Learning-Based Approach for Face Image Relighting

Ching-Ting Tu\textsuperscript{1, *}, Chin-Yu Chang\textsuperscript{1} and Yi-Chung Chen\textsuperscript{1}

\textsuperscript{1}Department of Computer Science and Information Engineering, Tamkang University, Taiwan, R.O.C.
*cttu@mail.tku.edu.tw

Abstract. In this study, we propose a system for face image relighting. Given an unseen testing image, the system automatically changes its light condition. Such system is difficult since the shadow and lighting areas of a particular lighting condition are highly depended on the lighting source position and the facial 3D geometry. To solve this problem, we introduce a learning-based approach to construct the appearance correlation between two lighting conditions in order, which is then used to transform the lighting condition from the source lighting domain to the target lighting domain. First of all, the discriminative facial features of lighting and personal characteristics are selected. Consequently, these discriminative features are then used to guide the light reconstruction process. According to the experimental results, the proposed system, where the discriminative features are integrated in the synthesis process, makes the synthesized results preserve the personal characteristics of system input.

1. Introduction

We proposed a system for facial image relighting, i.e. changing the lighting condition of the input facial image to another specified lighting condition. However, the lighting condition of a facial image is highly depended on the position and direction of the lighting sources. Furthermore, even under the same light condition, shadow and highlight facial areas of images from different individuals are not equal. To solve such problem, we proposed a learning-based algorithm to transfer the lighting condition of the input image (the source condition) to a target condition. Such learning process requires a training dataset, where two facial images of the same individual taken under the source and target lighting conditions, respectively, are required. The proposed learning approach can apply to other testing samples not observed in the training dataset. The key issues are (i) learn the complex appearance correlation of images taken under two different lighting conditions. (ii) Personal characteristics of input individual, such as the shapes of nose or eye, should be preserved after changing the lighting condition.

1.1. Related work

There are several works for facial image relighting. For example, Shihet al.’s \cite{1} proposed a lighting transformation system to relight facial image as a particular professional style. Such framework required a reference image first, and geometrically align any input testing image to the selected reference image. The lighting condition of input image is then mapped into the selected target lighting condition, e.g. the shadow and highlight areas of the reference image. However, the facial alignment process is generally not performed well for the shadow area, and the transformed result will be decreased. Work proposed in \cite{2} is also required the face alignment process. The idea of this work is
quite similar to work [1], it transfers the input lighting condition to that of the selected reference image. However, method [1] processed on the image lighting layer (the low frequency of facial image), on the hand, method [2] focused on the texture layer, which is the high frequency information containing the local detailed component. Approach in [3] calculated the lighting contrast between the input and reference images by the help of 3D facial model. In summary, these three methods are required a reference image. However, the personal geometry and the appearance characteristics of input and that of the selected reference may be totally different.

On the other hand, learning-based methods, such as [4] and [5], were proposed; method in [4] adopted kernel principal component analysis (kernel PCA) to learn the lighting correlation between two different lighting conditions from the training database. Method in [5] used Tensor PCA model to construct the subspace of different face individuals taken under various lighting conditions. Consequently, the constructed subspace is then used to predict the lighting of unseen input. However, according to the PCA property, the synthesized image of any unseen subject will be similar to appearances of training samples. In this study, we propose a learning-based approach for face image relighting, the aim is to preserve the personal characteristics of the input.

2. Learning-based synthesis system

A learning-based face relighting synthesis system is proposed in this study, where given a facial image with a particular lighting condition, a corresponding synthesized image under another lighting condition is produced. Such synthesis system is trained with the assistance of a training dataset containing pairwise lighting examples, where each one contains two images of the same individual taken under two particular lighting conditions. Especially, the proposed synthesis framework is cooperated with a feature extraction process, where facial regions that are invariant to cross-lighting conditions and significant for discriminating personal characteristics are extracted. The learned lighting-condition dependency feature map is then used to constrain the learning-based synthesis process, the proposed framework is illustrated in figure 1.

![Figure 1. Proposed example-based relighting system. L: the lighting stage, D: the detailed texture stage, x: images taken under source lighting condition, and y: image taken under target lighting condition.](image)

2.1. Pairwise dataset

Let the source (i.e. input) lighting condition as the style-X and the target (i.e. output) lighting condition as the style-Y. We collect N pairwise training samples from Yale B database [6], where each training individual contains two facial images taken under two specified lighting conditions X and Y, respectively. Denote the dataset \( X = [x_1, x_2, \ldots, x_N] \) as the \( N \) images of the style-X, and the dataset...
\( Y = [y_1, y_2, \ldots, y_N] \) as the images of style-\( Y \). Each image (either \( x_i \) or \( y_i \)) is decomposed into its lighting component (\( x_i^L / y_i^L \)) and the detailed texture component (\( x_i^P / y_i^P \)). For example, the decomposing process of image \( x_i \) is:

\[
x_i^L = x_i \otimes G(\sigma) \quad \text{and} \quad x_i^P = x_i - x_i^L
\]

Where \( G(\sigma) \) is the Gaussian blurring function with kernel size \( \sigma \); \( y_i \) can be decomposed as \( y_i^L \) and \( y_i^P \) in the similar way.

2.2. Significant feature extraction

Two training datasets, \( L = \{(x_i^L, y_i^L)\}_{i=1}^N \), and \( D = \{(x_i^P, y_i^P)\}_{i=1}^N \), are then used for extract the light and detailed texture features, respectively. For the light case, the feature selection process is to extract the significant facial patches that can classify the images taken under the lighting condition \( X \) (style-\( X \)) or from the lighting condition \( Y \) (style-\( Y \)). On the other hand, for the detailed texture case, the goal is to select patches that are lighting invariant for face identification in order to preserve the personal characteristics. The feature selections for lighting and detailed textures are processed by the Adaboost algorithm [7], respectively.

In the feature extraction process of the lighting case, the dataset \( L \) is used, where each images is divided into \( N_k \) overlapped patches. Accordingly, we adopted \( N_k \cdot N \) patches from the \( \{y_i^L\}_{i=1}^N \) as features for the feature pool of the Adaboost algorithm. In such classifier learning process (the feature selection process), \( \{y_i^L\}_{i=1}^N \) is used as positive samples and \( \{x_i^L\}_{i=1}^N \) as the negative one. Feature value of each sample (either positive \( y_i^L \) or negative \( x_i^L \)) for a particular feature is defined as the cosine similarity between this feature patch and the corresponding location of this feature patch in this sample. Figure 2(c) illustrates the selected feature maps. In this case, the negative samples (style-\( X \)) are taken under frontal source lighting condition, shown in figure 2(a); the positive samples (style-\( Y \)) are taken from the target left lighting condition as figure 2(b). We used the 40%-feature map for the following synthesis stage.

One the contrary, for the detailed texture case, dataset \( D \) is used, where the \( (x_i^P, y_i^P) \) pairwise sample from the same individual is denoted as the positive sample, on the other hand, \( (x_i^P, y_i^P) \) pairwise sample from two different individuals is denoted as the negative sample. In such case, the feature pool are collected from the \( N_k \cdot N \) patches in \( \{y_i^P\}_{i=1}^N \). Feature value of a pairwise sample \( (x_i^P, y_i^P) \) for a particular feature is defined as the cosine similarity between this feature patch and the corresponding patches at the same location in these two detailed images, respectively, and the product of these two cosine similarities is denoted as the feature value. Figure 2(d) illustrates the selected feature maps of detailed texture. We used the 60%-feature map for the following synthesis stage.

![Figure 2](image_url)

**Figure 2.** (a) example images of style \( X \); (b) corresponding images of style \( Y \); (c) lighting feature map selected by Adaboost algorithm; (d) detailed feature map. Note that, for (c) and (d), from left to right, are the top 20%, 40%, and 60% feature maps, respectively.

2.3. Synthesis stage
For any input image $x_{\text{test}}$ that taken under the source lighting condition (style-$X$) is decomposed as its lighting component ($x^l_{\text{test}}$) and the detailed texture component ($x^d_{\text{test}}$) by using (1) first. The goal of synthesis stage is to transfer (estimate) the lighting and detailed style of $x^l_{\text{test}}$ and $x^d_{\text{test}}$ to $y^l_{\text{test}}$ and $y^d_{\text{test}}$, respectively. The estimation process is based on the Markov random field (MRF) approach [8]. For the synthesis stage, the proposed MRF graph containing $N_k$ observation nodes, which are corresponded to the $N_k$ patches of each image, and each observation node has one corresponding hidden node, which represents the estimated patch appearance. In work [8], the candidate(s) of each hidden node are selected from the training dataset; however, in our approach, the candidates are either collected from the training dataset or is the same as the observation appearance. Functions of these two candidate types are different, where the former one makes the final result contain the target $Y$-style, while the later one makes the synthesized result preserve more personal characteristics of system input $x_{\text{test}}$.

The objective function of the proposed synthesis module is modelled by the MRF formulation:

$$p(x_{p1}, x_{p2}, \ldots, x_{pN_k}, y_{p1}, y_{p2}, \ldots, y_{pN_k}) \propto \prod_{(pi,pj)} \psi(y_{pi}, y_{pj}) \prod_{pk} \phi(x_{pk}, y_{pk})$$  \hspace{1cm} (2)

Where the joint probability function of the input, i.e. the observation layer $x = (x_{p1}, x_{p2}, \ldots, x_{pN_k})$, and the final result, i.e. the hidden layer $y = (y_{p1}, y_{p2}, \ldots, y_{pN_k})$, are modelled by using the product of data terms and the smoothness terms. The data term (or the observation probability), $\phi(x_{pk}, y_{pk})$, is used to evaluate the probability of using $y_{pk}$ as the state candidate of the $pk$’th hidden node given the $pk$’th observation node as $x_{pk}$. The smoothness term (or the transition probability), $\psi(y_{pi}, y_{pj})$, is used for the probability of using $y_{pi}$ as the state candidate of the $pi$’th hidden node given state candidate of its neighbour hidden node (the $pj$’th node) as $y_{pj}$. In the proposed framework, the observation function of the $pk$’th hidden node and the $pk$’th observation node is defined as:

$$\phi(x_{pk}, y_{pk}) = \exp\left(-\frac{\|x_{pk} - \theta_x(y_{pk})\|^2}{(\eta_1 \sigma^2_k)^2}\right).$$  \hspace{1cm} (3)

Where $\sigma^2_k$ is determined by the cross verification; $\theta_x(y_{pk})$ is the corresponding training patch of style-$X$ of $y_{pk}$. The transition function of the $pi$’th and $pj$’th hidden nodes is defined as:

$$\psi(y_{pi}, y_{pj}) = \exp\left(-\frac{od(y_{pi}, y_{pj})}{(\eta_1 \sigma^2_k)^2}\right).$$  \hspace{1cm} (4)

Where such probability value is used for measuring the appearance difference between the overlapped regions of the $pi$’th patch and its neighbour (the $pj$’th node), i.e. $od(y_{pi}, y_{pj})$. In short, the function of $\phi(\quad)$ is to select satiable candidates that with similar appearance as the observation data, while the function of $\psi(\quad)$ is used for confirming the smoothness (or the continuous appearance) between neighboring patches. Such formulation is applied for both the lighting and the detailed texture component synthesis, respectively.

In the optimization framework of the MRF method [8], the state candidates are selected from the training dataset. To be specified, using the lighting case as an example, the pairwise images of each example in the training dataset $L = ((x^l_i, y^l_i))_{i=1}^N$ are decomposed as $N_k$ patches, respectively, i.e. $x^l_i = (x^l_{ip1}, x^l_{ip2}, \ldots, x^l_{ipN_k})$ and $y^l_i = (y^l_{ip1}, y^l_{ip2}, \ldots, y^l_{ipN_k})$. Accordingly, for the $pi$’th hidden node, there are $N$ candidates from $\{y^l_{ipj}\}_{j=1}^N$, where each candidate has one corresponding patch from
\( \{x^L_{lpi}\}_{i=1}^N \), i.e. \( y^L_{lpi} = \theta_x(y^L_{lpi}) \) in (3). These candidates are then selected for maximizing the objective function formulated in (2), (3), and (4), by using the belief propagation algorithm. However, if all the state candidates are from the training samples as the original method in [8], the insufficient texture variance in the training dataset may cause unexpected synthesized results (as shown in figure 3(c)).

To solve such problem, in the proposed method, some specified observation nodes of the input, e.g. \( x^L_{test,pli} \), are also used as the state candidates of the corresponding hidden nodes. To be specified, these nodes are determined by the significant features learned by Adaboost algorithm in sec 2.2.

That is, the patch locations that are corresponding to significant feature maps are used to determine the state candidates should be from the training dataset, e.g. \( \{y^L_{lpi}\}_{i=1}^N \), or directly assigned as its observable patch, \( x^L_{test,pli} \). For example, for the lighting case, the location of \( pk \)’th patch that belongs to region of the selected features indicate that its candidates should be from the training dataset (i.e. \( \{y^L_{lpi}\}_{i=1}^N \) for inferring the appearance that are included in the input; on the contrary, candidates of other hidden nodes are from their corresponding observation \( x^L_{test,pli} \). A similar process is also applied for texture synthesis; however, the candidate nodes that corresponding to the location of selected are directly assigned as their observations, e.g. \( x^L_{test,pli} \), as these regions are invariant to lighting for face identification. Candidates of other nodes are form the training dataset (i.e. \( \{y^D_{lpi}\}_{i=1}^N \) ). Final synthesized results based on the selected feature map are shown in figure 3(d).

![Figure 3.](image)

Figure 3. (a) thee testing examples of style \( X \); (b) corresponding images of style \( Y \) (ground truth); results synthesized by (c) MRF model \[4\], (d) proposed method, (e) DCM model \[9\], (f) 2DDCM model \[10\], (g) LS method, and (h) CCA model \[11\].

3. Experimental results
We validated our method using the Yale B database \[6\]. There are 38 subjects in this dataset, where each subject has two pairwise image taken under two specified lighting conditions, i.e. the style \( X \) (as shown in figure 3(a)) and the style \( Y \) (as shown in figure 3(b)). In our setting, 32 subjects are selected for the training purpose, while the rest 6 while the rest 6 subjects are used for the testing. In figure 3, we also compare the proposed approach with other learning-based regression method for predicting facial structure under the style \( Y \), i.e. direct combined model (DCM) \[9\] (figure 3(e)), 2DDCM \[10\] (figure 3(f)), least squares regression (LS) (figure 3(g)), and canonical correlation analysis (CCA) \[11\] (figure 3(h)). In figure 4, we also evaluated the proposed approach for 4 cases of pairwise lighting conditions, synthesized results of three example testing subjects are provided.

4. Conclusion
We proposed an example-based face relighting system. Using Adaboost approach, the discriminative features are selected to boost the performance of synthesized result. The synthesis process is based on the MRF approach, where different from the existing approaches, the proposed optimization framework is modified by cooperating with the discriminative feature selection process in order to preserve the personal characteristics of the system input. Our method outperformed existing approaches in the current dataset.

![Figure 4. Synthesized results (style-Y) of testing examples, where training samples for (a) and (b) are from the same pairwise dataset, but style X and style Y are interchanged. Same for the results of (c) and (d). Note that, for each case, the first row: final synthesized results; the second row: synthesized lighting results; the third row: results of the detailed texture.](image)

References

[1] Shih Y., Paris S, Barnes C, Freeman W F, and Durand F 2014 Style transfer for headshot portraits ACM Trans. Graph, pp. 1-148

[2] Chen X, Wu H, Jin X, and Zhao Q 2013 Face illumination manipulation using a single reference image by adaptive layer decomposition. IEEE Transactions on Image Processing pp. 4249-4259

[3] Wang Y, Zhang L, Liu Z, Hua G, Wen Z, Zhang Z, and Samaras D 2009 Face relighting from a single image under arbitrary unknown lighting conditions IEEE Trans. Pattern Anal. Mach. Intell pp. 1968-1984

[4] Xie X, Zheng W S, Lai J, Yuen P C, and Suen C Y 2011 Normalization of face illumination based on large-and small-scale features IEEE Transactions on Image Processing pp. 1807-1821

[5] Geng X, Smith-Miles K, Zhou Z H, and Wang L 2011 Face image modelling by multilinear subspace analysis with missing values IEEE Transactions on System, Man, and Cybernetics, Part B (TSMC-B) pp.881-892

[6] Georgiades A S, Belhumeur P N, and Kriegman D J 2001 From few to many: illumination cone models for face recognition under variable lighting and pose IEEE Trans. Pattern Anal. Mach. Intell. 23(6) pp. 643-660

[7] Viola P A and Jones M J 2004 Robust real-time face detection International Journal of Computer Vision pp. 137-154

[8] Liu C, Shum H Y, Freeman W T 2007 Face hallucination: theory and practice International Journal of Computer Vision pp. 115-134

[9] Tu C T and Lien J J 2010 Automatic location of facial feature points and synthesis of facial sketches using direct combined model IEEE Trans. Systems, Man, and Cybernetics, Part B pp. 1158-1169

[10] Tu C T, Ho M C, and Lin M Y 2017 A new approach for face hallucination based on a two-dimensional direct combined model Pattern Recognition pp. 1-20

[11] Yang W, Yi D, Lei Z, Sang J, Li S Z 2008 2D-3D face matching using CCA FG pp. 1-6