Hierarchical Oil Painting Stylization with Limited Reference via Sparse Representation

Saboya Yang [1], Jiaying Liu [2]*, Shuai Yang [3], Sifeng Xia [4], Zongming Guo [5]

Institute of Computer Science and Technology, Peking University
No.5 Yiheyuan Road, Haidian District, Beijing, P. R. China, 100871

[1] yangsaboya@pku.edu.cn
[2] liujiaying@pku.edu.cn
[3] williamyang@pku.edu.cn
[4] xsfatpku@pku.edu.cn
[5] guozongming@pku.edu.cn

***Abstract*—Traditional image stylization is enforced by learning the mappings with an external paired training set. But in practice, people usually encounter a specific stylish image and want to transfer its style to their own pictures without the external dataset. Thus, we propose a hierarchical stylization model with limited reference particularly for oil paintings. First, the edge patch based dictionary is trained to build connections between images and limited reference, then reconstruct the structure layer. Due to the highly structured property of saliency regions, the saliency mask is extracted to integrate the structure layer and the texture layer with different weights. Hence, the advantages of both sparse representation based methods and example based methods are integrated. Moreover, the color layer and the surface layer are considered to make the output more consistent with the artist's individual oil painting style. Subjective results demonstrate the proposed method produces desirable results with state-of-art methods while keeping consistent with the artist's oil painting style.**

I. Introduction

Image stylization has attracted much attention from both researchers and users due to its practicality, adaptability and enjoyment. It aims to transform images from one style to another. Instead of a general image type, style here refers to a more specific individual drawing style. The input and output may have completely different visual perceptions while expressing the same contents. Nowadays, image stylization methods are widely used as facilities for image editing programs and camera applications of mobile phones. Therefore, it is of great importance to provide people with good stylization experience.

In the past decades, researchers made progress in image stylization problem. Hertzmann [1] proposed to synthesize stylish images by composing virtual brush strokes incrementally. This kind of methods is specialized for target styles, such as oil painting and crayon, to produce stylish images. Nevertheless, it is a hard job for users to seek the method that produces their favorite styles before stylization. This results in the difficulty of acquiring stylized images in specifically customized styles. Therefore, as introduced in [2], researchers proposed to do mapping in feature spaces for stylization. Sparse representation with a learned dictionary has been a popular research area recently due to its remarkable performance for many image restoration scenarios. The input image is adaptively decomposed, represented and reconstructed on the learned coupled [3] or semi-coupled [4] dictionaries. When it comes to image stylization, dictionaries are trained on paired source stylish images and target images. But in most cases, the paired training images are inaccessible as shown in the scenario of Fig.1, leaving the aforementioned methods unresolved with limited reference.

Fig. 1. The application scenario of image stylization.

To solve the above problem, we propose a hierarchical model with limited reference for image stylization. Due to the various characteristics of different types of drawings, we focus on oil paintings in this paper. The hierarchical model is made up of five layers: the structure layer, the texture layer, the saliency layer, the color layer and the surface layer. The edge feature [5] is utilized to train dictionaries to reconstruct the structure layer. The texture layer is subsequently generated by example based texture transfer. In addition, salient regions, which catch people's most attention, are extracted to help synthesize the texture layer with the structure layer [6]. Hence, the basic structures of the stylized image are well preserved while the textures are synthesized. At the same time, one of the

*Corresponding author

This work was supported by National High-tech Technology R&D Program (863 Program) of China under Grant 2014AA015205National Natural Science Foundation of China under contract No.61472011 and Beijing Natural Science Foundation under contract No.4142021.

978-1-4673-7478-1/15/$31.00

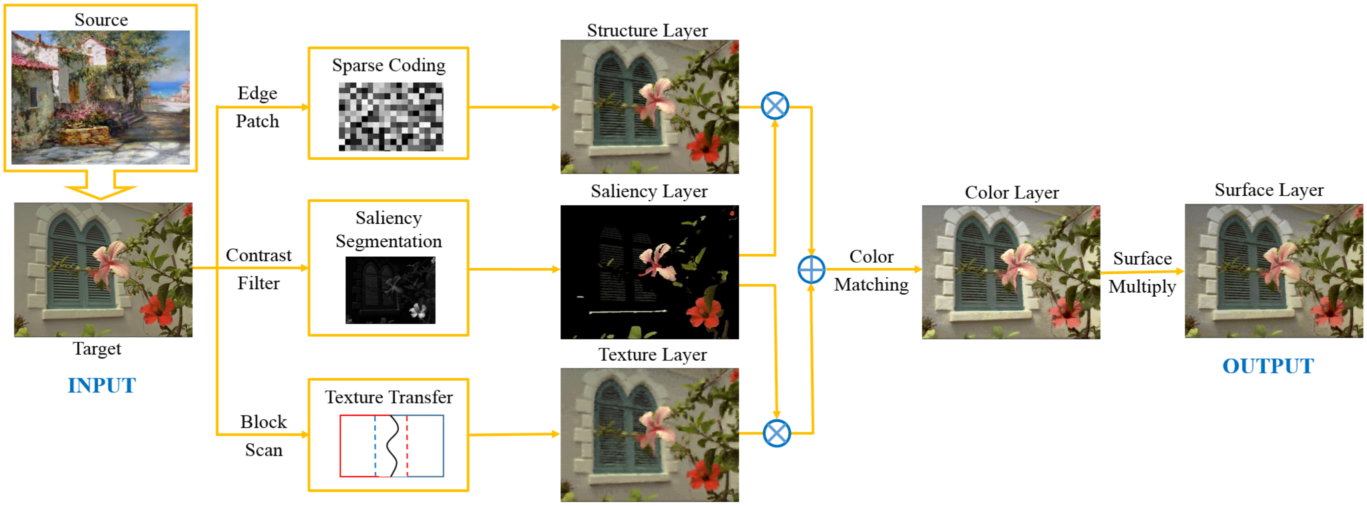

Fig. 2. Framework of sparse representation based hierarchical oil painting stylization with limited reference algorithm.

distinguishable features of oil paintings comes from a volume of colors expressed by used pigments and drawing surfaces. Therefore, we consider color and surface as supplementary features of artists' oil paintings to make the stylized image more consistent with the source style.

In conclusion, the contributions of this paper are :

- proposing a hierarchical image stylization model with limited reference;
- combining the advantages of both sparse representation based and example based methods by fusing the structure layer and the texture layer guided by the saliency mask;
- taking color and surface into consideration, which act as extensive features of artists' personalized styles.

The rest of the paper is organized as follows. In Sec. II, traditional sparse representation based image stylization algorithm is reviewed. Sec. III focuses on the proposed sparse representation based hierarchical oil painting stylization model with limited reference. Experiments are presented in Sec. IV. In the end, a brief conclusion is demonstrated in Sec. V.

II. General Sparse Representation Based Image Stylization

Traditional sparse representation based image stylization method stylizes the image by coupled dictionaries. The input image pairs consist of the stylish source images and the target images to be stylized. It is assumed that there exists a certain relationship between two styles in the sparse domain. The output is the stylized image Z which has the same contents as the target while in the same style as the source. Then the general framework is divided into two stages: the dictionary learning stage and the image reconstruction stage.

We assume D_i is the i-th dictionary base of an over-complete dictionary $D \in R^{m\times n}$. Each training sample y_i is corresponding to its own sparse coefficient α with the dictionary D. To solve the non-convex problem, the sparse constraint $\|\cdot\|_0$, which represents the number of nonzero elements in one vector, is replaced by $\|\cdot\|_1$ as follows

$$\begin{aligned} &\arg\min_{D,\alpha} \sum_i \|y_i - D\alpha\|_2^2 + \lambda\|\alpha\|_1, \\ &s.t. \|D_i\|_2^2 \leq 1, i = 1, 2, ..., n. \end{aligned} \tag{1}$$

In the dictionary learning stage, external source and target image pairs (e.g. oil painting-photo pairs) $\{S, T\}$ are segmented into small image patches and trained to acquire a coupled dictionary. The underlying mapping relations between training image pairs are learned in the sparse domain. In Eq.(2), D_s and D_t are coupled dictionaries.

$$\begin{aligned} D_s &= \arg\min_{D_s} \|S - D_s\alpha\|_2^2 + \lambda\|\alpha\|_1, \\ D_t &= \arg\min_{D_t} \|T - D_t\alpha\|_2^2 + \lambda\|\alpha\|_1. \end{aligned} \tag{2}$$

When reconstructing the images, image patches are represented by sparse coefficients. Moreover, it is assumed that the coupled dictionaries share the same sparse representations for each patch pair [7]. Hence, the transformed image Z is reconstructed by the source dictionary D_s with sparse coefficients α, coded by target image patches over the target dictionary D_t.

This method assumes the existence of an external paired training set which refers to a set of examples illustrating how images are stylized. However, people in the real world usually encounter a specific image and want to transfer its style to their own pictures with only the source accessible. Meanwhile, due to the imparity of different styles of paintings, we focus on oil paintings in this paper. Hence, we propose a sparse representation based hierarchical oil painting stylization algorithm with limited reference.

III. Sparse Representation Based Hierarchical Oil Painting Stylization with Limited Reference

Taking the notable properties of artist's oil painting works into account, we propose a hierarchical stylization algorithm with limited reference based on sparse representation. The framework of the proposed algorithm is illustrated in Fig.2.

It is separated into five layers: the structure layer, the texture layer, the saliency layer, the color layer and the surface layer. More details can be viewed in the following sections.

A. Sparse Representation Based Structure Layer

Oil paintings have to represent the main content of the image besides all esthetic treatments. The source stylish image and the target image have different contents and are not copies in different styles. Thus, it is very difficult to build mappings between them directly when training dictionaries with limited reference. This leads to the idea that we have to build a dataset of corresponding patch pairs derived from the input image pair first before dictionary learning. The corresponding patch pairs may have similar contents but in different styles. Therefore, it leaves us to seek a style-invariant feature to relate the corresponding patches together.

As a matter of fact, the edge feature is style-invariant in most cases [5]. We tend to utilize it to relate two stylish images and build the coupled dictionary. Therefore, the guided image filter [8] is applied on the images to build the edge patch based dictionary. The filtered images are subtracted from the original images to obtain edge maps. With the corresponding edge maps, we can implement patch matching on the input image pairs. While p acts as a patch in the source edge map, q is a patch in the target edge map. To evaluate the similarity $W(p,q)$ of different edge patches p and q, gradient mean squared error (GMSE) is utilized. It is necessary to emphasize that both the intensity similarity and the structure similarity are important during patch matching in order to maintain the image contents.

$$W(p,q) = \|p-q\|_2^2 + \eta\|\nabla p - \nabla q\|_2^2, \tag{3}$$

where η defines a weighting parameter and ∇ is the gradient operator.

Fig. 3. Edge features are used to map similar patches between different styles for coupled dictionary learning.

With the corresponding patch pairs shown in Fig.3, the coupled dictionary is trained as Eq.(2). On the basis of the learned dictionary, the target image is sparsely coded to get sparse coefficients α. Then the coefficients α are multiplied by the source style dictionary D_s, recovering the structure layer Z_{sl} of the target image.

$$Z_{sl} = D_s\alpha. \tag{4}$$

B. Example Based Texture Layer

One of the prime differences between the original image and the corresponding oil painting is texture, which stands for the style. But sparse coding process is compromised of some approximate solutions. Therefore, sparse representation based methods smooth many details and result in unapparent textures. Thus, we present a texture layer to supply more textures of the source image to the structure layer.

We implement [9], which synthesizes the texture layer from patches of the source image, to maintain more texture details. The example based texture transfer synthesizes images in units of block by raster scan order. For every location, the input texture is searched for a set of blocks that satisfy the overlap constraints within some error tolerance. The process of searching is optimized by nearest neighbor (NN) search [10]. The suitable block should match the target image at that spot to keep the scenario. Then, the chosen block is pasted into the resulting texture. It should fit in seamlessly with its neighbours after some cuts, which lead to ragged edges. The cuts between two overlapping blocks B_1 and B_2 are performed with dynamic programming [11] by pursuing the minimal cost path through the error surface $b_{i,j} = \|B_1 - B_2\|_2^2$.

$$E_{i,j} = b_{i,j} + \min(E_{i-1,j-1}, E_{i-1,j}, E_{i-1,j+1}), \tag{5}$$

where E is the cumulative minimum error for all paths. Therefore, the texture layer Z_{tl} is composed of these chosen blocks after some cuts as follows

$$Z_{tl} = \arg\min_{Z_{tl}} E. \tag{6}$$

As a matter of fact, the synthesized image should have the textures of the source locally but look like the target globally.

After obtaining the texture layer, it is then applied to the structure layer to enrich texture details. As discussed in Sec. III-C, different parameters are adopted during the fusion considering features of saliency regions.

C. Saliency Layer to Synthesis Information

In an oil painting, artists usually wish to emphasize some objects of interests while depicting other regions with fewer details, either less saturated or more blurred. Based on scientific analysis, human eyes are especially sensitive to structural information. Thus, salient regions, which attract most of observers' attention for the ease of recognition [12], tend to be highly structured. To simulate this phenomenon, we perform saliency detection [13] to identify the regions that are likely to be emphasized. Different parameters are then applied to corresponding regions during texture synthesis.

Salient regions are segmented by the contrast filter in CIELab color space [14]. The distance $d_{i,j}$ between pixels in the subregion and in the neighborhood is measured under different scales s to filtrate regions with larger contrasts. Additionally, the map is over-segmented by hill-climbing algorithm [15] and K-means. If the t-th region R_t with an average saliency value v_t exceeding the threshold θ, it is defined as a salient region.

$$v_t = \frac{1}{|R_t|} \Sigma_{i,j \in R_t} \Sigma_s d_{i,j}. \tag{7}$$

Then two different parameters λ and μ are adopted to the corresponding salient and unsalient regions. In this way, a saliency mask M_{sa} is obtained as follows

$$M_{sa}(R_t) = \begin{cases} \lambda, v_t > \theta \\ \mu, v_t <= \theta \end{cases} . \tag{8}$$

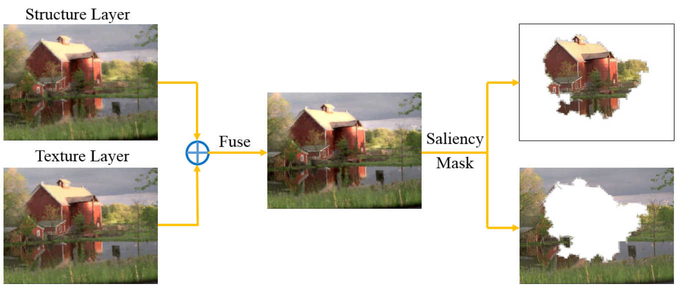

Fig. 4. The saliency layer provides a weighted mask for texture synthesis.

As illustrated in Fig.4 and Eq.(9), with the saliency layer M_{sa}, a weighted texture synthesis method is carried out. In the saliency mask M_{sa}, a weight λ with relatively high value is adopted to the salient region when attaching textures to the structure layer Z_{sl} to maintain the structure of the saliency regions. On the other hand, a relatively low weight μ is utilized for other regions to reflect the stroke feature of the source image from the texture layer Z_{tl}. Therefore, a fused stylized image Z_f is acquired.

$$Z_f = M_{sa} \cdot Z_{sl} + (\mathbf{1} - M_{sa}) \cdot Z_{tl}. \tag{9}$$

D. Color Layer for Color Adjustment

There exist a volume of colors in an oil painting which expresses both the original contents of the source image and the artist's feelings. When different artists draw the same scene, they utilize different pigments due to their personal preferences. Hence, color can be one of the distinguishable features of the artist's oil painting style. On account of this, the color layer is proposed to modulate the color of the stylized image to fit the color style of the source image.

When the image is represented in those common color spaces, such as RGB, HSV, there are many correlations between different channels' values. If pixels' colors in the stylized image are modified coherently, all channels must be adjusted in tandem to avoid distortions. Hence, we seek for an orthogonal color space without correlations to modify the color style.

The $l\alpha\beta$ color space [16] [17] is utilized in this paper to apply different operations in different channels without cross-channel artifacts. Eq.(10) shows how to transform the image from RGB to $l\alpha\beta$ color space.

$$\begin{gathered} \begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \log L \\ \log M \\ \log S \end{bmatrix}, \\ \begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \end{gathered} \tag{10}$$

$\bar{S}^l$ and $\bar{Z}_f^l$ are the mean of channel l in the source stylish image S and the stylized image Z_f. σ_{S^l} and $\sigma_{Z_f^l}$ refer to the standard deviations. The channel l in the stylized image Z_f is adjusted due to the channel l in the source image S.

$$Z_c^l = (Z_f^l - \bar{Z}_f^l) \times \frac{\sigma_{S^l}}{\sigma_{Z_f^l}} + \bar{S}^l. \tag{11}$$

Then the other two channels α, β are adjusted separately as l in Eq.(11). The image is transformed to RGB color space afterwards. In this way, the stylized image with color layer Z_c has the similar color style with the source as shown in Fig.5.

Fig. 5. An example of color layer. The color of one target image is transferred by two different source style and the results look pretty different.

E. Extra Surface Layer of Oil Paintings

Oil painting is drawn on surfaces like linoleum, wooden panel, paper, and canvas. The usage of different surfaces leads to different expressions even with the totally same brush strokes. It is mainly because of the textures and colors of different surfaces. Thus, in order to make our transformed images more similar to the artist's work, we consider a surface layer and assume that the transformed painting is drawn on the surface. In this paper, taking paper as an example of the surface layer, the texture of the paper surface Z_s is quantized directly from a scanned paper image and then the stylized image Z_c is adjusted in RGB color space as follows

$$Z = Z_c \cdot Z_s / 255. \tag{12}$$

In the way, the lightness, hue, and purity of the stylized image are adjusted to make it look like drawn on the paper surface.

With this aforementioned hierarchical model, the oil painting stylized image Z is obtained.

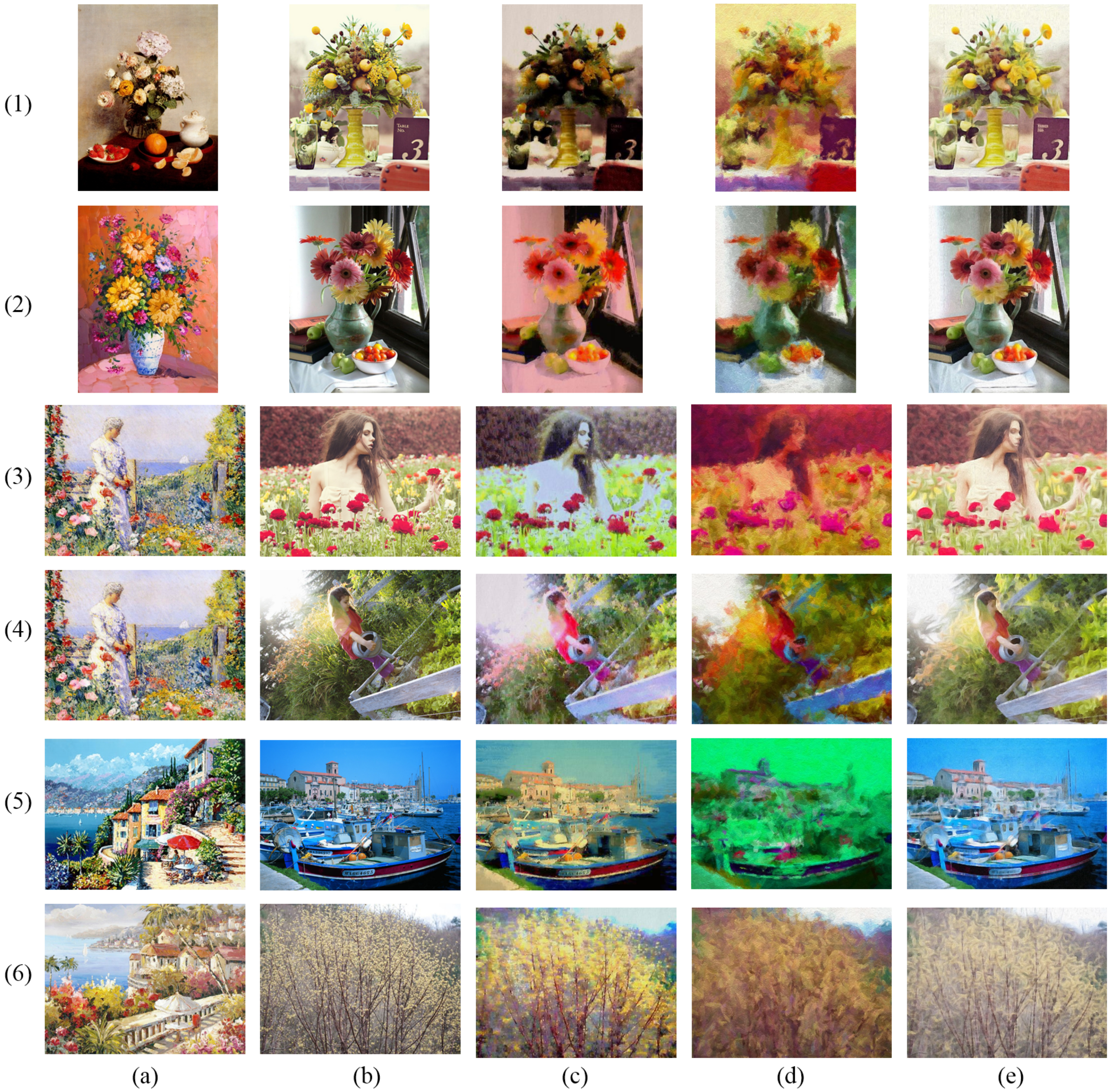

Fig. 6. Subjective experimental results. (a) Original source stylish image. (b) Target image. (c) Style transformation results by the proposed method. (d) Stylized image by Zhao's method [18]. (e) Stylized image by *BrushStroke* APP.

IV. Experimental results

To evaluate the effectiveness of the proposed method, we conduct experiments on several test images, gathered from the Kodak dataset and the Internet. These images have been released on our website[1]. In experiments, we input an oil painting image as the source and an original image as the target to be transformed. And for each input, the coupled dictionary is learned independently. The patch size is 7×7, and the overlap is $[5, 5]$. We compare the proposed algorithm with Zhao's method [18] and the oil painting filter of the *BrushStroke* APP [19]. And the subjective results are illustrated in Fig.6, Table I and Fig.7.

Fig.6(c) shows the image stylization results using the proposed hierarchical method. We pick three types of the common oil paintings: still-life, people and scenery, and compare the performances separately. The top two rows of Fig.6 show the stylization results of still-life. Compared with *BrushStroke* APP, the texture of the surface layer and the adjustment of color sufficiently embody the characteristics of the artist's

[1]http://www.icst.pku.edu.cn/course/icb/Projects/HOPS.html

source oil painting. The results in column(e) look similar to the original target image. Rows(3)(4) belong to the type of people. Our methods keep the details of people's faces while the other two methods lose them. Also, strokes in the stylized images *BrushStroke* APP creates have odd textures and look more like watercolor drawings. Moreover, the last two rows are the results of scenery. Our method synthesises the details of textures while preserving the fundamental structures of the original image. The coloring of the images Zhao's method produces is unwarranted.

At the same time, to ensure the credibility of our method, we invited 30 testees with different ages, different genders, and different backgrounds, to finish the survey we made. In the survey, we ask testees to score the similarity of the style between the source image and the stylized images created by the mentioned three methods from 1 to 5 respectively. To avoid the testees guessing which is our method, the orders of three methods are randomly changing every round. As shown in Table I, while 5 refers to the most similar one, our method acquires the highest score in each round. And the six rounds actually relate to the six images in Fig.6.

TABLE I
SCORES OF DIFFERENT METHODS IN IMAGES

Images	Proposed	Zhao's	*BrushStroke*
1	**3.50**	2.21	3.29
2	**4.04**	2.79	2.75
3	**3.11**	2.36	2.29
4	**3.67**	1.97	2.90
5	**3.33**	1.47	2.77
6	**4.07**	2.40	2.60
Average	**3.62**	2.20	2.76

We also ask the testees to choose the method which creates the most similar one among the three methods in each round. And Fig.7 shows that more than 60% testees think that ours is the best for each round. The average result is 76.79% which demonstrates that our proposed method outperforms the other two methods in most people's eyes.

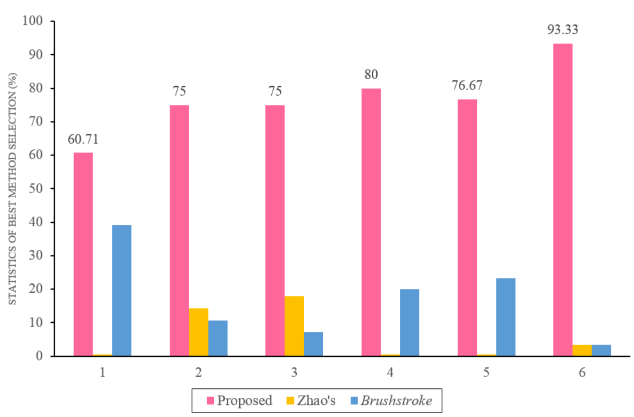

Fig. 7. Statistics of best method selection with different images.

In fact, both Zhao's method and the *BrushStroke* APP can only convert the target image into the general oil painting style without considering the individual drawing style of the source stylish image. But the proposed method is applicable to the specific oil painting stylization problem with limited reference, which is more useful. Moreover, experimental results demonstrate that the proposed method produces better stylized images than these aforementioned methods do.

V. CONCLUSIONS

In this paper, based on the sparse representation image stylization method, we propose a hierarchical model to achieve the oil painting stylization with limited reference. The structure layer which maintains the structure is fused with the texture layer which reflects the stylish textures based on the saliency mask. At the same time, owing to the features of artist's oil painting works, the color layer and the surface layer are considered to make it more similar to the source stylish image. Experimental results indicate the proposed method outperforms the state-of-the-art algorithms.

REFERENCES

[1] A. Hertzmann. Painterly rendering with curved brush strokes of multiple sizes. *Proc. of SIGGRAPH*, pp.453-460, Jan. 1998.

[2] K. Jia, X. Wang, and X. Tang. Image transformation based on learning dictionaries across image spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.35, No.2, pp.367-380, Feb. 2013.

[3] J. Yang, J. Wright, T. Huang, Y. Ma. Image super resolution via sparse representation. *IEEE Transactions on Image Processing*, Vol.19, No.11, pp.2861-2873, Nov. 2010.

[4] S. Wang, L. Zhang, Y. Liang, and Q. Pan. Semi-coupled dictionary learning with applications to image super resolution and photo-sketch synthesis. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.2216-2223, Sep. 2012.

[5] H. Bhujle and S. Chaudhuri. Novel speed-up strategies for non-local means denoising with patch and edge patch based dictionaries. *IEEE Transactions on Image Processing*, Vol.23, No.1, pp.356-365, Jan. 2014.

[6] N. G. Sadaka and L. J. Karam. Efficient superresolution driven by saliency selectivity. *Proc. of IEEE International Conference on Image Processing*, pp.1197-1200, Sept. 2011.

[7] Q. Liu and X. Tang. A nonlinear approach for face sketch synthesis and recognition. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.1005-1010, Jun. 2005.

[8] K. He, J. Sun, and X. Tang. Guided image filtering. *Proc. of European Conference on Computer Vision*, pp.1-14, Springer, 2010.

[9] A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. *Proc. of SIGGRAPH*, pp.341-346, 2001.

[10] L. Liang, C. Liu, , Y. Xu, B. Guo, and H.-Y. Shum. Real-time texture synthesis by patch-based sampling. *Technical Report* MSR-TR-2001-40, Microsoft Research, March 2001.

[11] J. Davis. Mosaics of scenes with moving objects. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.

[12] L. Itti, C. Koch, and E. Niebur. A model of saliencybased visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20, No.11, pp.1254-1259, Nov. 1998.

[13] R. Achanta, F. Estrada, P. Wils, S. Susstrunk. Salient region detection and segmentation. *Proc. of International Conference on Computer Vision Systems*, pp.66-75, 2008.

[14] R. W. G. Hunt. Measuring Color. Fountain Press, 1998.

[15] T. Ohashi, Z. Aghbari, A. Makinouchi. Hill-climbing algorithm for efficient color-based image segmentation. *Proc. of IASTED International Conference On Signal Processing, Pattern Recognition, and Applications*, 2003.

[16] D. L. Ruderman, T. W. Cronin, and C. C. Chiao. Statistics of Cone Responses to Natural Images: Implications for Visual Coding. *J. Optical Soc. of America*, Vol.15, No.8, pp.2036-2045, 1998.

[17] E. Reinhard, M. Ashikhmin, B. Gooch, P. Shirley. Color Transfer between Images. *IEEE Computer Graphics and Applications*, Vol.21, No.5, pp.34-41, 2001.

[18] M. Zhao and S. C. Zhu. Sisley the Abstract Painter. *Proc. of the 8th International Symposium on Non-Photorealistic Animation and Rendering*, pp.99-107, 2010.

[19] https://itunes.apple.com/us/app/brushstroke/id824421012