Artistic Illumination Transfer for Portraits

Multimedia Attachments

Xiaowu Chen, Xin Jin*, Qinping Zhao* and Hongyu Wu

State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China
*Corresponding authors: {jinxin, zhaoqp}@vrlab.buaa.edu.cn

Paper ID 1013

In the multimedia attachments, we show additional transferred results (including the additional implementation details), the whole matching results of Yale Face Database B and a video demo. Please zoom-in to examine all figures.

1. Additional transfer results

We show additional transferred results and implementation details complementing the section 5 in the paper. We use the method of [JZC10] to predict indicative scores from 0 to 1 of the artistic lighting usage of the input and the result portraits by our system.

1.1 Reference from the Database (Section 5.1)

1.2 User-provided Reference (Section 5.2)

1.3 Applications (Section 5.4)

1.3.1 Numerical Assessment and One-key Transfer

1.3.2 Paper-cut

1.3.3 Sketch

2. The whole matching results of Yale Face Database B

The whole illumination matching results of Yale Face Database B [GBK01] is shown in a zip file named "yaleBResults.zip", which contains the "yaleBResults.csv" and a "readme.txt" file.

3. Video demo

The video demo shows our work and the software developed by us.

Reference

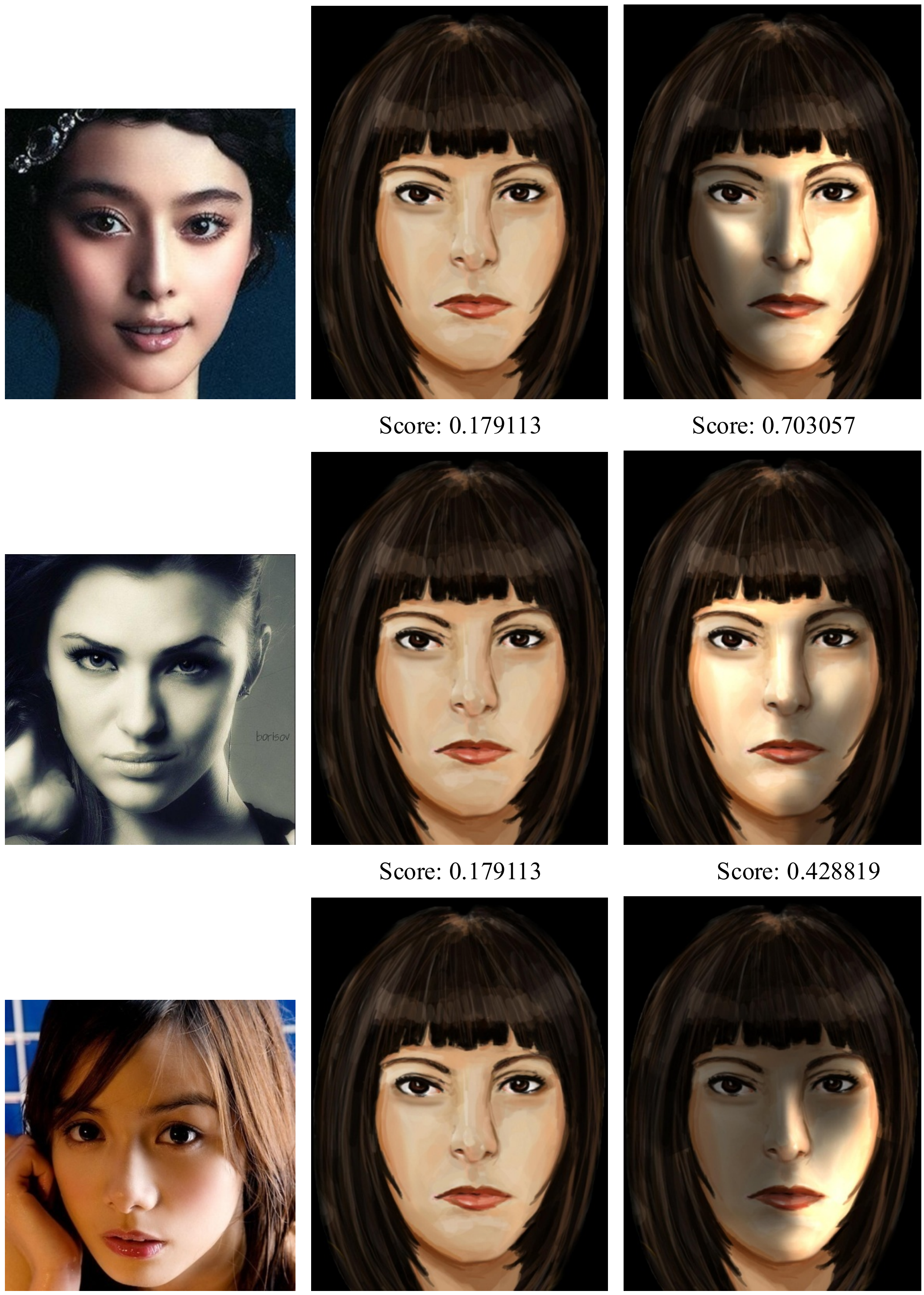
[GBK01] GEORGHIADES A. S., BELHUMEUR P. N., KRIEGMAN D. J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001), 643–660.

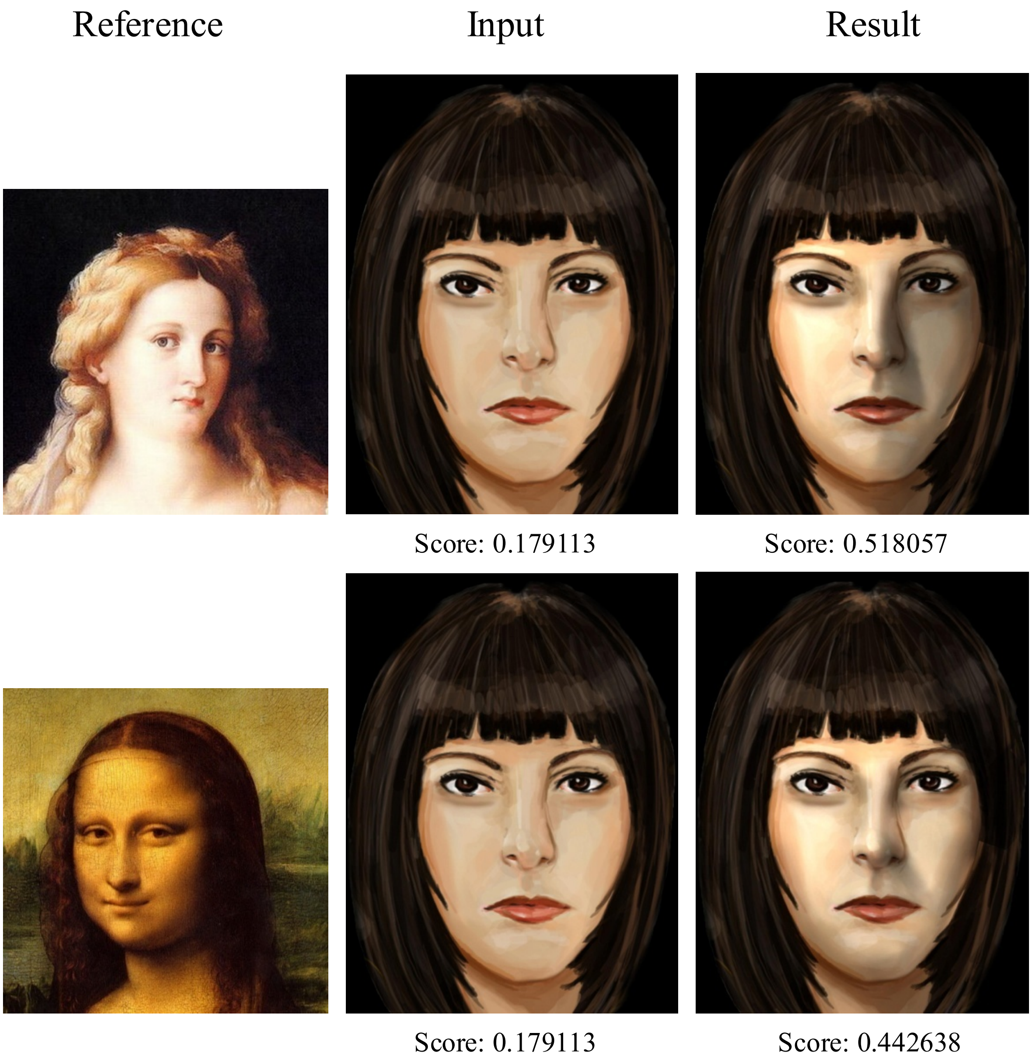
[JZC10] JIN X., ZHAO M.-T., CHEN X.-W., ZHAO Q.-P., ZHU S.-C.: Learning artistic lighting template from portrait photographs. In Proc. ECCV (2010).

1.1 Reference from the Database

(Section 5.1)

Input: Painting


Reference
Input
Result
Score: 0.179113
Score: 0.703057
Score: 0.179113
Score: 0.428819
Score: 0.179113
Score: 0.477923


Reference
Input
Result
Score: 0.179113
Score: 0.518057
Score: 0.179113
Score: 0.442638

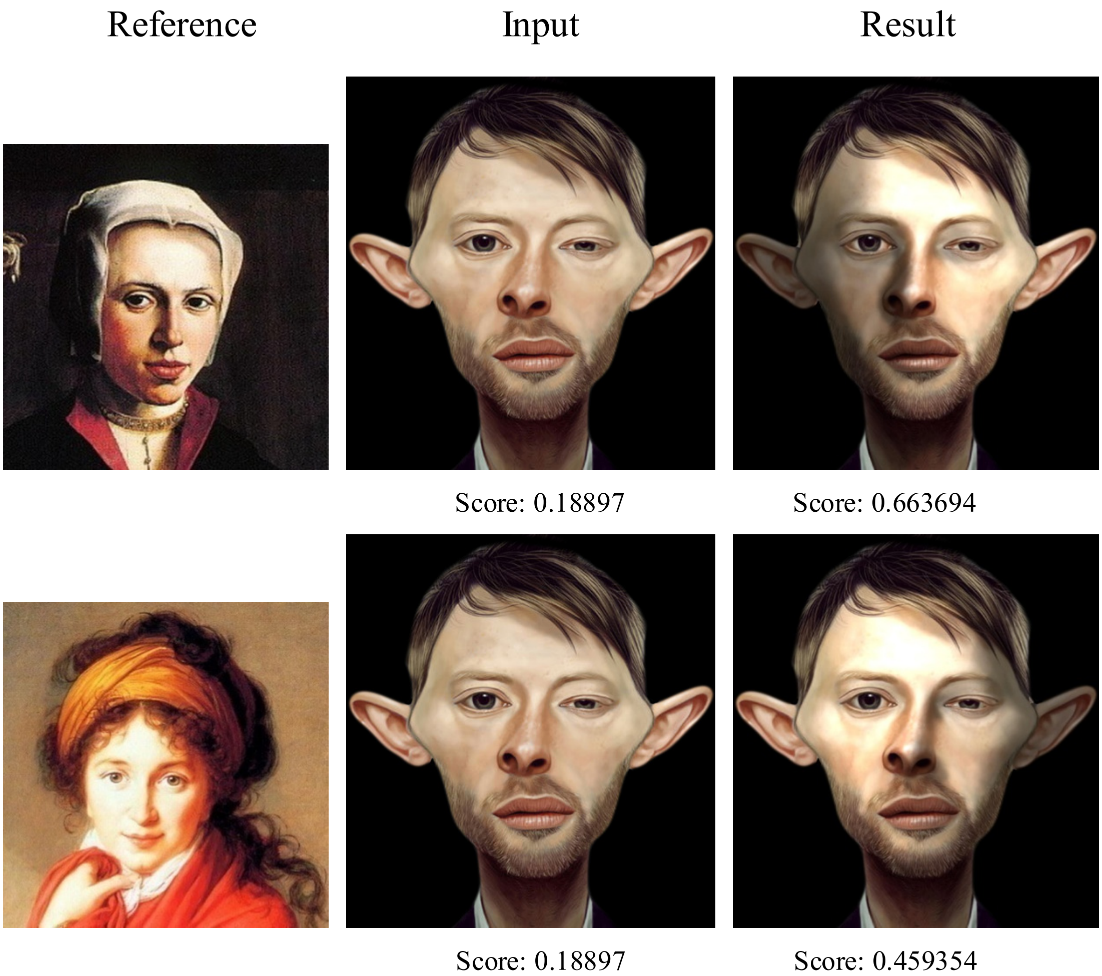
Reference Input Result

Score: 0.18897 Score: 0.555821

Score: 0.18897 Score: 0.411126

Score: 0.18897 Score: 0.520186


Reference
Input
Result
Score: 0.18897
Score: 0.423763
Score: 0.18897
Score: 0.443777
Score: 0.18897
Score: 0.474395


Reference
Input
Result
Score: 0.18897
Score: 0.663694
Score: 0.18897
Score: 0.459354

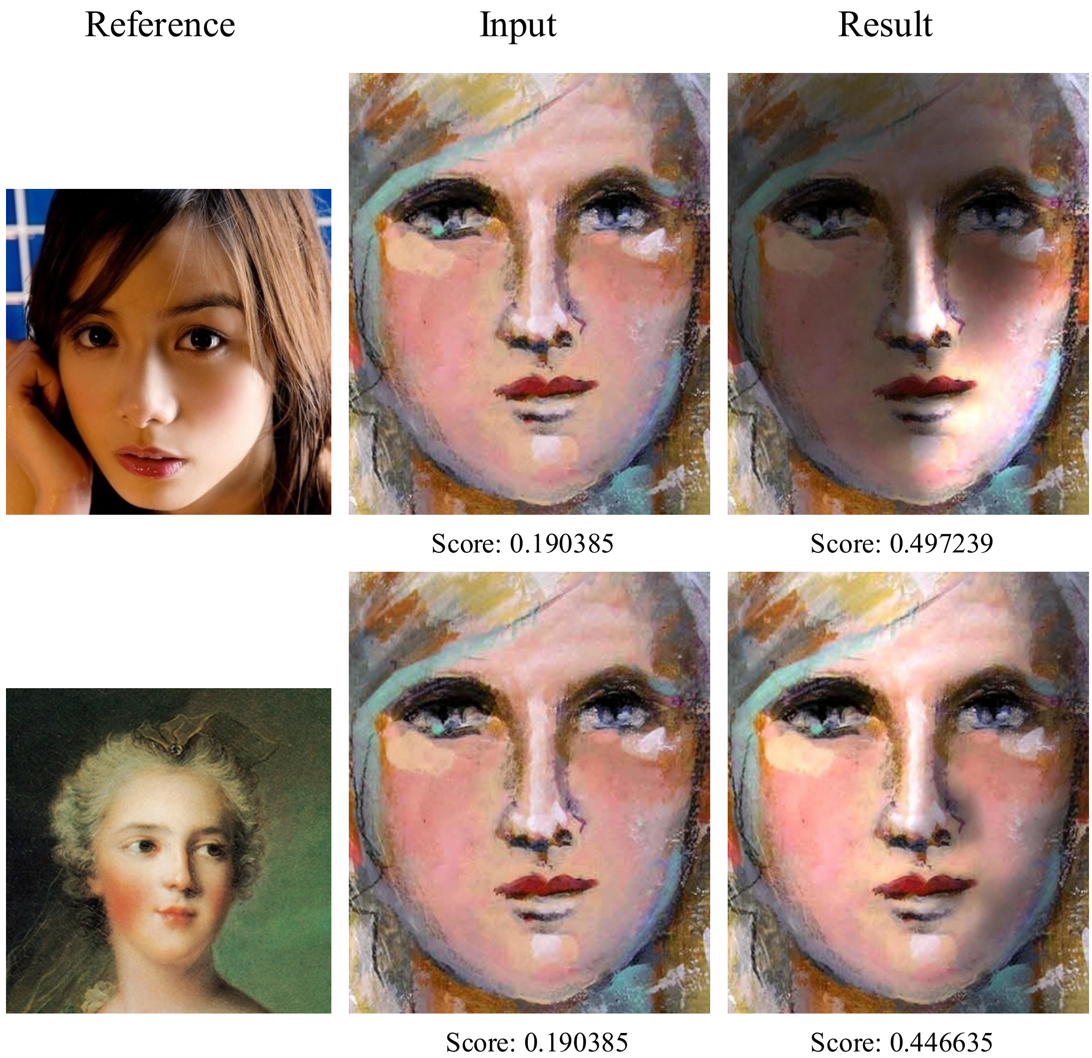


Figure 1: *Additional painting results. First column: reference artistic portraits. Second column: input paintings. Third column: the transferred results.*

1.1 Reference from the Database

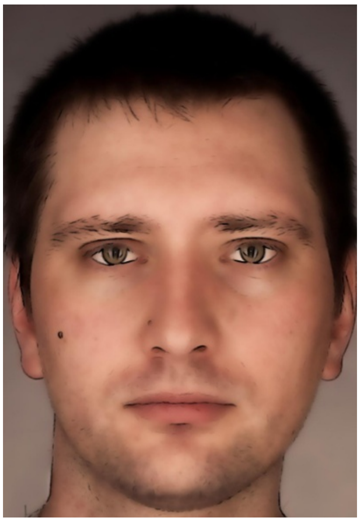
(Section 5.1)

Input:

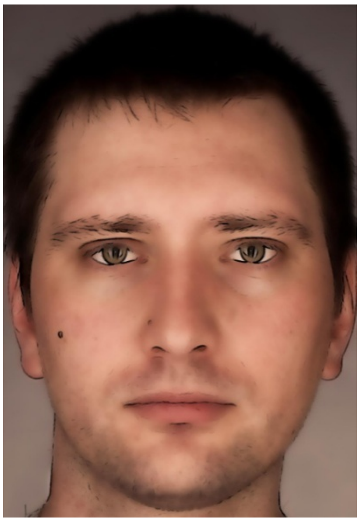
Computer Generated Painting

(CGP)

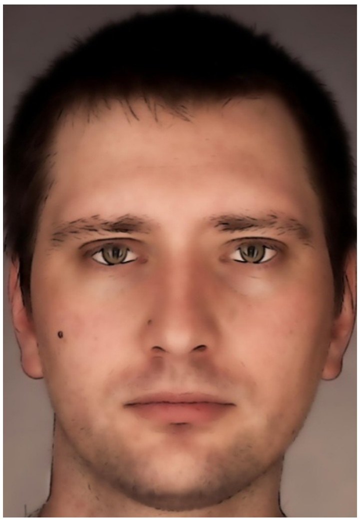
Reference | Input | Result


Score: 0.182544 Score: 0.616133

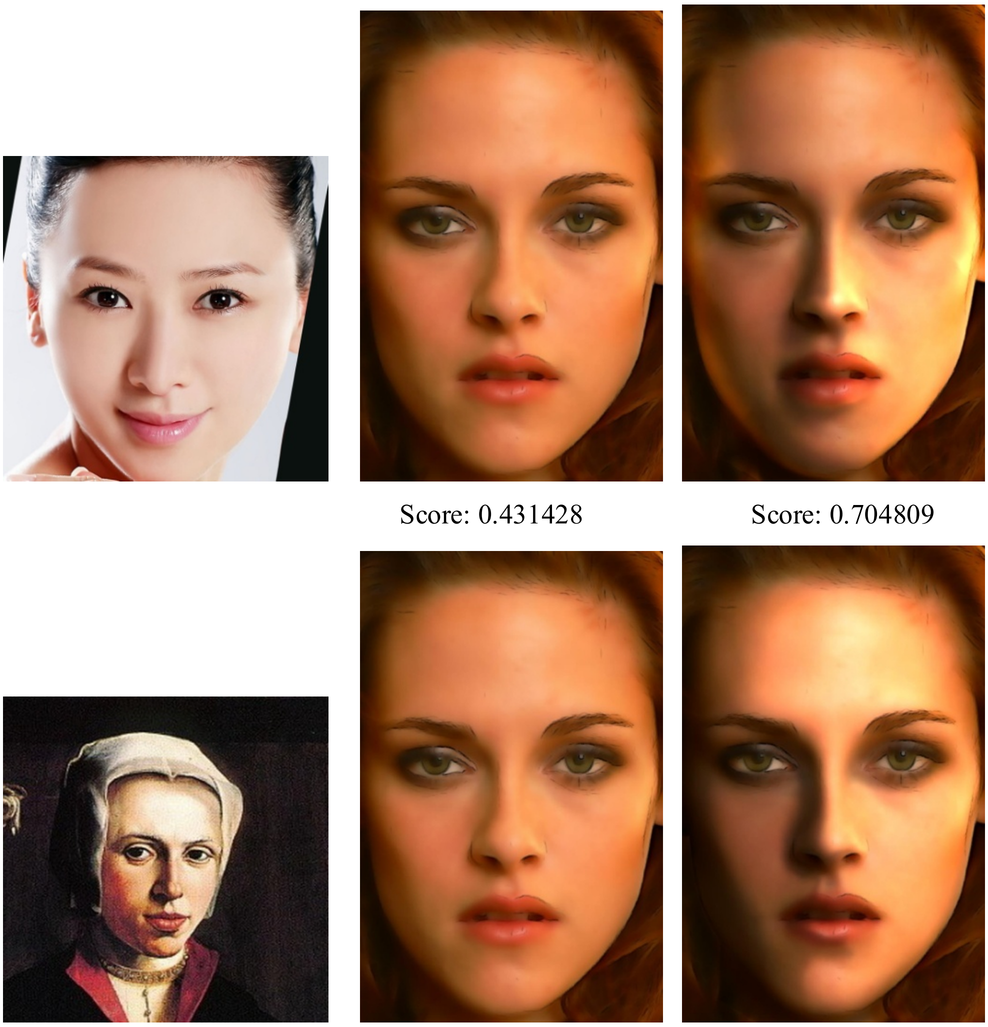




Score: 0.182544 Score: 0.532916


Score: 0.182544 Score: 0.527497

Reference Input Result



Score: 0.431428 Score: 0.704809

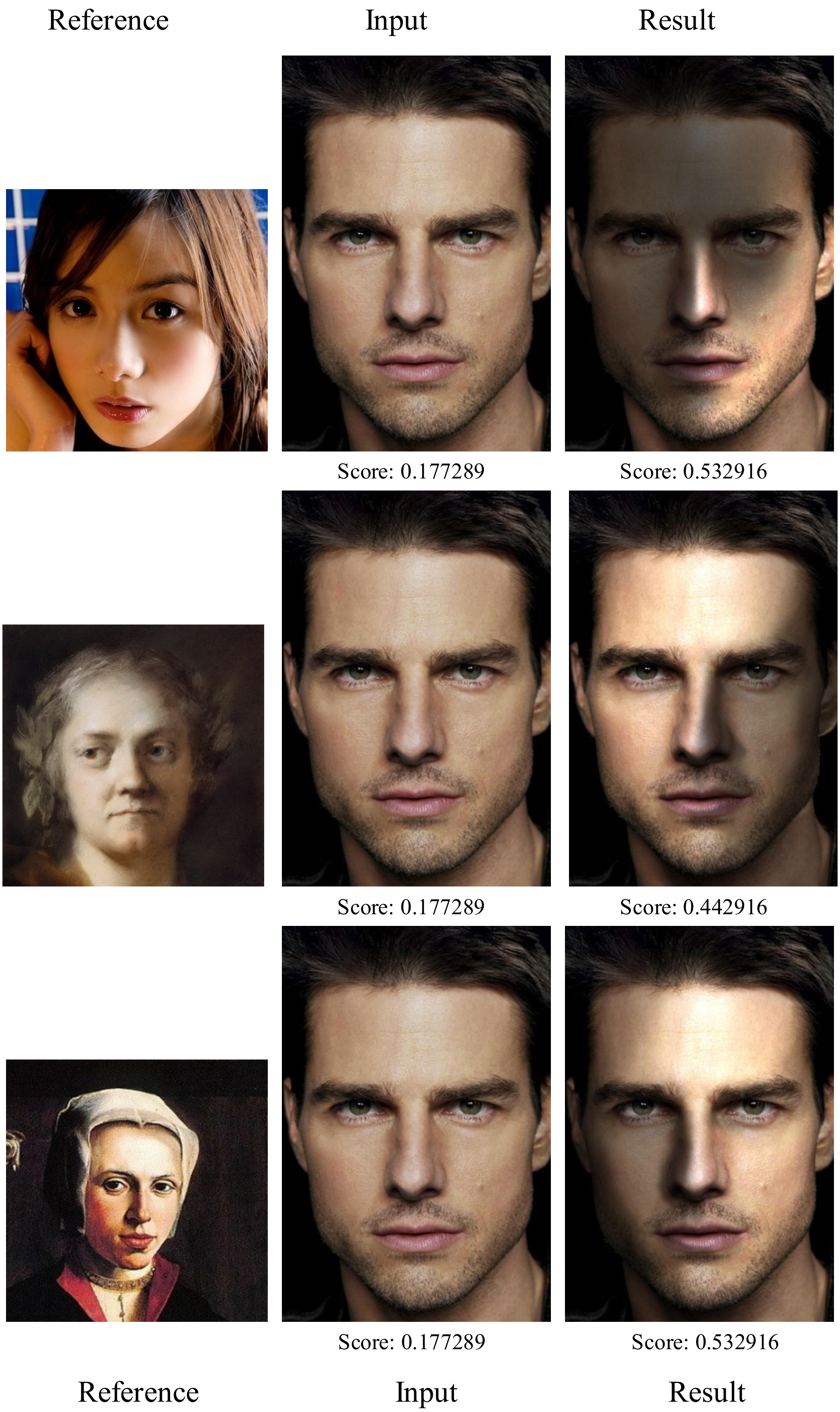
Score: 0.431428 Score: 0.814819

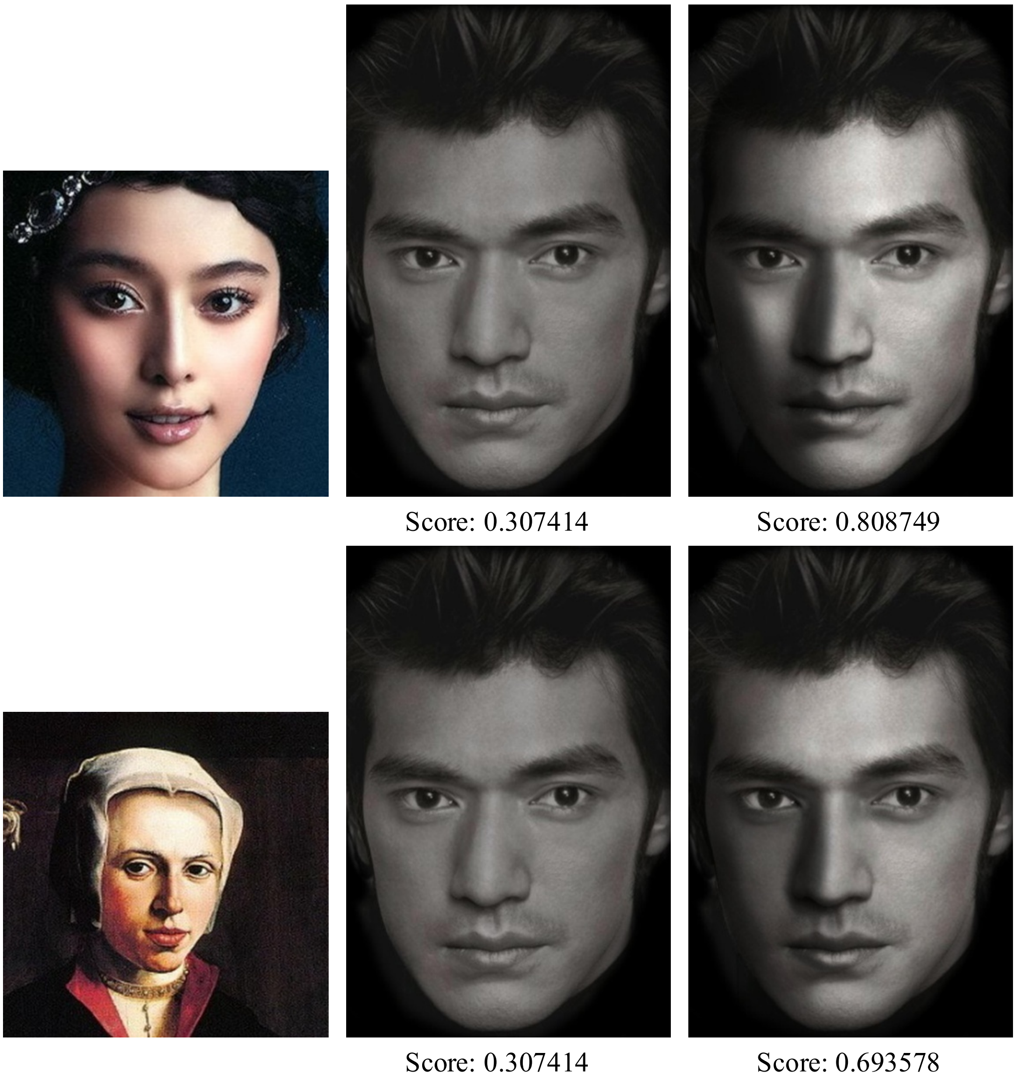
Figure 2: *Additional CGP results. First column: reference artistic portraits. Second column: input CGPs. Third column: the transfer results.*

1.1 Reference from the Database

(Section 5.1)

Input: Fine Art Photo


Reference
Input
Result
Score: 0.177289
Score: 0.532916
Score: 0.177289
Score: 0.442916
Score: 0.177289
Score: 0.532916
Reference
Input
Result


Score: 0.307414
Score: 0.808749
Score: 0.307414
Score: 0.693578

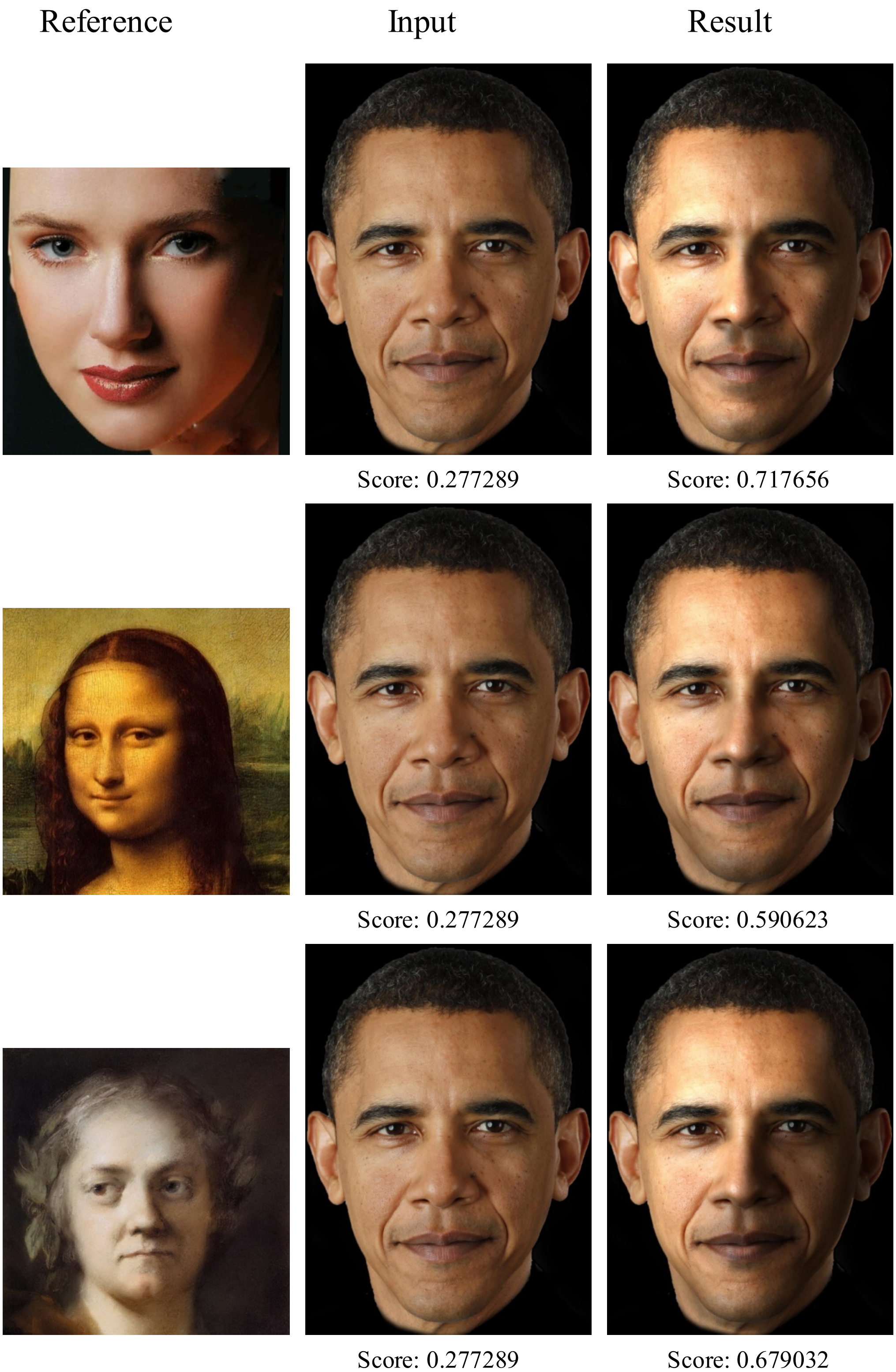


Figure 3: *Additional fine art photo results. First column: reference artistic portraits. Second column: input fine art photos. Third column: the transfer results.*

1.2 User-provided Reference
(Including the Implementation Details)

To extend the ability of our template database, for the user-provided reference artistic portrait out of our database, the one with the most similar illumination to that of the user-provided reference one is selected from our database. Then we make some adjustment of the corresponding illumination template to make the illumination effects of the results more similar to that of the desired reference ones. We design a Face Illumination Descriptor (FID) to describe the illumination effects of a portrait, which can be used for both illumination matching and illumination template adjusting.

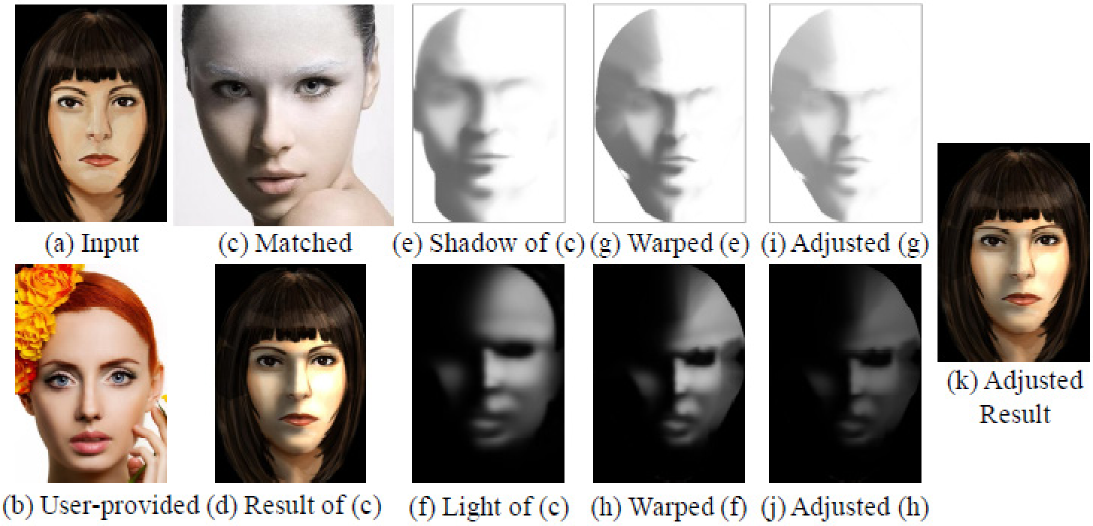


Figure 4: *The extended transfer workflow (User User-provided Reference). (a) is the input portrait; (b) is the user user-provided reference outside our database; (c) is the matched result obtained by our illumination matching method; (d) is the synthesized result by referencing (c); (e) and (f) is the shadow template and the light template of (c); (g) and (h) are the warped templates from (e) and (f); (i) and (j) are the automatic adjusted templates from (g) and (h); (k) are the adjusted result of (d). Illumination in (k) is more similar to that of (b) than that of (c).*

Face Illumination Descriptor. The local lighting contrast features has been extracted from 16 pre-defined regions in frontal faces and selected to form an artistic lighting template for classification and assessment of portrait photos in lighting usage [JZC*10]. We extend the rectangle region to non-regular region with a mask of each parts such as forehead, nose, eyebrows, mouths, etc. (see Figure 5 (b) and (c)). Jin et al. [JZC*10] define the lighting contrast features in each region with 3 directions: left-vs-right, top-vs-bottom and center-vs-periphery. They extract the contrast features using the contrasts between various statistics in various channels of each 2 sub-regions. Then 224 features are exacted to form a candidate feature set. In the learned average template of [JZC*10], 8 out of 12 selected features are on the L channel in *"CIE1976 (L*;a*;b*)"* color space. As for target statistics, all selected features are on means.

Thus we only extract the local lighting contrast of each region by calculate the difference between the mean pixel values of the 2 sub-regions on the L channel.

Considering the regions with mixed lighting contrast type of the 3 basic directions, we compute the local contrast in the vertical, horizontal and center-periphery directions in each region. Then 48 lighting contrast features (with 3 lighting contrast features in 16 regions) form the Face Illumination Descriptor to describe the illumination effects of a portrait.

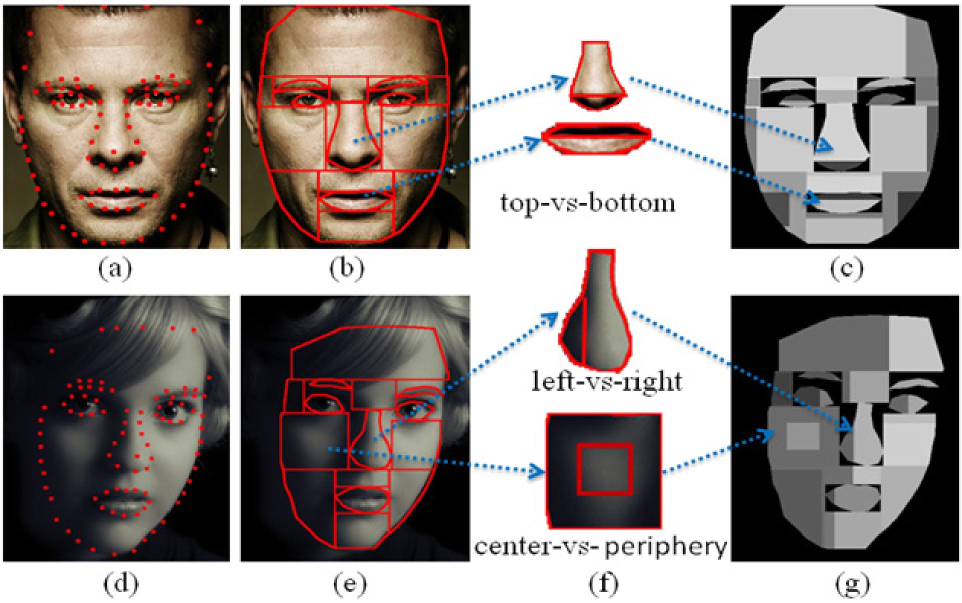


Figure 5: *Face illumination descriptor. (a) and (d): 90 facial feature points; (b) and (e): 16 regions according to facial feature points, which can cover most areas on a face; (c) and (g): the corresponding illumination descriptor of (a) and (d). We select the largest contrast direction in each region for visualization. (f): the 3 basic lighting contrast directions. The lines within regions are the separators.*

Each region Ri of the 16 regions $\{R_1, R_2, \ldots, R_{16}\}$ is divided into 2 sub-regions R_{i0} and R_{i1} by a separator (a vertical line for left-vs-right lr, a horizontal line for top-vs-bottom tb and a rectangle for center-vs-periphery cp) in position p_d, $d \in \{lr, tb, cp\}$, as illustrated in Figure 5. The mean pixel values of the 2 sub-regions are $\mu_{i1}(p_d)$ and $\mu_{i2}(p_d)$. We look for the maximum difference between the mean pixel values of the 2 sub-regions in each direction of each region to describe the local lighting contrast. Therefore, the local lighting contrast values in 3 directions of region R_i are:

$$p_d^* = \arg\max_{p_d} |\mu_{i1}(p_d) - \mu_{i2}(p_d)|$$
$$r_i^d = \mu_{i1}(p_d^*) - \mu_{i2}(p_d^*)$$

where $d \in \{lr, tb, cp\}$. R_{i1} is the left, top or center region, and R_{i2} is the right, bottom or periphery region for the directions lr, tb, and cp, respectively. p_d is determined by an exhaustive search with a certain steps in each region. The illumination descriptor $\mathrm{FID}(P)$ of a face photo P is a 48d vector, which consists of local lighting contrast values in the 3 directions of the 16 regions in the face:

$$\mathrm{FID}(P) = \{r_i^{lr}, r_i^{tb}, r_i^{cp} | i = 1,2, \ldots 16\}$$

Face Illumination Matching. We define the difference between 2 face illumination descriptors by using the Euclidean distance as:

$$\mathrm{ED}(fid_1, fid_2) = \sqrt{\sum_{\mathrm{i}=1}^{16} \sum_{\mathrm{d}} \left(r_{1i}^{d} - r_{2i}^{d}\right)^2}$$

where $\mathrm{d} \in \{\mathrm{lr}, \mathrm{tb}, \mathrm{cp}\}$, and $\mathrm{r}_{1\mathrm{i}}^{\mathrm{d}}$ and $\mathrm{r}_{2\mathrm{i}}^{\mathrm{d}}$ are the elements in fid_1 and fid_2, respectively.

As shown in Figure 5, we use just 90 facial feature points for FID computation. The 90 points cover most of the areas in the face, which are sufficient for both illumination matching and shadow template adjusting.

For artistic portrait photos out of the database, the one with the most similar illumination is selected. We assume that the faces have similar illumination effects under the same illumination condition. We thus use the illumination descriptor to approximate the lighting effects for illumination matching of faces. The one with the minimum difference on the FID under a threshold is considered as the matched result. For more general cases, the face illumination matching can be denoted as $\mathrm{FIM}(\mathrm{P}, \Omega)$, which means finding the one under the most similar illumination to that of photo P in a portrait set Ω.

Artistic Illumination Template Adjusting. The illumination effects of the matched reference portrait still have a little difference from those of the user user-provided one. Since our database covers most of the lighting directions within each illumination type, the one with the similar lighting directions of multiple light sources can always be matched. Based on the observations of artistic portrait images and according to the painters, one of the biggest differences within an illumination type is the intensity ratio between the shadow and light regions, which is mainly due to the ratio between the key and the fill light sources in studios. The contrasts between the light and the shadow areas are often with a little difference in the matched portrait and the user user-provided one. This can be automatically adjusted based on FID.

During the drawing procedure of the painters, some intermediate light and shadow layers are darker or lighter than those of the reference artistic portrait. Then they adjust the whole light and shadow layer by the “level" tool, which in fact adjusts the input and output graylevel of light and shadow layer pixels. Making the shadow template darker or lighter will influence the ratio. The adjustment of the light template is similar. Since FID consists of local lighting contrast features, we define the whole contrast of a portrait as

$$\mathrm{CT}(fid) = \sum_{\mathrm{i}=1}^{16} \sum_{\mathrm{d}} r_i^d, d \in \{lr, tb, cp\}$$

The matched reference portrait from our database and the user user-provided portrait are denoted as P_m and P_c, respectively. The transferred result of the input portrait $\mathrm{P_i}$ is denoted as P_r. If the whole lighting contrast of P_r is higher than that of P_c, we make the light template darker and the shadow template lighter by controlling the output and input graylevel and vise versa. Then the result P_r is re-rendered according to the adjusted light and shadow template. The illumination template adjusting scheme is described as

$$\text{reduce } h_L, \text{increase } l_S, \mathrm{CT}(\mathrm{FID}(P_r)) > CT(\mathrm{FID}(P_c))$$
$$\text{increase } l_L, \text{reduce } h_S, \mathrm{CT}(\mathrm{FID}(P_r)) \le \mathrm{CT}(\mathrm{FID}(P_c))$$

where $\mathrm{h_L}$ and $\mathrm{l_L}$, $\mathrm{h_S}$ and $\mathrm{l_S}$ are the highest output graylevel and lowest input graylevel of the light template, the shadow template, respectively. According to the painters, the adjustment of the graylevel is always within 20. Thus when $\mathrm{CT}(\mathrm{FID}(P_r)) > CT(\mathrm{FID}(P_c))$, we limit the h_L value in [235, 255) and the l_S value in (0, 20]. Then we find the combination with the lowest difference between the FID of P_r and P_c in all the 400 combinations of h_L and l_S and vise versa. Then the adjusted synthesized result is considered to have the minimum difference with respect to the illumination effects from that of the user user-provided portrait, which is out of our database.

After the illumination template is adjusted, with face alignment, the artistic illumination effects of the user user-provided reference artistic portrait, which is out of our database, are transferred to the input portrait. This last step is the same as described in Section 3 of the submitted paper. The results will be shown below.

Experimental Results

Illumination Matching Results. We test the face illumination matching in the Yale Face database B and the Extension [GBK01], which contain frontal face photos of 38 person under the same 64 lighting conditions (64 point light sources in 64 directions). We denote the frontal face database as $\{\Omega^{y_i} | i = 1,2, \ldots ,38\}$, $\Omega^{y_i} = \{P_i^j | j = 1,2, \ldots ,64\}$. For each P_i^j, we run the illumination matching $\mathrm{FIM}(P_i^j, \Omega^{y_k})$ to search the same lighting direction in each $\Omega^{y_k}, k = 1,2, \ldots ,38, k \neq i$. Due to the dense light directions of Yale database, even human can not distinguish the illumination effects of two light directions with small angle. We thus relax the correct matching criterion, and set that if the same lighting condition as that of P_i^j is in the top 3 matched results, we consider it as correct matching. The accuracy rate for each P_i^j is the times of correct matching divided by 37. The average accuracy rate for each lighting condition j is the average over the accuracy rate of $\{P_i^j | j = 1,2, \ldots ,38\}$.

The accuracy rates are shown in Figure 6 and Figure 7. Most lighting conditions achieve above 70% accuracy except the ones with small angle with respect to camera axis, which are really difficult to distinguish even by human. This demonstrates how well the face illumination matching performances in such a strict testing. Some matched results of the real artistic portraits and our database are shown in Figure 8.

The query and the matched results returned by our method have the similar illumination effects, which show the performance of the method. More matched results of the real artistic portraits and our database are shown in Figure 8.

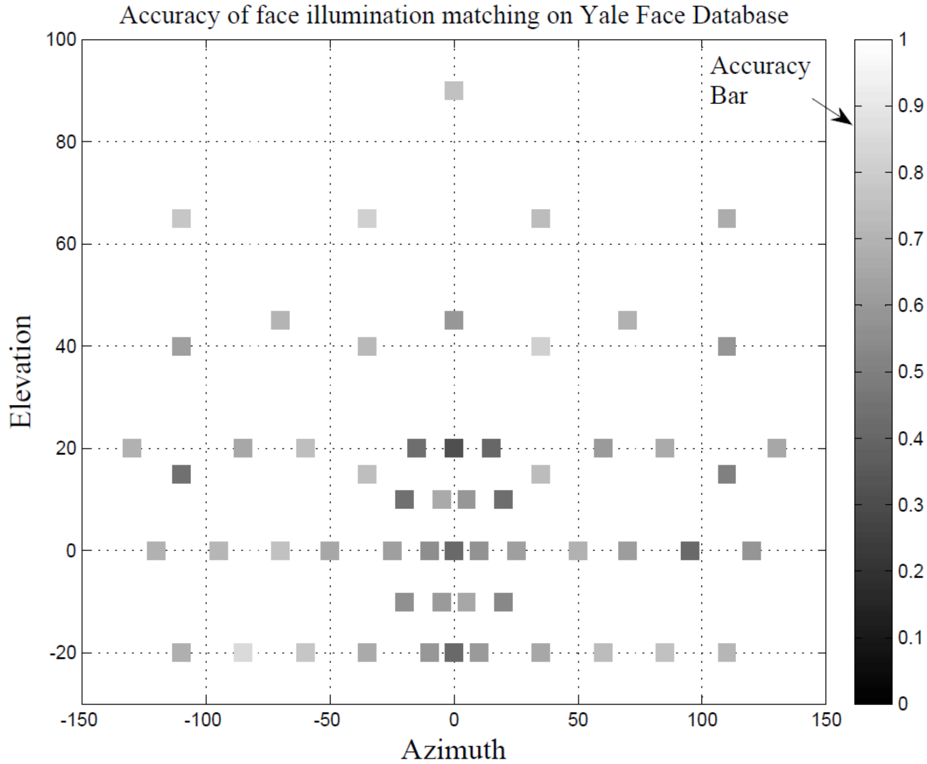


Figure 6: *Accuracy of face illumination matching on Yale Face database. The distribution of all the 64 light sources is shown with each direction as a square. The average accuracy of each direction is shown as a small squares. The whiter the squares are, the more accurate they are considered to be.*

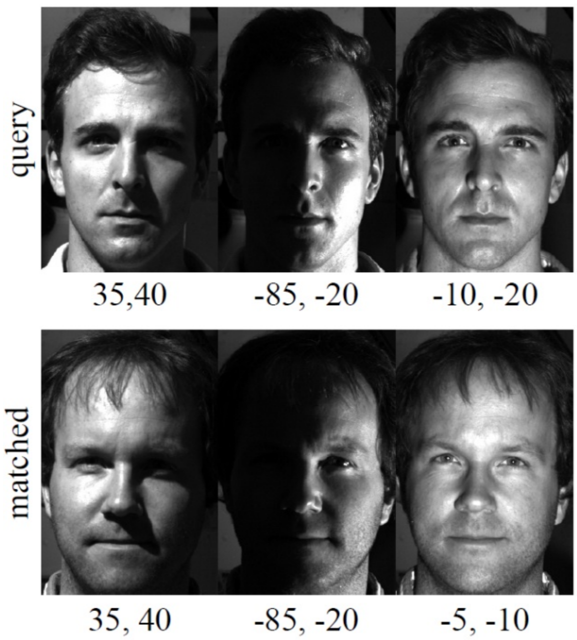
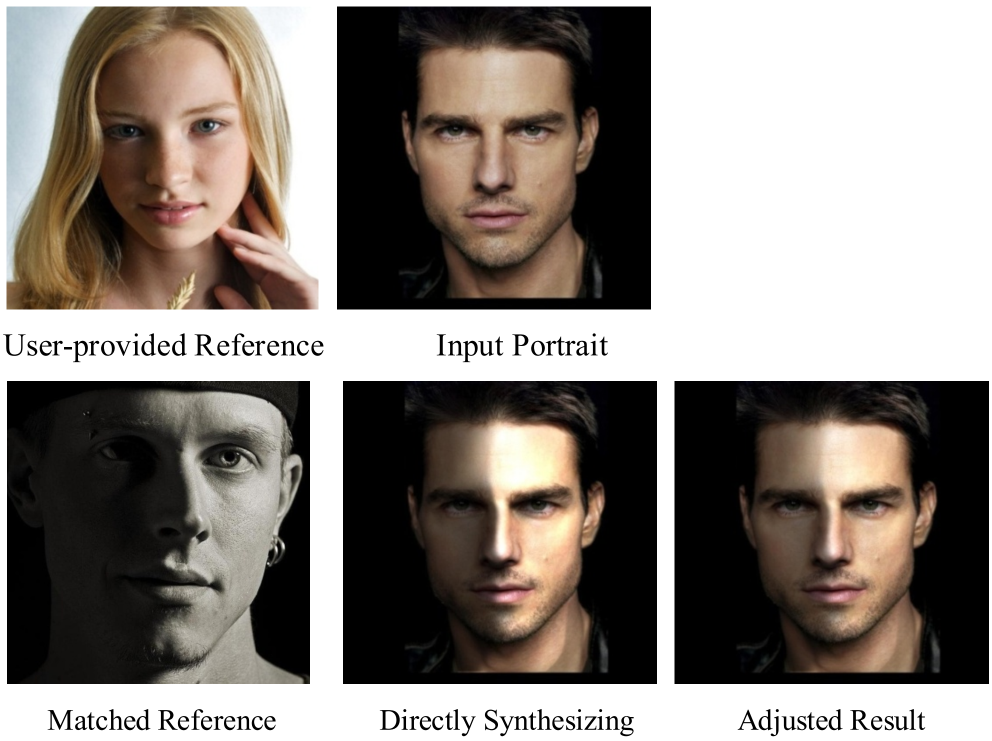


Figure 7: *We show 3 matched results of querying the person 10 of YaleB in the person 3's images. The light source directions are with respect to the camera axis, which is perpendicular to human faces for frontal view photos. The azimuth is the horizontal angle, while the elevation is the*

vertical angle. The numbers below are azimuths and elevations of the point light sources.

Figure 8: *More matched results. Top line: the queried art portrait out of our database. Bottom line: the matched results obtained by our face illumination matching method. The queried art portraits shown in this figure cover most of the artistic illumination styles.*

Illumination Template Adjusting. Figure 4 shows an example of illumination template adjusting. The user user-provided reference portrait is not in our database. Through the illumination matching, the one with the most similar illumination in our database to that of the user-provided reference is matched. By our adjusting method, we obtain the adjusted illumination template that minimizes the distance between the FID of the transferred result and the user-provided artistic portrait. More results are shown in Figure 9.




User-provided Reference

Input Portrait

Matched Reference

Directly Transfer

Adjusted Result

User-provided Reference

Input Portrait

Matched Reference

Directly Transfer

Adjusted Result

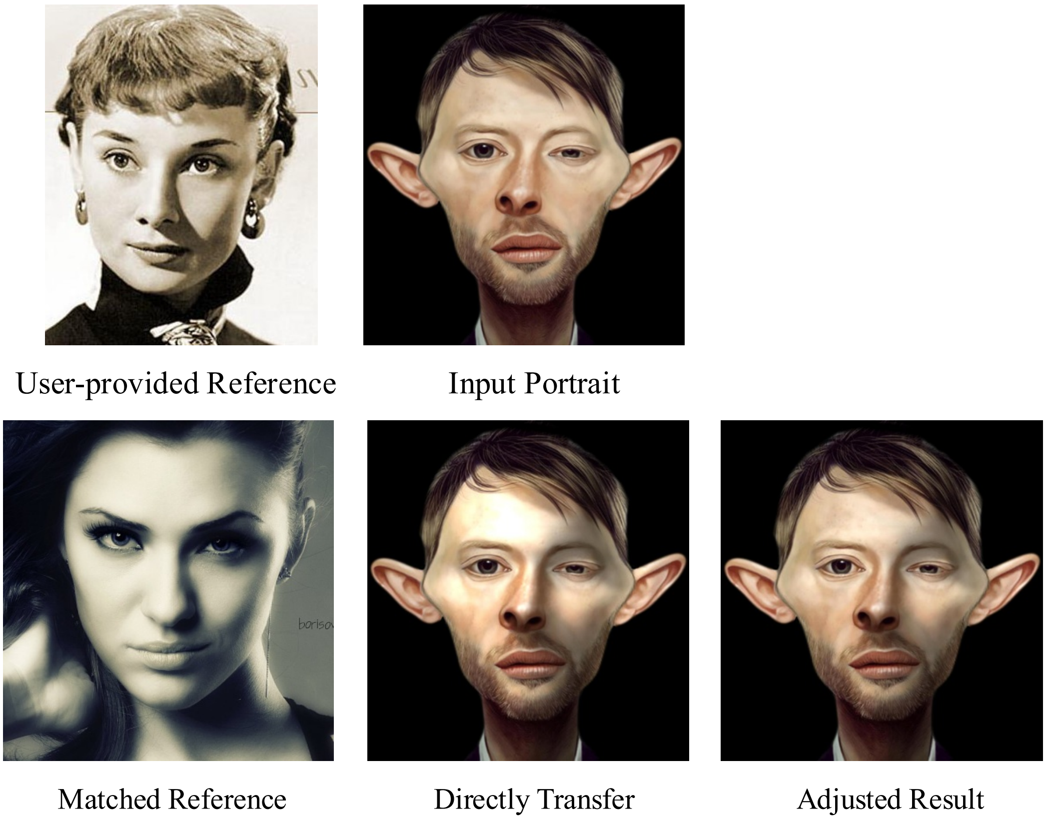


Figure 9: *For a user-provided reference not in our database, the one with the most similar illumination effects with that of the user-provided reference is matched from our database. Then the illumination template of the matched one is adjusted according to the difference of illumination effects of the user-provided reference and the transferred result.*

1.3 Applications

(Section 5.4)

1.3.1 Numerical Assessment and One-key Transfer

We use the method of [JZC 10] to predict indicative scores from 0 to 1 of the artistic lighting usage of the input and the result portraits by our system. Jin et al. [JZC 10] randomly choose 50 photographs (either artistic or daily) as the training examples. 10 graduate students of various majors are chosen as test subjects to do the comparisons between photographs. With the learned average artistic lighting template from 350 artistic photos, the classification and assessment tasks are achieved under probability ratio test formulations. An artistic illumination score $s \in (0, 1)$ can be predicted of a test portrait image. In our work, we adopt similar method to that of [JZC 10]. The differences are: (1) For the training examples, we randomly choose 100 portrait images from the portrait photos and paintings (both with and without artistic illumination effects) selected by professional artists; (2) Our subjects are 15 professional artists from professional art studios and the department of fine arts; (3) We learn the average artistic lighting template from not only portrait photos but also masterpieces by famous artists (with total number 550). To some extent, these scores are only indicative. We test on 50 input portraits and produce 500 results by randomly choosing reference portraits from our database. The results show that Most of the transferred results by our system can achieve higher scores than the input portraits which have not obvious illumination effects. (see Figure 10)

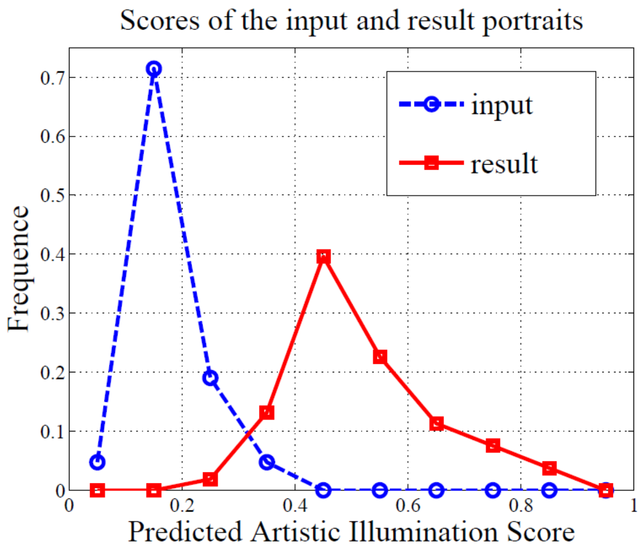
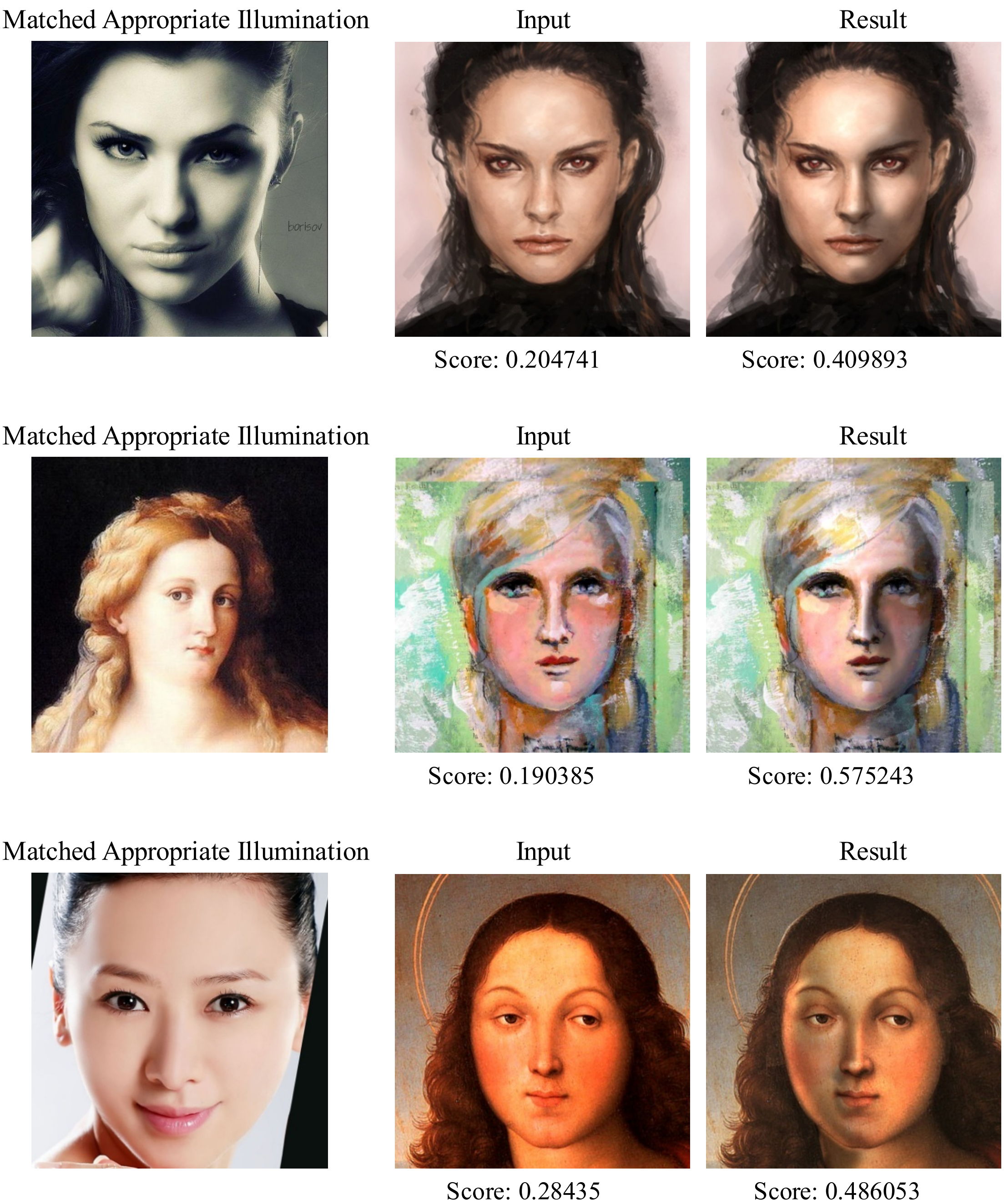


Figure 10: *Artistic illumination scores of the input and the result portraits. The results show that our system can synthesize portraits with higher artistic illumination scores than those of the input ones.*

[JZC10] JIN X., ZHAO M.-T., CHEN X.-W., ZHAO Q.-P., ZHU S.-C.: Learning artistic lighting template from portrait photographs. In Proc. ECCV (2010).

Based on the predict scores, our system can automatically recommend an *appropriate* reference portrait and transfer its illumination template to the input portrait. Artists often design lighting according to the shape of the face and the distribution and shapes of facial parts (eyes, nose, etc.). Thus our system automatically searches a set of faces (10 in our system) with similar shapes and facial part distributions to those of the input portraits from our database. Then the result with the highest score is selected as the proposed appropriate result of the input portrait. For the face matching, we connect the facial feature points to form the face contour and the shapes of the parts in the face. Then the face shape similarity is calculated in terms of the shape context distance metric [BMP02] by using both the face contour and the shapes of the facial parts. More results are shown in Figure 11.

[BMP02] BELONGIE S., MALIK J., PUZICHA J.: Shape matching and object recognition using shape contexts. TPAMI 24, 4 (April 2002), 509–522.


Matched Appropriate Illumination
Input
Result
Score: 0.204741
Score: 0.409893
Matched Appropriate Illumination
Input
Result
Score: 0.190385
Score: 0.575243
Matched Appropriate Illumination
Input
Result
Score: 0.28435
Score: 0.486053

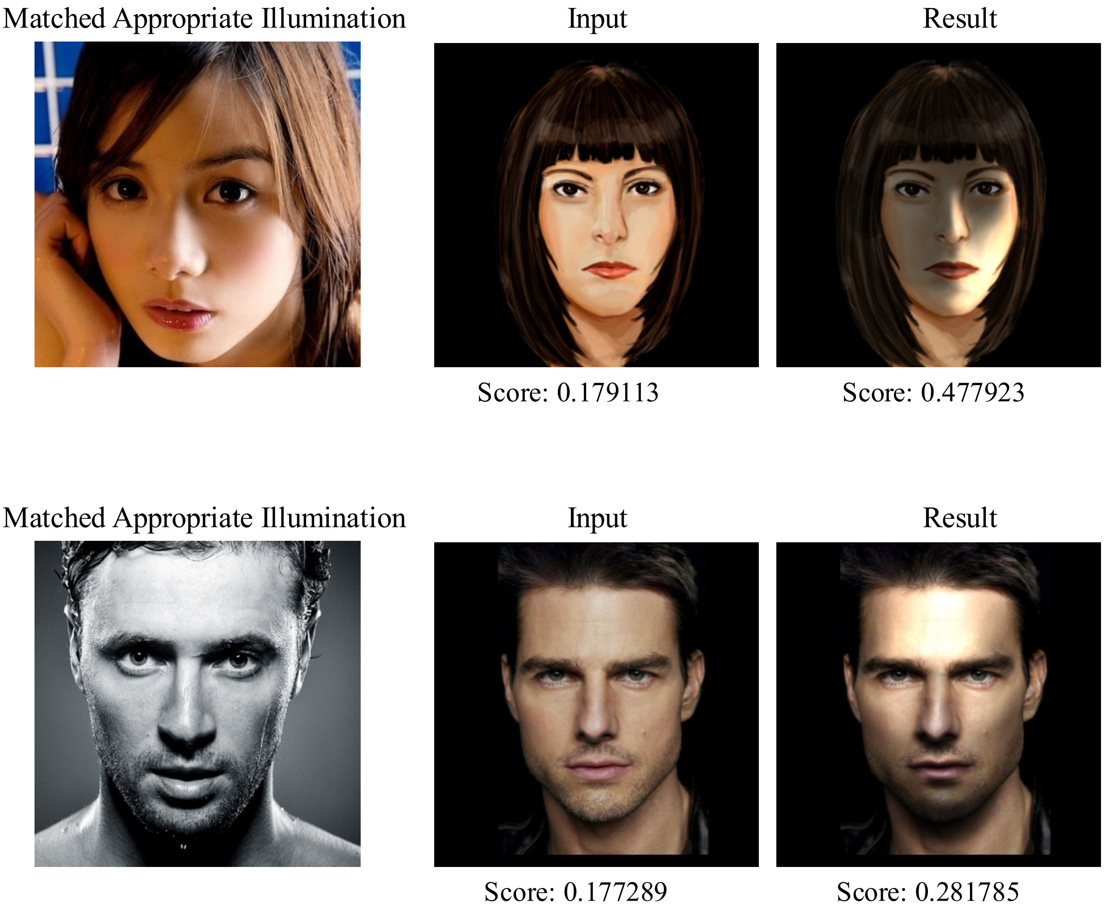


Figure 11: *One-key Transfer. Based on the predict scores, our system can automatically recommend an appropriate reference portrait and transfer its illumination template to the input portrait.*

We have the observation that our basis illumination templates are suitable for oil painting, watercolor styles, etc. However, in some artistic portrait styles such as paper-cut and sketch, illumination effects are drawn in special effects of two-tone form and hatching form. Thus the styles of our illumination templates should be adapted to the styles of the underlying portrait images. Based on the separated light and shadow templates, the styles of the illumination templates can be easily adapted to paper-cut style and sketch style.

1.3.2 Input: Paper-cut

Recently, Meng et al. [MZZ10] have proposed a method of rendering paper-cut images from portrait photos based on a pre-collected representative paper-cut templates created by artists. However, they focus on the representation of facial features and their paper-cut templates are created from portrait under normal illumination condition. This limits their results in the representation of illumination effects, which is a critical factor in producing visual appealing paper-cut portraits. While our database consists various separated artistic light and shadow templates drawn by professional artists. The artistic illumination effects can be easily added in their output portrait images by adapting the styles of our illumination templates to the paper-cut styles.

Paper-cut portraits are usually in a very concise two-tone (red foreground and white background) form. The illumination effects in professional paper-cut are expressed in such a two-tone form. The light areas are often in white, while the shadow areas red. Since the reference portrait photos are often taken under normal illumination condition, the outputs of [MZZ10] are paper-cut portraits without obvious shadow effects. Most areas of the face are already in white, thus we can only add shadow templates to the face with appropriate thresholds. We observe that Otsu's method [Ots79] can select such an appropriate binarization threshold on the warped shadow template to render continuous shadow effects pasted into the input paper-cut portrait.

Connectivity. Meng et al. [MZZ10] pre-defined a few possible curves for enforcing the connectivity, which is an important characteristic of traditional paper-cut. After the shadow effects being synthesized in the input paper-cut portrait, some of the pre-defined connectivity curves are not necessary. Thus, in the implementation, we remove these curves firstly. After the shadow effects being synthesized, we detect the connectivity and add back the needed curves.

The rendered illumination effects in paper-cut and sketch portraits are shown in Figure 12. We obtain similar binarization result to that of real paper-cut artwork by artists. More results are shown in Figure 13.

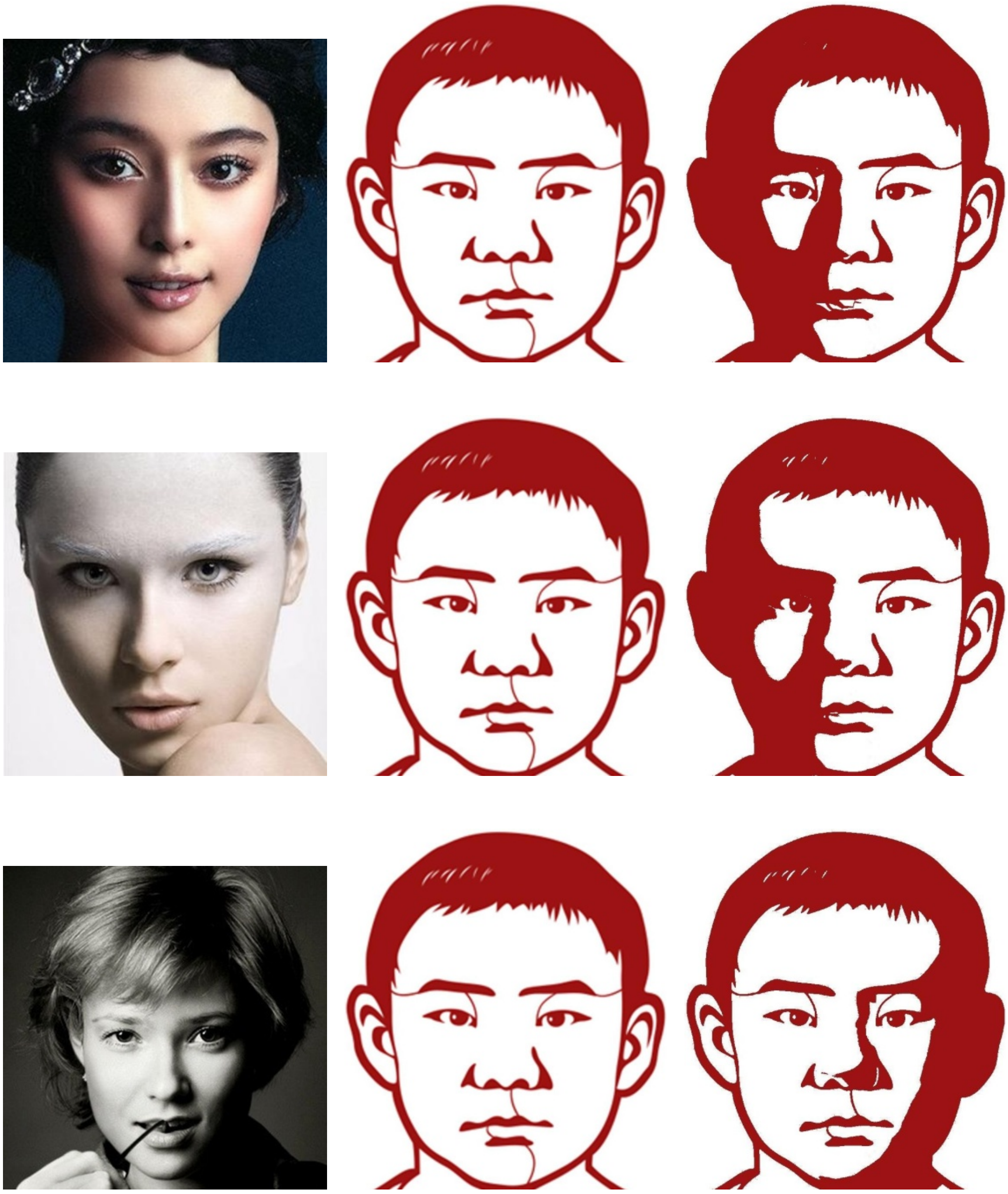
(a) Paper-cut by artists (b) Input Paper-cut (c) Result

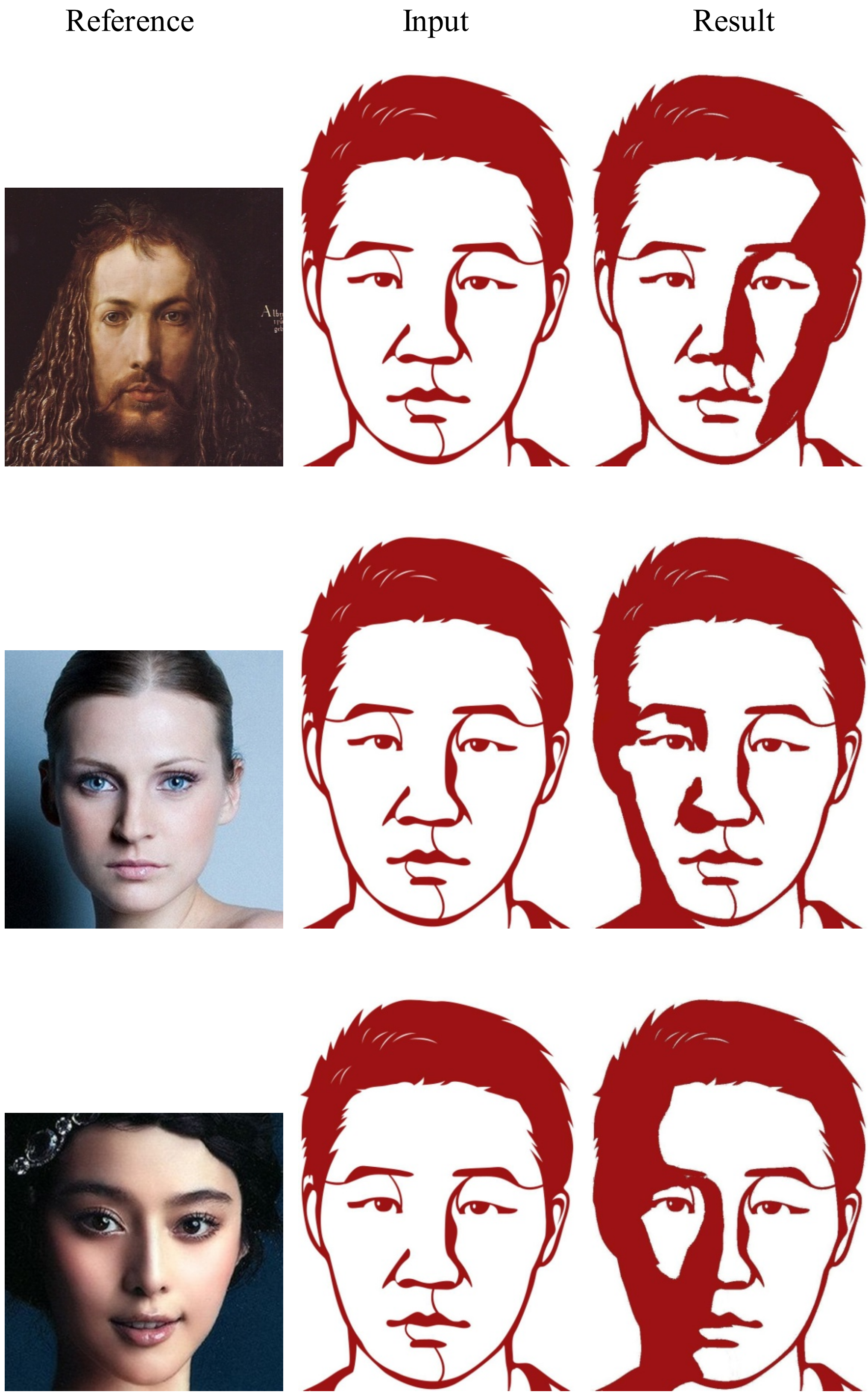
Figure 12: *Paper-cut Result. (a) is a paper-cut portrait created by an artist. (b) is the input sketch created by [MZZ10]. (c) is the result by adding shadow effects by our work.*

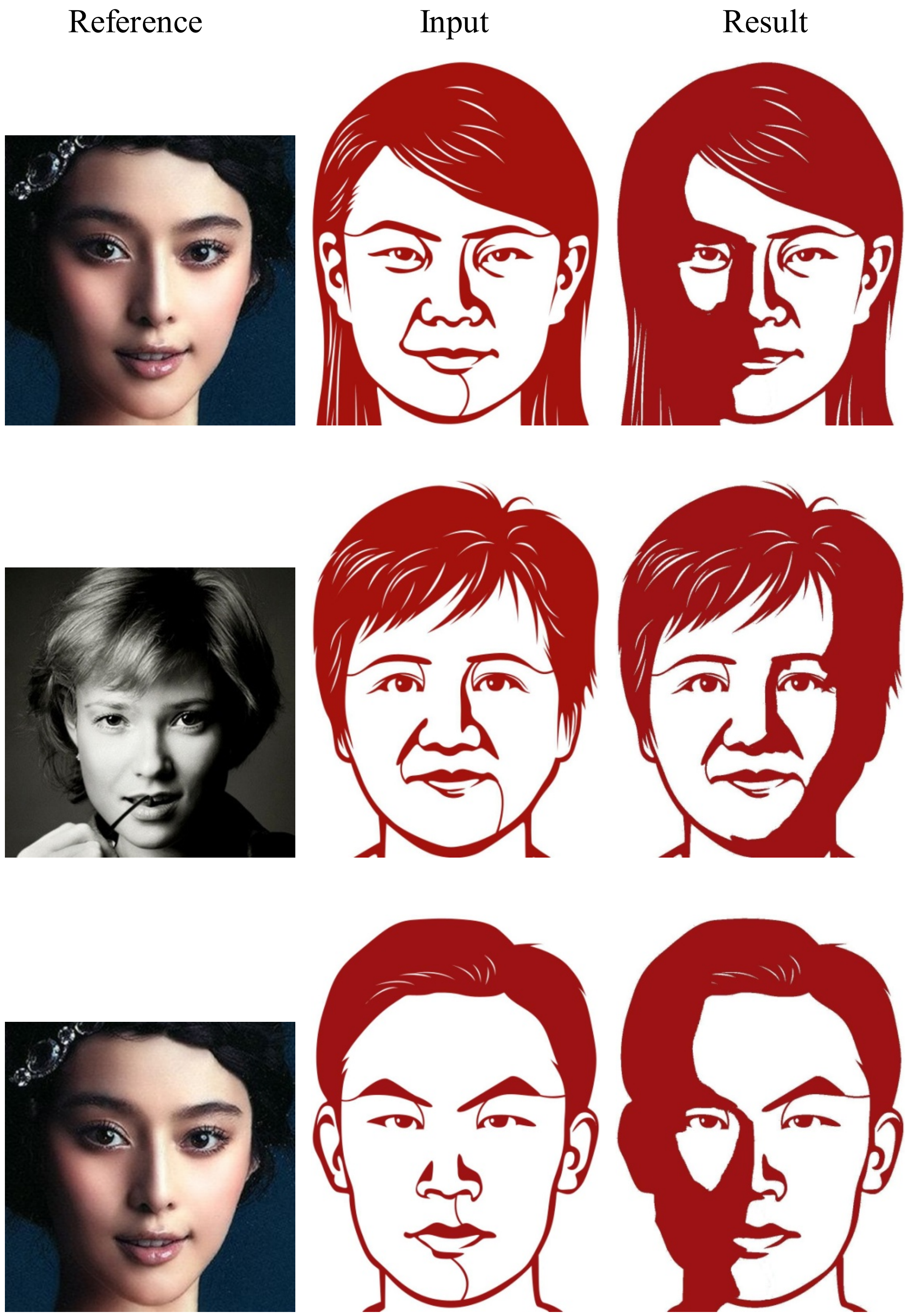
[MZZ10] MENG M., ZHAO M., ZHU S.-C.: Artistic paper-cut of human portraits. In ACM Multimedia (2010), pp. 931–934.

[Ots79] OTSU N.: A Threshold Selection Method from Graylevel Histograms. TSMC 9, 1 (1979), 62–66.

Reference Input Result




Reference
Input
Result


Reference
Input
Result

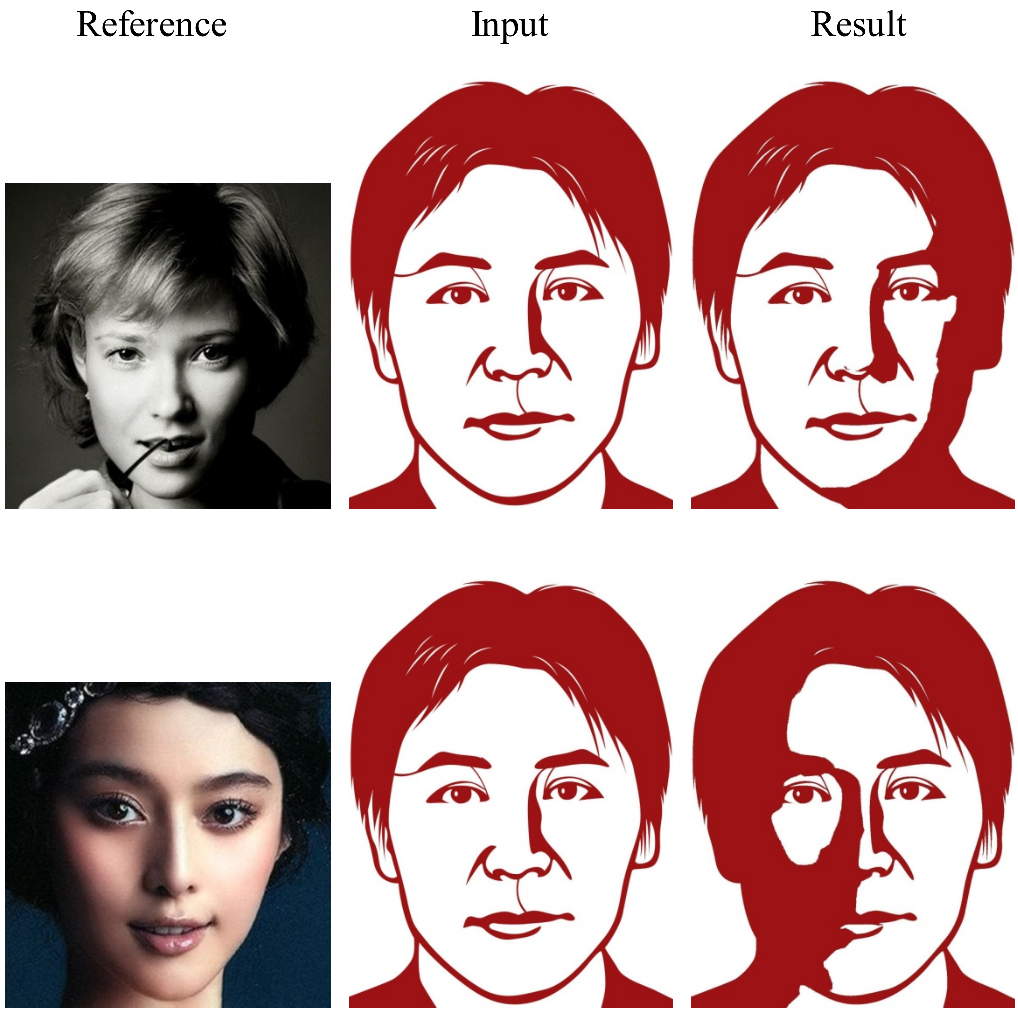


Figure 13: *Additional paper-cut results. First column: reference artistic portraits. Second column: input paper-cuts. Third column: the transferred results.*

1.3.3 Input: Sketch

Illumination effects in sketch portraits are usually expressed by hatching. Learning hatching texture from sketch artworks is out of the scope of this article. Thus, after an artist creating a sketch portrait under normal illumination, we ask him/her to provide a Hatching Art Map (see the figure in the next page) which contains discrete key hatching texture samples at key tones from dark to light. Then we use our illumination template to guide the synthesis of various illumination effects in the normal illumination sketch portraits.

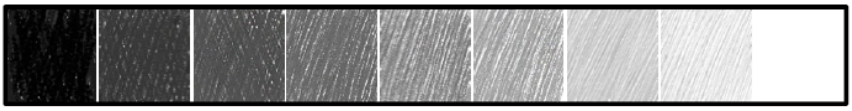
The hatching effects rendering is a modified texture transfer technique [EF01] guided by illumination, which is similar to what Kulla's work in [KTBG] of synthesizing illumination effects in 3D meshes from scanned paint samples that represent dark to light transitions. To render coherent illumination effects, the hatching texture samples between two adjacent key hatching samples in the hatching tonal map are synthesized by the linear combination of the two adjacent key hatching samples. The guide image for the texture transfer is generated by the basic transfer workflow described in Section 3 of the submitted manuscript, which takes the normal illumination sketch as the input image. Then the hatching effects are rendered by hatching texture transfer according to the intensity of the guide image. The transferred results of sketch portraits are shown in Figure 14.

[EF01] EFROS A. A., FREEMAN W. T.: Image quilting for texture synthesis and transfer. In SIGGRAPH (2001), SIGGRAPH '01, pp. 341–346.

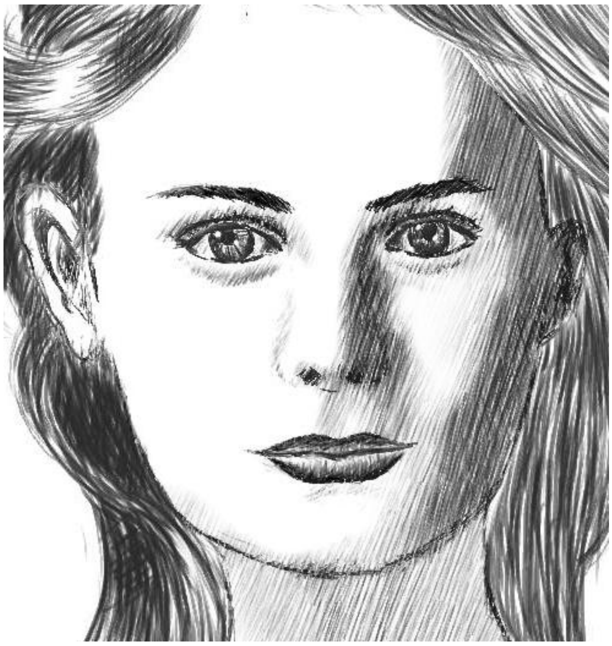
[KTBG] KULLA C., TUCEK J., BAILEY R., GRIMM C.: Using texture synthesis for non-photorealistic shading from paint samples. In Pacific Graphics (2003), pp. 477–481.

Reference: 

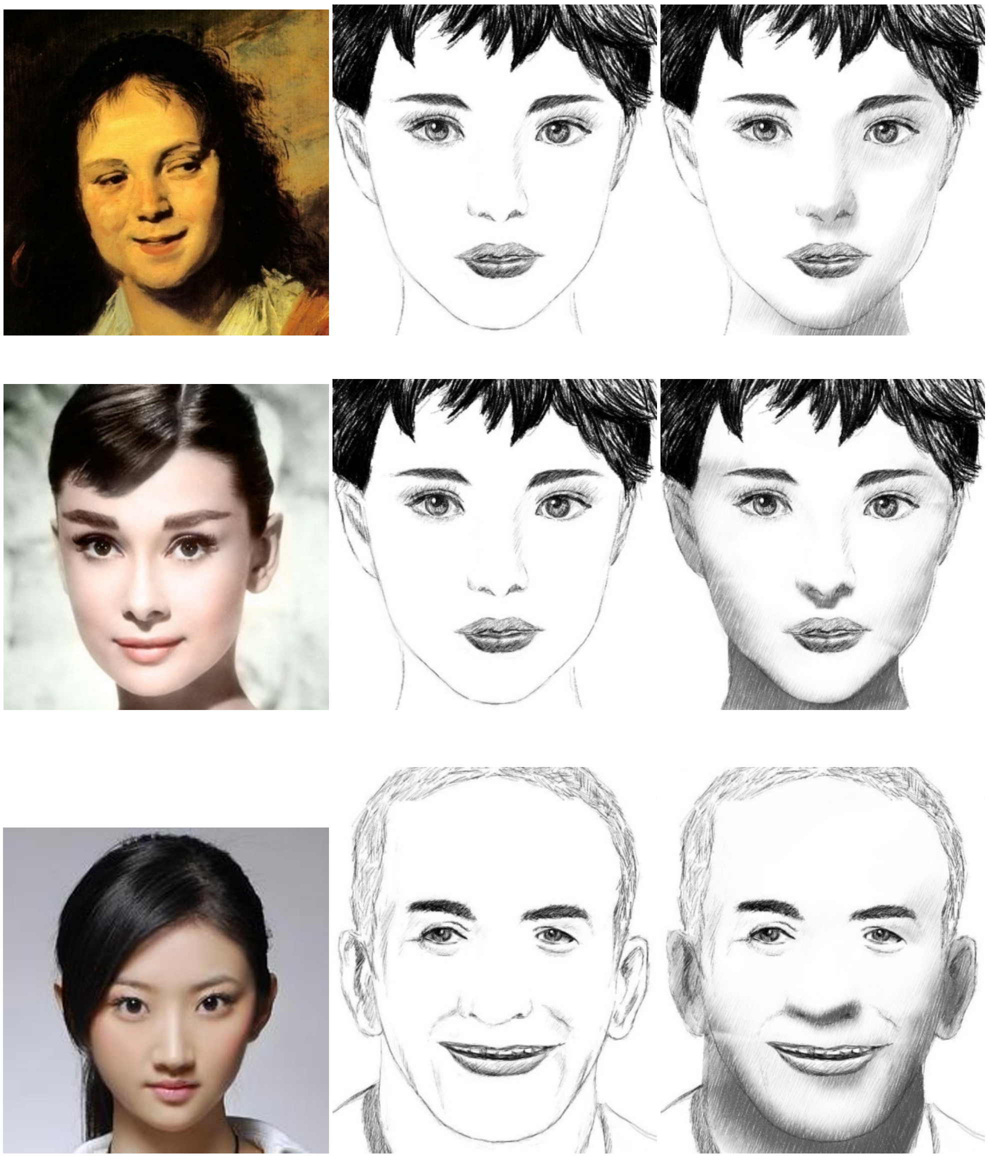
Hatching Art Map



Input: 

Result: 

Reference Input Result



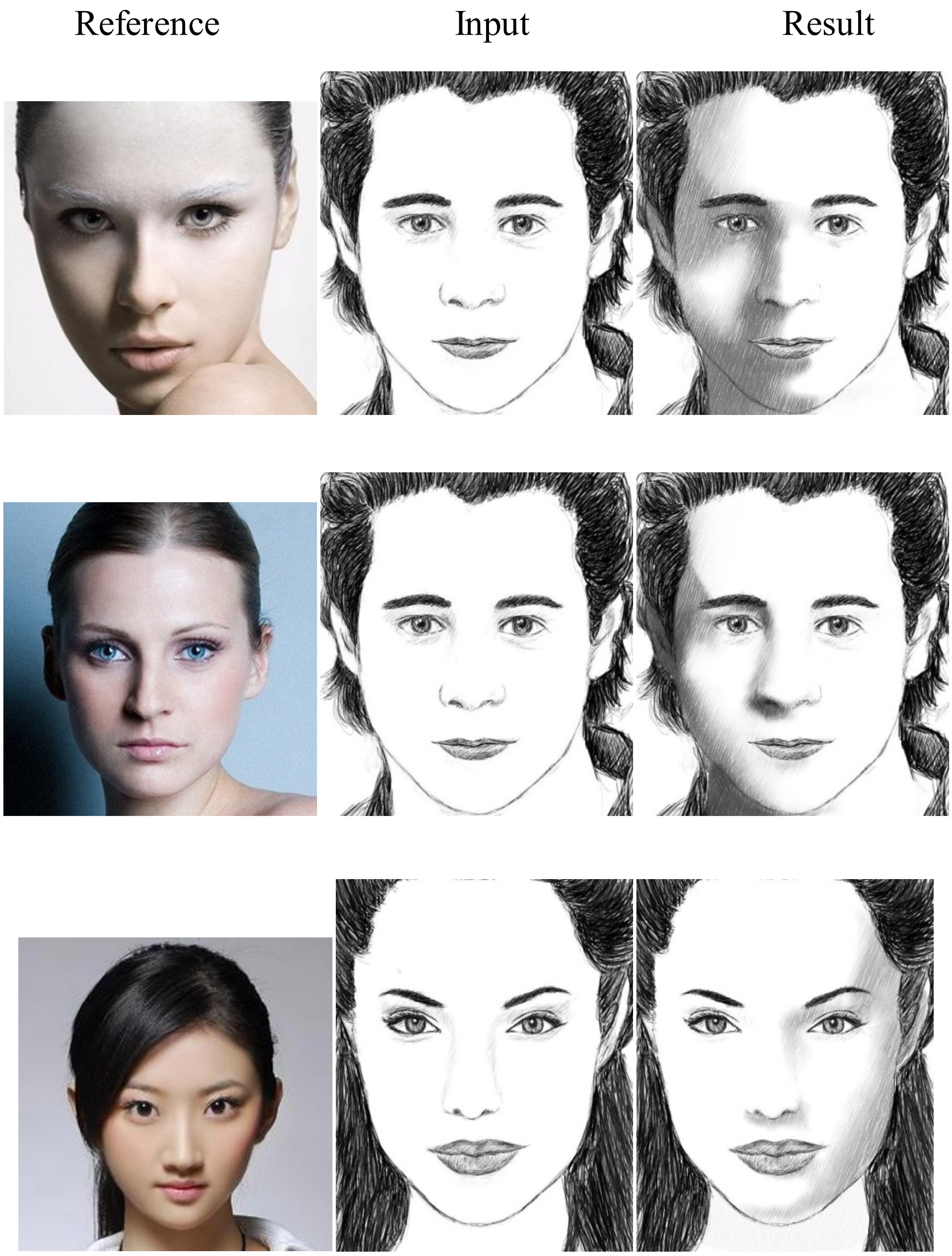


Figure 14: *Additional sketch results. First column: reference artistic portraits. Second column: input sketches. Third column: the transferred results.*