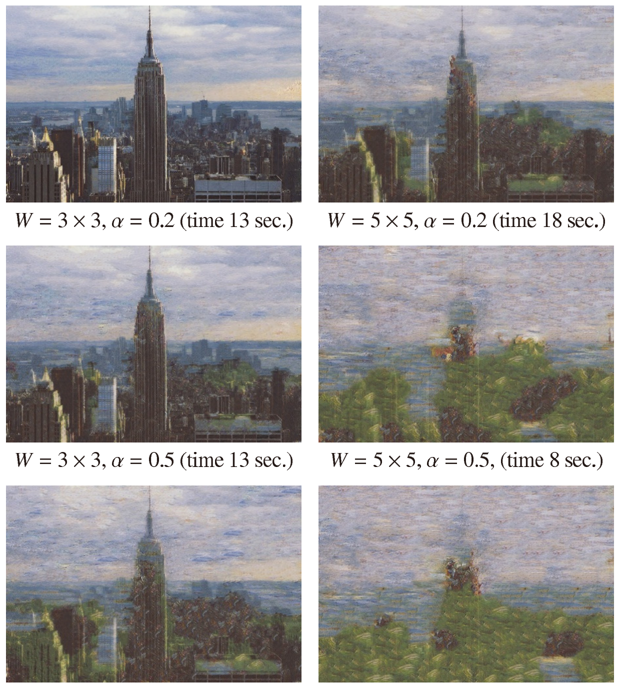
Figures 15 and **16** show results for target image T4, which has little texture and low contrast because of snowy scene. From the results, we can confirm that the structure of textures in the target image is not preserved when the patch size and the weight become even slightly large. Since the contrast of the target image is low and the image contains similar texture patterns throughout the entire region, some texture patterns in the reference images are intensively used as exemplars. As a result, the conversion lost detailed structure in the target image.

As for the overall tendency with respect to patch size and weight α, as the patch size and weight α become larger, the stroke in a reference image gets reflected more, but the details of texture in a target image get lost. Considering an ideal conversion that



$W = 3 \times 3$, $\alpha = 0.2$ (time 22 sec.) $W = 5 \times 5$, $\alpha = 0.2$ (time 9 sec.)
$W = 3 \times 3$, $\alpha = 0.5$ (time 10 sec.) $W = 5 \times 5$, $\alpha = 0.5$ (time 8 sec.)
$W = 3 \times 3$, $\alpha = 0.8$ (time 7 sec.) $W = 5 \times 5$, $\alpha = 0.8$ (time 8 sec.)

Fig. 11 Results with different parameters for target image T2 and reference image R1.

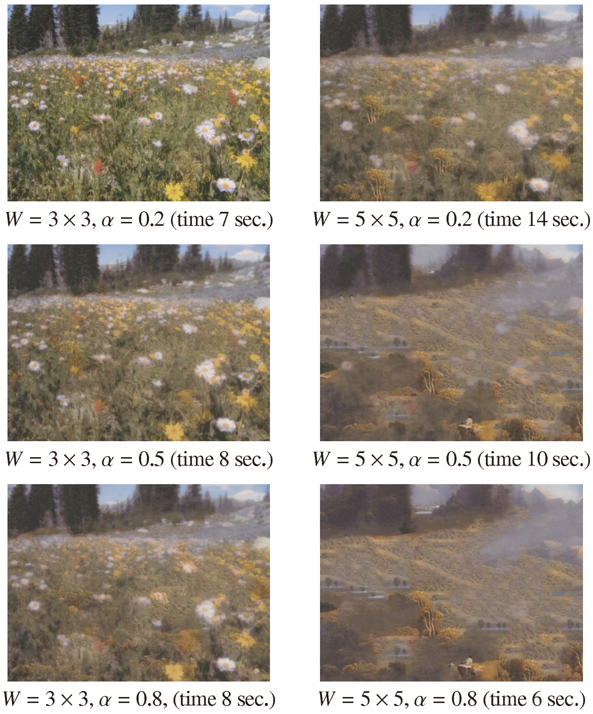


$W = 3 \times 3$, $\alpha = 0.2$ (time 13 sec.) $W = 5 \times 5$, $\alpha = 0.2$ (time 18 sec.)
$W = 3 \times 3$, $\alpha = 0.5$ (time 13 sec.) $W = 5 \times 5$, $\alpha = 0.5$, (time 8 sec.)
$W = 3 \times 3$, $\alpha = 0.8$ (time 10 sec.) $W = 5 \times 5$, $\alpha = 0.8$ (time 7 sec.)

Fig. 12 Results with different parameters for target image T2 and reference image R4.

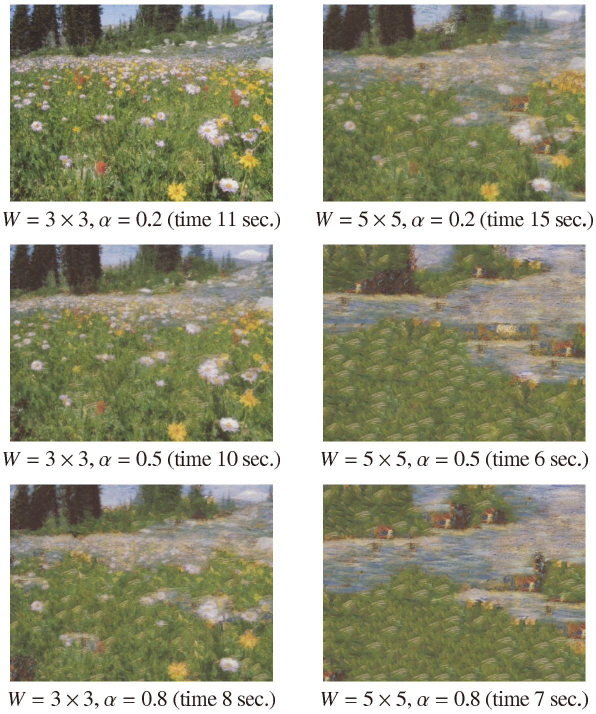
preserves the structure of textures in target images and reflects the paintings style of reference images as much as possible, it may be good to employ a large patch size for regions with little and homogeneous texture in target images such as sky, sea and mountains for reflecting the strokes in reference images and employ a small patch size for regions with detailed textures in target images such as buildings for preserving the structure of textures. To achieve this, we should consider a framework to adaptively change patch sizes according to the characteristics of textures in

© 2016 Information Processing Society of Japan



$W = 3 \times 3$, $\alpha = 0.2$ (time 7 sec.) $W = 5 \times 5$, $\alpha = 0.2$ (time 14 sec.)
$W = 3 \times 3$, $\alpha = 0.5$ (time 8 sec.) $W = 5 \times 5$, $\alpha = 0.5$ (time 10 sec.)
$W = 3 \times 3$, $\alpha = 0.8$, (time 8 sec.) $W = 5 \times 5$, $\alpha = 0.8$ (time 6 sec.)

Fig. 13 Results with different parameters for target image T3 and reference image R2.

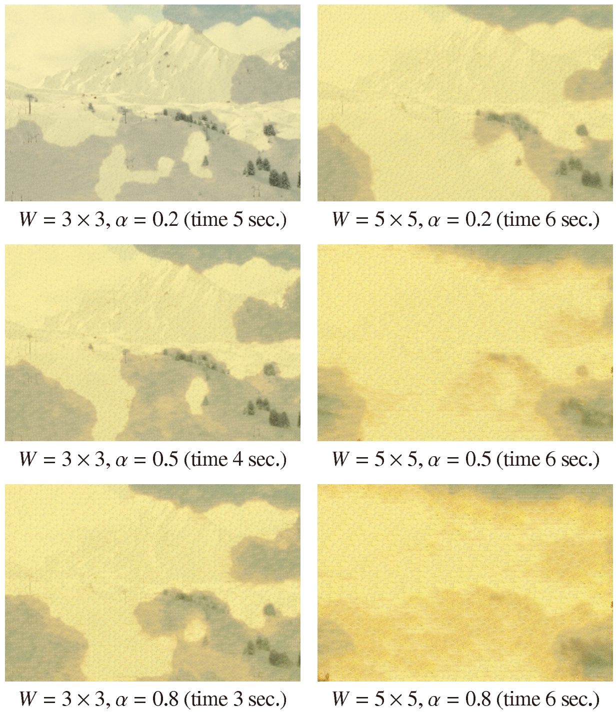


$W = 3 \times 3$, $\alpha = 0.2$ (time 11 sec.) $W = 5 \times 5$, $\alpha = 0.2$ (time 15 sec.)
$W = 3 \times 3$, $\alpha = 0.5$ (time 10 sec.) $W = 5 \times 5$, $\alpha = 0.5$ (time 6 sec.)
$W = 3 \times 3$, $\alpha = 0.8$ (time 8 sec.) $W = 5 \times 5$, $\alpha = 0.8$ (time 7 sec.)

Fig. 14 Results with different parameters for target image T3 and reference image R4.

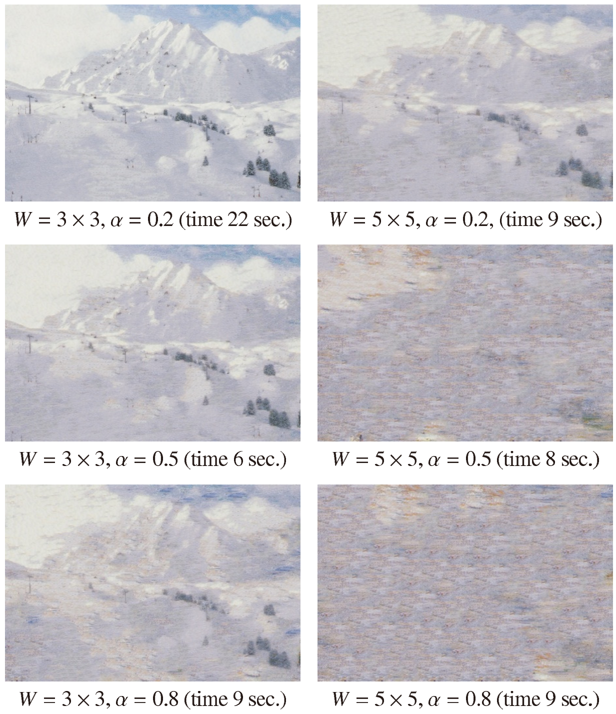
target images in future work. Currently, although the quality of the results depend on the patch size and weight, we could obtain averagely good results with 3×3 pixels patch and $\alpha = 0.5$. Therefore, we employ these values for the next experiments with intensity adjustment.

As the processing time for each conversion is shown in each figure, we confirmed that the processing time becomes shorter as weight α becomes larger. This is because larger weights speed up the convergence of energy. We also confirmed that the time is



$W = 3 \times 3$, $\alpha = 0.2$ (time 5 sec.) $W = 5 \times 5$, $\alpha = 0.2$ (time 6 sec.)
$W = 3 \times 3$, $\alpha = 0.5$ (time 4 sec.) $W = 5 \times 5$, $\alpha = 0.5$ (time 6 sec.)
$W = 3 \times 3$, $\alpha = 0.8$ (time 3 sec.) $W = 5 \times 5$, $\alpha = 0.8$ (time 6 sec.)

Fig. 15 Results with different parameters for target image T4 and reference image R3.



$W = 3 \times 3$, $\alpha = 0.2$ (time 22 sec.) $W = 5 \times 5$, $\alpha = 0.2$, (time 9 sec.)
$W = 3 \times 3$, $\alpha = 0.5$ (time 6 sec.) $W = 5 \times 5$, $\alpha = 0.5$ (time 8 sec.)
$W = 3 \times 3$, $\alpha = 0.8$ (time 9 sec.) $W = 5 \times 5$, $\alpha = 0.8$ (time 9 sec.)

Fig. 16 Results with different parameters for target image T4 and reference image R4.

sufficiently short for practical use.

4.2 Experiments with Intensity Adjustment

This section discusses experimental results obtained by the proposed method with intensity adjustment. In this experiment, we employed 3×3 pixels as patch size and 0.5 as weight α as mentioned above.

For intensity adjustment, we modified the intensities of the target image so that the histogram of the target image can correspond

© 2016 Information Processing Society of Japan

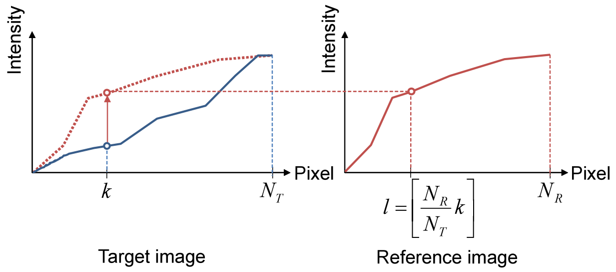


Fig. 17 Intensity adjustment of target image.

to that of the reference image. Specifically, we first convert the colors of target and reference images from RGB to HSV color space and sort pixels of target and reference images in ascending order of their V values, respectively. We then change the V value of k-th pixel in the target image to that of l-th pixel in the reference image, as shown in **Fig. 17**, where the relationship between k and l is represented as follows:

$$l = \left[\frac{N_R}{N_T} k \right], \tag{8}$$

where $[\cdot]$ rounds down a value to an integer, and N_T and N_R indicate the resolution of images T and R, respectively. Finally, the color space of the target image is converted from HSV to RGB. In this experiment, we regarded the intensity-adjusted target image or the original target image as target image T.

Figures 18 to 21 show target images with intensity adjustment and converted results for target images T1 to T4 with two reference images, respectively. In the following, we discuss the results.

Figure 18 shows the results of target image T1 converted with reference images R1 and R2. The left images in Fig. 18 were obtained by intensity adjustment so that the histogram of the target image becomes almost the same with those of the reference images. From the results, we confirmed that the results with intensity adjustment were worse than those obtained using the same parameter $W = 3 \times 3$ and $\alpha = 0.5$ without intensity adjustment as shown in Figs. 6 and 7. Since the adjustment enhanced the contrast of intensities in sky and mountain regions, the high contrast caused inappropriate texture correspondences, resulting in implausible converted images.

Figure 19 shows the results of target image T2 converted with reference images R1 and R4. When reference image R1 was used, the target image after intensity adjustment became dark. As a result, the converted result provides users with an impression different from the result without intensity adjustment shown in Fig. 11. When reference image R4 was used, the intensity adjustment brightened the entire target image. The converted image with the brighter target image has low contrast and pale colors. Since the color tones and strokes in the reference images are reflected to the results for both cases when intensity adjustment is applied or not, the evaluation may depend on the user's preference.

Figure 20 shows the results of target image T3 converted with reference images R2 and R4. The target image was brightened when reference images R2 and R4 were used. The strokes in the reference images were reflected to the converted results than

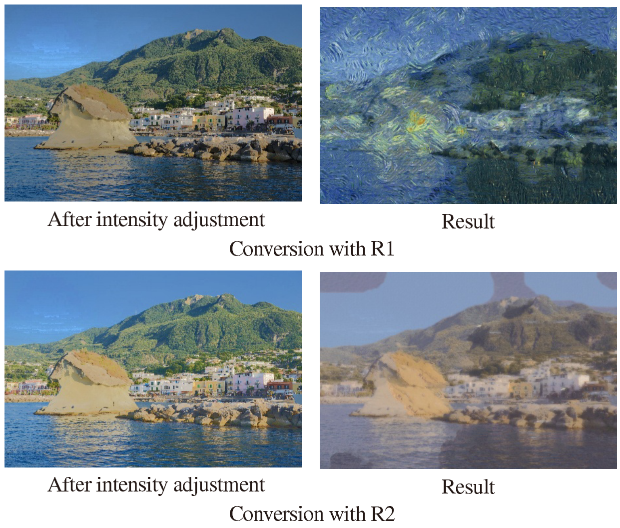


Fig. 18 Results of T1 with intensity adjustment.

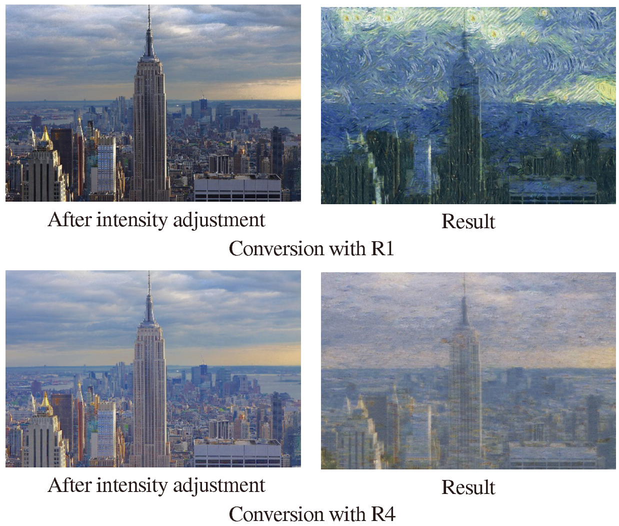


Fig. 19 Results of T2 with intensity adjustment.

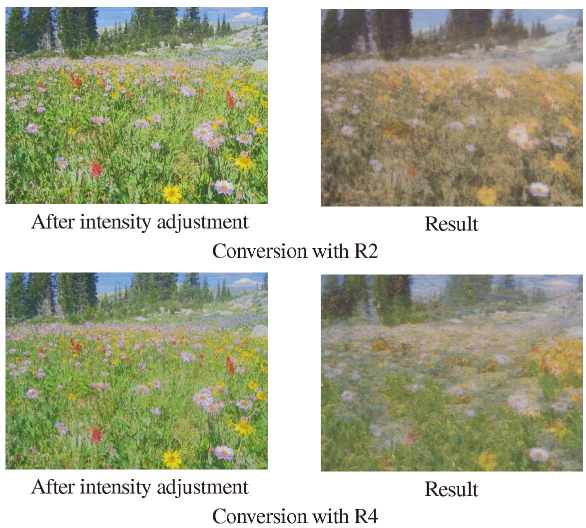


Fig. 20 Results of T3 with intensity adjustment.

those without intensity adjustment shown in Figs. 13 and 14; however, the blurry effects were also generated for both results.

Figure 21 shows the results of target image T4 converted with reference images R3 and R4. Since target image T4 has low con-

© 2016 Information Processing Society of Japan

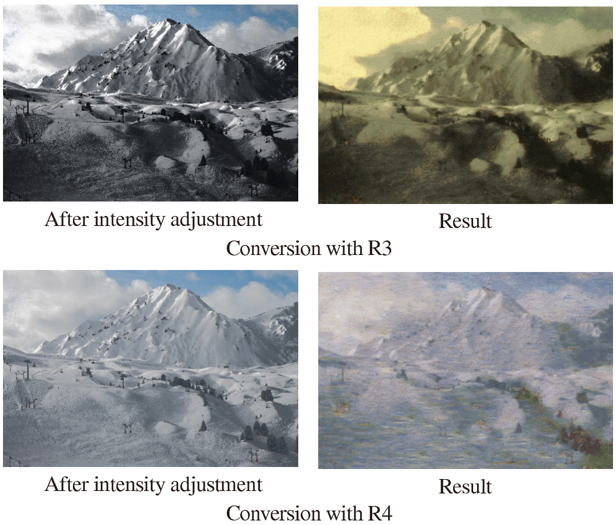


Fig. 21 Results of T4 with intensity adjustment.

trast, the intensity adjustment enhanced the contrast. As a result, the color tones and strokes in the reference images R3 and R4 are well reflected and more natural resultant images were obtained compared with the results without intensity adjustment shown in Figs. 15 and 16.

In summary, when the contrast of the target image is low, the intensity adjustment by making the histogram of target image comparable to that of the reference image often enhances the contrast and tends to reflect the characteristics of textures in a reference image to the converted result more clearly. In other cases, the intensity adjustment does not always make converted results better; however, it may generate additional converted results that have different strokes and brightness from those obtained by the proposed method without intensity adjustment. For practical use, since we cannot know the users' preferences, a system in which users can select their desired results by outputting converted results both with and without intensity adjustment may be useful.

4.3 Comparison

This section compares the results obtained by the proposed method with those obtained by one of the state-of-the-art methods [17]. **Figure 22** shows a target image and three reference images, and **Fig. 23** shows results by the proposed method and the method in Ref. [17].

Reference image A is an oil painting with clear strokes and distinct use of colors. In the resultant image converted by the proposed method, the reflectance of strokes is slight but actually confirmable, and the colors of the entire resultant image are surely influenced by those in the reference image. In particular, the combination of blue and brown strokes appears in the entire resultant image. On the other hand, in the resultant image obtained by the conventional method, clear stroke shapes similar to those in the reference image are generated in the resultant image. However, the color tones and the combination of blue and brown strokes are not reflected because the conventional method does not intend to reflect color tones.

Reference image B is a water painting of fruits that are drawn on a piece of bumpy paper. In the resultant image converted by

Fig. 22 Target and reference images for comparison.

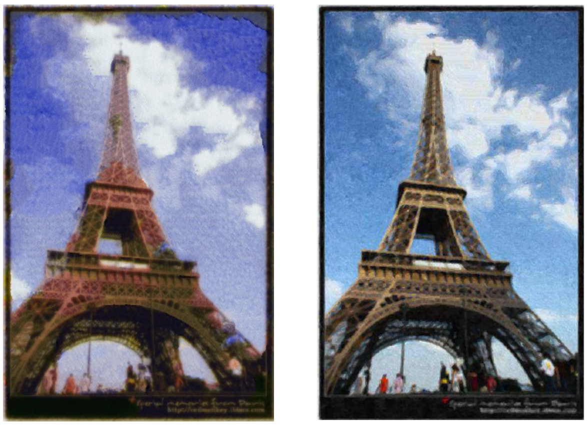
the proposed method, color tones used in the reference image appears over the entire image. The bumps on the paper are also reflected. However, since the proposed method does not have the effect of preserving edges, the bleeding of black ink can be observed. On the other hand, the bumps also seem to be reflected to the resultant image by the conventional method, but it may not be easy to clearly find the effect.

Reference image C is an image with pastel colors. The proposed method reflects the strokes observed in the surface of the water and the bird to the sky and cloud regions and also reflects color tones to the entire target image. In the resultant image obtained by the conventional method, it may be difficult to recognize the effect of the reflectance because of the difference in color tones.

From the overall results, it seems to be easier to recognize the effect of the proposed method than the conventional method because the proposed method reflects not only strokes but also color tones. However, for detailed textures such as the framework of the tower, the proposed method tends to generate blurrier textures than the conventional method. The difference is remarkable especially for the results for Reference A. Since pixel values are determined by the uniform average of pixel values of patches for energy minimization, the generated textures tend to blur if the reference image does not contain the texture whose structure is quite similar to that in the target image. In addition, the reflection of textures with large scales is still insufficient. Although we may reflect textures with large scales if we use larger patches, large patches often do not preserve the structure of the target image. To overcome these problems, we should consider a method that automatically determines the position-dependent patch size according to the fineness of the texture in the target image and the existence of similar textures in the reference image because

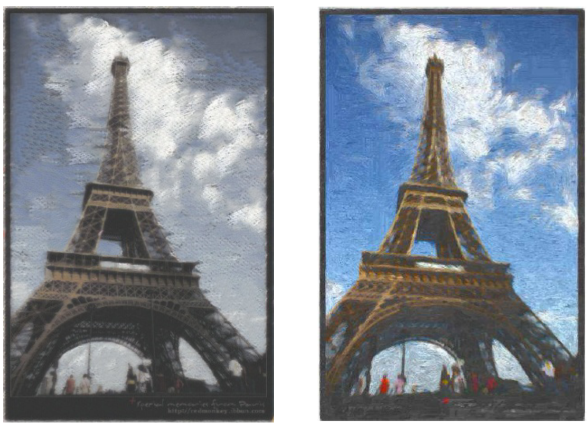
Result by proposed method ($W = 3 \times 3$, $\alpha = 0.8$, no intensity adjustment) Result by Ref. [17]

Reference A



Result by proposed method ($W = 3 \times 3$, $\alpha = 0.6$, no intensity adjustment) Result by Ref. [17]

Reference B



Result by proposed method ($W = 5 \times 5$, $\alpha = 0.2$, no intensity adjustment) Result by Ref. [17]

Reference C

Fig. 23 Comparison with the method in Ref. [17].

large patches reflect textures with large scales and small patches preserve the detailed structure of the target image without blurs even if the reference image contains few similar textures. Considering the level M of the coarse-to-fine approach also may improve the reflection of textures and the preservation of structures because starting from larger level M (smaller resolution of the image) with a certain patch size is equivalent to using patches which have a larger size, and vice versa.

5. Conclusion

This paper has proposed an exemplar-based method that transfers a painterly style in a reference image to a photograph. The texture transfer is achieved by minimizing an energy function that consists of two terms: intensity differences between target and converted (output) images, and pattern dissimilarity between reference and converted images. In experiments, we examined the influence of parameters and intensity adjustment on transfered images. We also compared converted images obtained by the proposed method with those obtained by a conventional method. From the results, we confirmed that the proposed method does not always reflect the strokes sufficiently, but handling both strokes and color tones is effective to convert a target image into a painterly style. In future work, we should develop a method that adaptively adjusts the patch size in energy definition considering the fineness of textures and the existence of similar textures.

Acknowledgments This work was partially supported by JSPS KAKENHI No. 15K16039.

References

[1] Ashikhmin, M.: Synthesizing Natural Textures, *Proc. Symposium on Interactive 3D Graphics*, pp.217–226 (2001).

[2] Ashikhmin, M.: Fast Texture Transfer, *IEEE Computer Graphics and Applications*, Vol.23, No.4, pp.38–43 (2003).

[3] Barnes, C., Shechtman, E., Finkelstein, A. and Goldman, D.B.: PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing, *ACM Trans. Graphics*, Vol.28, No.3, pp.24:1–24:11 (2009).

[4] Berlin, B. and Kay, P.: *Basic Color Terms: Their Universality and Evolution*, University of California Press (1969).

[5] Chang, Y., Saito, S., Uchikawa, K. and Nakajima, M.: Example-based Color Stylization of Images, *ACM Trans. Applied Perception*, Vol.2, No.3, pp.322–345 (2005).

[6] Darabi, S., Shechtman, E., Barnes, C., Goldman, D.B. and Sen, P.: Image Melding: Combining Inconsistent Images Using Patch-based Synthesis, *ACM Trans. Graphics*, Vol.31, No.4, pp.82:1–82:10 (2012).

[7] Davis, J.: Mosaics of Scenes with Moving Objects, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.354–360 (1998).

[8] Efros, A.A. and Freeman, W.T.: Image Quilting for Texture Synthesis and Transfer, *Proc. SIGGRAPH*, pp.341–346 (2001).

[9] Guo, C.-E., Zhu, S.-C. and Wu, Y.N.: Primal Sketch: Integrating Structure and Texture, *Computer Vision and Image Understanding*, Vol.106, No.1, pp.5–19 (2007).

[10] Haeberli, P.: Paint by Numbers: Abstract Image Representations, *ACM SIGGRAPH Computer Graphics.*, Vol.24, No.4, pp.207–214 (1990).

[11] Hertzmann, A.: Painterly Rendering with Curved Brush Strokes of Multiple Sizes, *Proc. SIGGRAPH*, pp.453–460 (1998).

[12] Hertzmann, A.: Paint by Relaxation, *Proc. Computer Graphics International 2001*, pp.47–54 (2001).

[13] Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B. and Salesin, D.H.: Image Analogies, *Proc. SIGGRAPH*, pp.327–340 (2001).

[14] Kang, H., Lee, S. and Chui, C.K.: Coherent Line Drawing, *Proc. International Symposium on Non-photorealistic Animation and Rendering*, pp.43–50 (2007).

[15] Kovács, L. and Szirányi, T.: Painterly Rendering Controlled by Multiscale Image Features, *Proc. Spring Conference on Computer Graphics*, pp.177–184 (2004).

[16] Kyprianidis, J.E., Collomosse, J., Wang, T. and Isenberg, T.: State of the "Art": A Taxonomy of Artistic Stylization Techniques for Images and Video, *IEEE Trans. Visualization and Computer Graphics*, Vol.19, No.5, pp.866–885 (2013).

[17] Lee, H., Seo, S. and Yoon, K.: Directional Texture Transfer with Edge Enhancement, *Comput. Gr.*, Vol.35, No.1, pp.81–91 (2011).

© 2016 Information Processing Society of Japan

[18] Tang, Y., Shi, X., Xiao, T. and Fan, J.: An Improved Image Analogy Method Based on Adaptive CUDA-accelerated Neighborhood Matching Framework, *The Visual Computer*, Vol.28, No.6, pp.743–753 (2012).

[19] Tu, Z., Chen, X., Yuille, A.L. and Zhu, S.-C.: Image Parsing: Unifying Segmentation, Detection, and Recognition, *International Journal of Computer Vision*, Vol.63, No.2, pp.113–140 (2005).

[20] Tu, Z. and Zhu, S.-C.: Parsing Images into Regions, Curves, and Curve Groups, *International Journal of Computer Vision*, Vol.69, No.2, pp.223–249 (2006).

[21] Wang, B., Wang, W., Yang, H. and Sun, J.: Efficient Example-based Painting and Synthesis of 2D Directional Texture, *IEEE Trans. Visualization and Computer Graphics*, Vol.10, No.3, pp.266–277 (2004).

[22] Zeng, K., Zhao, M., Xiong, C. and Zhu, S.-C.: From Image Parsing to Painterly Rendering, *ACM Trans. Graphics*, Vol.29, No.1, pp.2:1–2:11 (2009).

Yuta Tsunematsu received his B.E. degree in information and computer sciences from Osaka University in 2012. He received his M.E. degree in information science from Nara Institute of Science and Technology (NAIST) in 2015. He has been working for Ricoh IT Solutions Co., Ltd. from 2015.

Norihiko Kawai received his B.E. degree in informatics and mathematical science from Kyoto University in 2005. He received his M.E. and Ph.D. degrees in information science from Nara Institute of Science and Technology (NAIST) in 2007 and 2010, respectively. He was a research fellow of the Japan Society for the Promotion of Science and a postdoctoral fellow at the University of California at Berkeley in 2010–2011. He has been an assistant professor at NAIST since 2011. He is a member of IEEE, IEICE, IPSJ and VRSJ.

Tomokazu Sato received his B.E. degree in computer and system science from Osaka Prefecture University in 1999. He received his M.E. and Ph.D. degrees in information science from Nara Institute of Science and Technology (NAIST) in 2001 and 2003, respectively. He was an assistant professor at NAIST in 2003–2011. He was a visiting researcher at Czech Technical University in Prague in 2010–2011. He has been an associate professor at NAIST since 2011. He is a member of IEEE, IEICE, IPSJ, VRSJ and ITE.

Naokazu Yokoya received his B.E., M.E., and Ph.D. degrees in information and computer sciences from Osaka University in 1974, 1976, and 1979, respectively. He joined the Electrotechnical Laboratory (ETL) of the Ministry of International Trade and Industry in 1979. He was a visiting professor at McGill University in Montreal in 1986–87 and has been a professor at Nara Institute of Science and Technology (NAIST) since 1992. He has also been a vice president at NAIST since April 2013. He is a fellow of IPSJ, IEICE and VRSJ and a member of IEEE, ACM SIGGRAPH, JSAI, JCSS and ITE.

© 2016 Information Processing Society of Japan