Exemplar-based Portrait Photograph Enhancement as informed by Portrait Paintings

Xiaoyan Zhang$^{1,3}$ and Kap Luk Chan$^1$ and Martin Constable$^2$

$^1$School of Electrical & Electronic Engineering, NTU, Singapore
$^2$School of Art, Design and Media, NTU, Singapore
$^3$Institute for Media Innovation, NTU, Singapore

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 and 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 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 are effectively improved and the face and the figure become more salient.

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

1. Introduction

Figure 1: One result produced by our proposed approach. (a) Original photograph, (b) Reference painting, (d) Result.

Photo enhancement has received considerable interest with the prevalence of digital photography. For the influence of the quality and settings of the camera, the skill of the operator, and the external light condition, the captured photographs are not always so ‘desired’. Most existing approaches in image improvement enhance the appearance of images globally or locally in the same way for all image categories, no matter 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 different kinds of photos have different subject focus and different formal needs, therefore they 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 ends up shooting portraits. Portrait photos constitute an important part of our personal photo collections.

In a portrait, the figure, especially the face is predominant. However, the key to controlling the strength of the attraction in a portrait is not by just controlling the value of the center of interest, but rather it is the contrast among the face and skin areas of the foreground figure, the non-face/skin part of the foreground figure with accessories, and background to lead the eye of the viewer to concentrate on the expression of the person [Per06]. The contrast among all these regions makes the face stand out. Therefore, our
portrait photograph enhancement does not focus on the enhancement simply on the face, but on the relative relationship of the face and skin areas of the foreground figure, the non-face/skin part of the foreground figure with accessories, and background. A content-aware automatic photo enhancement method is proposed in [KLW12] to enhance the face, sky and saliency area differently. However it only processes these regions separately and does not consider the relative relationship of these regions. In addition, the exposure and color of the face and sky are adjusted to fixed values for all the images. As a result, the contrast on the face is sometimes reduced and also the relative contrast between the face and background is may be reduced. Although the Adobe® Photoshop® provides a powerful tool for users to manually edit their images, the required skills to achieve this region specific improvement are beyond the casual users. The users have to mask regions and make decisions based on the relative relationship of these regions. Additionally, “how the relative relationship of these regions should be defined 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 another group of portrait representations, supply us with good exemplars to modify the relative relationship of these regions. Compared with photography, which can be regarded as a projection of physical nature, traditional portrait painting is a very carefully constructed thing. 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 the color contrast values [Arn04]. Almost all portrait paintings are easily divisible into strong foreground (FG)/background (BG) zones (Figure 2). The FG is comprised of the figure with the accessories, and the remainder part is 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 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 and 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) 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’), the lighting, colorfulness and the regional contrasts are not well organized (e.g. the two photographs in Figure 3(c)(d)).

In order to better understand the regional contrast organization of the paintings, a statistical comparison on the contrast values of lightness and saturation is conducted between photographs and paintings. Hue is not touched in the paper to preserve the natural appearance of the photos. Differences between the paintings and the snap-shot photographs are observed with evidence that the lightness and saturation intra-contrasts and inter-contrasts of the paintings are more purposefully organized. This contrast organization in paintings concentrates on making the figure especially the face stand out in the image, while preserving the contrast within each region for the visual appeal of the whole image. This observation inspires us to enhance portrait photographs through contrast adjustments to their regions according to the 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.](image)

![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].](image)
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 example result is shown in Figure 1.

2. Related Works

**Portrait rendering.** With recent development of painterly rendering, portrait rendering has become an interesting topic of the computer vision and computer graphics research communities. Colton propose to use a Non Photorealistic Rendering (NPR) system to automatically produce portraits based on the recognized emotions [CVP08]. The artistic styles in the NPR system are based on the painting materials, color palette and brush model. Some recent works attempt to render a portrait photo to an artistic style [CW08, ZGZ09, MZZ10, YLLea10]. However, the styles are limited to simulate the abstraction, line drawing styles or organic models. Face relighting is another popular research on portrait. In [JZCe10], artistic face lighting templates are learned from a dataset of professional and amateur portrait photographs. The lighting template is based on the light distribution and shading of the face, e.g. left weighted or right weighted. Chen et al. develope an algorithm to relight the face based on one reference face image [CShea10]. The work in [CIZW12] applies the face lighting template generated by artist on faces of photographs. All the mentioned research works above about portrait only focus on the manipulation of face. Differently, our method not only considers the face, but also the relationship of the face with the BG and other parts.

**Image enhancement.** Based on the strategy, the methods on image enhancement can be classified into three categories: single image manipulation method, exemplar-based method and learning-based method. While most of the single image manipulation methods only automatically enhance the image globally or locally without consideration on the content, several content-aware methods are proposed to process regions differently. Rivera et al. propose to adjust the contrast differently in the dark, middle and bright regions [RRC12]. The contrast in each region is mapped separately and the mapping function is generated based on the analysis of the content. Finally the mapping results are merged using a weighting function. This context-aware method aims to reduce the artifacts and other unnatural effects in the resulting images. The considered regions are split based on the intensity value and have no high level semantic meaning. In addition, only lightness is adjusted. Differently, the content-aware automatic photo enhancement method proposed in [KLW12] aims to enhance the face, sky and saliency area differently. The faces and sky in all the photos are shifted to the fixed exposure and color. This process reduces the variation of scenes from different environments. In addition, the method only processes these regions separately and has no consideration on the relative relationship of these regions. In some cases, the contrast on the face and the contrast between the face and background are reduced. Apart from automatic methods, interactive local contrast adjustment methods [LFUS06, DGV09] can apply change to the interest regions or objects through the drawing of brush strokes. These interactive adjustments need time and skill to master. The users need to make decisions based on the relative relationship of selected regions.

For the difficulty in defining the enhancement level for different images in single image manipulation method, learning-based methods are proposed to enhance images to specific styles. Bychkovsky et al. [BPCD11] propose to learn the photographic global tonal adjustment function from photographers to personalize the tone adjustment. The global tone adjustment function is trained on input-output images pairs, where the output images are enhanced by professional photographers. Similarly, Wang et al. apply the learned color and tone mapping to stylize the image enhancement [WYX11]. These methods mostly rely on the lower-level image statistics (e.g. color, intensity gradient) and do not exploit object-level semantics except the face in [BPCD11]. Even the face is specifically considered in [BPCD11] in the tone mapping learning, but only global tone mapping function is learned.

Dale et al. [DJS’09] propose 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 is performed between corresponding regions that are segmented according to their visual context, only the global restoration is explored. Joshi et al. propose to improve the quality of faces in personal photo collection by leveraging better photos of the same person [JMAK10]. This work focuses on performing both global and face-specific corrections. Kang et al. [KKL10] describe such an exemplar-based method. A metric is first learned for enhancement parameter similarity, then it is used to find the image in the training set that is closest to the given test image in terms of enhancement parameters. The enhancement parameter associated with the “closest” image is then used for the input image. This work only performs the global tone contrast adjustment and color correction. Furthermore, Hwang et al. [HKK12] propose to locally correct the color and tone of the image by searching for the best transformation for each pixel. The search is done by finding the best candidate images and then finding the best matched pixels from these candidate images for each input pixel. The search is based on the local scene descriptors and context. Then, the mapping on the matched pixels is applied on the input pixel. These exemplar-based methods mentioned above need the priori enhancement parameters of the exemplar. For pre-modern styles, it is not possible to know the priori enhancement parameters, therefore the application is limited.
Differently, Zhang et al. propose to enhance the landscape photographs in the depth view by adjusting the contrast organization inter- and intra-depth planes [ZCC11]. This enhancement is based on an example landscape painting which is good at contrast organization in the depth view. Applying similar consideration, we conduct contrast adjustment separately on FO, FS and BG regions of the portrait photograph based on an example portrait painting.

3. Paintings vs Photographs

In this section, a statistical comparison on the inter-contrast values and intra-contrast values of lightness and saturation is conducted 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 clearly show the difference of value distribution in regions. Hue is not touched in the paper to preserve the natural appearance of the photos.

Portrait paintings collected for this statistics are chosen from artists in the 17-18 centuries such as: Thomas Gainsborough, Fransisco Goya, Edouard Manet, and Jean-Auguste-Dominique Ingres. These artists are all maintain a high dynamic range in their paintings which is similar with photographs. The database covers a variety of environments and relationships of FG and BG. In total, 300 portrait paintings are collected to constitute 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 the Weber’s law, the intra-contrast of one region is defined as $\frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{mean}}}$. The inter-contrast between two regions with mean values $M_1$, $M_2$ is defined as $\frac{|M_1 - M_2|}{\text{mean} M_1, M_2}$. This definition is similar to the weber contrast [SW89], where the region with low value is considered as the BG. The mean values, intra-contrasts and inter-contrasts of lightness and saturation 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.

Boxplot of the mean values and contrast values is in Figure 4. The mean values show that the FS region is generally the lightest part of the painting and much colorful than the FO and BG. This is coherent with the objective of portraiture which is to focus on the expression of the figure, especially the face. However, we can not observe the tendency in lightness of photographs. Although the saturation of photographs has this tendency, the difference is not so clear as that in paintings. The inter-contrast expresses the relative difference of mean values. From the Figure 4(b) we can clearly see that the lightness inter-contrasts between the FS and FO, FS and BG in paintings are much higher than those in photographs. The saturation inter-contrasts between FS and BG, FS and FO in paintings are also higher than those in photographs.

In the intra-contrast comparison, the paintings and photographs are classified into two classes based on the complexity of the BG. One class is with complex BG (intra-contrasts are in Figure 4(c)), the other class is with simple BG (intra-contrasts are in Figure 4(d)). In the paintings with complex BG, the lightness intra-contrasts are more likely higher in FO and BG than photographs. However in the inter-contrast of saturation, there is no clear difference in paintings and photographs. For the simple BG class, in both paintings and photographs, the intra-contrast in FO is bigger than BG. However, the intra-contrasts in FO and

![Image](https://example.com/image.png)
BG of paintings are smaller than those in photographs. One interesting point is that, the lightness and saturation intra-contrasts of FO in the photographs with simple BG are larger than those with complex BG. The reason is that the figure is more focused in the image with simple BG. These differences in intra-contrast organization in paintings and photographs show that artists exaggerate the contrast of regions in paintings. The intra-contrast is exaggerated to be bigger in complex BG, and exaggerated to be smaller in simple BG.

In summary, the regional contrasts in paintings are more purposely organized than those in photographs. Higher inter-contrasts are used by painters to concentrate the focus of attention on the face in portrait paintings. The intra-contrasts in regions are also specially organized by the painters to serve the requirement of the content for the visual appeal.

4. Overview

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

![Figure 5: Framework of the proposed approach.](image)

Our exemplar-based approach starts with the segmentation of the FG/BG and face, skin areas detection. Although a lot of works have been done on human detection, it is still a challenging work to extract the human with accurate shape in 2D images. Additionally, in half body portrait where only the face and part of the upper body are visible, the human detection methods can not work. In the paper, we propose to use the GrabCut segmentation [RKB04] to segment the FG/BG automatically based on an initialized window. One thing that is consistent in portrait is that, the front face is generally always facing the viewer. Therefore, we use the face as the important information to detect the position of the figure. The Haar detector in OpenCV is used to detect the face. Then based on the priori knowledge of the size of figures, an enlarged window, which is more likely to cover the FG area, is defined to initialize the GrabCut. When the detected single face is around the horizontal center of the image, the enlarged window is centering at the face window. When the face window is biased to the right or left of the image, more space is given to the left or right side correspondingly of the enlarged window. For an image with complex BG, the segmented FG may not be so accurate. As in [RKB04], we provide a refine interface 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 morphology post-processing. Some results are shown in Figure 6.

![Figure 6: FG/BG segmentation and face, skin areas detection.](image)
saturation is conducted separately in the same way. Finally, the transition boundary between regions is smoothed.

5. Reference Selection

The main task of reference selection is to select reference paintings that can create more pleasing results comparing to the original input, while not pushing the original image too far away from its original natural property. Because of our portrait enhancement is to enhance the relative relationship of regions based on their original values, therefore the reference paintings should be selected by comparing their value organization in regions with that of the photograph. The relationship of regions can be modeled as a graph as shown in Figure 7. The features within regions $\Psi$ and the relationship of regions $\Phi$ are used to select the reference painting. After we get the feature vector $V = (\Phi, \Psi)$, the task is to determine the similarity of the features between input photographs and paintings. Particularly, the paintings that can create more pleasing results comparing to the original input photograph should be match well with the input photograph. Therefore, we propose to use distance metric learning to determine the similarity of paintings and input photographs according to their features.

![Figure 7: The graph model.](image)

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

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

where $\alpha$ is the parameter to linearly combine the distance of features, $\Psi_k(i)$ is the $k$th feature in the feature vector of photograph $I_i$, and $\Psi_k(j)$ is the $k$th feature in the feature vector of painting $R_j$. Our objective requires the parameter $\alpha$ to be set such that paintings which can create more pleasing results are closer to the input photograph than others. As in [KKL10, HKK12], we learn the parameter $\alpha$ using a target distance function $D_t(i, j)$. The parameter $\alpha$ is set by minimizing the objective function

$$\text{arg min}_{\alpha} \sum_{i,j} \|D(i, j) - D_t(i, j)\|^2$$

The target distance $D_t(i, j)$ is gained from the feedback for the enhancement results. Small distance is given for pleasing results and larger distance is given for unacceptable results. Minimizing this objective function leads to finding 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.

In our implementation, the feature within each region $\Psi$ is constructed by the mean values $l_{\text{mean}}$ of lightness and saturation, the 10-bins histograms of lightness and saturation, the maxima $l_{\text{max}}$ (95th percentile value), minima $l_{\text{min}}$ (5th percentile value) values of lightness and saturation distributions, and the inter-contrasts of lightness and saturation. The BG global contrast factor $G_B$ is also calculated as one feature to express the complexity of the BG. Detailed and variation-rich image BG has a high global contrast factor and the simple BG has a low global contrast factor [MNNea05]. So the contrast factor can represent the difference of the simple and complex BG. The global contrast factor is calculated as the weighted average of local contrasts at various resolution levels (more details can be seen in [MNNea05]). The relationship feature $\Phi$ is constructed by the lightness and saturation inter-contrasts between FS and FO, FS and BG, FO and BG.

To learn the distance metric, the collected 300 portrait paintings (180 half body portraits and 120 full body portraits) are applied as references to generate enhancement results for the training photographs. The target distance function $D_t(i, j)$ is the assigned score for the enhancement result by an expert who has lot of experience on photo adjustment. It is a challenge action to rate a score for an image for the difference of subjective preference. However, it is more consistent for classifying the results to be good, acceptable and unacceptable. Therefore, instead of directly giving a score to each result, the expert is asked to classify the enhancement results for 32 half body portrait photographs and 30 full body portrait photographs to three classes: good, acceptable and unacceptable. Then the largest score is given to good results and smallest score is given to unacceptable results. For each half body portrait photograph, 180 enhancement results are generated while 120 enhancement results are generated for each full body portrait photograph. In all the enhancement results, the expert only classifies ones that he can confidently find a class for them, otherwise they are discarded. Finally, we get 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 spread range and mean value 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 regions in pho-
to graphs referring to those of the reference paintings to map the contrast. Contrast stretching is commonly used to change the spread range to enhance contrast when avoiding artifacts. The spread range can be expressed as \( \text{maxima} - \text{minima} \). Therefore, we can map the maxima and minima values instead of mapping the spread range directly.

Given the target mean value \( m'_t \), maxima \( l'_{\text{max}} \), minima \( l'_{\text{min}} \), the task of the contrast mapping is

\[
I(m_s, I_{\text{max}}, I_{\text{min}}) \xrightarrow{f} I'(m'_t, l'_{\text{max}}, l'_{\text{min}}) \quad (3)
\]

Piecewise contrast stretching separately on the lower value zone and upper value zone separated by the mean value has been used in [ZCC12] to adjust the mean value and spread range. This piecewise linear transformation is not continuously differentiable across histogram regions which may cause contouring defects [DGV09]. Additionally, the method cannot handle the values out of the main range in a reasonable way. To more effectively perform contrast mapping, an interpolating transformation curve which is piecewise defined, monotonically increasing and continuously differentiable is applied. This interpolating transformation curve is used as the histogram warping in [GD04, DGV09].

Given some 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]:

\[
f(x) = b_{k-1} + \left( \frac{r_k x^2 + d_{k-1} (1-t) y}{r_k + (d_k + d_{k-1} - 2r_k) (1-t) y} \right) (b_k - b_{k-1})\]

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 \((a_1, b_1, d_1) = (I_{\text{min}}, l'_{\text{max}}, d_1)\), \((a_2, b_2, d_2) = (m_s, l'_s, d_2)\) and \((a_3, b_3, d_3) = (I_{\text{max}}, l'_{\text{max}}, d_3)\), and two endpoints are \((0, 0, d_0)\) and \((1, 1, d_4)\). The definition of the contrast adjustment at each point will controls the shape of the transformation curve. To be more closely fitting at the 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 \( d_0 = \frac{d_1 + d_2}{2} \), \( d_3 = \frac{d_0 + d_2}{2} \), the contrast adjustments \( d_k \), \( k=1, 2, 3 \) are

\[
d_k = \left( \frac{b_k - b_{k-1}}{a_k - a_{k-1}} \right)^{\frac{1}{2}} \left( \frac{b_{k+1} - b_k}{a_{k+1} - a_k} \right)^{\frac{1}{2}}
\]

where \( g_k = \frac{F(a_k) - F(a_{k-1})}{F(a_{k+1}) - F(a_{k+1})} \) and \( g_k = \frac{F(a_k) - F(a_{k-1})}{F(a_{k+1}) - F(a_{k+1})} \) are the slope weights. \( F(x) \) is the cumulative distribution function.

One example transformation curve is shown in Figure 8(a).

Good contrast in image is not simply the higher the better. Artists are skilled at controlling the perceptual 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 in the reference. This means, the reference painting provides the template to change the photograph, but how close to match to the template depends on the original image and the user’s preference. Thus, two parameters \( \alpha, \beta \) are used to control the inter-contrast and intra-contrast mapping. For the inter-contrast is the relative difference of mean values, and the mean lightness and saturation values also influences the appearance of the image, therefore the inter-contrast mapping is conducted by moving the mean value of each region to some where 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 \( I_r \) in the reference, the mean value is shifted toward to \( m_r \) and the shifting amount is controlled by a weighting factor \( \alpha \in [0 - 1] \).

\[
m'_t = m_s + \alpha (m_r - m_s)
\]

This adjustment on the mean values implies that the inter-contrast of the input is moved close to that of the reference controlled by \( \alpha \). Given the intra-contrast of \( I \) and the intra-contrast of \( I_r \) which are \( c_s \), \( c_t \), the target intra-contrast is

\[
c'_s = c_s + \beta (c_t - c_s)
\]

where \( \beta \) is a weighting factor in \([0 - 1]\). Therefore, the enhancement gain of the spread range will be \( \lambda = \frac{c'_s}{c_s} \). Then we get the target maxima and minima values as

\[
l'_{\text{max}} = \lambda (l'_{\text{max}} - m_s) + m'_t
\]

\[
l'_{\text{min}} = \lambda (l'_{\text{min}} - m_s) + m'_t
\]
mean, maxima and minima are mapped exactly 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 of 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.

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

The smoothing is to keep 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 $F$. $\theta$ controls the sensitivity to the gradient of $F$. $\epsilon$ is a small constant. The weight $w_s$ is

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

$\tau_1$ is larger than $\tau_2$ for smoothing the boundary area between regions. In our implementation, we use the parameters $\theta=0.0001$, $\tau_1=0.3$ 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

$$ F^* = F + D $$  

7. Experiments and Discussion

In this section, we will analyze the performance of the proposed portrait photograph enhancement by manipulating regional contrasts as informed by portrait paintings. 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.

Reference influence. The enhancement results of two photographs using five selected top ranked reference paintings are shown in Figure 9. The results using the selected reference paintings by the reference selection method are encouraging. By applying the proposed enhancement method, more details of the figure especially on the face are visible and the figure stands out in the images. However, if we use one painting which is far away from the input photograph using the learned distance metric as reference (as the paintings in the seventh column of Figure 9), the face and figure also stand out in the images. However, the change is too far away from its original natural property (e.g. the pink cloth becomes black). Therefore, the reference selection is demonstrated the importance for the portrait enhancement. Figure 10 gives another two examples with indoor bright and dark FG. We can see the selected reference paintings have similar natural property with the input photograph. In other words, the photograph with a dark FG and bright BG is more likely to be improved using a reference painting with also dark FG and bright BG.

Parameter analysis. In the contrast mapping, we use the same $\alpha$, $\beta$ for the three regions FO, FS, and BG. Parameters of lightness are expressed as $\alpha_l$, $\beta_l$, and for saturation $\alpha_s$, $\beta_s$. When $\alpha_l = \alpha_s$, $\beta_l = \beta_s$, we indicate them as $\alpha_l$. $\alpha_s$, $\beta_s$ control the matching degree to the contrasts of the reference. Figure 11 shows the appearance change when changing the $\alpha_l$, $\beta_s$ from 0 to 1. When $\alpha_l$ increases from 0 to 1, the

Figure 10: The first column: original photographs. The second to third columns: two of the selected reference paintings for the two photographs and the corresponding results.
Figure 9: 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 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 $\alpha_l = 0.6$, $\alpha_s = 0.6$, $\beta_l = \beta_s = 0.5$.

inter-contrast is moved towards to that of the reference (see Figure 11(c)). On the appearance, it is that the figure stands out to be more bright in the example in Figure 11. When $\beta_l$ changes from 0 to 1, the lightness intra-contrasts of the three regions are moved to those of the reference correspondingly (see Figure 11(d)). More local details in the BG are visible when $\beta_l$ increases. The influence of $\alpha_l$, $\beta_l$ are in the similar way with $\alpha_s$, $\beta_s$.

In the reference selection, the results used for training are generated 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 method, we can get acceptable results 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 $\alpha_l$ is in the range [0.3, 0.7], and $\alpha_s$ is in [0.2, 0.6]. $\beta_l$ and $\beta_s$ are suggested to be set as the same value, generally are in the range [0.3, 1].

Comparison. To evaluate the advantage of our proposed portrait photograph enhancement referring to paintings, our proposed approach is compared to the content-aware enhancement (CAE) method in [KLW12] which performs better than the global learning-based enhancement method in [BPCD11] and resent popular softwares (Google’s Picasa, Microsoft’s Office Picture Manager and Photoshop). In addition, our results are also compared with the results enhanced using the average mean and contrast values from the statistics (AVE method). Figure 12 lists 5 results using the three methods. We can clearly see that the faces and figures in our results become more salient comparing to those in the re-
results generated by CAE method [KLY12]. Additionally, the light on the face in our results is more naturally presented. However, in the results of the second and third rows using CAE method, the light on the face becomes unnatural because of the lightness is differently processed on the left and right side face. The enhancement results using the average values from the statistics in the first to third rows are comparable with the proposed exemplar-based method. However for the photographs in the forth and fifth rows, the proposed exemplar-based method performs better than using the average values. The exemplar-based method can handle the variations of the scenes better.

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

For the first user study, we randomly selected 40 of the test photographs, including variety of scenes. This user study has two objectives. One is to show that the results using the proposed approach become more pleasing and the figure and face become more salient. The other is to show that the results using the top ranked reference paintings are more desirable than 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. The results generated using the 5 reference paintings were all compared with the original photograph. Each result and its corresponding original 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, the participants were asked to choose the answers for two questions. The first question Q1 was “In which image, the figure and face is more salient?” and the second Q2 was “Which is more pleasing?”. The participants could choose ‘left’ or ‘right’ as response to each question according to their first impression. No further information on the goals of the experiment or the origin of the images was provided to the participants.

For the second objective, the results using the reference paintings from the middle and bottom in the painting ranking were compared with those generated using the top ranked reference paintings. 5 paintings from the ranking in [51, 65] for full body portrait and ranking in [65, 81] for half body portrait were randomly selected as the middle reference paintings. 5 paintings from the ranking in [106, 120] for full body portrait and ranking in [166, 180] for half body portrait were randomly selected as the bottom reference paintings. The three group results using the top, middle and bottom reference paintings were aligned to three rows. The order of the group arrangement was randomly generated. Participants were asked to select the group that they prefer more. The participant can 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 generated using the top ranked reference painting with the original side by side and the comparison of the three group results using the top, middle and bottom reference paintings. In the first set, there were 24 comparisons and in other 6 sets, there

Figure 11: Influence of $\alpha, \beta$. (e)-(j): $\alpha_l = 0.4$, $\beta_l = \beta_s = 0.5$. (j)-(n): $\alpha_l = 0.5$, $\alpha_s = 0.4$, $\beta_s = 0.5$. 
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 the comparison. Finally there were 38 participants, 21 were males and 17 were females. The age was ranging from 20 to 45. We totally collected 2070 samples for the comparison of result vs original. The average selection rate of each result vs original is shown in Figure 13(a). The result 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 photographs is effectively improved and the face and the figure become more salient. In the comparison of the three group results generated by using top, middle and bottom reference paintings, more than 80% of the responses are in favor of the results generated 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.

The second user study was conducted to compare our results with those by CAE method in [KLW12] and by AVE method. The CAE method in [KLW12] provided about 55 results for images with face. 36 of them which can be considered as portrait were used for this user study. Our result and the corresponding result using other method 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 choose the answers for the same two questions as in the first user study. Differently, the participants could choose ‘Left’, ‘Right’, or ‘They look the same.’ as response to each question according to their first impression. The 36 images were divided into 2 sets. Each participant can select to complete one or two sets. There were total 29 participants aging from 20 to 35, 18 of them...
were male and 11 of them were female. For each comparison, there were 18 rates in average. The user study result is shown in Table 1. It shows that our results are preferred in a significant higher percentage of the responses than results using other methods. This result is not surprising, since the method in [KLW12] can not handle the relationship of regions and the method using average statistical values does not consider the variation of scenes.

**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 segmentation and face and skin areas detection. Although the interactive segmentation has reduced the segmentation errors and the skin area detection in the segmented FG also is much 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. As in the example shown in Figure 14(a), the color and lightness of the clothes are very close with those of the skin areas, as a result some regions of the clothes are detected as skin (see Figure 14(b)). After regional contrast manipulation using the proposed approach, these regions differ slightly from the rest of the clothes (see Figure 14(c)).

Another limitation is that artifacts around the transition boundary between brightened FG region and dark BG may be visible. This is because the brightness of the FG region is extended to the neighboring BG region through the boundary smoothing. This brightness change in the dark BG is particularly more visible than on brighter BG. As in the result in Figure 14(e), the artifact is produced around the boundary of the right lower arm when the face and skin areas are brightened. Some noise my also become visible in face and skin areas after being brightened using the regional contrast manipulation.

**Application on RGBD images.** Along with the development of techniques and commercial cameras, the RGB color image with depth (D) is more easily captured. While the Xbox 360 Kinect depth sensor by Microsoft can only capture depth for indoor scenes, the light field camera (e.g raytrix camera) can capture the depth in both indoor and outdoor. The depth of scenes can also be calculated from videos. With depth, the figure can be easily and automatically extracted from the image. The framework of our proposed application in RGBD image is in Figure 15.

The main element that influences the FG extraction from depth is the ground plane. Therefore, we first detected the ground plane using the filter in [XCA11] and removed the ground plane. Then, region grouping was applied to extract

---

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

<table>
<thead>
<tr>
<th>Questions</th>
<th>Our</th>
<th>ACE</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>62.0%</td>
<td>30.6%</td>
<td>7.4%</td>
</tr>
<tr>
<td>Q2</td>
<td>66.8%</td>
<td>29.6%</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions</th>
<th>Our</th>
<th>Ave</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>68.4%</td>
<td>19.6%</td>
<td>12.0%</td>
</tr>
<tr>
<td>Q2</td>
<td>70.0%</td>
<td>22.1%</td>
<td>7.9%</td>
</tr>
</tbody>
</table>
the FG by using a seed point as the center of the detected face area.

More experiment results are shown in Figure 16.

8. Conclusion

This paper proposes a portrait photograph enhancement method 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 becomes more salient and the photograph is more pleasing by applying the proposed approach.

References


submitted to COMPUTER GRAPHICS Forum (8/2013).


submitted to COMPUTER GRAPHICS Forum (8/2013).
Figure 16: More experiment results. The first and fourth rows: reference paintings. The second and fifth rows: original photographs. The third and sixth rows: results.

submitted to COMPUTER GRAPHICS Forum (8/2013).