Depth-based Reference Portrait Painting Selection for Example-based Rendering

Xiaoyan Zhang, Kap Luk Chan and Martin Constable

Abstract— **The task objective concerned in this paper is to preserve the natural attributes of a photograph and only enhance its aesthetic perceptual feeling. The paper proposed to select reference portrait paintings based on depth for example-based photograph rendering to improve its aesthetic appeal. Hence, the rendered photograph acquires the aesthetic style as informed by the selected reference paintings. The depth attributes are based on the notions of foreground/background or figure/non-figure relationship. The analysis on portrait paintings suggests that the natural attributes can be measured by the lightness/saturation/hue distributions or contrasts within and between foreground and background. By segmenting the photograph and paintings based on the depth information, and computing the intra layer distributions and inter layer contrasts as the features for the similarity measurement, references are then selected from the top ranked paintings. Some rendering examples are presented in this paper for evaluating the performance of the selection method.**

I. INTRODUCTION

Portrait is the representation of a person to display the likeness, personality and mood. Portrait painting plays an important role in art history. The expression of the sitter depends on the projection of the feeling of the artist rather than on physically accurate depiction [1]. Comparing with photography, which is the projection of the physical nature, the portrait artist is able to filter and manipulate the subject scene based on the visual perception with an artistic perspective [2]. They are good at making their portrait subjects seem to be 'live' by using techniques such as emotion, gesture, composition, lighting and shadow. Can we learn these from portrait paintings and transfer them to portrait photographs?

With recent development in computer vision and computer graphics, some studies on portrait have been reported. Sablatnig et al. developed an classification system to attribute works of art to a particular artist based on the facial and brush stroke characteristics [3]. Albuquerque et al. proposed to select the 'good' portrait photography from a series of image based on the eye and mouth models [4]. Ye Ning reported the work of changing the emotion of the artist's portrait based on the face emotion of the person captured by the camera [5]. With recent development of painterly rendering, portrait rendering has become an interesting topic of the computer vision and computer graphics research communities. Colton proposed to use a Non Photorealistic Rendering(NPR) system to automatically produce portraits based on the recognized emotions [6]. The artistic styles in the NPR system are based on the painting materials, color palette and brush model. Some recent works tried to render a portrait photo to an artistic style [7] [8] [9] [10]. However the styles are limited to simulate the abstraction, line drawing styles or organic models. In [11], artistic face lighting templates were learned from a dataset of professional and amateur portrait photographs. The lighting template is based on the light distribution and shading on the face, e.g left weighted or right weighted. Chen et al. developed an algorithm to relight the face in image rendering [12].

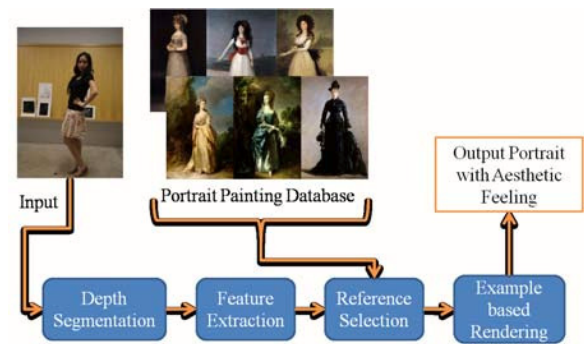


Fig. 1. Framework of example-based portrait rendering.

Even though face is an important part in the portrait, however, the composition, shadow, lighting organization, edge, background contrast also play other important roles to enhance the visual impact of the portrait. So, different from the researches above, we explore to learn the light and color contrast organizations in the portrait painting and transfer these to the photograph. We need to understand and learn the organization of these aesthetic aspects of painting that influence the visual appeal. One of the frustrating difficulties in analyzing these paintings is that though general rules of formal behavior can be laid down they are very far from being absolute. There exists many local variants of these rules and though to an experienced human eye they are still visually coherent, but to a systematic computational approach they are not so. It is therefore a challenge to capture the rules of each artist or an style. Therefore, we propose to render the portrait photograph based on an example painting. The frame work of the rendering process is shown in the Fig.1. This paper focuses on the selection of references for example-based rendering. The references are selected based on the contrasts between and within the foreground(FG) and background(BG) as seen by an artist.

X. Zhang is with School of Electrical & Electronic Engineering, and Institute for Media Innovation, Nanyang Technological University, Singapore 639798 zhan0292@ntu.edu.sg

K.L. Chan is with School of Electrical & Electronic Engineering,Nanyang Technological University, Singapore 639798 eklchan@ntu.edu.sg

M. Constable is with School of Art, Design and Media,Nanyang Technological University, Singapore 639798 MConstable@ntu.edu.sg

First, we analyzed the portrait paintings in the direction of scene depth for half and full body portraits. Then, the depth based segmentation was introduced to segment the FG from the BG for reference selection and rendering. Based on the segmented FG and BG, the similarity between the portrait photograph and the painting was calculated using inter and intra contrast features and ranked by descending similarity. References were selected from the top n ranked paintings. The user can choose one of them as the example to render the photograph.

II. Portrait painting analysis

.

The portrait paintings collected for the references are chosen from the famous artists: Thomas Gainsborough (1727-1788), Fransisco Goya (1746 -1828), Edouard Manet (1832-1883), Jean-Auguste-Dominique Ingres (1780-1867). The portraits from the four artists are stylistically different and have a nice dynamic range which is pleasing for human vison.

There are two geometries in a painting: 2D and 3D. 2D geometry consists of center, corner and edge values. 3D geometry is usually understood by painters in terms of the traditional zones: FG, middle-ground (MG), BG and sky separation. As a rule of thumb things of high pictorial value are placed somewhere in the FG of the painting. These general rules of placement are particular evident in the tradition of portrait painting (Fig.2). Almost all portraits are easily divisible into strong FG / BG zones (aka figure-not figure), and, in the case of full body portraits, upper BG, lower BG and Ground (this gives us a clear feet-meets-ground zone necessary for maintaining the stability of the figure).

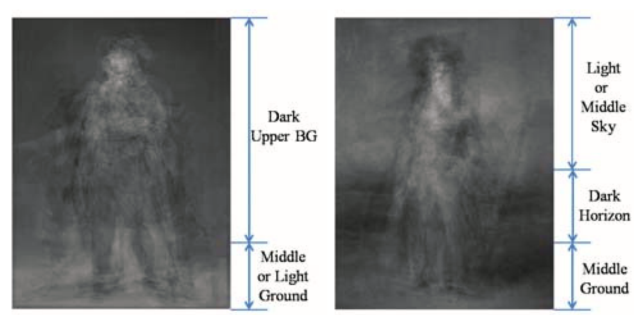
Fig. 2. Left: Francisco Goya (1746 - 1828), “Portrait of King Ferdinand VIII” 1803, Right: the portrait divided into clear FG / BG.

The artists render the FG and BG differently to highlight the expression of the figure. Based on the contrast relationship of the FG and BG, the half body portraits can be classified to three cases as shown in Fig.3. In the first portrait, the FG is darker than the BG, in the second they are approximately the same and in the third the BG is darker than the FG. In all the three different cases, the artist can organize contrasts within and between FG and BG to enhance the expression of the figure.

Fig. 3. Three portraits by Fransisco Goya (1746 -1828).

The full body portraits also have these three kinds of contrast styles as half body portraits. However, based on the complexity of the BG, the full body portraits can be considered in indoor and outdoor styles as shown in Fig.4. In both of them, the BGs have clearly contrast with FG to highlight the figure. The average lightnesses of the two kinds of portraits are shown in Fig.5, which present the difference of the BGs between these two kinds of portraits.

Fig. 4. Two full body portraits. The left: “portrait of the Countess of Chincon”, by Fransisco Goya (1746 -1828). The right: “Mrs. Peter William Baker”, by Thomas Gainsborough (1727-1788).



(a) indoor (b) outdoor

Fig. 5. The average lightness of the full body portrait paintings.

The contrast organization in different natural conditions are different. We need to select suitable references. As the principle of rendering is to retain the natural property of the photograph and only enhance its aesthetic perceptual feeling, the reference paintings should be selected as these having similar natural property. For example, the dark figure with a light BG should be rendered referring to the painting having a dark figure and light BG. This natural property can be measured by the contrasts within and between FG and BG.

III. Depth segmentation

The segmentation is achieved based on the depth information. The full body portrait contains the Ground, upper BG and figure(as shown in Fig.6). For the continuity of the Ground, we cannot segment the figure using depth threshold. However, the Ground plane has a clearly different normal vector from the figure. Surface normals have been used for range image segmentation [13]. Similarly, we use the normal vector and depth as feature to segment the figure/non-figure.

(a) photograph (b) depth

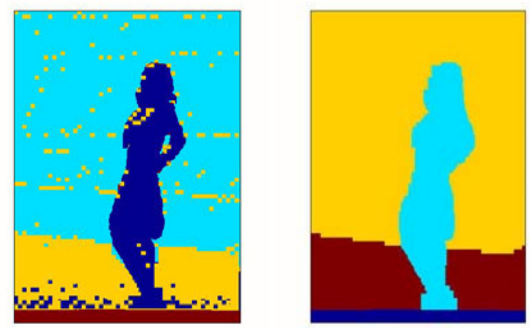
Fig. 6. One portrait photograph with depth information

The normal vector of flat plane is calculated by the cross product of two intersected vectors lying on the plane. We consider every $(d+1)$ by $(d+1)$ square forming one surface plane. Given the depth z, the normal vector of the plane is calculated as

$$\begin{aligned} &\overrightarrow{N} = \overrightarrow{P_1} \times \overrightarrow{P_2} \\ &\overrightarrow{P_1} = (x+d, y, z(x+d, y)) - (x, y, z(x, y)) \\ &\overrightarrow{P_2} = (x, y+d, z(x, y+d)) - (x, y, z(x, y)) \end{aligned} \tag{1}$$

where (x, y) is the top-left corner pixel of the $(d+1)$ by $(d+1)$ square. To handle the noise in real depth data, we choose $d = 6$. Then the normal vector of each pixel in the square will be expressed as $\frac{\overrightarrow{N}}{|\overrightarrow{N}|} = (a, b, c)$.

After calculating the normal vector for each pixel, the feature (a, b, c, z) is used for the region clustering. In this work, k-means method is chosen for its simplicity and speed in clustering. The initial cluster centers and the number of clusters are defined based on the priori knowledge of the data that the bottom part is the Ground, the top part is the upper BG and the figure is in the center. So three initial cluster centers are defined as the feature of the points in the bottom left corner, top left corner and image center which correspond to the Ground, upper BG and figure. For half body portrait, the ground is not in the scene. In this case, the depth values of the bottom left corner and top left corner are very near. If the difference of the two depth values is smaller than a threshold(10% depth range), it will be considered as no Ground contained, and the number of cluster will be 2. For another case that the bottom left corner is part of the person, the depth value difference between the image center and bottom left corner will be very close. Similarly, the number of cluster reduces to 2. The segmentation result of the example in Fig.6 is shown in Fig.7(a). Then post-processing was performed using morphological operations. The result after post-processing (Fig.7(b)) better represents the Ground, upper BG and figure.



(a) before post-processing (b) after post-processing

Fig. 7. Segmentation result of the portrait depth in Fig.6

IV. Reference Selection

Though lightness contrast plays the most important role in the FG and BG difference, the saturation and hue contrasts still influence the similarity of the scene. So lightness, saturation and hue contrasts are considered at the same time in the LCH space (Luminance, Chroma (considered as saturation), and hue), which is a polar transformation of the CIE L*a*b* color space. The L channel best matches the human perception of the lightness of colors and the Chroma is indicative of saturation. The intra contrasts within the figure and BG are expressed by the histogram distributions. The inter contrasts are calculated by contrast ratios as follows.

The lightness inter contrast is calculated by the difference in logarithms which is a function of contrast [14].

$$C_l = \lg_{10}(L_F) - \lg_{10}(L_B) \tag{2}$$

where L_F is the average lightness of the FG, L_B is the average lightness of the BG.

The saturation inter contrast is

$$C_s = S_F - S_B \tag{3}$$

where S_F is the average saturation of the FG, S_B is the average saturation of the BG.

The inter hue contrast is defined as the average hue difference between FG and BG. The saturation weighted average hue is calculated using circular statistics [15] as follows.

$$\begin{aligned} A &= \textstyle\sum_{j=0}^{n} S_j \cos(H_j) \\ B &= \textstyle\sum_{j=0}^{n} S_j \sin(H_j) \end{aligned} \tag{4}$$

$$\overline{H} = \begin{cases} \frac{1}{2\pi}\arctan(\frac{B}{A}) & \text{if } A > 0; \\ \frac{1}{2\pi}\arctan(\frac{B}{A}) + 0.5 & \text{if } A < 0. \end{cases} \tag{5}$$

Given the average hue H_F, H_B of FG and BG, the hue contrast is expressed as

$$C_h = \begin{cases} |H_F - H_B| & \text{if } |H_F - H_B| \leq 0.5; \\ 1 - |H_F - H_B| & \text{if } |H_F - H_B| > 0.5. \end{cases} \tag{6}$$

Besides the lightness, saturation and hue inter contrasts, the background global contrast factor G_B is calculated to express the complexity of the background. Detailed and variation-rich image background has a high global contrast factor and the simple background has a low global contrast factor [16]. So the contrast factor can represent the difference of the indoor and outdoor background. The global contrast factor is calculated as the weighted average of local contrasts at various resolution levels(more details can be seen in [16]).

$V = (L_F, L_B, C_l, S_F, S_B, C_s, C_h, G_B)$ is defined as the feature vector describing the contrast ratios of the portrait. The similarity of the normalized feature vectors V_1 and V_2 is calculated as

$$\Gamma_V = e^{-\|W_0 \cdot (V_1 - V_2)\|} \tag{7}$$

where W_0 is the weight vector.

The intra lightness, saturation and hue contrasts of FG and BG are expressed by the 10-bins normalized histogram distributions, which are expressed as T_{lf}, T_{lb}, T_{sf}, T_{sb}, T_{hf}, T_{hb}. The hue distribution of the figure is dependent on the clothes of the subject. It is not a factor that influence the contrast within the FG, so it is not considered in the similarity measure. The hue histogram T_{hb} is saturation weighted. The similarity of histogram is calculated by histogram intersection as

$$\Gamma_{T_{1,2}} = \sum_{i=1}^{n} \min(T_1(i), T_2(i)) \tag{8}$$

where n is the number of bins, $n = 10$. The similarity of the histograms is

$$\Gamma_T = W_1[\Gamma_{T_{lf}}, \Gamma_{T_{lb}}, \Gamma_{T_{sf}}, \Gamma_{T_{sb}}, \Gamma_{T_{hb}}]^T \tag{9}$$

where W_1 is the weight vector. The final similarity of two portrait images is

$$\Gamma = (1 - w)\Gamma_T + w\Gamma_V \tag{10}$$

where w is the weight scale to fusion the two parts of contrast similarity. For each portrait photograph, portrait paintings having top n similarity values with the photograph are recommended as the references. The user can choose one of them as example template to render the photograph portrait.

V. Rendering based on selected reference

The principle of the rendering is to keep the natural relationship of FG and BG and the general colors of the photograph, and only enhance its perceptual contrast for aesthetic feeling. Currently, the features considered in the rendering are the lightness and saturation contrasts within and between FG and BG. The face and skin areas (e.g hands and arms) are almost the lightest part of the body and have a high contrast with the BG and other part of the FG to highlight the expression of the emotion and posture. So the face and skin areas, which are detected based on the skin color in the FG, are rendered separately from other part of the FG. The contrast mapping was achieved using standard deviation scaling and weighted mean value shifting on the lightness and saturation channels.

VI. Experiments and Discussion

Totally 50 half body portrait paintings and 34 full body portrait paintings were collected forming the painting database. The half body database contains all the three contrast styles as shown in Fig.3, and also contains some portraits with outdoor BG. Half of the full body database are with outdoor BG and half are with indoor BG. The figures in the portrait paintings were outlined manually for the feature extraction. The depth of the portrait photograph was captured by the Xbox 360 Kinect depth sensor by Microsoft.

For an input portrait photograph, the figure/non-figure was partitioned using depth segmentation. The center cluster is considered as the figure. Then features were extracted for the similarity measure. The weights in the similarity measure control the influence of each feature component. The weights can be set according to special purpose and application. For the half body portrait with simple background, the lightness has a more influence on the FG/BG contrast. So the lightness histogram features have a larger weight than hue and saturation features. In the feature vector V, the inter contrast components have a larger weight than the average features and BG contrast factor. In the experiment the weights were set as $W_0 = (1, 1, 2, 1, 1, 2, 2, 1)/11$, $W_1 = (3, 3, 1, 1, 1)/9$, $w = 0.5$. Fig.8 shows the reference selection results of three indoor portrait photographs. The first photograph portrait has similar FG and BG, the second has a black FG and light BG, and the third one is with light FG and black BG. The top 5 ranked references generally have the similar contrast properties with the input photograph portraits.

In the outdoor portrait, the BG is more likely to contain the nature color green and more details. So larger weights need to be set for the hue histogram feature and BG contrast factor in the outdoor portrait reference selection. For the example in Fig.9, the weights were set as $W_0 = (1, 1, 2, 1, 1, 2, 2, 2)/12$, $W_1 = (1, 1, 1, 1, 4)/8$, $w = 0.5$. From the results, we can see that the selected references all have a green BG. When the portraits with a outdoor green BG are selected, some indoor portraits having green BG also have a high similarity. This is a limitation of the similarity measure used in the current work.

The BG contrast factor plays a more important role in the full body portrait reference selection to select portraits with similar BG complexity. The references of the two examples (one is the indoor portrait and one is outdoor portrait) in Fig.10 were selected using weights $W_0 = (1, 1, 2, 1, 1, 2, 2, 3)/13$, $W_1 = (3, 3, 1, 1, 1)/9$, $w = 0.5$. Generally, the selected references have similar contrast styles with the input photograph.

Fig. 8. Reference selection of half body indoor portrait. The left column: the original portrait photographs. The second column: the segmented depth. Column 3-7: the first 5 top ranked references.

Fig. 9. Reference selection of half body outdoor portrait. The left column: the original portrait photograph. The second column: the segmented depth. Column 3-7: the first 5 top ranked references.

Fig. 10. Reference selection of full body portrait. The left column: the original portrait photographs. The second column: the segmented depth. Column 3-7: the first 5 top ranked references.

Fig. 11. Rendering results of the photograph in the first row of Fig.8 using selected references. The first 5 results are corresponding to the reference paintings in the first row of Fig.8. The last result is using the fourth reference painting in the third row of Fig.8.

The rendering results of the photograph portrait in the first row of Fig.8 are shown in Fig.11. The results using the selected references by the reference selection method are encouraging. However, if we use one reference with different natural relationship as the fourth reference painting in the third row of Fig.8, the result(the last image in Fig.11) is unnatural. Therefore, the reference selection is demonstrated important for the rendering. For the contrast mapping was implemented separately in the face and skin areas, other part of FG and BG, so there are some artifacts near the boundary transition. Better boundary smoothing method is needed.

VII. CONCLUSIONS

The paper proposed a portrait reference selection method based on depth for example-based rendering. The paper focuses on the example portrait painting selection based on the contrasts within and between FG and BG. First we analyzed the portrait paintings with respect to depth layers. FG and BG were segmented based on the depth information using normal vector and depth value for clustering. The intra histograms and inter contrast ratios were extracted as the feature for the similarity measure of the natural property. The top n ranked paintings were recommended as references for the example-based rendering. The user can choose one of them as the example to render the photograph. Generally, the proposed method can successfully select the references having similar natural property with the input portrait. The rendering results using the selected references are more encouraging compared with the result using the painting having a low similarity. This demonstrates the importance of the reference selection.

VIII. ACKNOWLEDGMENTS

The authors would like to acknowledge the seed fund and Ph.D. grant from the Institute for Media Innovation, Nanyang Technological University, Singapore.

REFERENCES

[1] C. Freeland, "Portraits in painting and photography," *Philosophical Studies*, vol. 135, no. 1, pp. 95–109, 2007.
[2] S. DiPaola, "Exploring a parameterized portrait painting space," *International Journal of Art and Technology*, vol. 2, no. 1, 2009.
[3] R. Sablatnig, P. Kammerer, and E. Zolda, "Hierarchical classification of paintings using face- and brush stroke models," in *Proceedings of Fourteenth International Conference on Pattern Recognition*, vol. 1, 1998, pp. 172–174.
[4] G. Albuquerque, T. Stich, and M. Magnor, "Qualitative portrait classification," in *Proceedings of Vision, Modeling, and Visualization*, 2007.
[5] Y. Ning and T. Sim, "Interactive portrait art," in *IEEE Workshop on Applications of Computer Vision*, 2008, pp. 1–6.
[6] S. Colton, M. F. Valstar, and M. Pantic, "Emotionally aware automated portrait painting," in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*, 2008, pp. 304–311.
[7] Y. Cao and Z. X. Wang, "Automatic generation of portrait," in *International Conference on Audio, Language and Image Processing*, 2008, pp. 735–738.
[8] Y. Zhang, L. Gao, and S. Y. Zhang, "Feature-based automatic portrait generation system," in *2009 World Congress on Computer Science and Information Engineering*, 2009, pp. 6–10.
[9] M. Meng, M. T. Zhao, and S. C. Zhu, "Artistic paper-cut of human portraits," in *Proceedings of ACM Multimedia 2010 International Conference*, 2010, pp. 931–934.
[10] M. Yang, S. Lin, P. Luo, and et al., "Semantics-driven portrait cartoon stylization," in *Proceedings of 2010 IEEE 17th International Conference on Image Processing*, 2010, pp. 1805–1808.
[11] X. Jin, M. T. Zhao, X. W. Chen, and et al., "Learning artistic lighting template from portrait photographs," in *Proceedings of the 11th European Conference on Computer Vision(ECCV)*, 2010, pp. 101–114.
[12] J. S. Chen, G. D. Su, J. P. He, and et al., "Face image relighting using locally constrained global optimization," in *Proceedings of the 11th European Conference on Computer Vision(ECCV)*, 2010, pp. 44–57.
[13] K. Pulli and M. Pietikinen, "Range image segmentation based on decomposition of surface normals," in *Scandinavian Conference on Image Analysis (SCIA)*, 1993.
[14] T. R. CORWIN, "Contrast and the difference-of-iogarithms," *Perception & Psychophysics*, vol. 29, no. 3, pp. 289–290, 1981.
[15] A. Hanbury and J. Serra, "A 3d-polar coordinate colour representation suitable for image analysis," in *Pattern Recognition and Image Processing Group, Technical Report 77*, 2003.
[16] K. Matkovic, L. Neumann, A. Neumann, and et al., "Global contrast factor - a new approach to image contrast," in *Computational Aesthetics in Graphics, Visualization and Imaging*, 2005, pp. 159–168.