Generating Pseudo-3D Painting Based on Visual Saliency and Composition Rules

zhibin zheng[1,2] yan zhang[†1,2] zhengxing sun[1,2]

[1]State Key Lab for Novel Software Technology, Nanjing University, Jiangsu, nanjing, 210093
[2]Department of Computer Science and Technology, Nanjing University, Jiangsu, nanjing, 210093

Abstract

Pseudo-3D painting shows stereoscopic artistic effect in the 2D plane and displays artistic conception as the known verse "Far, near, high, low—no two parts alike", it brings a true feeling of stereo space to observers. This paper proposes a cropping method based on visual saliency and composition rules to get pseudo-3D painting from a 2D given image. On the basis of segmenting the given image into foreground and background, we utilize visual saliency to determine the highlighted region and combine existing composition rules to automatically crop the image using PSO optimization algorithm. Additionally, we add the border and pseudo-perspective transformation to enhance stereoscopic effect of our results. Experimental results demonstrate our method could turn a reference image into a pseudo-3D painting effectively, bringing a stereoscopic feeling to observers.

Categories and Subject Descriptors (according to ACM CCS): I.4.0 [Image Processing and Computer Vision]: General—Image displays

1. Introduction

In recent years, pseudo-3D painting works of art are increasingly appearing, they draw virtual 3D space in 2D plane and show diverse pictures of the same image. Besides, they highlight the focus of the image effectively, presenting stereoscopic feeling in simple 2D space.

Starting from a simple 2D image, achieving pseudo-3D effect has attracted researchers' attention in computer graphics. Gooch et al. [GG04] optimize the depth information of original scene by adding artistic mattes, enhancing stereoscopic impression of original image. Lopez-Moreno et al. [LMJH10] create new illumination image by simple depth estimation, based on which to carry on non-photorealistic rendering, and get different drawing effects. Ritschel [RTMS12], with the known depth information of the image, crop the background using automatic or manual cropping method to achieve pseudo-3D effect, our research content is similar with their goal.

In traditional photography and painting, authors could usually utilize certain composition skills to complete a pleasant work. Recently, many researchers have used related composition rules to modify an image to get better pictures in visual effect [LCWCO10] [GLGW12]. These methods get excellent quality of the image through corresponding image restructuring based on some basic composition rules. Such methods make it possible for computers to complete an excellent work by combining high-level composition rules with low-level methods in image processing.

Researchers take different ways to get depth information of image in 3D effect displaying, on which basis to take different processing, but as we know there exists no methods to precisely acquire depth information of a single 2D image. However, speaking from the human visual perception, some visual sensitive significant regions frequently exist in our observed information, these regions are important areas to which observers pay attention, they are similar to information that the depth information of image provides. Inspired by Ritschel [RTMS12] and [LCWCO10], we combine visual saliency and basic high-level composition rules to provide a generation method of pseudo-3D painting.

The specific contributions of our work include:

- Starting from a simple 2D image, in condition of unable to get the accurate depth information of the image, we

† zhangyannju@nju.edu.cn

DOI: 10.2312/conf/EG2013/short/049-052

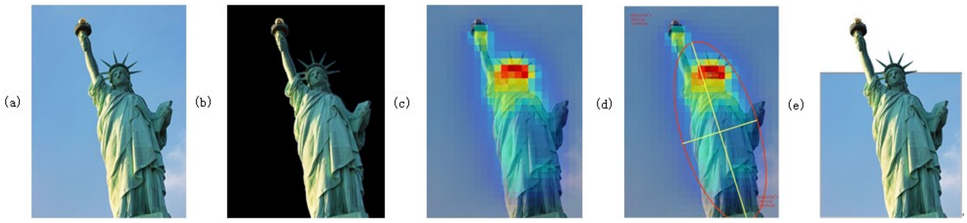

Figure 1: *Related images involved in our method.(a):input image; (b):segmentation result; (c):saliency map; (d):fitting ellipse for cropping; (e):our result*

use visual saliency instead to provid important basis for subsequent cropping.

- Taking advantage of basic composition rules in painting and photography with visual saliency to automatically crop the image using PSO optimization algorithm, presenting pseudo-3D effect visually.

2. Salient Region Detection

When comparing a 2D image with pseudo-3D painting drawn by artists, the prominent regions are usually the significant foreground regions that our visual system has noticed. We estimate the reliable visual saliency of the image to find out these prominent regions which need to be highlighted.

We use "Global Contrast based Salient Region Detection" method [CZM*11] to segment an image into foreground and background region. Then we combine it with "Graph-based visual saliency" method [HKP07], using local multi-scale image features and global color contrast [CZM*11] to get the saliency map of input image, in order to determine the prominent areas which need to be highlighted in foreground region. The red high luminance area of the foreground region in Figure 1(c) is the saliency region that will be highlighted during cropping process.

3. Optimally Cropping

We use Particle Swarm Optimization (PSO) algorithm [KE95] to optimize the fitness function. Each particle consists four attributes $(X_{top}, Y_{top}, X_{down}, Y_{down})$, where (X_{top}, Y_{top}) indicates the top endpoint of diagonal line, and (X_{down}, Y_{down}) means the down endpoint of diagonal line, both determine the final borders of the image. Firstly, we set particles' initial positions randomly according to the overall trend of the foreground object, then particles move to optimize the fitness function and finally we get the cropped borders based on the best particle.

3.1. Preprocess of Cropping

In order to accelerate convergence, we put initial particles randomly near the two endpoints of the image's diagonal line which has a similar trend with the foreground object. Firstly we extract the outer contour of the foreground object, the outer contour is consist by a series of points as $P = \{p_i\}_{i=1}^{n}$; then we use the curve fitting method to fit these points as an ellipse [FF95]. We use the long axis of the ellipse to represent the overall trend of the foreground object. The elliptic curve needed to be fit is defined in vector form as follows:

$$\begin{aligned} F(\alpha, x, y) &= [x^2, y^2, xy, x, y, 1] \cdot [A, B, C, D, E, G] \\ &= \chi \cdot C(\alpha) \end{aligned} \tag{1}$$

Where $\chi = (x^2, y^2, xy, x, y, 1)$ corresponds each contour point p_i, and $C(\alpha) = (A, B, C, D, E, G)$ is the unknown coefficient. We use the least square method to fit $C(\alpha)$ to get the final circumscribing ellipse calculated as follows:

$$\varphi^2(\alpha) = \sum_{i=1}^{n} \delta(C(\alpha), p_i) \tag{2}$$

Where $\delta(C(\alpha), p_i)$ is the closet distance (Euclidean distance) from point p_i to the ellipse $C(\alpha)$, defined as the minimal distance from p_i to point on the ellipse. The goal is to minimize $\varphi^2(\alpha)$. Figure 1(d) is the result of ellipse fitting corresponding to Figure 1(a). As is shown in Figure 1(d), since the direction of the long axis is close to the diagonal line which is from left-top to right-down, we put initial particles' positions near the endpoints of the diagonal line, as the red words near the left-top and right-down show.

3.2. Fitness Function Definition

To define a proper fitness function, we consider three factors. First of all, we should satisfy the requirement of pseudo-3D paintings that part of the foreground region must be highlighted outside of the border. The rest factors are two basic high-level composition rules that are widely used in photography and painting. One is the rule of thirds, which trisects the two vertical borders of the image to get four golden section points and four intersected lines, and the focus of performance is usually put on one of the four golden section points, and strong horizontal and vertical components in the image should be aligned with one of the four intersected lines to present distinctive and harmonious picture. The other is the diagonal dominance, based on which foreground objects are usually distributed along the diagonal direction to strengthen the sense of space and the sense of perspective as well. We combine the three factors to define a proper fitness function.

Firstly, due to the requirements of pseudo-3D painting, we should expose certain pixels outside of the border to highlight part of the foreground region. However, because of the different sizes of foreground region in different sizes of images, we expose appropriate ratio of the foreground region to highlight the foreground object. We define the fitness of exposed pixels as follows:

$$S_{pixel} = e^{\frac{-D^2(ratio)}{2\sigma_1}} \tag{3}$$

Where $D(ratio) = P_{ratio} - thresh$, and P_{ratio} is the proportion of exposed pixels to all foreground pixels, $thresh$ is the threshold of exposed pixels proportion. Through experiments, we find $thresh = 8\%$ and $\sigma_1 = 0.13$ are proper. The closer the P_{ratio} and $thresh$ are, the stronger of the effect of the pseudo-3D, so our goal aims at maximizing S_{pixel}.

Then, in order to attract the viewer's attention, we deem that the centralized portion of the visual salient region of the foreground object is the part which needs viewer to pay close attention to, that is, the focus of the performance we need to highlight, so we hope that the focus position could be put on the golden section point possibly based on the rule of thirds. We utilize the evaluation function for putting the focus of performance near one of the four golden section points in [LCWCO10] to define the fitness function of the intersected points as follows:

$$S_{point} = \frac{1}{\sum_i M(S_i)} \sum_i M(S_i) e^{\frac{-D^2(S_i)}{2\sigma_2}} \tag{4}$$

Where $D(S_i) = \min_{j=1,2,3,4} d(C(S_i), G_j)$, $C(S_i)$is the center of the visual saliency region, G_j is the golden section point of the cropping image; $M(S_i) = A(S_i)I(S_i)$, and $A(S_i)$ is the area of the visual saliency region, $I(S_i)$ is the saliency value of the visual saliency region, d is the Euclidean distance of two points, and usually $\sigma_2 = 0.17$.

Finally, we consider the overall trend of the foreground object. We combine the diagonal dominance with the factor that strong horizontal and vertical components in the image should be aligned with the four intersected trisection lines as two high-level factors. We use ellipse to fit the foreground object, we deem that the direction of the long axis of the foreground object could represent the overall trend of the foreground object. And if the ratio of long axis of the ellipse relative to short axis of the ellipse is bigger than certain threshold, we think the trend of foreground object is "significant", at which time we will consider the above two high-level factors. When the trend of the foreground object is nearer to the direction of one of the two diagonal lines of the image, we consider the diagonal dominance factor; otherwise, we consider the other factor as the trend is nearer to vertical or horizontal. Based on the above analysis, we define the fitness of the trend of foreground object as follows:

$$S_{line} = \begin{cases} S_{diagonal} = e^{\frac{-Angle_{cos}^2}{2\sigma_3}}, (30^\circ \le angle \le 60^\circ) \\ S_{lineDistance} = e^{\frac{-D_{line}^2}{2\sigma_4}}, \text{Otherwise} \end{cases} \tag{5}$$

Where $S_{diagonal}$ measures how close the trend of the foreground object is to the direction of diagonal line, and $S_{lineDistance}$ measures how close the prominent component lies to the four intersected lines. $angle$ is the acute angle between the long axis of the ellipse and the horizontal, its range is ($0^\circ \le angle < 90^\circ$) ; when $angle$ is between 30° and 60°, the trend of the foreground object is nearer to the direction of one diagonal line of the image, so we consider diagonal dominance factor and define $S_{line} = S_{diagonal}$, otherwise we define $S_{line} = S_{lineDistance}$. $Angle_{cos}$ is the cosine of the angle between the trend of the foreground object and its nearer diagonal line, and usually $\sigma_3 = 0.15$. $D_{line} = \min_{j=1,2} d(L, L'_j)$, L is the long axis of the fitting ellipse, L'_j is the trisection line of the image, $d(L, L'_j)$ is the distance between the long axis and trisection line, defined as the sum of vertical distance of each point of the long axis to trisection line, and generally $\sigma_4 = 0.15$.

Base on the two high-level composition rules our goal aims at maximizing S_{point} and S_{line}. Considering the above three factors we define fitness function as follows:

$$F_{value} = \frac{\alpha \cdot S_{pixel} + \beta \cdot S_{point} + \gamma \cdot S_{line}}{\alpha + \beta + \gamma} \tag{6}$$

Where α, β, γ are the corresponding weights, through experiments we set $\alpha = 0.2, \beta = 0.4$; $\gamma = 0.8$ when the ratio of long axis of the ellipse relative to short axis is bigger than $threshold$ (we set $threshold = 2.5$), otherwise $\gamma = 0$. The higher the value of F_{value} is, the stronger the stereoscopic effect shows.

In this paper, PSO and GPU are used to optimize the fitness function F_{value} , after a certain number of iterations, particles eventually converge to an appropriate position and then determine the corresponding cropping border, we keep the foreground pixels outside the border and original pixels inside the border to get pseudo-3D artistic effect.

4. Results And Analysis

Our method is not applicable to all types of images, but only to those images that the foreground region is on the large proportion and the detected saliency region is near the border. The quality of the simulation result depends on the degree of satisfaction of human visual perception. We implement the whole process in C++, and for a general size 800×600 and 400×700 of image, the best particle tends to lie proper constant position after about 40 iterations as we deem the fitness function achieve convergence, which takes about 4 $\sim$ 6 minutes. Part of our results are shown in Figure 2. In order to get more realistic simulation of pseudo-3D effects, we take some post-processing such as adding borders as is shown in Figure 3. Besides, we also make pseudo-perspective transformation by stretching the border of the non-highlighted region, as the example of fish, frog and car in Figure 2.

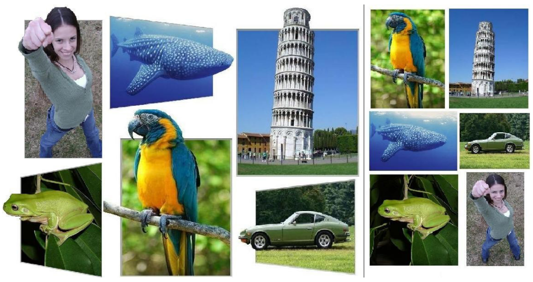

Figure 2: *Left:our results; right:input images.*

Figure 3: *Left:input image; right:our result by adding border.*

We also compare our results with "Virtual Passepartouts" method proposed by Ritschel [RTMS12] since we focus on the similar research content, the comparison is shown in Figure 4. Compared with "Virtual Passepartouts" method, we don't need the depth information of the image and consider visual saliency instead, since it is easier to get visual saliency for a single 2D image our method is suitable for a wider scope of applications relatively. Besides, instead of exposing certain numbers of pixels of forground region, we expose certain proportion of the forground region and take into account high-level composition rules as guidelines for automatic cropping which is in conformity with the principle of pseudo-3D painting and get more harmonious results.

Figure 4: *Comparison with "Virtual Passepartouts". left: input image; middle: our result; right: result from [RTMS12].*

5. Conclusion And Future Work

In this paper, we combine visual saliency and high-level composition rules to propose a pseudo-3D painting generation method which could be easily used in the domains of advertising design and art drawing, and our results show stereoscopic effect similar to real pseudo-3D paintings.

Since our method determines highlighted region by visual saliency, which is not suitable for all images, in future work, we will focus on combining other features of the image with visual saliency to determine highlighted regions to solve the above restriction.

6. Acknowledgement

This paper is supported by the grants from "the National Natural Science Foundation of China" [Grants No.61021062, 61100110 and 61272219], "the Science and Technology Program of Jiangsu Province" [Grants No.BE2010072, BE2011058 and BY2012190].

References

[CZM*11] CHENG M.-M., ZHANG G.-X., MITRA N. J., HUANG X., HU S.-M.: Global contrast based salient region detection. In *IEEE CVPR* (2011), pp. 409–416. 2

[FF95] FITZGIBBON A., FISHER R. B.: A buyer's guide to conic fitting. In *In British Machine Vision Conference* (1995), pp. 513–522. 2

[GG04] GOOCH A. A., GOOCH B.: Enhancing perceived depth in images via artistic matting. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization* (New York, NY, USA, 2004), APGV '04, ACM, pp. 168–168. doi:10.1145/1012551.1012590. 1

[GLGW12] GUO Y. W., LIU M., GU T. T., WANG W. P.: Improving photo composition elegantly: Considering image similarity during composition optimization. *Computer Graphics Forum 31*, 7pt2 (2012), 2193–2202. doi:10.1111/j.1467-8659.2012.03212.x. 1

[HKP07] HAREL J., KOCH C., PERONA P.: Graph-based visual saliency. In *Advances in Neural Information Processing Systems 19* (2007), MIT Press, pp. 545–552. 2

[KE95] KENNEDY J., EBERHART R.: Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on* (nov/dec 1995), vol. 4, pp. 1942 –1948 vol.4. doi:10.1109/ICNN.1995.488968. 2

[LCWCO10] LIU L., CHEN R., WOLF L., COHEN-OR D.: Optimizing photo composition. *Computer Graphics Forum 29*, 2 (2010), 469–478. doi:10.1111/j.1467-8659.2009.01616.x. 1, 3

[LMJH10] LOPEZ-MORENO J., JIMENEZ J., HADAP: Stylized depiction of images based on depth perception. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering* (New York, NY, USA, 2010), NPAR '10, ACM, pp. 109–118. doi:10.1145/1809939.1809952. 1

[RTMS12] RITSCHEL T., TEMPLIN K., MYSZKOWSKI K., SEIDEL H.-P.: Virtual passepartouts. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering* (Aire-la-Ville, Switzerland, Switzerland, 2012), NPAR '12, pp. 57–63. URL: http://dl.acm.org/citation.cfm?id=2330147.2330160. 1, 4