Scene-Aware Style Transferring using GIST

Masahiro Toyoura Noriyuki Abe Xiaoyang Mao

University of Yamanashi
Kofu, Yamanashi, Japan

Abstract. This paper proposes a new method of transferring style between images by considering scene matching between the source image and the target image. Artists often employ different colors and brushwork for individual subjects. Likewise, the connections between various subjects in a work also affect the colors and brushwork used. Our method begins with input images, searches an example database for paintings with scenes similar to that in the input image, and transfers the color and brushwork of the paintings to the corresponding target images to generate painterly images that reflect specific styles. Our method applies a GIST approach to the process of searching for paintings with similar scenes before performing style transfers. The spatial correspondence between the source image and the target image is also used to ensure close correlation between various elements in order to reproduce styles faithfully.

1 Introduction

Painting is a challenge for most people. In the past decade, many commercial or free image processing software have been developed allowing users to create painterly images with ease. In most cases, however, these types of software simply process the entire input image with some kind of filter and the end result is thus usually far removed from the look of an actual painting. More sophisticated filtering algorithms, using modern computer vision and image processing techniques, have been developed to create various stylized images [9]. However, the representation that can be created by those techniques is inherently limited to an abstraction of some particular style. Working to generate painterly images that more closely resemble real paintings, researchers have conducted physical simulations of various painting materials and painting methods, such as colored pencils, watercolors, and ink, among others [7]. Using these techniques to create painterly images requires making adjustments to a vast array of different parameters, such as pen tip shape, brush pressure, pigment viscosity, and translucency. A user thus often needs to know not only how to use the software in question but also how to do the actual painting.

Recently, an example-based approach that uses learning technology to transfer the artistic style of an example work onto an arbitrary image is attracting a large amount of attention [2, 5, 17]. While the filtering or physical simulation–based approaches usually encode a set of heuristics to mimic or simulate particular predefined styles, the largest advantage of the example-based approach is that it applies to any style in principle given the example image in the style. The majority of example-based painterly image generation techniques apply texture synthesis technology for transferring the appearance of the brush strokes in the example painting onto a target image. As style of brush stroke is one of the main attributes contributing to a unique painting style, such approaches have high potential to generate painterly images reflecting the style of individual artists.

Turn one's eye to real paintings, however, and artists usually use different colors and brush strokes for different objects. Even the same subjects may be painted differently based on where the subjects are and what else is in the scene. Therefore, transferring a style from a painting depicting a completely different scene may fail to reproduce the intended style. One such example is shown in Figure 1. The results are obtained by applying Image Analogies, a well-known example-based technique proposed by Hertzmann et al. [8], to transfer the texture of the brush stroke from the reference painting to the input photograph. The stroke texture of the white German iris in the painting was transferred into the sky in the photograph because of the similarities in color and position of the two subjects, resulting in the unnatural depiction of the sky. The example shows the importance of the correspondence of subjects and compositions between the source painting and the target image.

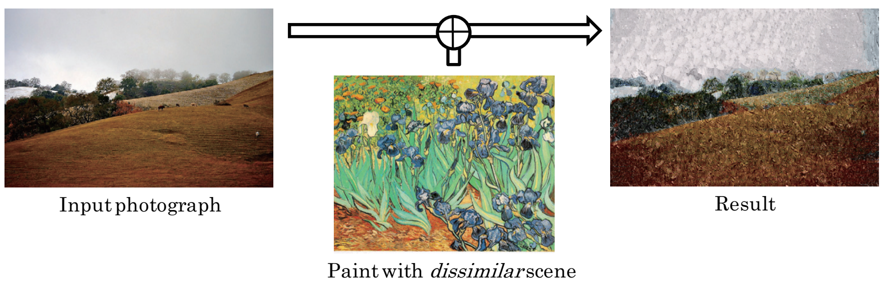


Fig. 1. Transferring style from a painting with a *dissimilar* scene.

In this paper, we propose a new scene-aware style transferring technique to automatically generate painterly images. Given an input image, our method first searches the example painting database for a painting depicting a similar scene and then transfers the color and texture between the corresponding regions of the painting and the input image. Establishing perfect correspondence between the elements of two images requires comprehensive understanding of the contents of the image, which cannot be achieved even with the most advanced computer vision technologies. Fortunately, it is known that artists, especially the majority of impressionists, are likely to paint a scene perceptually without constructing a semantical model of the scene. Based on this assumption, we combine a GIST feature [6], which describes the overall composition of a scene as initially perceived by a human, normally within 100 msec of seeing an image or scenery, with a color feature for searching for paintings of similar scenes. The result generated by applying the proposed method to the photograph is shown in Figure 2. The algorithm first obtained painting with a similar scene from the example painting database and then transferred the color and texture of this painting to the input image by considering the local correspondence between the two images. Compared with the synthesized result image in Figure 1, the result in Figure 2 appears to better reflect the style of the original painting. We have conducted subject studies to evaluate the effectiveness of our new painting retrieval technique and we investigate whether the proposed technique can generate painterly images that well preserve the styles of the example paintings.

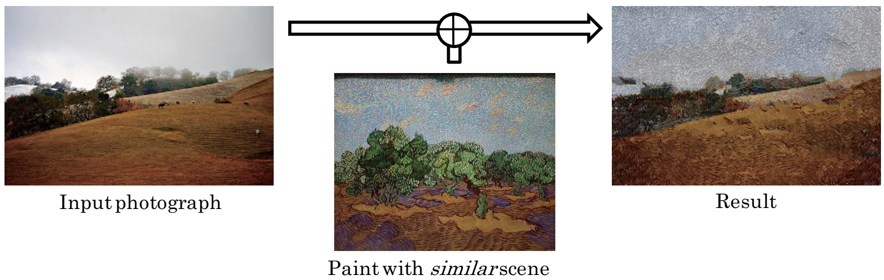


Fig. 2. Transferring style from a painting with a *similar* scene.

Our major contributions can be summarized as follows:

(1) The novel idea of introducing a scene-aware approach to the texture synthesis–based style transferring method.
(2) Improvements to the conventional GIST-based image retrieval method by incorporating color feature and local distance.
(3) The new scene-aware style transferring algorithm obtained by adding position-based constraints to the existing color and texture transfer technique.

The remainder of this paper is organized as follows: Section 2 reviews related works from the literature. After giving an overview of the proposed method in Section 3, Section 4 presents the technique for searching for paintings with similar scenes. Section 5 describes the algorithm for transferring styles. Section 6 explains the results of the conducted experiments, and Section 7 concludes the paper.

2 Related Works

Image Analogies [8], mentioned in the Introduction, is a pioneering work of example-based style transferring techniques. The method learns the mapping between an exemplar pair: a source image and an artist's rendering of that image. The learned mapping can then be applied to render arbitrary images in the exemplar style. As depicted in Figure 1, although Image Analogies provided an elegant framework for learning arbitrary artistic style using non-parametric texture synthesis, it may fail to reproduce the desired style if the scene depicted by the painting and the scene captured by the photograph are too different. Our proposed method employs Image Analogies for local texture transferring but ensure the quality of the result by finding a matching painting and performing scene aware transferring. Chang et al. [2] used the mean-shift method to divide the input image and paintings into color-specific regions. A region of a sample painting is cropped to a rectangular patch and used as the sample texture for the region with similar average color in the input image. As the method uses patch–based texture synthesis for filling each region with the sample texture, it fails to represent the subtle tone changes in each region. Furthermore, the method can incorrectly link regions with similar average colors, for example blue skies and water, which should have brush strokes with different feels.

Several methods have been proposed to learn the directional appearance of brush strokes in a painting. Wang et al. [2][17] and Guo et al. [5] proposed methods assuming painting styles are represented as one or more blocks of sample textures selected by the user from the example painting. Those samples are superimposed over regions of the input image following directions defined either by the shape of the region or by the brightness gradient. These two methods require the user to specify regions manually and the quality of the resulting image may depend on the user's skill.

Lee et al. [10, 11] extended Ashikhmin's fast texture synthesis technique [1] to generate directional effect in the resulting image. Either in Image Analogies or Ashikhmin's algorithm, two different searches are combined to find the best matching pixels. One is an approximate search that attempts to find the closest-matching pixel according to the neighborhoods in the source image and the other is a coherence search that attempts to preserve coherence with the neighboring synthesized pixels in the resulting image. Lee et al 's method generates the directional effect by adding a third coherence search item which search for the pixel in the source image with small difference from the average value of the pixels along the direction perpendicular to the gradient of the resulting image. Since the directional effect is produced by favoring the coherence in the resulting image along a particular direction, their method may fail to transfer the original filtering effect.

Xie et al. [18, 19] also tackled to improve the quality of the resulting image by trying to better preserve the structure information in the synthesized image. They proposed to adaptively vary the coherence parameter according to their distance from the structural features such edges and boundaries. Chang et al. [3] proposed to adaptively changing the patch size when performing patch based texture synthesis. Lee et al. [12] tried to improve the quality of texture synthesis by using the structure information in similarity searching. Same as in the very basic Image Analogies technique, all of these methods perform texture transfer without aware of the scene correspondence between the example painting and the input image. Our technique can be easily combined with those sophisticated texture transferring method for achieving more impressive results.

The above–mentioned style transferring approaches mainly rely on non-parametric techniques to directly manipulate the pixel representation of an image. Gatys et al. [4] attempted to use deep neural networks to carry out manipulations in feature spaces that explicitly represent the high–level content of an image. The feature space is built on top of the filter responses in multiple layers of the network. Then, by mixing the higher layer of the input image and the lower layers of the example images, they succeeded in generating results that present the content of the input image in the style of the example input image. Although several plausible results are shown in [4], the selection of appropriate layers in the multi-scale feature spaces can be image dependent and difficult.

However, the idea of combining data–driven approaches with the rich data resources of the web has attracted a great deal of attention. Liu et al. [14] proposed a technique that stylizes a user's photo by transferring style from a collection of images returned by a web search for a particular keyword. However, their methods mainly focus on transferring color and contrast, and the expected results can be obtained only when the retrieved images have attached keywords matching the image style.

As a content–aware approach, Shih et al. [16] developed a technique for transferring style between headshot portraits. Their method establishes the correspondence between the local features of the example stylized portrait and the target photograph and robustly transfers the local statistics from the example image to the target image. However, their technique is tailored for portraits and cannot be applied to other photographs.

3 Overview

Figure 3 shows a schematic overview of our method, which comprises the example painting database construction and runtime painterly image generation stages. To construct the database, feature vectors for describing the scene depicted by each painting example are computed and stored together with the paintings. The examples are divided into different sets by artist name, materials, and techniques. At runtime, the user inputs an image and keywords (artist name, materials, and/or techniques). The system then computes the scene feature vector for the input image and uses the feature vector to search the corresponding image set in the database for a painting that depicts a similar scene as the input image. Upon finding a similar painting, the system transfers the color and brushwork from the painting to the input image to generate a painterly image. The details of each step of the algorithm are introduced in the following sections.

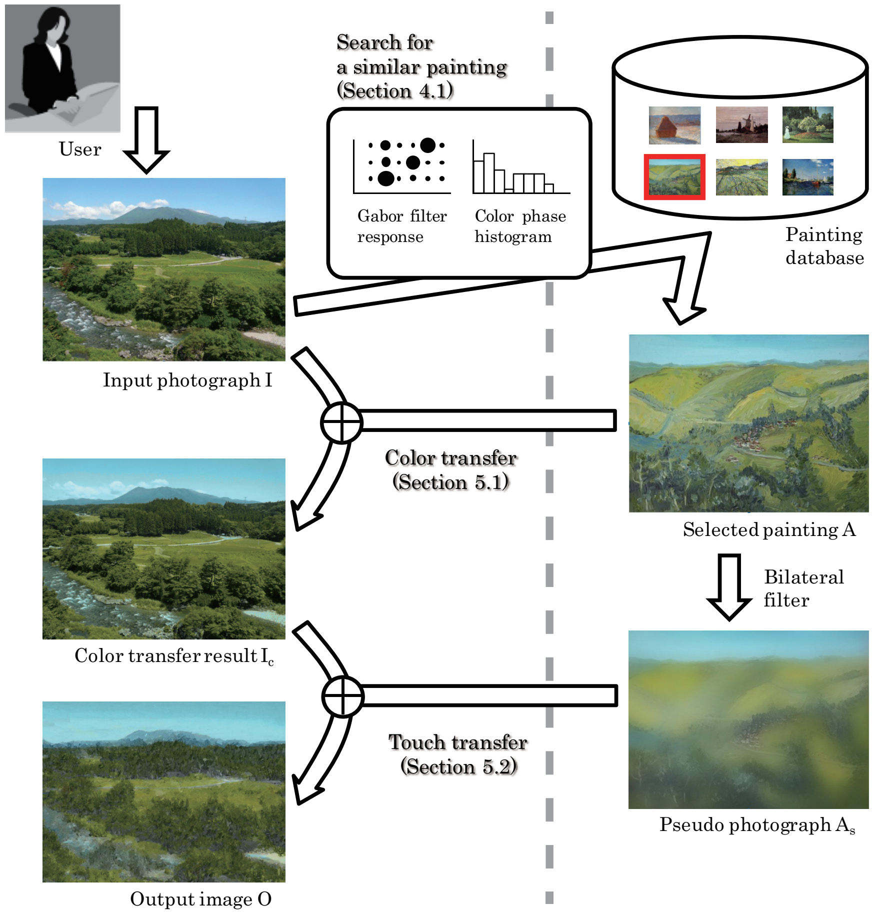


Fig. 3. Overview of our proposed method.

4 Searching for Paintings with Similar Scenes

Composition and color are two main elements that help to determine the similarity of two scenes. The GIST feature, which models the rapid initial recognition of human visual perception, is an effective gauge of a scene's composition. We combine a GIST feature with a color feature to automatically select from a sample painting set a painting that depicts a scene similar to the scene in the input picture.

4.1 Feature vector

GIST GIST perception is the visual perception that a human experiences within 100 ms of seeing an image or scenery. In this perceptional period, humans can recognize the overall composition of what they are looking at. Oliva and Torralba [15] proposed the following multidirectional, multifrequency Gabor filter banks that could be used as a computational model for GIST: Passing an image through Gabor filters gives the viewer a basic idea of the shapes of the subjects in the image. Formula 1 is the formula for Oliva and Torralba's multidirectional, multifrequency Gabor filters.

$$
\begin{aligned}
&g(x, y : \lambda, \theta, \psi, \sigma, \gamma) \\
&\qquad = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda} + \psi\right), \\
&x' = x\cos\theta + y\sin\theta,\ \ y' = x\sin\theta + y\cos\theta.
\end{aligned} \tag{1}
$$

Here, λ is the frequency parameter. Controlling this value makes it possible to create Gabor filters for various frequencies. θ is the parameter for direction, the control of which makes it possible to create Gabor filters with different orientations. Figure 4 shows the results of passing an image through multidirectional, multifrequency Gabor filters.

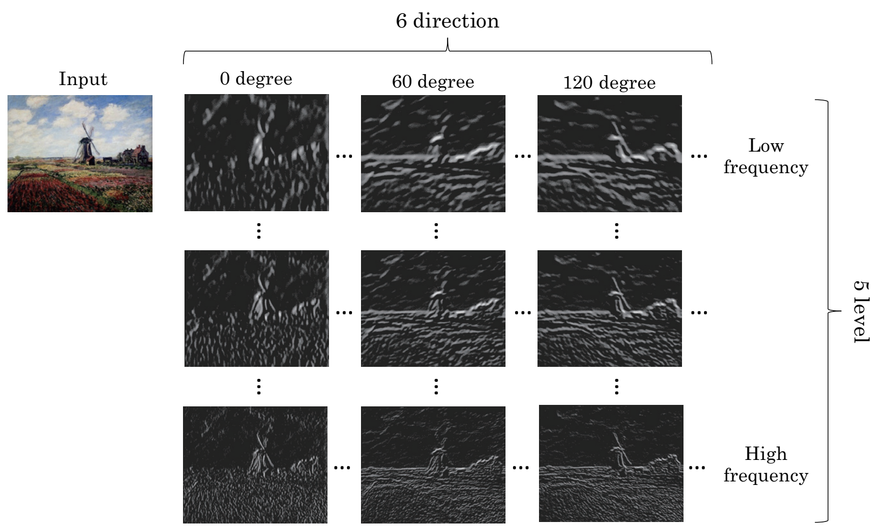


Fig. 4. Structure information obtained via multidirectional, multiresolutional Gabor filters.

A low-frequency Gabor filter reacts to and detects the larger, wider structures in the image, while a high-frequency Gabor filter reacts to and detects the smaller, finer structures in the image. The filter also detects structures extending in specific directions, depending on the angle. The GIST feature is calculated by applying multidirectional, multifrequency Gabor flters to the image, dividing the results from each filter into a grid of regions, and computing the sum of each region. Such a GIST feature represents the composition of an image or view and provides a general structure of the subjects in each region. The method proposed in this paper involves passing the input image through a total of 30 Gabor filters (six directions at five frequency levels) to compute the GIST feature. GIST feature quantity is usually calculated by applying multidirectional, multiresolutional Gabor filters to the image, dividing the results from each filter into a grid of regions, and finding the sum of each region. The GIST feature has been primarily used in performing high-speed searches of large-scale image databases for images with similar compositions. Hoping to interpolate partially damaged images, Hays and Efros [6] achieved some good results using a GIST feature to search the massive volume of images available on the Internet for images with compositions similar to that of an input image. There are many other cases where GIST has helped researchers rapidly sift through the innumerable images on the Internet to find and use images with similar compositions.

Color Users can use a GIST feature to get a basic idea of the structure of the subjects in an image and then search for an image with similar composition. However, these features are generally based on image edge information. While this allows for image composition detection, it does not provide enough information to determine whether two scenes are actually similar; a tree with colored leaves and a tree with green leaves can end up indistinguishable. To rectify this problem, our method uses a GIST feature in combination with a color feature. We use color feature corresponding to a histogram of the hue component of the HSV color model. We quantize the hue wheel into 12 equal parts and compute the histogram for the input image as shown in Figure 5. Combining the color feature obtained via this method with the corresponding GIST feature makes it possible to search for images with perceptually matching scenes.

4.2 The rule of thirds

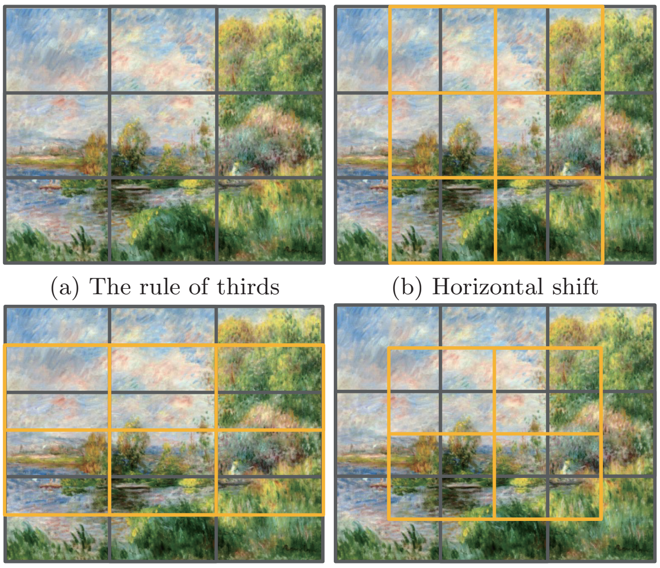
The above discussion examined the methods for calculating the GIST feature and color feature. These features are computed by passing an image through filters, dividing the resulting image into a grid, and then adding up the corresponding pixel values in individual regions. Oliva and Torralba [15] used a 4×4 grid to compute the GIST feature. For the method proposed in this paper, we use the "rule of thirds," a classic technique in the artistic disciplines, based on the consideration that the images to be searched for are artistic works. The rule of thirds, used to compose paintings, pictures, and other visual images, proposes that an image should be divided by two equally spaced horizontal lines and two

(a) Input image (b) Results of quantizing the image into 12 colors using color phases

Fig. 5. Image quantized into 12 colors using color phases.

equally spaced vertical lines. The boundaries between land and sky, etc., in the image should be placed along these lines, and the most important subjects should be situated on the intersections between the lines. The Monet painting shown in Figure 6(a) is an illustration of the rule of thirds. The boundary between the water and sky aligns with a horizontal dividing line, while the important subjects – the yacht and the house – lie where the lines cross. This placement guideline stabilizes the composition, creating a more attractive paintings or picture. The rule of thirds has also been used as a criterion for a compositional placement optimization technique designed to improve the aesthetic quality of a painting (Liu et al. [13]). The method proposed in this paper adopts the rule of thirds to create nine block regions, divided by equally spaced horizontal and vertical lines, when calculating the GIST feature and color feature. As the boundaries and objects tend to fall on these lines, we shift phases and use block regions containing lines to capture this information (Figures 6(b) and 6(c)). The important subjects in an image also tend to lie on the intersections between the lines representing the rule of thirds. Information about these objects can be captured by using block regions that contain these intersections, which can be obtained by shifting phases both vertically and horizontally as shown in Figure 6(d). In summary, our method calculates the GIST and color features based on 25 total block regions: the 9 block regions created by breaking the image into a 3×3 grid and the 16 regions created by shifting phases to include the lines and intersections of the rule of thirds.

The feature calculation is performed after normalizing the sizes of paintings. Using the results of passing the image through the Gabor filter and the results of quantizing the image into 12 colors, the method then calculates the GIST feature and color feature of each of the 25 block regions to produce a feature vector with 1,050 dimensions ((6 directions $\times$ 5 frequency levels + 12 colors) $\times$ 25 block regions).



(a) The rule of thirds (b) Horizontal shift

(c) Vertical shift (d) Horizontal + vertical shift

Fig. 6. Rule of thirds–based composition.

4.3 Distance

We evaluate the similarity of the scenes of input image i and the searched painting s using the following global distance $G_{g(i,s)}$.

$$G_{g(i,s)} = \sum_{n=1}^{25} |F_n^i - F_n^s|. \tag{2}$$

Here, $|\cdot|$ expresses the L_2 norm, while F_n^i and F_n^s are the 42-dimension (6 directions $\times$ 5 frequency levels + 12 colors) feature vectors for the image i and painting s, respectively. However, this formula is not ideal. Imagine, for instance, two scenes depicting cars. Provided that there are no major differences in the backgrounds, a person would think the scenes are essentially the same even if the cars are in slightly different positions. If the positions of the cars in the images are at least 1 block different, Eq. (2) would produce a large distance value and classify the images as dissimilar. To rectify this problem, our method introduces a component that incorporates local mapping into the distance: it compares regions that are the most similar rather than regions in the same positions. This component is called the local distance value G_l and is expressed as follows:

$$G_{l(i,s)} = \sum_{n=1}^{25} \min_m |F_n^i - F_m^s|. \tag{3}$$

F_m^s represents the region with the feature vector that most closely approximates region F_n^i in the input image. Independent of position, Eq. (3) produces a small distance value as long as the painting contains regions similar to those of the picture. As the resulting value will be low in any case where two images have similar elements, using this local distance by itself creates the risk of returning a search result with a completely different composition. Thus, our method calculates a final difference value G as the weighted sum of the global distance in Eq. (2) and the local distance in Eq. (3) in order to measure the similarity in terms of both composition and elements.

$$G_{(i,s)} = (1.0 - \rho)G_{g(i,s)} + \rho G_{l(i,s)}. \tag{4}$$

One can change the focus between composition and elements by modifying coefficient $\rho(0.0 \leq \rho \leq 1.0)$. Increasing ρ makes the search procedure more likely to return a painting with the same sorts of subjects as the picture, even if the scenes are different.

As local distance values are calculated using the most similar regions from the two images, the theoretical range may be the same as that of the global distance value, but the actual value will be much smaller. In order to ensure the effectiveness of Eq. (4) in an actual usage setting, our method uses statistical methods to predict and normalize the actual ranges of both distance values. After calculating the averages and distributions of the local and global distances for the input image and all the painting images in the painting database, the method uses Eqs. (5) and (6) to normalize both distance values to within the range of 0.0 to 1.0.

$$G'_{g(i,s)} = \frac{1}{1 + exp(-(G_{g(i,s)} - \bar{G}_g)/\sigma_{Gg})}, \tag{5}$$

$$G'_{l(i,s)} = \frac{1}{1 + exp(-(G_{l(i,s)} - \bar{G}_l)/\sigma_{Gl})}. \tag{6}$$

Here, $G'_g(i,s)$ and $G'_l(i,s)$ are the normalized local distance and global distance, respectively. $\bar{G}_g$ and $\bar{G}_l$ are the statistically calculated averages of the respective distance, while σ_{G_g} and σ_{G_l} are the respective distributions. In our method, we calculate these values for every input image and normalize accordingly based on the pictures. The method then uses distance, calculated by replacing $G_g(i,s)$ and $G_l(i,s)$ from Eq. (4) with $G'_{g(i,s)}$ and $G'_{l(i,s)}$ from Eqs. (5) and (6), to measure the similarity of the images' compositions and elements.

5 Transferring Painting Features

The painting used for feature transfer purposes shares a similar composition with the picture in question, so there is a good chance that both images will feature similar elements in corresponding positions. Given this assumption, our method takes position linking into account when transferring features. Before transferring color, we remap the colors to align the overall color of input image i with that of example painting s. Color remapping is performed using Eq. (7). Here, *color* is the pixel value of each component of input image i in YIQ color space; σ_i and σ_s are the standard deviations of each component of the input image and example painting, respectively; and $\overline{i}$ and $\overline{s}$ are the averages of each component of the input image and example painting, respectively.

$$color^{'} = \frac{\sigma_s}{\sigma_i}(color - \overline{i}) + \overline{s}. \tag{7}$$

5.1 Transferring color

Color transferring is done by setting the color of a pixel p in input image i to that of the pixel q that has the smallest distance $C(i,s,p,q)$ from p among all other pixels in painting s. The distance $C(i,s,p,q)$ is given by Eq. (8). ΔEab is the color distance in $L^*a^*b^*$ color space between p and q, and σ is the standard deviation of the brightness in their neighborhood. The differences of both the color and brightness deviation are divided by the available maximum to normalize them within the range of 0.0 to 1.0.

$$C_{(i,s,p,q)} = (1-\gamma)(\frac{\Delta Eab}{2} + \frac{|\sigma_p - \sigma_q|}{2}) + \gamma \Delta F_{l(i,s,p,q)}, \tag{8}$$

$$\Delta F_{l(i,s,p,q)} = |F_p^i - F_q^s|.$$

The third term $\Delta F_{l(i,s,p,q)}$ is the local distance measuring the similarity between the block regions consisting of pixel p in input image i and pixel q in painting s. It is computed as the L_2 norm of the 42-dimension GIST and color feature vector of the two regions. Because $\Delta F_{l(i,s,p,q)}$ is computed for each region, discontinuity can be observed at the boundaries between two regions. To solve this problem and prevent $\Delta F_{l(i,s,p,q)}$ from changing suddenly at region boundaries, our method computes $\Delta F_{l(i,s,p,q)}$ through the bi-linear interpolation of its values at the four surrounding block regions. As shown in Figure 7, we denote the four surrounding closest regions as LT, RT, LB, and RB and the distance from the pixel to the center of each block as D_T, D_L, D_R, and D_B, respectively, from which $\Delta F_{(i,s,p,q)}$ can be computed with Eq. (9).

$$\Delta F_{l(i,s)} = \\ (D_L \ D_R) \begin{pmatrix} |F_p^{iLT} - F_q^s| & |F_p^{iLB} - F_q^s| \\ |F_p^{iRT} - F_q^s| & |F_p^{iRB} - F_q^s| \end{pmatrix} \begin{pmatrix} D_T \\ D_B \end{pmatrix}. \tag{9}$$

In Eq. (8), a constant $r(0.0 \leq r \leq 1.0)$ is used to control the weight of color and brightness distribution over the GIST feature and can be specified by the user. Figure 8 shows the color transfer results. Figure 8(c) shows the results generated when the overall color of the painting in Figure 8(b) was remapped onto

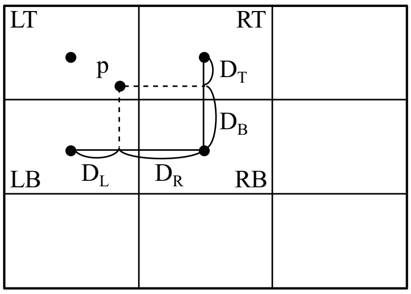


Fig. 7. Bilinear interpolation of region difference.

the picture in Figure 8(a) using Eq. (7). Figure 8(d) shows the results generated by transferring the color of pixels in Figure 8(b) with the smallest distance given by Eq. (8) without considering the discontinuity at the boundaries. Figure 8(e) shows the results when using bilinear interpolation to compute $\Delta F_{(i,s,p,q)}$. The predominant red tones of Figure 8(b) were transferred to Figure 8(c), making the trees with slightly colored leaves look red. Considering that there are no red trees in the painting, however, the process of transferring the style of the painting should not result in the production of red trees. For Figures 8(d) and 8(e), the color from Figure 8(b) was transferred in a more detailed manner by including the cost of local distance. We can see that the trees that were red in the last comparison have turned green and now bear the color of the trees in the painting. The sky also features yellow, bringing it closer to the color of the sky in the painting. Figure 8(d) does not use interpolation and therefore shows sharp changes at regional boundaries, creating discontinuity in the transfer of the sky colors. Meanwhile, the discontinuity is eliminated in Figure 8(e) because the use of biliear interpolation prevents marked changes in feature similarity near boundaries.

5.2 Brushwork transfer

Our method uses Image Analogies [8] for brushwork transfer. With Image Analogies, one can transfer the relationship between texture features in example images A and A' to B and then create B', an image with the same relationship between texture features. In other words, a user with a picture of a scene that appears in an existing painting could apply Image Analogies to the picture painting pair (the example images) and transfer the stroke textures of the painting to the input image. However, there are rarely any actual pictures of the scenes depicted in existing paintings. We create a substitute picture by using a bilateral filter on the source painting. As painters often use brushstrokes to evoke various colors and tones, regions of pictures where colors and brightness change in a smooth fashion sometimes also contain high-frequency stroke texture components. Using a

(a) Input image (b) Painting with a similar scene (c) After overall color transfer (d) Local color transfer (without bilinear interpolation) (e) Local color transfer (with bilinear interpolation)

Fig. 8. Color transfer results.

bilateral filter, one can eliminate these in-region high-frequency components but still protect regional boundaries. The Image Analogies approach copies brightness from A' to B on a pixel-by-pixel basis, comparing the colors of A and B in window regions and copying the pixels from the most similar regions. Our method brings $\Delta F_{l(i,s,p,q)}$, which Eq. (9) uses to gauge the similarity of feature vectors across regions, into this operation. The following formula is used to calculate distance $S_{(i,s,p,q)}$, which figures into the transfer of pixel q from painting s to pixel p of input image i.

$$S_{(i,s,p,q)} = (1-\gamma)(|W_p^B - W_q^A| + |W_p^{B'} - W_q^{A'}|) + \gamma \Delta F_{l(i,s)}. \qquad (10)$$

Here, W_qs and W_ss are the window regions around the pixels p and q, respectively, and $\Delta F_{l(i,s,p,q)}$ is the local distance and is the same as for Eq. (8). For the transfer to pixel p, the method finds the pixel q with the smallest $S_(i,s,p,q)$ in Eq. (10) and transfers the Y value of pixel q in YIQ color space to pixel p.

6 Results and Experiments

6.1 Results

We used Google Images to search for paintings by 12 different painters, including Van Gogh and Monet, using the search string "(painter name)" and "landscape-photo." We retrieved approximately 500 images of paintings and stored them in a database together with the precomputed feature vectors.

Figures 9 and 10 show several results generated with the proposed method. The left column shows the input pictures. The images shown in the middle column are the paintings automatically retrieved from the database. The right column shows the results obtained by transferring the color and textures of the corresponding paintings to the input pictures. We can see that paintings with similar scenes have been retrieved and their color and brush work styles were successfully transferred to the input images.

6.2 Evaluation

Painting retrieval with improved GIST feature We conducted a subjective evaluation experiment to determine whether the method proposed in this paper could search the database for images with similar scenes and properly rank the images by similarity. We recruited 12 university students as the subjects. For the experiment, we applied our method to eight pictures with different scenes and searched the database for ranked results of similar painting images. We set the ρ coefficient for weighting local distance and global distance to 0.5. Then, we selected the highest-ranked image search result and 11 other randomly selected search results to create a set of 12 painting images for each picture.

When presenting the 12 painting images to the participants, we randomly shuffled the highest-ranked search result with the 11 other randomly selected search results and displayed all the images simultaneously on a single web page. From these 12 choices, the subjects were asked to select the painting image with the scene that most closely resembles that of the picture. Our null hypothesis was that the 12 painting images would all be chosen with equal probability. Under that hypothesis, we used a binomial test to determine whether the painting image with the highest ranking by the proposed method received a significant number of votes from the subjects. If our method proved to be an effective means of searching for painting images with similar scenes, there would be a statistically significant correlation between the highest-ranking image selected by the system and the images selected by the experiment participants. Table 1 shows the results.

Table 1. The number (and ratio) of people (out of 12 subjects in total) who chose the method's highest-ranking image as the image with the most similar scene (**: $p \leq 0.01$, *:$p \leq 0.05$, +: $p \leq 0.1$).

Dataset	A	B	C	D	E	F	G	H
Number of subjects	3	10	3	8	9	4	1	8
Ratio	0.25^{+}	0.83^{**}	0.25^{+}	0.67^{**}	0.75^{**}	0.33^{*}	0.08	0.67^{**}

The subjects selected the highest-ranking image with statistical significance for 7 of the 8 pictures. These results suggest that, for most pictures, the search

Fig. 9. Results. Left: input images, middle: retrieved example paintings, right: synthesized images. Corresponding painters, from top to bottom: (5) Gustave Caillebotte, (7) Jacob Isaacksz van Ruisdael, (8) Jean-Baptiste Armand Guillaumin, and (10) Claude Monet.

method proposed in this paper is capable of retrieving images that are perceptually similar to the picture. The findings also imply that the automatic process of selecting the highest ranking painting image from the database as the transfer source is effective in the majority of cases.

Fig. 10. Results. Left: input images, middle: retrieved example paintings, right: synthesized images. Corresponding painters, from top to bottom, (1) Berthe Morisot, (3) Eugene Henri Paul Gauguin, (4) Vincent van Gogh, and (12) Pierre-Auguste Renoir.

Style transferring We also conducted a subjective evaluation experiment to determine whether the method could reflect the styles of specific painters. Example paintings by 12 artists were used in the experiment.

(1) Berthe Morisot
(2) Edgar Degas
(3) Eugene Henri Paul Gauguin
(4) Vincent van Gogh
(5) Gustave Caillebotte

(6) Jacob Camille Pissarro
(7) Jacob Isaacksz van Ruisdael
(8) Jean-Baptiste Armand Guillaumin
(9) Jean-Baptiste-Camille Corot
(10) Claude Monet
(11) Paul Cezanne
(12) Pierre-Auguste Renoir

Subjects were shown an image generated by transferring a specific painter's style onto a picture and a set of 12 painting images, including the specific painter's work used for the transfer and 11 other paintings not used as the example painting. The 11 paintings not used as the example (hereafter called reference paintings) were obtained by searching the database for paintings by the other 11 painters and choosing the works that had the strongest similarities with the example painting in terms of both the scene and the painting style. We displayed each of the 12 paintings together with the generated image one at a time for 10 seconds each in random order. We then explained the concepts of "brushwork" and "color" to the subjects, who proceeded to rate each painting's degree of similarity to the generated image (with 1 being the lowest degree of similarity and 7 being the highest). The subjects are eight university students. Table 2 shows the results of the experiment. We computed the average degree of similarity of the eight subjects for each of the 12 paintings and then used the Wilcoxon signed-rank test to test the presence of any significant difference between the example painting and each of the 11 reference paintings. Table 2 shows the number of paintings with an average degree of similarity significantly lower (at significance level of 5%) than that of the example painting used to generate the image.

Table 2. Number of paintings (out of 11 in total) evaluated as having significantly lower average degree of similarity (at a significance level of 5%) than the example painting.

ID	1	2	3	4	5	6	7	8	9	10	11	12
Vote	7	9	11	11	11	11	9	7	11	11	11	4

Among the 12 images evaluated, 7 images were evaluated to be more similar to the example painting than all 11 of the other reference paintings. Among the remaining 5 images, 2 were evaluated to be more similar to the example painting than 9 of the other paintings and 2 were evaluated to be more similar to the example painting than 7 of the other paintings. Only 1 image was evaluated to be less similar to the example painting than 7 other paintings. The fact that for 5 images, the example image was not always the one selected as most similar does not necessarily mean that the method failed to transfer the style. In some cases, an image by an artist with a style more similar to that of the artist of the example painting was present.

We showed the resulting images to two professional painters. The following are their comments on the three images with the best, worst, and middle scores in Table 2. The images are also shown in Fig. 10.

ID 3 (11 votes): The generated image is good as a painting. Colors are well transferred. The characteristics of Gauguin are not well represented in the image. Fine textures are lost. By blending the generated image and the original input, it may be possible to represent the textures in the result.

ID 1 (7 votes): The result is acceptable though not so good. Coloring and shading look good. The touch is beautiful, but does not resemble that of the original painting. The touch of tiny regions such as woods and small objects is well transferred in the image. Colors are also transferred, but lack depth. The image well reflects the atmosphere of the original painting on the whole. We can feel hazy air in the space.

ID 12 (4 votes): Objects are more vividly colored in the painting than they actually are, but such exaggeration is not achieved in the generated image. The vertical lines are transferred, but not enough. One can see the flow of touch in the original painting, while such effect is not well represented in the generated image. The inconsistency of directions between shadows and touch can be perceived. The aliasing artifacts can be seen on the thin branches of trees. The thinnest ones are not preserved at all. A general watercolor effect of commercial painting systems can generate similar images.

Overall impression: All images are good in the sense that they resemble real paintings, but they are not of the quality of professionally painted ones.

7 Conclusion

Grounded in the belief that painters change their painting parameters based on their scenes and subjects, we proposed a painterly image generation method that involves searching for images with similar scenes and transferring the color and brushwork from the painting to the target picture in order to reflect the specific painter's style. We also showed through subjective evaluation experiments that the method can successfully generate a painterly image that reflects the specific painter's style. The spread and popularity of mobile devices with high-performance cameras has allowed people to photograph and capture striking scenes quickly and easily, regardless of the person's artistic ability. The Internet, meanwhile, has given people direct access to a wealth of painterly images and artistic work in a diverse array of styles. Content prosumers should thus benefit greatly from the method proposed in this paper, which makes use of pictures and Internet-based resources. Although the method works with essentially any type of painting material or painting style, the experiments in this paper used only landscape paintings by several prominent artists for copyright-related reasons. Moving forward, we plan to experiment with paintings of different subjects and styles.

Oliva and Torralba [15] have already proven that GIST feature vector, calculated by dividing an image into a grid of regions, is an effective means of ascertaining the broad, general composition of an image. By complementing GIST with the rule of thirds, we have successfully improved the accuracy of searches for similar scenes.

We now plan to develop a method that will break an image into regions based on features like color, texture, and frequency and then perform matching on the various regions. Given that painters use rough strokes and fine strokes to give their scenes basic outlines and depict the smaller details of their subjects, there is also the potential to devise a method for transferring styles to a given picture by region. It may be possible to appropriate the background and basic outlines of the painting with the most similar scene and then, instead of using the same base painting for the remaining aspects of the image, pick and choose from a selection of all the available painting samples to find the paintings that most closely match the individual regions. Deep neural network is another promising tool for painting style transferring [4]. Further improvement can be expected by combining our scene aware approach with newest deep neural network such as CNN.

The spread and popularity of mobile devices with high-performance cameras has allowed people to photograph and capture striking scenes quickly and easily, regardless of the person's artistic ability. The Internet, meanwhile, has given people direct access to a wealth of painterly images and artistic work in a diverse array of styles. Content prosumers should thus benefit greatly from the method proposed in this paper, which makes use of pictures and Internet-based resources. Although the method works with essentially any type of painting material or painting style, the experiments in this paper used only landscape paintings by several prominent artists for copyright-related reasons. Moving forward, we plan to experiment with paintings of different subjects and styles.

Acknowledgement

This work was partially supported by JSPS KAKENHI (JP16K12459, JP26560006, JP25280037).

References

1. Ashikhmin, M.: Fast texture transfer. IEEE Computer Graphics and Applications **23**(4), 38–43 (2003)
2. Chang, C., Peng, Y., Chen, Y., Wang, S.: Artistic painting style transformation using a patch-based sampling method. Journal of Information Science and Engineering **26**(4), 1443–1458 (2010)

3. Chang, W.H., Cheng, M.C., Kuo, C.M., Huang, G.D.: Feature-oriented artistic styles transfer based on effective texture synthesis. Journal of Information Hiding and Multimedia Signal Processing **6**(1), 29–46 (2015)
4. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
5. Guo, Y., Yu, J., Xu, X., Wang, J., Peng, Q.: Example based painting generation. Journal of Zhejiang University SCIENCE A **7**(7), 1152–1159 (2006)
6. Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Transactions on Graphics **26**(3) (2007)
7. Hegde, S., Gatzidis, C., Tian, F.: Painterly rendering techniques: a state-of-the-art review of current approaches. Computer Animation and Virtual Worlds **24**(1), 43–64 (2013)
8. Hertzmann, A., Jacobs, C., Oliver, N., Curless, B., Salesin, D.: Image analogies. In Proc. SIGGRAPH pp. 327–340 (2001)
9. Kyprianidis, J.E., Collomosse, J., Wang, T., Isenberg, T.: State of the "art": A taxonomy of artistic stylization techniques for images and video. IEEE Transactions on Visualization and Computer Graphics **19**(5), 866–885 (2013)
10. Lee, H., Seo, S., Ryoo, S., Yoon, K.: Directional texture transfer. In: Proceedings of International Symposium on Non-Photorealistic Animation and Rendering, pp. 43–48 (2010)
11. Lee, H., Seo, S., Yoon, K.: Directional texture transfer with edge enhancement. Computers and Graphics **35**(1), 81–91 (2011)
12. Lee, T.Y., Yan, C.R.: Feature-based texture synthesis. In: Proceedings of International Conference on Computational Science and Its Applications (2005)
13. Liu, L.G., Chen, R.J., Wolf, L., Cohen-Or, D.: Optimizing photo composition. EUROGRAPHICS **29**(2), 469–478 (2010)
14. Liu, Y., Cohen, M., Uyttendaele, M., Rusinkiewicz, S.: Autostyle: automatic style transfer from image collections to users' images. In: Computer Graphics Forum, vol. 33, pp. 21–31. Wiley Online Library (2014)
15. Oliva, A., Torralba, A.: Building the gist of a scene: The role of global image features in recognition. In Visual Perception, Progress in Brain Research **155**, 23–36 (2006)
16. Shih, Y., Paris, S., Barnes, C., Freeman, W.T., Durand, F.: Style transfer for headshot portraits. ACM Transactions on Graphics **33**(4) (2014). Article 148
17. Wang, W., Yang, H., Sun, J.: Efficient example-based painting and synthesis of 2d directional texture. IEEE Transactions on Visualization and Computer Graphics **10**(3), 266–277 (2004)
18. Xie, X., Tian, F., Seah, H.S.: Feature guided texture synthesis (fgts) for artistic style transfer. In: Proceedings of International Conference on Digital Interactive Media in Entertainment and Arts, pp. 44–49 (2007)
19. Xie, X., Tian, F., Seah, H.S.: Style learning with feature-based texture synthesis. Computers in Entertainment (CIE) **6**(4) (2008)