

Exemplar-based Portrait Photograph Enhancement as informed by Portrait Paintings

Xiaoyan Zhang^{1,3} and Martin Constable² and Kap Luk Chan^{1*}

¹School of Electrical & Electronic Engineering, NTU, Singapore

²School of Art, Design and Media, NTU, Singapore

³Institute for Media Innovation, NTU, Singapore

Email: {xiaoyan.zhang02@gmail.com; MConstable@ntu.edu.sg; eklchan@ntu.edu.sg}

Abstract

This paper proposes an approach to enhance the regional contrasts in snap-shot style portrait photographs by using pre-modern portrait paintings as aesthetic exemplars. The example portrait painting is selected based on a comparison of the existing contrast properties of the painting with those of the photograph. The contrast organization in the selected example painting is transferred to the photograph by mapping the inter- and intra-regional contrasts of the regions, such as the face and skin areas of the foreground figure, the non-face/skin part of the foreground, and the background region. A piecewise nonlinear transformation curve is used to achieve the contrast mapping. Finally, the transition boundary between regions is smoothed to achieve the final results. The experimental results and user study demonstrate that, by using this proposed approach, the visual appeal of the portrait photographs is effectively improved, and the face and the figure become more salient.

Keywords: Image manipulation, regional contrast, portrait photograph enhancement, portrait painting.

Categories and Subject Descriptors (according to ACM CCS): [Computing Methodologies]: Computer graphics—Image manipulation

1. Introduction

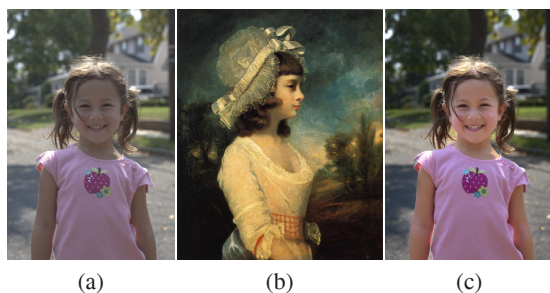


Figure 1: One of the results produced by our proposed approach. (a) Original photograph, (b) Reference painting, (c) Result.

Photo enhancement has received considerable interest with the prevalence of digital photography. The captured photograph may not always be as “appealing” as we hope, due to such influencing factors as: the quality and settings

of the camera, the skill of the operator and the external light condition. Most existing approaches in image improvement enhance the appearance of images globally or locally in the same way for all image categories, regardless whether it is a portrait picture or a natural scene. Similar to the strategy of using different modes or lenses to capture landscapes or portraits, we believe that photos with different subject focus need to be enhanced in different ways. In this paper, we address a specific genre of photographs: portrait photographs, which are photos focusing on the depiction of a person. Whether you are traveling, enjoying a family event, or recording your glory moment, everyone very often ends up shooting portraits. Portrait photos constitute an important part of our personal photo collections.

In a portrait, the figure, especially the face, is visually predominant. However, the key to controlling the strength of the attraction of a portrait is not by simply controlling the value of the center of interest, but rather by managing the contrast between the significant regions of the portrait: the face and skin areas of the foreground figure, the non-face/skin

part of the foreground figure and its accessories, and the background [Per06]. This is done in order to lead the eye of the viewer to the expression of the figure. The management of contrast among all these regions makes the figure and the face stand out. Therefore, our portrait photograph enhancement does not focus on the enhancement simply on the face, but on the relative relationships of the face and skin areas of the foreground figure, the non-face/skin part of the foreground figure and its accessories, and the background. A content-aware automatic photo enhancement method is proposed in [KLW12] in order to enhance the face, sky and saliency areas differently. However it only processes these regions separately and does not consider the relative relationships between these regions. As a result, the contrast between the face and the background may be reduced. Although Adobe Photoshop provides a powerful tool for users to manually edit their images, the skills required to achieve this region-specific improvement are beyond the scope of casual users. The user has to mask regions and make decisions based on the relative relationships of these regions. Additionally, defining the relative relationships for these regions in order to make the figure, especially the face more attractive, is a skill that is not easy to acquire.

Fortunately, portrait paintings, which are a particular class of the portrait image, supply us with good exemplars for modifying the relative relationship of these regions. Compared with a photograph, which can be regarded as a projection of the physical nature, a traditional portrait painting is a very carefully constructed artefact. Portrait artists are able to filter and manipulate the subject scene based on their visual perception [DiP09]. Key to this is the consideration given to the organization of color contrast values [Arn74]. Almost all portrait paintings are easily divisible into strong foreground (FG)/background (BG) regions (Figure 2). The FG is comprised of the figure with its accessories, with the remainder being the BG. In the example paintings shown in Figure 2 and Figure 3(a), the FG figure is darker than the outdoor BG. In the painting in Figure 3(b), the outdoor BG is darker than the figure. Despite the difference in strategies that have been employed, in all the three paintings, the artists have maintained a visual emphasis on the figure. The face and skin areas (FS) are almost the brightest parts of the figure. They have a high contrast with the non-face/skin part of the FG (FO) and BG regions. Despite the contrast between these regions (inter-contrast), there is also contrast organization within each region (intra-contrast) that serves to express the emotion on the face, the pose of the body, and also the richness of the BG. Differently, in snap-shot style photographs which are more casually framed ("point and shoot"), their lighting, colorfulness, and the regional contrasts are typically not well organized (e.g. the two photographs in Figure 3(c)(d)).

In order to better understand the regional contrast organization of paintings, a statistical comparison on the contrast values of lightness and saturation has been conducted

between photographs and paintings. Hue is not touched in this paper to preserve the natural appearance of photographs. Differences between paintings and snap-shot photographs are observed with evidence that the intra-contrast and inter-contrast values of lightness and saturation in paintings are more purposefully organized. This contrast organization purposely makes the figure, especially the face, stand out in the image, while preserving the contrast within each region in order to serve the visual appeal of the whole image. This observation inspires us to enhance portrait photographs through contrast adjustments to their regions according to the corresponding contrast organization in paintings.



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

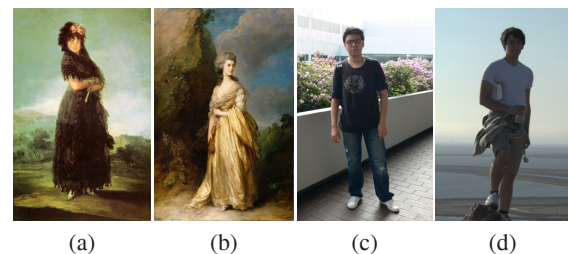


Figure 3: (a) "Portrait of Mariana Waldstein", by Francisco de Goya (1746-1828). (b) "Mrs. Peter William Baker" by Thomas Gainsborough (1727-1788). (c) and (d) are photographs. (d) is cropped from an image in the database [BPCD11].

Therefore, in this paper, we explore a method by which the high-level lightness and saturation contrast organization of portrait paintings can be transferred to portrait photographs in order to enhance their visual appearance. This is done with respect to the contrast values of the three regions FS, FO, and BG. These contrast values are considered both inter- and intra-regionally. One difficulty in applying regional contrast organization in paintings is that though general rules of regional contrast organization can be observed,

there exists many local variants of these rules. Therefore, instead of a rule-driven approach, an exemplar-based approach is proposed to enhance the regional contrasts of photographs. A selected reference painting is used as an exemplar to guide the enhancement of the portrait photograph. One result produced by our proposed approach is shown in Figure 1. The visual appearance of the photograph is improved and the figure stands out in the photograph.

2. Related Work

Portrait rendering. With the recent development of painterly rendering, portrait rendering has become an interesting topic in the computer vision and computer graphics research communities. Colton proposed to use a Non-Photorealistic Rendering (NPR) system to automatically produce artistic stylized portraits [CVP08]. The artistic styles in this NPR system were based on painting materials, color palette and brush models. Some recent work also attempted to render a portrait photo in an artistic style [CW08, ZGZ09, MZZ10, YLLea10, Rez12]. However, the styles were limited to simulating abstraction, line drawing styles or organic models. Face relighting is another popular research topic. A face lighting template or an example face image is used as reference to guide the face relighting of the input image [JZCea10, CSHea10, CJZW12]. All these research works on the subject of portraits only focus on the manipulation of the face. Differently, our method not only considers the face, but also the relationships of the face with the BG and other parts of the figure.

Image enhancement. Based on the strategy that has been employed, methods on image enhancement can be classified into three categories: single image manipulation, learning-based and exemplar-based methods. While most common single image manipulation methods only automatically enhance the image globally or locally without consideration of the content, some content-aware methods have been proposed to process regions differently. Rivera et al. proposed to adjust the contrast differently in the dark, middle and bright regions [RRC12]. However, the considered regions were split based on their intensity values and had no high-level semantic meaning. In addition, only lightness was adjusted. Differently, the content-aware method in [KLW12] aimed to enhance the face, sky and saliency areas differently. The faces and sky in all the photos were shifted to a pre-defined exposure or color. The pre-defined face exposure was from the analysis of a set of well-exposed faces, and the pre-defined sky color was obtained from examining a set of professional manually-enhanced photographs. This kind of method that was based on average values from statistics could generate improved photographs simply and fast. However, without considering the content of the original photograph, this method may adjust the photograph too far away from its original appearance and reduce the diversity of scenes. In addition, the method in [KLW12] only pro-

cesses regions separately and does not consider their relative relationships. Apart from automated methods, interactive local contrast adjustment methods [LFUS06, DGV09] can apply change to interest regions or objects through the drawing of brush strokes. These interactive adjustments need time and skill to master. Users need to make decisions based on the relative relationships of selected regions.

To circumvent the difficulty in defining the enhancement level for different images in single image manipulation methods, learning-based methods have been proposed to enhance images so that they match specific styles. Bychkovsky et al. [BPCD11] proposed to learn the global tonal adjustment function from photographers to personalize the tone adjustment. The global tone adjustment function was learned from input-output image pairs, where the output images were those enhanced by professional photographers. Similarly, Wang et al. applied learned color and tone mapping to stylize the image enhancement [WYX11]. These methods mostly rely on lower-level image statistics (e.g. color, intensity gradient) and do not exploit object-level semantics other than the face in [BPCD11]. Though the face is specifically considered in [BPCD11] for learning the tone mapping, only a global tone mapping function is learned.

Dale et al. [DJS*09] proposed to explore three global restorations: white balance, exposure correction, and contrast enhancement, based on a set of "semantically" similar images selected from a database. Though color transfer was performed between corresponding segmented regions, only the global restoration was explored. Joshi et al. proposed to improve the quality of faces in a personal photo collection by leveraging better photos of the same person [JMAK10]. This work focused on performing both global and face-specific corrections. Kang et al. [KKL10] described such an exemplar-based method where the enhancement parameter associated with the exemplar image was used to enhance the input image. The exemplar image was selected based on a learned metric. This work only performed global tone contrast adjustment and color correction. Furthermore, Hwang et al. [HKK12] proposed to locally correct the color and tone of the image by searching for the best transformation for each pixel. The transformation function, which was derived from the best few candidate image pairs on the matched pixel, was applied to the input pixel. These exemplar-based methods need the enhancement parameters for the exemplar image pair to be known as *a priori*. For pre-modern painting styles, it is not possible to know the enhancement parameters, therefore the method is not applicable.

Reinhard et al. proposed a color transfer method between images [RAGS01]. It was conducted by transferring the mean and variance of the value distribution in $l\alpha\beta$ color space from the example image to the target image. This color transfer method was used for local color transfer by performing color transfer between corresponding segmented regions or swatches [TJT07, WAM02]. Reinhard's method is a linear

stretching process. It has the disadvantage that it may extend the value out of range and over-compress or over-exaggerate the visual contrast in scene details [XM09].

Differently, Zhang et al. developed a method to enhance landscape photographs by adjusting the contrast organization inter- and intra- depth planes based on an example landscape painting [ZCC14]. With a similar consideration, we conduct contrast adjustment separately on FO, FS and BG regions of a portrait photograph based on an example portrait painting.

3. Paintings vs Photographs

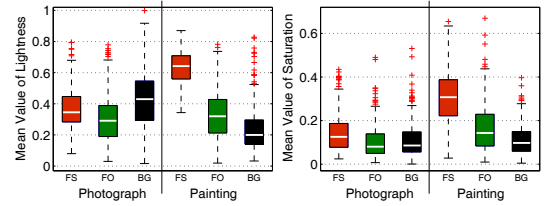
In this section, a statistical comparison of the inter-contrast and intra-contrast values of lightness and saturation is performed between a set of photographs and a set of paintings. In addition, the mean values of lightness and saturation of regions are also compared to show the difference of their distribution in different regions. Hue is not touched so that the natural appearance of the photograph will be preserved.

Portrait paintings used for this statistical study are collected from artists of the 17 – 18th century such as: Thomas Gainsborough, Francisco Goya, Edouard Manet, and Jean-Auguste-Dominique Ingres. These artists usually maintained a high dynamic range in their paintings which was similar to that of professional photographs. In total, 300 portrait paintings are collected to form the dataset. 180 of them are half body portraits and 120 of them are full body portraits. 250 portrait photographs from the MIT-Adobe FiveK Dataset [BPCD11] and 50 from our personal collection are collected to form the photograph dataset.

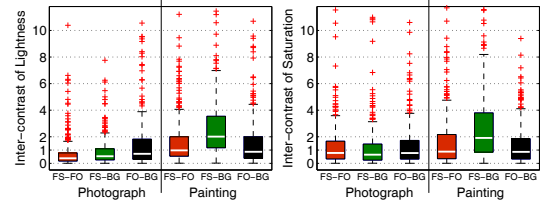
Based on Weber’s law, the intra-contrast of one region is defined as $\frac{I_{max} - I_{min}}{I_{mean}}$, where I_{max} and I_{min} are the maxima and minima values of the region respectively, and I_{mean} is the mean value of the region. The inter-contrast between two regions with mean values M_1, M_2 is defined as $\frac{|M_1 - M_2|}{\min(M_1, M_2)}$. This definition is similar to the Weber contrast definition [SW89], where the region with the lowest value is considered as the BG. The mean and contrast values are calculated in LCH color space. The L channel best matches the human perception of the lightness of colors and the Chroma is indicative of saturation.

The boxplots of the mean values and contrast values are shown in Figure 4. The distribution of mean values shows that the FS region is generally the brightest part of the painting and much more colorful than the FO and BG regions. This is coherent with the objective of portraiture which is to focus on the expression of the figure, especially on the face. However, we do not observe this tendency in the lightness of photographs. Although the saturation of photographs has this tendency, the difference is not as clear as that of paintings. The inter-contrast represents the relative difference of mean values. Figure 4(b) clearly shows that the inter-contrast values of lightness and saturation between FS and FO, FS

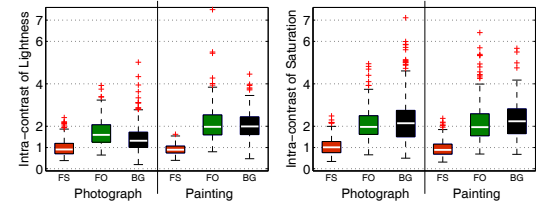
and BG in paintings are much larger than those of the photographs.



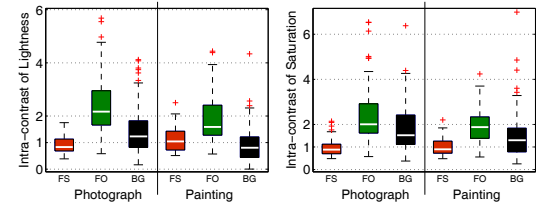
(a) Mean values



(b) Inter-contrasts



(c) Intra-contrasts (complex BG)



(d) Intra-contrasts (simple BG)

Figure 4: (a) and (b) Boxplots of mean values and inter-contrasts of 300 paintings and 300 photographs. The central line in the box is the median of the distribution, and the edges of the box are the 25th and 75th percentiles. The two vertical lines extending from the central box indicate the remaining data outside the central box, which extend maximally to 1.5 times the height of the central box. The red crosses are the remaining data (outliers). (c) Boxplots of intra-contrasts of paintings and photographs with a complex BG. (d) Boxplots of intra-contrasts of paintings and photographs with a simple BG.

For the intra-contrast comparison, paintings and photographs are classified into two classes based on the complexity of the BG. One class has a complex BG (intra-contrast is shown in Figure 4(c)), the other class has a sim-

ple BG (intra-contrast is shown in Figure 4(d)). In paintings with a complex BG, the intra-contrast values of lightness in FO and BG are more likely to be larger than those of the photographs. However for the intra-contrast of saturation, we cannot observe clear difference between paintings and photographs. In both paintings and photographs with a simple BG, the intra-contrast of the FO is larger than those of the BG and FS regions. However, the intra-contrast values of the FO and BG in paintings are smaller than those of photographs. One more interesting point is that the intra-contrast values of the lightness and saturation of the FO region in photographs with a simple BG are larger than those of photographs with a complex BG. The reason is that the figure that has been set against a simple BG is more focused while capturing the photograph. These differences in the intra-contrast organization between paintings and photographs show that artists exaggerate the contrast organization of regions in paintings. The intra-contrast of a complex BG is exaggerated to be larger, and the intra-contrast of a simple BG is reduced to be smaller.

In summary, the regional contrasts of paintings are more purposely organized than those of photographs. Larger inter-contrast is used by painters to attract the focus of attention on the face or the figure in portrait paintings. The intra-contrast values of regions are also specially organized by painters to serve the visual appeal of the painting.

4. Overview

Statistics in the preceding section has shown the purposely created lightness and saturation inter-contrast and intra-contrast organization in paintings. This section attempts to use this contrast organization as reference to enhance photographs. The framework of our exemplar-based portrait photograph enhancement is shown in Figure 5.

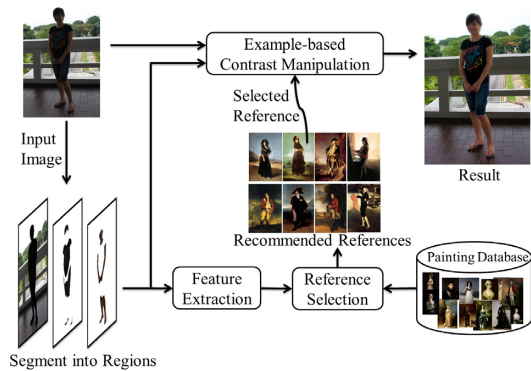


Figure 5: Framework of the proposed approach.

The proposed exemplar-based approach starts with the

segmentation of FG and BG regions plus face/skin area detection. Although a lot of research has been done on human detection, it is still a challenge to accurately extract the human shape from 2D images. In this paper, we propose to use GrabCut [RKB04] to segment FG and BG regions with an initialization window. The technique in [FMJZ08] is used to define the initialization window. Instead of using the upper-body detector in [FMJZ08], the upper-body region is estimated based on the location and size of the detected face region. The Haar detector of OpenCV is used to detect the face. Given the location of the left top corner (x, y) , the width W , and height H of the face window, an upper-body window is defined as $R = F + A \circ B$, where $F = [x, y, W, H]$, $B = [W, H, W, H]$, A is the scale vector to enlarge the face window. In our implementation, $A = [-1.8, -0.5, 3.6, 3.5]$ is used for faces around the vertical center area, $A = [-2.4, -0.5, 4.1, 3.5]$ is used for faces located in the right side of the photograph, while $A = [-1.7, -0.5, 4.1, 3.5]$ is used for faces located in the left side. For an image with a complex BG, the segmented FG may not be so accurate. As in [RKB04], a user interface is provided for users to draw some strokes to refine the results. The face mask and skin areas are detected based on the skin color in the figure using the method in [Fai10] with our morphological post-processing. Two results are shown in Figure 6.

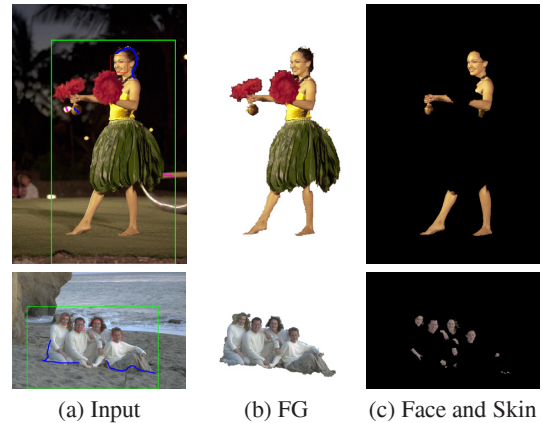


Figure 6: FG/BG segmentation and face/skin area detection. The red window in the input photograph is the face window, the green window is the initialization window for Grabcut. The blue and purple lines are the FG and BG brushes for refining the FG/BG segmentation.

Given the segmented BG region, detected face and skin areas, and the FO region, the portrait painting database is then searched for the best matching reference paintings. The similarity of paintings and photographs is judged based on their original contrast values.

Then, using a user-selected painting among the top ranked reference paintings as an exemplar, the regional contrasts in the portrait photograph are enhanced. The operation on

lightness and saturation is performed separately in the same way. Finally, the transition boundary between regions is smoothed.

5. Reference Selection

The main objective of reference selection is to select reference paintings that can produce more pleasing results as compared to the original photograph, while not pushing the photograph too far away from its original natural property. Because our portrait enhancement approach enhances relative relationships of regions, therefore reference paintings should be selected based on these relative relationships. The relationships among regions can be modeled as a graph as shown in Figure 7. The features within regions $\Psi = \{\Psi_S, \Psi_O, \Psi_B\}$ and the relationships among regions $\Phi = \{\Phi_{SO}, \Phi_{SB}, \Phi_{OB}\}$ are used to select reference paintings, where Ψ_S , Ψ_O , and Ψ_B are the features within FS, FO and BG respectively, and Φ_{SO} , Φ_{SB} , and Φ_{OB} are the relationship features between FS and FO, FS and BG, FO and BG respectively. After calculating the graph $V = (\Phi, \Psi)$ for each image, the task is to determine the similarity of the graphs between input photographs and paintings. Particularly, paintings that can produce more pleasing results to the original input photograph should be ranked higher in the selection process in terms of graph similarity according to some distance metric. Therefore, distance metric learning is proposed to determine the similarity of paintings and input photographs based on the graph model.

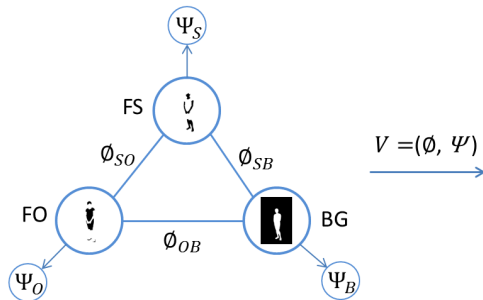


Figure 7: The graph model to show the relationships of regions.

The features within each region are mean values of lightness and saturation, 10-bins histograms of lightness and saturation, maxima (95th percentile value), minima (5th percentile value) values of lightness and saturation distributions, and intra-contrast values of lightness and saturation. The BG global contrast factor G_B is also calculated as a feature to describe the complexity of the BG. A highly detailed and variation-rich BG will have a high global contrast factor and a simple BG will have a low global contrast factor [MNNea05]. The global contrast factor is calculated as the weighted average of local contrasts at various resolution

levels (more details can be found in [MNNea05]). The relationships among regions Φ are defined by the inter-contrast values of lightness and saturation between FS and FO, FS and BG, and FO and BG.

Given the features, the distance metric between the input photograph I_i and painting R_j is

$$D(i, j) = \sum_{k=1}^N \alpha_k (v_i^k - v_j^k)^2 \quad (1)$$

where α is the parameter to linearly combine the distance of features. v_i^k is the k^{th} feature in the feature vector formed by concatenating the ordered graph attributes of photograph I_i , and v_j^k is the k^{th} feature in the feature vector, formed in the same way, of painting R_j . The objective here is to determine the parameter α such that paintings which are more likely to produce more pleasing results are closer to the input photograph than others. As in [KKL10, HKK12], we learn the parameter α by using a target distance function $D_t(i, j)$. The parameter α is determined by minimizing the objective function

$$\arg \min_{\alpha} \sum_{i, j} \|D(i, j) - D_t(i, j)\|^2 \quad (2)$$

The target distance $D_t(i, j)$ is defined based on the assigned score by a human expert for an enhanced photograph from a photograph-painting pair (i, j) . Smaller distance signifies a more pleasingly enhanced photograph and larger distance signifies an unacceptably enhanced photograph. Minimizing this objective function returns an appropriate distance function that reflects how far the photograph and painting should be in terms of their enhancement results. This objective function is convex and the unique optimum can be easily found by running a gradient descent procedure.

To learn the distance metric, the collected 300 portrait paintings (180 half body portraits and 120 full body portraits) are used as references to generate enhancement results for each training photograph. The target distance function is $D_t(i, j) = \frac{1}{S(i, j)}$, where $S(i, j)$ is the assigned score by an expert for the enhanced photograph I_i by using the painting R_j as reference. The expert has lot of experience on photo adjustment. It is a challenge to score an image directly for its aesthetic value due to subjective preferences. However, it might be more consistent for classifying enhancement results to be good, acceptable or unacceptable. Therefore, instead of directly giving a score, the expert is asked to classify the enhancement results for 32 half body portrait photographs and 30 full body portrait photographs to three categories: good, acceptable and unacceptable with pre-set score values. The highest score is given to results judged as good, the lowest score is given to results judged as unacceptable, and the score for results judged as acceptable is in the middle. In our implementation, the assigned values for $S(i, j)$ are 1, 3, and 5. For each half body portrait photograph, 180 enhancement results are obtained while 120 enhancement results are obtained for each full body portrait photograph.

The expert only classifies the ones that can be confidently assigned to one of the three categories, otherwise they are discarded. Finally, we collected 2776 samples to learn the distance metric for half body portraits and 1381 samples for full body portraits.

6. Contrast Mapping

Based on the definition of intra-contrast, adjusting the value spread range and mean can change the intra-contrast. Meanwhile, adjusting the mean value of each region changes the inter-contrast of regions. More formally this means that we could adjust the mean and spread range of values in regions of photographs with reference to those of the reference paintings to map the contrast to that of the latter. Contrast stretching is a commonly used technique to change the spread range to enhance contrast while avoiding artifacts. The spread range can be expressed as *maxima* - *minima*. Therefore, we can map the maxima and minima values instead of mapping the spread range directly.

Given the target mean value m_s^t , maxima I_{max}^t , minima I_{min}^t , the task of the contrast mapping is

$$I(m_s, I_{max}, I_{min}) \xrightarrow{f} I'(m_s^t, I_{max}^t, I_{min}^t) \quad (3)$$

Piecewise linear contrast stretching has been used in [ZCC12] to adjust the mean, maxima and minima values of regions. This piecewise linear transformation is not continuously differentiable across a histogram span, which may cause contouring artifacts [DGV09]. Moreover, this method may produce out of range mapping. To more effectively perform contrast mapping, an interpolating transformation curve is piecewisely defined, and it is monotonically increasing and continuously differentiable. This interpolating transformation curve is known as histogram warping in [GD04, DGV09]. For a given set of control points (a_k, b_k, d_k) , where $b_k = f(a_k)$, $d_k = f'(a_k)$ which is the contrast adjustment at the key value a_k , the transformation curve is generated using a piecewise rational quadratic interpolating spline [SAMA97]:

$$f(x) = b_{k-1} + \frac{(r_k t^2 + d_{k-1}(1-t)t)(b_k - b_{k-1})}{r_k + (d_k + d_{k-1} - 2r_k)(1-t)t} \quad (4)$$

where $r_k = \frac{b_k - b_{k-1}}{a_k - a_{k-1}}$ and for $x \in [a_{k-1}, a_k]$, $t = \frac{x - a_{k-1}}{a_k - a_{k-1}}$.

In [DGV09], the control points are specified by users while the control points are generated automatically based on the analysis of the histogram in [GD04]. Differently, in our application, we have three control points which are automatically defined based on the original and target mean, minima and maxima values. Specifically, the three control points are $(a_1, b_1, d_1) = (I_{min}, I_{min}^t, d_1)$, $(a_2, b_2, d_2) = (m_s, m_s^t, d_2)$ and $(a_3, b_3, d_3) = (I_{max}, I_{max}^t, d_3)$, and two endpoints are $(0, 0, d_0)$ and $(1, 1, d_4)$. The definition of the contrast adjustment at each point controls the shape of the transformation

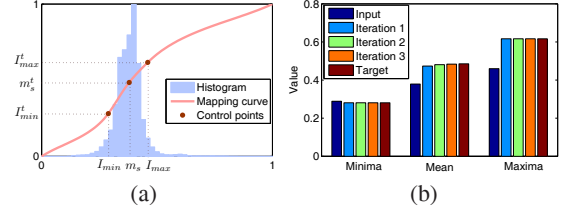


Figure 8: Contrast mapping. (a) Transformation curve, (b) Convergence of maxima, minima and mean values.

curve. To fit a piecewise linear curve, as the work in [GD04], the contrast adjustments at the control points are defined as the geometric mean of the slopes weighted by the probability mass of the slopes' intervals. Given $d_0 = \frac{b_1}{a_1}$, $d_4 = \frac{1-b_3}{1-a_3}$, the contrast adjustments d_k , $k=1, 2, 3$ are

$$d_k = \left(\frac{b_k - b_{k-1}}{a_k - a_{k-1}} \right)^{g_k^l} \left(\frac{b_{k+1} - b_k}{a_{k+1} - a_k} \right)^{g_k^h} \quad (5)$$

where $g_k^l = \frac{F(a_k) - F(a_{k-1})}{F(a_{k+1}) - F(a_{k-1})}$ and $g_k^h = \frac{F(a_{k+1}) - F(a_k)}{F(a_{k+1}) - F(a_{k-1})}$ are the slope weights. $F(x)$ is the cumulative distribution function. An example of the transformation curve is shown in Figure 8(a).

Good contrast within an image is not simply a case of the higher being the better. Artists are skilled at controlling the perceptually high contrast values of their paintings. However, giving high computational contrast to the photographs cannot ensure a good perceptual look. Therefore, the target inter- and intra-contrasts may not exactly move to those of the reference. This means that the reference painting provides the template by which the photograph will be changed, but how close the match is to this template depends on the original image and also the user's preference. Thus, two parameters α , β are used to control the inter-contrast and intra-contrast mapping. Since the inter-contrast is the relative difference of mean values, and the mean lightness and saturation values also influence the appearance of the image, the inter-contrast mapping is then performed by moving the mean value of each region to somewhere close to that of the reference. Given the mean value m_s of one region I of the photograph and the mean value m_r of the corresponding region I_r of the reference, the mean value is shifted towards to m_r and the magnitude of shift is controlled by a weight α in $[0, 1]$.

$$m_s^t = m_s + \alpha(m_r - m_s) \quad (6)$$

This adjustment on the mean value implies that the inter-contrast of the input is moved close to that of the reference controlled by α . Given the intra-contrast values of I and I_r as c_s , c_r respectively, the target intra-contrast is

$$c_s^t = c_s + \beta(c_r - c_s) \quad (7)$$

where β is a weight in $[0, 1]$. Based on the definition of intra-

contrast, the degree of enhancement on the spread range will be $\lambda = \frac{c_r^t m_s^t}{c_r m_s}$. Then we have the target maxima and minima values as

$$I_{max}^t = \lambda(I_{max}^r - m_r) + m_s^t \quad (8)$$

$$I_{min}^t = \lambda(I_{min}^r - m_r) + m_s^t \quad (9)$$

where I_{min}^r and I_{max}^r are the minima and maxima values of I_r . The target maxima and minima values are limited in the range [0.01, 0.99]. After one iteration of the process, the mean value has a shift from the target mean value as the bars show in Figure 8(b). By using three iterations, the mean, maxima and minima usually converged to the target values.

Finally, to smooth the boundary between regions and keep the change spatially coherent, edge-preserving smoothing [FFLS08] is performed on the difference image between the input F (L or C channel in LCH space) and contrast mapping result \tilde{F} , $d = \tilde{F} - F$. For the output difference image D , the energy function is defined as follows.

$$E = \sum_{p \in D} \{[D(p) - d(p)]^2 + w_s(p)h(\nabla D, \nabla F)\} \quad (10)$$

$$h(\nabla D, \nabla F) = \frac{|D_x(p)|^2}{|g_x(p)|^\theta + \epsilon} + \frac{|D_y(p)|^2}{|g_y(p)|^\theta + \epsilon} \quad (11)$$

The smoothing is to make the gradients of the output D as small as possible, unless the input F has significant gradient. (D_x, D_y) is the gradient of D , (g_x, g_y) is the gradient of F . θ controls the sensitivity to the gradient of F . ϵ is a small regularizing constant. The weight w_s is

$$w_s(p) = \begin{cases} \tau_1 & \text{if } p \in \text{boundary area} \\ \tau_2 & \text{otherwise} \end{cases} \quad (12)$$

τ_1 is larger than τ_2 for smoothing the boundary area between regions. In our implementation, we use the parameters $\theta=1.2$, $\epsilon=0.0001$, $\tau_1=0.3$ and $\tau_2=0.05$. The minimization of the energy cost function E can be achieved using standard or weighted least-squares techniques like the conjugate-gradient method. The result after smoothing is

$$F^* = F + D \quad (13)$$

7. Experiments and Discussion

In this section, we will analyze the performance of the proposed portrait photograph enhancement approach. The algorithm was implemented in MATLAB on a PC with an Intel 2.93GHz processor and 3GB RAM. Among the 300 photographs, 62 of them are used for the distance metric training in the reference selection, the others are used as test photographs.

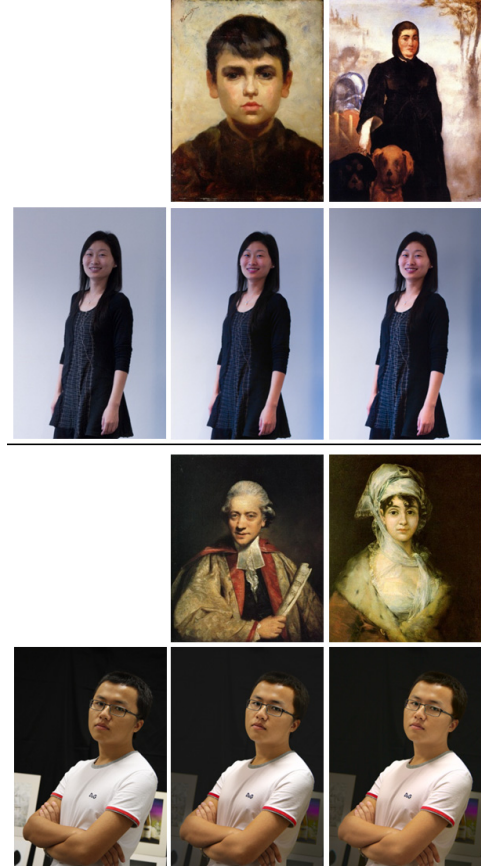


Figure 10: Experiments on two indoor photographs with simple BGs. The first column: original photographs. The second and third columns: two of the selected reference paintings for the two photographs and the corresponding results. The parameters are $\alpha_l = 0.5$, $\alpha_s = 0.4$, $\beta_l = \beta_s = 0.5$.

7.1. Influence of reference

The enhancement results of two photographs using 5 selected top ranked reference paintings are shown in Figure 9. One photograph is full body portrait and the other is half body portrait. The two photographs both have a complex outdoor BG. The selected reference paintings for the two photographs also have a complex BG. Results obtained using reference paintings selected by the reference selection method are encouraging (the second to sixth columns of Figure 9). More details of the figure, especially around the face area, are visible and the figure stands out in the enhancement results. However, if a painting not selected by the reference selection method is used as the reference (as the paintings in the seventh column of Figure 9), the face and figure also stand out in the enhancement result. However, the change is too far away from its original natural property (e.g. the pink T-shirt becomes black). This demonstrates the importance

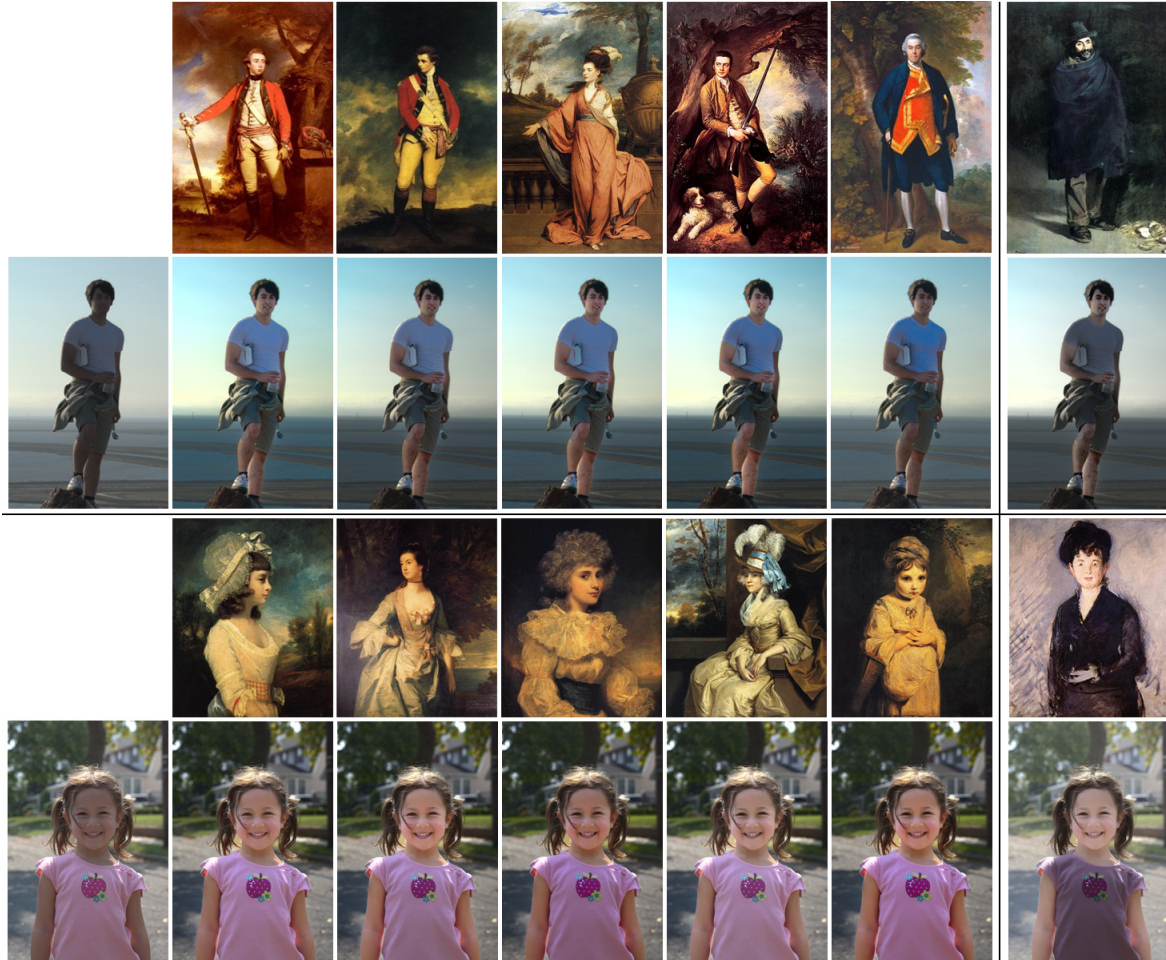


Figure 9: Experiments on two outdoor photographs with complex BGs. The first column: original photographs. The second to six columns: 5 of the selected reference paintings for the two photographs and the corresponding results. The seventh column: paintings not recommended as reference and the corresponding results. The parameters for the first example $\alpha_l = 0.5$, $\alpha_s = 0.4$, $\beta_l = \beta_s = 0.5$. The parameters for the second example $\alpha_l = 0.6$, $\alpha_s = 0.6$, $\beta_l = \beta_s = 0.5$.

of the reference selection for the portrait enhancement and the effectiveness of the proposed reference selection method. While the two example photographs in Figure 9 are of outdoor scenes with complex BGs, Figure 10 shows two indoor photographs with simple BGs. The first one is with bright BG and dark FG, and the other is with dark BG and bright FG. The selected reference paintings have similar natural properties with the input photograph. In other words, a photograph with a dark FG and bright BG is more likely to be improved using a reference painting with similar dark FG and bright BG.

7.2. Influence of parameters

In the contrast mapping, we use the same α , β for the three regions FO, FS, and BG. Parameters for mapping lightness

are expressed as α_l , β_l , and for saturation are α_s , β_s . When $\alpha_l = \alpha_s$, $\beta_l = \beta_s$, they are indicated as α , β . α , β control the degree of match to the contrasts of the reference. Figure 11 shows the effect on the appearance of the enhancement results for changing α_l , β_l from 0 to 1. When α_l increases from 0 to 1, the inter-contrast of lightness is moved towards to that of the reference (see the comparison of inter-contrast in Figure 11(c)). The figure stands out more brightly as α_l increases in Figure 11. When β_l changes from 0 to 1, the intra-contrast values of lightness of the three regions are moved to those of the reference (see the comparison of intra-contrast in Figure 11(d)). More local details in the BG become visible when β_l increases and the contrasts on the face and figure are also enhanced. The influences of α_s , β_s are similar to those of α_l , β_l .

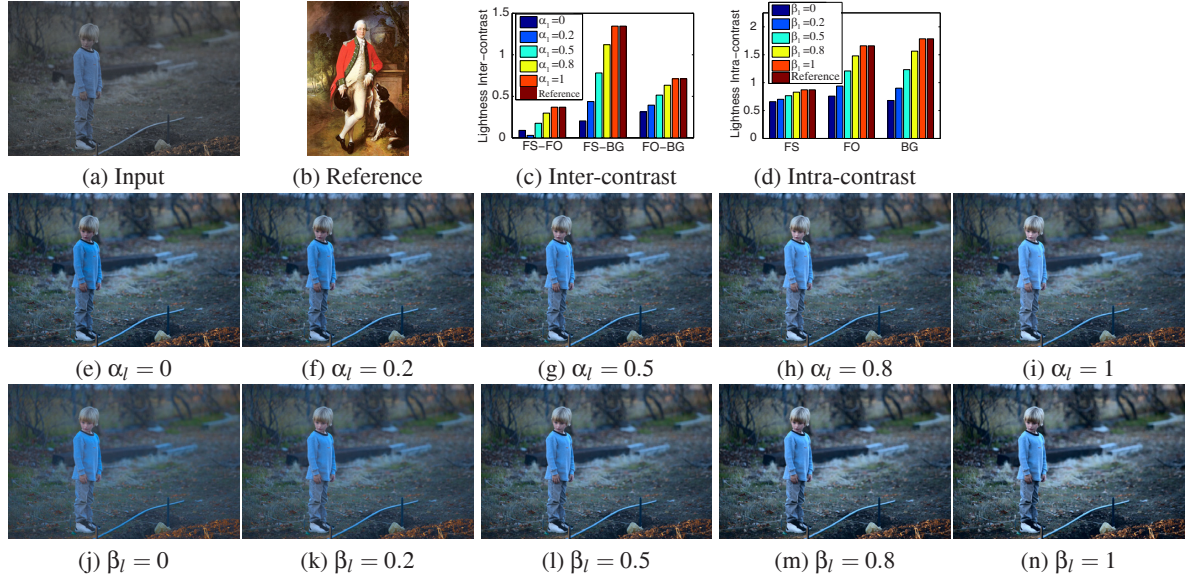


Figure 11: Influence of α, β . (e)-(i): $\alpha_s = 0.4, \beta_l = \beta_s = 0.5$. (j)-(n): $\alpha_l = 0.5, \alpha_s = 0.4, \beta_s = 0.5$.

In the reference selection, the enhanced photographs used for distance metric learning are obtained by using the default parameters $\alpha_l = 0.5, \alpha_s = 0.4, \beta_l = \beta_s = 0.5$. Therefore, given the selected reference paintings using the proposed reference selection method, acceptable results can be obtained using these default parameters. The user can also adjust the parameters around the default values to change the appearance according to the personal preference. Generally α_l is in the range [0.3, 0.7], and α_s is in [0.2, 0.6]. β_l and β_s are suggested to be set in a similar way, and generally they are in the range [0.3, 1].

7.3. Comparison with related methods

To assess the performance of our proposed portrait photograph enhancement method, it is compared to the content-aware enhancement (CAE) method in [KLW12] which has been reported to perform better than the global learning-based enhancement method in [BPCD11] and recent popular software tools (“I’m Feeling Lucky” of Google’s Picasa, “Auto Correct” of Microsoft’s Office Picture Manager, and “Auto Smart Fix” of Photoshop). In addition, our results are also compared with results enhanced by using the average mean and contrast values (AVE method). The average mean and contrast values are obtained from the study on contrast organization in paintings in Section 3. Figure 12 shows the enhancement results by using the three methods on 5 different scenes. We can clearly see that faces and figures in our results become more salient as compared to those in the results obtained by the CAE method [KLW12]. Moreover, the light falling on the face in our results is more naturally presented. However, in the results on the second and third

rows obtained by the CAE method, the light falling on the face becomes unnatural because the lightness has been processed differently on the left side of the face, as compared to the right. The enhancement results using the AVE method in the first to third rows are comparable with the proposed exemplar-based method. However for the photographs in the fourth and fifth rows, the proposed exemplar-based method performs better than using the average values. The exemplar-based method can handle the diversity of scenes better.

7.4. User studies

In order to significantly evaluate the effectiveness and advantage of the proposed approach, two user studies were conducted. The first user study was to evaluate the effectiveness of our proposed reference selection method and the effect of the proposed portrait photograph enhancement method. The second user study was to compare the effect of the proposed portrait photograph enhancement method with other methods.

For the first user study, we randomly selected 40 of the test photographs from a variety of scenes. This user study had two objectives. The first was to show that results obtained using the proposed approach become more pleasing and the figure and face become more salient. The other was to show that results obtained using the top ranked reference paintings were more desirable than those produced using other paintings. For the first objective, 5 reference paintings were randomly selected from the top ranked 15 paintings for each of the 40 selected test photographs. Results obtained using the 5 reference paintings were all compared with the original photograph. Each result and its corresponding original

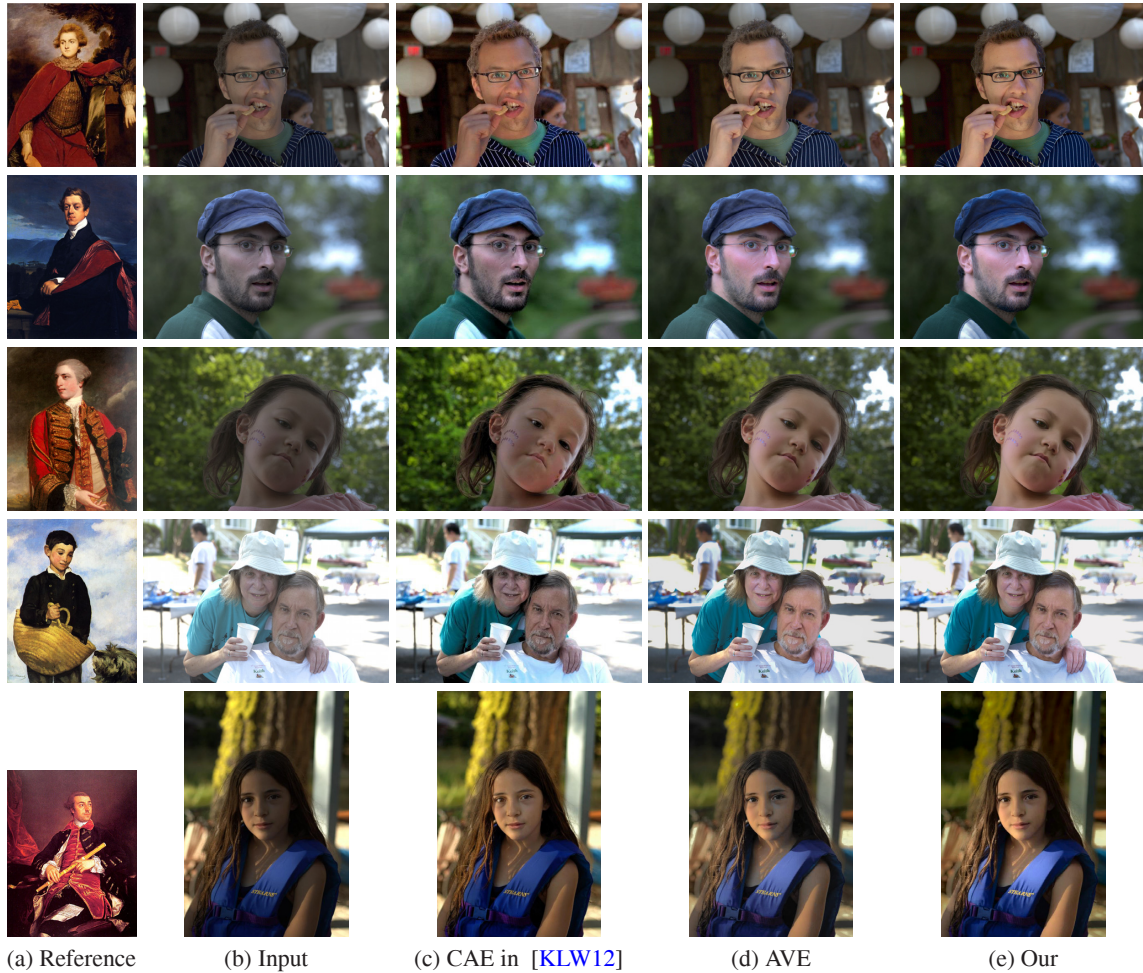


Figure 12: Comparison with other methods. The photograph in the fifth row is from [KLW12].

photograph were shown as a pair side by side. The left/right ordering of the images in each pair was randomly generated. For each image pair, participants were asked to answer two questions. The first question Q1 was “In which image, are the figure and face more salient?” and the second Q2 was “Which is more pleasing?”. The participants could choose “left” or “right” as a response to each question according to their first impression. No further information on the goal of the experiment or the origin of the images was provided to the participants.

For the second objective, results using reference paintings selected from the middle and bottom ranges of the ranking were compared with those obtained using the top ranked reference paintings. 5 paintings from the ranks in the range of 51-65 for full body portrait and 5 paintings from the ranks in the range of 65-81 for half body portrait were randomly selected as the middle ranked reference paintings. 5 paintings from the ranks in the range of 106-120 for full body

portrait and 5 paintings in the ranks in 166-180 for half body portrait were randomly selected as the bottom ranked reference paintings. The three groups of results obtained using the top, middle and bottom ranked reference paintings were arranged in three rows. The order of the group arrangement was randomly generated. Participants were asked to select the group that they prefer the most. The participant could select “Group 1”, “Group 2” or “Group 3”.

The 40 photographs were divided into 7 sets. The first set had 4 photographs while the other sets had 6. The comparison in each set contains the comparison of the result obtained using the top ranked reference painting with the original side by side and the comparison of the three groups of results obtained using the top, middle and bottom ranked reference paintings. In the first set, there were 24 comparisons and in the other 6 sets, there were 36 comparisons. We used Google Form to create the webpage for the user study. The participants connected to the webpage using their own computers

and monitors. Each participant could select to complete one, two or three sets of comparisons. There were 38 participants. 21 were males and 17 were females. The age ranged from 20 to 45. In total, 2070 samples for the comparison of enhancement result vs original were collected. The average selection rate of each enhancement result vs original is shown in Figure 13(a). The bar graph shows that a significant majority of the responses (more than 90%) from the participants are in favor of our results in both the two attributes (pleasing and salient). This demonstrates that, by using our proposed approach, the visual appeal of the portrait photograph is effectively improved and the face and the figure become more salient. In the comparison among the three groups of enhancement results obtained by using top, middle and bottom ranked reference paintings, more than 80% of the responses are in favor of the results obtained using the top ranked reference paintings (see Figure 13(b)). This significantly shows the effectiveness of the proposed reference selection method and the importance of reference selection.

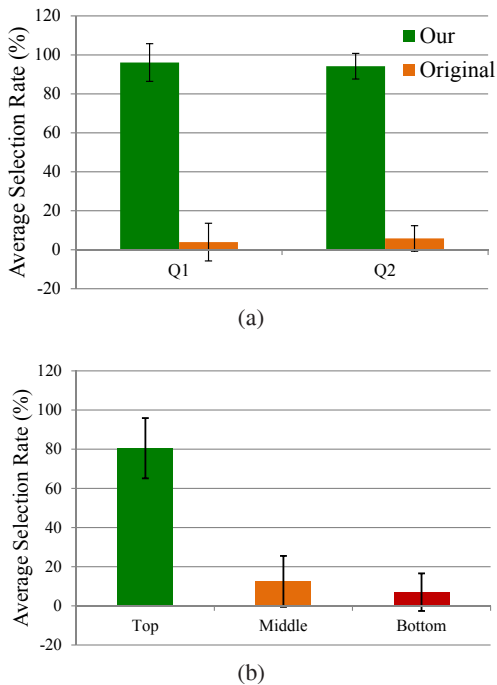


Figure 13: User study result for the evaluation of the proposed approach. (a) Selection rate for the comparison between our results and originals. (b) The comparison of the selection rates among the three groups of results obtained by using top, middle and bottom ranked reference paintings.

The second user study was conducted to compare our results with those obtained by the CAE method in [KLW12] and by the AVE method. The CAE method in [KLW12] provided about 55 results for images with faces. 36 of them which could be considered as portraits were used for this user study. The results obtained by our proposed method and

Table 1: User study result for the comparison of the proposed approach with the CAE method in [KLW12] and with the AVE method.

Our vs ACE	Questions	Our	ACE	Same
	Q1	62.0%	30.6%	7.4%
	Q2	66.8%	29.6%	3.6%
Our vs Ave	Questions	Our	Ave	Same
	Q1	68.4%	19.6%	12.0%
	Q2	70.0%	22.1%	7.9%

one of other methods were shown as a pair side by side. The left/right ordering of the images in each pair was randomly generated. For each image pair, the participants were asked to answer the same two questions as in the first user study. Differently, the participants could choose “Left”, “Right”, or “They look the same.” as a response to each question according to their first impression. The 36 images were divided into 2 sets. Each participant could select to complete one or two sets. There were a total of 29 participants aged from 20 to 35. 18 of them were male and 11 of them were female. For each comparison, there were on average 18 responses. The result of the user study is shown in Table 1. It shows that our results are preferred by a significantly higher percentage of responses than those by other methods. This is not surprising, since the method in [KLW12] cannot handle the relationships between regions and the AVE method does not consider the diversity of scenes.

7.5. Limitations

While the experiments have demonstrated the effectiveness of the proposed approach, we observe a few limitations. The quality of the produced results relies on the success of FG/BG segmentation and face/skin area detection. Although interactive segmentation can reduce the segmentation errors and the skin area detection in the segmented FG also is much more robust than the detection in the whole image, the skin area detection based on color may still be disturbed by the color of clothes in some situations.

Another limitation is that artifacts near the transition boundary between the brightened FG region and the dark BG region may be visible. This is because the brightness of the FG region is extended to the neighboring BG region by the boundary smoothing if there is no clear boundary between FG and BG in the original photograph. The technique in [FFLS08] was designed to smooth an image while preserving edges with large gradients. In the photograph in Figure 14(a), there is no clear edge between the right lower arm and the ground (see the area in the red box in Figure 14(a)). After interactively segmenting the FG/BG and adjusting the contrast between regions, the right arm is brightened and a sharp boundary is produced between the right lower arm and ground. However, after edge preserving smoothing, the sharp boundary is smoothed. Therefore the brightness on the

lower arm is extended to the neighbouring BG region. The smoothing on the boundary region is stronger than that of the non-boundary region. Hence, artefacts may be visible. Some noise may also become visible in dark face and skin areas after getting brightened.

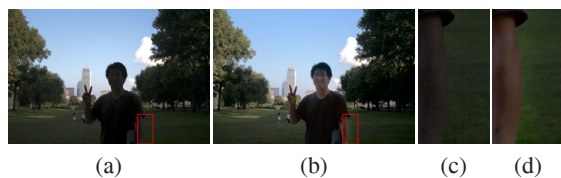


Figure 14: One example demonstrating the limitation of the proposed approach. (a) Original photograph, (b) Result by proposed approach, (c) Close-up of the red box region of (a), (d) Close-up of the red box region of (b).

8. Conclusion

This paper proposes a portrait photograph enhancement approach by manipulating the regional contrasts. This enhancement is guided by a selected portrait painting as an exemplar. The experimental results and user study significantly show that the figure and face become more salient and the photograph becomes more pleasing by applying the proposed approach. In future studies, we will study the lightness shading on the face features (eye, mouth, nose) and the shape of these features in portrait paintings for possible application on portrait photograph enhancement.

Acknowledgment

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

References

- [Arn74] ARNHEIM R.: *Art and visual perception: a psychology of the creative eye*, 2 ed. University of California Press, Berkeley and Los Angeles, CA, USA, 1974. 2
- [BPCD11] BYCHKOVSKY V., PARIS S., CHAN E., DURAND F.: Learning photographic global tonal adjustment with a database of input / output image pairs. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* (Colorado Springs, CO, USA, 2011), pp. 97–104. 2, 3, 4, 10
- [CJZW12] CHEN X., JIN X., ZHAO Q., WU H.: Artistic illumination transfer for portraits. In *Proc. of Eurographics Symposium on Rendering* (Paris, France, 2012), vol. 31, pp. 1425–1434. 3
- [CSHea10] CHEN J. S., SU G. D., HE J. P., ET AL.: Face image relighting using locally constrained global optimization. In *Proc. of the 11th European Conference on Computer Vision* (Crete, Greece, 2010), pp. 44–57. 3
- [CVP08] COLTON S., VALSTAR M. F., PANTIC M.: Emotionally aware automated portrait painting. In *Proc. of the 3rd international conference on Digital Interactive Media in Entertainment and Arts* (Athens, Greece, 2008), pp. 304–311. 3

- [CW08] CAO Y., WANG Z. X.: Automatic generation of portrait. In *Proc. of International Conference on Audio, Language and Image Processing* (Shanghai, China, 2008), pp. 735–738. 3
- [DGV09] DODGSON N. A., GRUNDLAND M., VOHRA R.: Contrast brushes: Interactive image enhancement by direct manipulation. In *Proc. of Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging* (British Columbia, Canada, 2009), pp. 107–114. 3, 7
- [DiP09] DiPAOLA S.: Exploring a parameterized portrait painting space. *International Journal of Art and Technology* 2, 1-2 (2009), 82–93. 2
- [DJS*09] DALE K., JOHNSON M. K., SUNKAVALLI K., MATUSIK W., PFISTER H.: Image restoration using online photo collections. In *Proc. of IEEE International Conference on Computer Vision* (Kyoto, Japan, 2009), pp. 2217–2224. 3
- [Fai10] FAIZAN: *face detection using skin color*. <http://www.mathworks.com/matlabcentral/fileexchange/26849-face-detection-using-skin-color>, 2010. 5
- [FFLS08] FARBMAN Z., FATTAL R., LISCHINSKI D., SZELISKI R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics* 27, 3 (2008). 8, 12
- [FMJZ08] FERRARI V., MARIN-JIMENEZ M., ZISSERMAN A.: Progressive search space reduction for human pose estimation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* (Anchorage, AK, USA, 2008), pp. 1–8. 5
- [GD04] GRUNDLAND M., DODGSON N. A.: Automatic contrast enhancement by histogram warping. *Computational Imaging and Vision* 32 (2004), 293–300. 7
- [HKK12] HWANG S. J., KAPOOR A., KANG S. B.: Context-based automatic local image enhancement. In *Proc. of European Conference on Computer Vision* (Firenze, Italy, 2012), vol. 1, pp. 569–582. 3, 6
- [JMAK10] JOSHI N., MATUSIK W., ADELSON E. H., KRIEGMAN D. J.: Personal photo enhancement using example images. *ACM Transactions on Graphics* 29, 2 (2010), 12:1–12:15. 3
- [JZCea10] JIN X., ZHAO M. T., CHEN X. W., ET AL.: Learning artistic lighting template from portrait photographs. In *Proc. of the 11th European Conference on Computer Vision* (Crete, Greece, 2010), pp. 101–114. 3
- [KKL10] KANG S., KAPOOR A., LISCHINSKI D.: Personalization of image enhancement. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* (San Francisco, CA, USA, 2010), pp. 1799–1806. 3, 6
- [KLW12] KAUFMAN L., LISCHINSKI D., WERMAN M.: Content-aware automatic photo enhancement. *Computer Graphics Forum* 31, 8 (2012), 2528–2540. 2, 3, 10, 11, 12
- [LFUS06] LISCHINSKI D., FARBMAN Z., UYTENDAELE M., SZELISKI R.: Interactive local adjustment of tonal values. *ACM Transactions on Graphics* 25 (2006), 646–653. 3
- [MNNea05] MATKOVIC K., NEUMANN L., NEUMANN A., ET AL.: Global contrast factor - a new approach to image contrast. In *Proc. of Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging* (Girona, Spain, 2005), pp. 159–168. 6
- [MZZ10] MENG M., ZHAO M. T., ZHU S. C.: Artistic paper-cut of human portraits. In *Proc. of ACM Multimedia 2010 International Conference* (Firenze, Italy, 2010), pp. 931–934. 3
- [Per06] PERKINS M.: *Professional Portrait Lighting: Techniques and Images from Master Photographers*. Amherst Media, Inc., Buffalo, NY, USA, 2006. 2

- [RAGS01] REINHARD E., ASHIKHMIN M., GOOCH B., SHIRLEY P.: Color transfer between images. *IEEE Computer Graphics and Applications* 21, 5 (2001), 34–41. [3](#)
- [Rez12] REZAEI M.: Artistic rendering of human portraits paying attention to facial features. In *Proc. of International Conference ArtsIT 2011, LNICST* (Esbjerg, Denmark, 2012), vol. 101, pp. 90–99. [3](#)
- [RKB04] ROTHER C., KOLMOGOROV V., BLAKE A.: "grab-cut" - interactive foreground extraction using iterated graph cuts. In *Proc. of ACM SIGGRAPH* (Los Angeles, CA, USA, 2004), vol. 23, pp. 309–314. [5](#)
- [RRC12] RIVERA A. R., RYU B., CHAE O.: Content-aware dark image enhancement through channel division. *IEEE Transactions on Image Processing* 21, 9 (2012), 3967–3980. [3](#)
- [SAMA97] SARFRAZ M., AL-MULHEM M., ASHRAF F.: Preserving monotonic shape of the data using piecewise rational cubic functions. *Computers & Graphics* 21, 1 (1997), 5–14. [7](#)
- [SW89] SPILLMANN L., WERNER J. S.: *Visual Perception: The Neurophysiological Foundations*. Academic Press, Inc, San Diego, CA, USA, 1989. [4](#)
- [TJT07] TAI Y.-W., JIA J., TANG C.-K.: Soft color segmentation and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 9 (2007), 1520–1537. [3](#)
- [WAM02] WELSH T., ASHIKHMIN M., MUELLER K.: Transferring color to greyscale images. In *Proc. of ACM SIGGRAPH* (New York, NY, USA, 2002), pp. 277–280. [3](#)
- [WYX11] WANG B., YU Y., XU Y.-Q.: Example-based image color and tone style enhancement. *ACM Transactions on Graphics* 30, 4 (2011). [3](#)
- [XM09] XIAO X., MA L.: Gradient-preserving color transfer. *Computer Graphics Forum* 28, 7 (2009), 1879–1886. [4](#)
- [YLLea10] YANG M., LIN S., LUO P., ET AL.: Semantics-driven portrait cartoon stylization. In *Proc. of IEEE 17th International Conference on Image Processing* (Hong Kong, 2010), pp. 1805–1808. [3](#)
- [ZCC12] ZHANG X., CONSTABLE M., CHAN K. L.: Example-based contrast enhancement for portrait photograph. In *Proc. of International Conference on Pattern Recognition* (Tsukuba, Japan, 2012), pp. 943–946. [7](#)
- [ZCC14] ZHANG X., CHAN K. L., CONSTABLE M.: Atmospheric perspective effect enhancement of landscape photographs through depth-aware contrast manipulation. *IEEE Transactions on Multimedia* 16, 3 (2014), 653–667. [4](#)
- [ZGZ09] ZHANG Y., GAO L., ZHANG S. Y.: Feature-based automatic portrait generation system. In *Proc. of World Congress on Computer Science and Information Engineering* (Los Angeles, CA, USA, 2009), pp. 6–10. [3](#)