Learning Templates for Artistic Portrait Lighting Analysis

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

***Abstract*—Lighting is a key factor in creating impressive artistic portraits. In this paper, we propose to analyse portrait lighting by learning templates of lighting styles. Inspired by the experience of artists, we first define several novel features which describe the local contrasts in various face regions. The most informative features are then selected with a stepwise feature pursuit algorithm to derive the templates of various lighting styles. After that, the matching scores which measure the similarity between a testing portrait and those templates are calculated for lighting style classification. Furthermore, we train a regression model by the subjective scores and the feature responses of a template to predict the score of a portrait lighting quality. Based on the templates, a novel Face Illumination Descriptor (FID) is defined to measure the difference between two portrait lightings. Experimental results show that the learned templates can well describe the lighting styles, while the proposed approach can assess the lighting quality of artistic portraits as human being does.**

***Index Terms*—Portrait Lighting Analysis, Contrast Feature, Template Learning, Lighting Style Classification, Quantitative Assessment, Face Illumination Matching.**

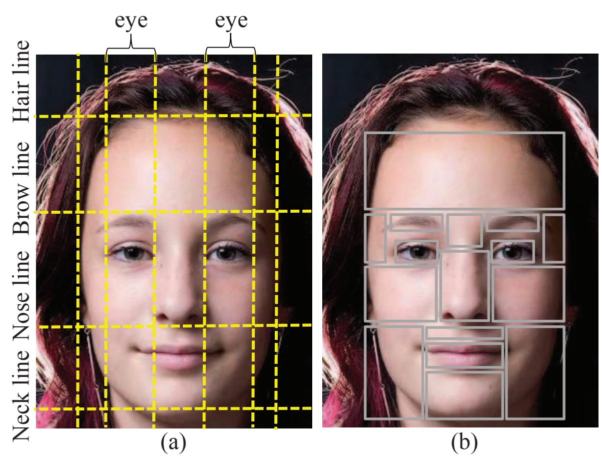


Fig. 1. The face regions definition. (a) The artists' common practice to analyze portrait lighting quality. The face is coarsely divided into regions [11]. (b) Inspired by the artists, we divide the face into 16 rectangular parts.

I. Introduction

Lighting is one of the most important factors of portrait. Artists usually use professional lighting equipment in producing artistic portraits. An ingenious lighting can highlight the shape of face and make the portraits more impressive and interesting. In many applications such as image searching and face illumination transfer [1, 2, 3], it is of significance that the portraits can be searched by lighting, or lighting quality can be automatically assessed.

Several algorithms related to lighting analysis are proposed. Environmental irradiance is estimated from images [4, 5, 6, 7]. Portrait lighting is the product of lighting and facial geometry. It is impractical to analyze portrait lighting from environmental irradiance since recovering the facial geometry from images is a difficult task. Various features are designed for the visual aesthetic quality assessment [8, 9, 10]. Those features are efficient for aesthetic quality analysis, but are not designed for portrait lighting.

On the other hand, artists have rich experience in analyzing portrait lighting. They often check the lighting effects in local face regions to analyze portrait lighting. They firstly divide the face coarsely into 3×5 regions (Fig. 1 (a)). Then, the shadow and non-shadow areas in those regions, with their relative locations, area ratios, etc. are used to assess the lighting quality or classify the lighting styles: The *Rembrandt* style is featured by a triangular highlight region below one of the eyes. The *Paramount* style has a butterfly shape for the shadow between the nose and the mouth. The *Loop* style is named after the loop-shaped shadow below the nose. The *Split* style has half of the face in the shadow [11, 12, 13, 14].

According to the artists' experience, we propose to analyze the portrait lighting from the lighting effect in the local face region. The local lighting contrast features are inspired by the Haar-like features [15, 16, 17] used in face recognition. We divide the frontal face into 16 rectangular regions (Fig. 1 (b)). For each region, we define a set of local lighting contrast features with the help of artists. In order to obtain more local lighting contrast information, the local lighting contrast types are performed on various image channels using various image statistics. The templates (Fig. 2) are learned by selecting the most informative features using a stepwise feature pursuit algorithm [18]. The information gain of each feature is calculated by a log-linear model.

The templates are further used to analyze portrait lighting. For classification, we calculate the matching score between a testing portrait and a lighting template. If the matching score is greater than a threshold, the testing portrait photo is classified to that lighting style. For quantitative assessment,

The authors are with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: chen@buaa.edu.cn; jinxin@buaa.edu.cn; whyvrlab@buaa.edu.cn; zhaoqp@vrlab.buaa.edu.cn, Corresponding authors: Xiaowu Chen and Hongyu Wu).

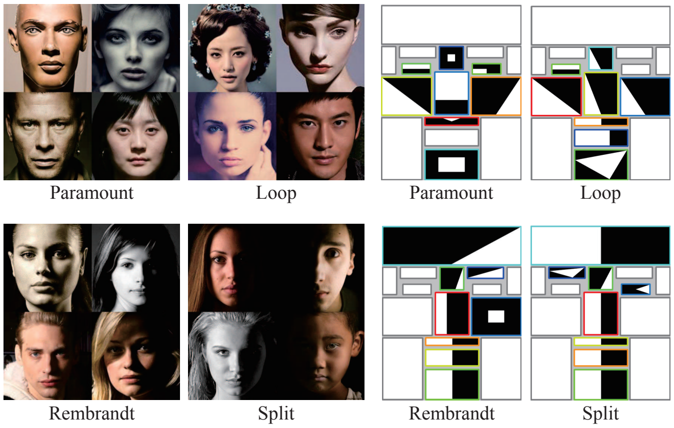


Fig. 2. The templates of four lighting styles. The left shows part of the training portraits. The right is the learned templates of each lighting style.

we first select a fixed number of portraits with lighting quality scores graded by artists and feature responses of a template. A regression model is then trained by the lighting quality score and the feature responses of a template to predict the lighting quality of a testing portrait. For face illumination matching, we select the local lighting contrast features which can well describe the local contrast to form the FID for measuring the lighting difference between two portraits.

The contributions of this paper include: (1) learning artistic lighting templates for artistic portraits through a set of local lighting contrast features, (2) portrait lighting analysis including lighting style classification and quantitative lighting assessment using learned artistic lighting templates, and face illumination matching using local lighting contrast features.

The remainder of this paper is organized as follows: we describe related work in Section II. The design of lighting contrast features is described in Section III. The learning of lighting templates is described in Section IV. We analyze portrait lighting in Section V. The experimental results are shown and discussed in Section VI. Finally, we conclude our paper with discussions in Section VII.

II. Related Work

In this paper, we focus on artistic portrait lighting analysis including lighting style classification and quantitative assessment. Roughly, the methods related to our topic can be divided into two aspects: image-based lighting analysis and photo aesthetic assessment.

A. Image-based Lighting Analysis

Lalonde et al. [4] estimated outdoor illumination from only a single outdoor image. They used a dataset of 6 million images to train the illumination inference model, and estimate a sun and sky dome model that is especially for outdoor images. The three most evident appearance cues (i.e. the sky, shadows on the ground and the varied intensities of the vertical surfaces to estimate the direction of light) are directly employed to estimate the illumination in a scene.

Xing et al.[19] estimated the dynamic outdoor illumination of an video sequence. The vertical and horizontal plans in the video are interactively selected to recover the sun and sky light in real time. Chen et al. [20] used the known scene geometry and shading information to estimate the illumination. A sparse set of 3D surfaces is selected by normal and semantic constraints. The coarse shading image of those 3D surfaces, together with the 3D surfaces, are used for illumination estimation.

Panagopoulos et al. [21] jointly recovered the illumination environment and estimate the cast shadows in a scene from a single image. A higher-order Markov Random Field (MRF) illumination model was used, which combines low-level shadow evidence with high-level prior knowledge for the joint estimation of cast shadows and illumination environment.

Zhu et al. [22] recognized shadows of monochromatic natural images. Both shadow-variant and shadow-invariant cues from illumination, textural and odd order derivative are used to train a classifier from a decision tree, and are integrated into a conditional random Field. Shadowed areas of an image can be identified using proposed monochromatic cues.

Johnson et al. [5] used both model-independent methods (cast-shadow analysis, occluding-contour analysis) and model-based methods (physical models of the eyes etc.) to infer the lighting direction. A Bayesian evidence integration scheme was used for the two disparate sources of information.

Nishino et al. [6, 23] estimated the lighting condition from eyes. The eye was treated as a natural light probe, and an environment map of the scene was computed from the image. The tasks such as inserting virtual objects into an image, face relighting, lighting robust face recognition benefits from the environment map from the eye.

Spherical harmonics [24] provided a way to approximate the environment map by spherical harmonic basis functions. Wen et al. [7] used a linear combination of spherical harmonic basis to approximate the environment map for any given image of a face, and Bitouk et al. [25] used the spherical harmonics approximation to adjust lighting for face image swapping. Han et al. [26] used spherical harmonics lighting model for improved face recognition performance.

Those works focus on environmental irradiance estimation, are not suitable for our task. We analyse the portrait lighting from the local contrast feature of 2D images to avoid environmental irradiance and facial geometry estimation from images.

B. Photo Aesthetic Assessment

Luo el al. [8] collected $17,613$ photos with manually labelled ground truth. They divided the photos into seven categories based on the photo contents, and developed a set of subject area extraction methods and visual features, which are specially designed for different categories. Which greatly improves photo quality assessment performance.

Another line of research in the field of aesthetic quality assessment is by using generic features. Marchesotti et al. [9] proposed to use generic image descriptors to assess aesthetic quality of a photo. The generic image descriptors which aggregate the statistics computed from low-level local features, implicitly encode the photo aesthetic properties.

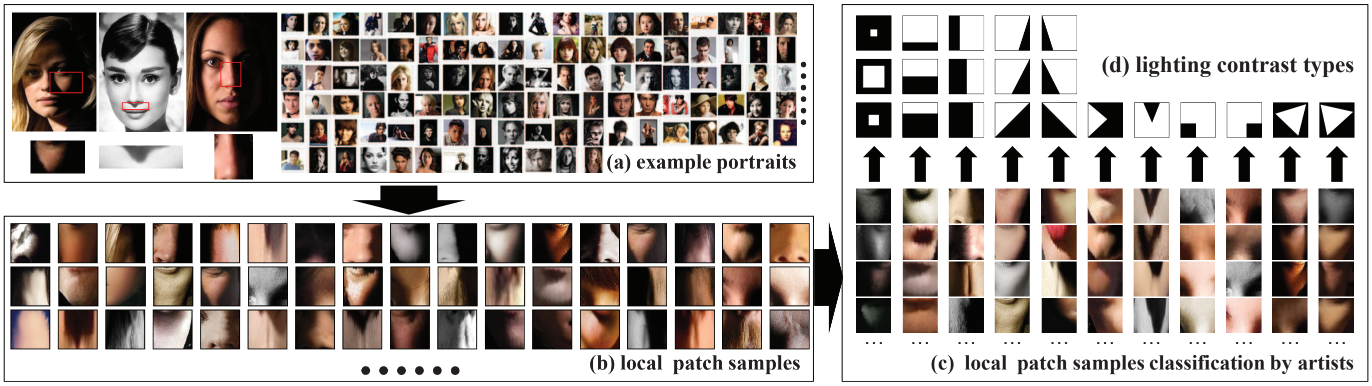


Fig. 3. Local lighting contrast type design with the help of artists. We first collect a dataset of artist lighting portraits. Then, the portraits are divided into local patches (the patches of faces are regularized to square). Finally, the local patches are classified by artists and the lighting contrast types are derived from the local patches.

Khan et al. [10] developed a small set of classification features to evaluate the visual aesthetic in photographic portraitures. They proposed features for spatial composition and highlight and shadow composition. Their lighting features are computed in the global face, or cannot well describe the complex artistic lighting effectsand find informative local lighting patterns in portraits .

Those features are efficient in aesthetic quality analysis, but are not designed for portrait lighting. This paper proposes a set of local contrast features specifically for portrait lighting.

III. Local Lighting Contrast Features

The local lighting contrast features are designed to capture the local lighting contrast characteristics. Each feature F_k includes 4 dimensions: $F_k : \{E, H, C, S\}$, where E, H, C, S are defined as:

Rectangular region E on the image lattice. As shown in Fig. 1, we divide the area of a frontal face into 16 rectangular parts: nose, left eye, right eye, left eyebrow, right eyebrow, mouth, region between mouth and nose, forehead, left up cheek, right up cheek, left down cheek, right down cheek, region left to left eye, region right to right eye, and chin.

Local contrast type H. In our previous work [27], we adopted 3 local lighting contrast types to approximate some local lighting patterns of relative brightness and darkness. In order to describe the local lighting contrast more precisely, we define a new set of local lighting contrast types with the help of artists.

Fig. 3 shows the design of local lighting contrast types. We collected 350 example portraits with artistic lighting styles (Fig. 3 (a)). We first divided all the 350 example portraits of our database into local patches, and the local patches with no obvious shadow or highlight were deleted (Fig. 3 (b)). Then, artists were invited to classify the local patches according to the local lighting effect, and divide the local patches into 11 kinds. For each kind of local patch, the artists designed a contrast type to represent the lighting contrast property (Fig. 3 (c)). In order to cover more local lighting effect, the first 5 contrast types were extended to 15 contrast types by moving the split line. Finally, we got 21 local contrast types which have the most potential to be selected as the representative local lighting patterns of artistic portrait lighting (Fig. 3 (d)).

According to the local lighting contrast type, the rectangular region is divided into two subregions A and B (black and white areas). In Section VI, we will show that the 21 local contrast types outperform the 3 ones in portrait lighting analysis.

Target channel C of the image. Our 8 adopted channels include the graylevel, the 3 channels L, a and b in the CIE 1976 (L^*, a^*, b^*) color space (L differs slightly from graylevel although the 2 channels are heavily correlated), and the hue and saturation channels in the HSV color space. In order to capture the effect of staggered highlight under sidelight, we also involve the channel of gradients in the graylevel channel, as well as an edge channel including all edge pixels generated by a Canny edge detector using OpenCV library with default parameters. (Canny et al. [28], OpenCV is the Open Source Computer Vision library, opencv.org).

Target statistic S of the image. We use 3 types of statistics: mean value μ, histogram h, and density ρ (proportion of edge pixels) which is specially designed for the edge channel.

The feature response of local contrast feature k between the two sub-regions A and B of image $\mathbf{I}$ on channel C is defined as

$$r_k(\mathbf{I}) = \begin{cases} |\mu_A - \mu_B| & \text{if } S \text{ is mean} \\ \mathrm{JS}(h_A||h_B) & \text{if } S \text{ is histogram} \\ |\rho_A - \rho_B| & \text{if } S \text{ is density, } C \text{ is edge,} \end{cases} \quad (1)$$

where, $\mathrm{JS}(\cdot||\cdot)$ denotes the discrete Jensen-Shannon divergence [29].

IV. Learning Artistic Lighting Templates

The lighting template consists of a set of local lighting contrast features. We learn a template for one lighting style using a stepwise feature pursuit algorithm. The learning algorithm is built on a log-linear model of the probability distributions of the lighting style, with the feature responses as factors.

Observing that the rectangular regions have no or very little overlap with each other, we assume independence of features in different rectangular regions and model the distributions of

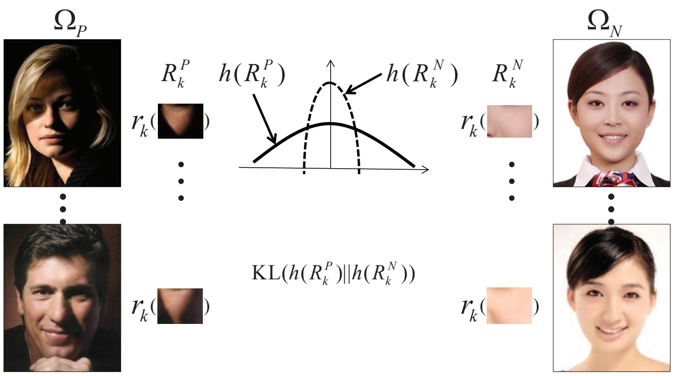


Fig. 4. Information gain of a feature. For feature F_k in the candidate feature set, its responses R_k^P and R_k^N are calculated for all images in Ω_P and Ω_N, and two histograms $h(R_k^P)$ and $h(R_k^N)$ are obtained for the positive examples and negative examples, respectively. The KL divergence between the two histograms approximates the information gain of F_k on our dataset.

portrait images of the same artistic style in a generative log-linear form. With the set of portrait images Ω_P with a lighting style P (positive examples), and against the portrait images Ω_N with other lighting styles (negative examples, including the portrait images with no lighting effect on their faces), we are able to pursue the most representative features with their weight parameters. We would like to build a lighting template for Ω_P against Ω_N, as well as a probability distribution upon this template. The template is a group of features characterizing the portrait lighting style P.

In this work, Ω_P can include all the portraits with artistic lighting styles, while Ω_P contain the portraits without obvious artistic lighting styles. A common template for all the artistic lighting styles can be learned. Ω_P can also include a subset of artistic lighting styles, such as *Rembrandt, Paramount, Loop, and Split*. The template is specific for that artistic lighting style. For example, when Ω_P includes the *Rembrandt* lighting style, Ω_N includes all the other styles (*Paramount, Loop, Split*) and daily photos. Then, the learned template is specific for *Rembrandt*.

Suppose the template is composed of a set of features $\{F_1, \cdots, F_K\}$, a probabilistic model for each image $\mathbf{I} \in \Omega_P$ can be defined in a log-linear form [30, 18]

$$p(\mathbf{I}) = q(\mathbf{I}) \prod_{k=1}^{K} \frac{1}{z_k} \exp\{\lambda_k r_k(\mathbf{I})\} \ , \tag{2}$$

where, $q(\mathbf{I})$ is the null distribution of $\mathbf{I}$ without any knowledge of the feature responses r_k, λ_k is the weight parameter, and z_k is normalizing constant for the factors. In our case, we use the template to describe the new information of photographs in Ω_P compared with that in Ω_N, and leave the rest information in $q(\mathbf{I})$. In this way, $q(\mathbf{I})$ models the negative samples.

We select an informative subset of the features for the template, rather than using the whole feature set, for two reasons: (1) features tend to be correlated, and (2) we prefer a simple template for both capability of generalization and computational efficiency. Since selecting an optimal subset of features simultaneously is a non-trivial task, we adopt a stepwise feature pursuit algorithm [31, 18] to select one feature at each step to construct $p(\mathbf{I})$.

Algorithm 1 Stepwise pursuit for lighting contrast features

Input:

Ω_P, positive samples; Ω_N, negative samples;

$F_C : \{F_1, \cdots, F_L\}$, candidate feature set;

Output:

T: the learned lighting template;

Compute feature responses and information gain:

for each feature in F_C **do**
  Compute its feature response on each image in Ω_P and Ω_N by 1;
  Compute its λ and $\log z$ by 4;
end for

Pursuit lighting contrast features:

Let lighting template $T \leftarrow \phi$, $t \leftarrow 1$
while $t < Thresh_t$ **do**
  Select the max-gain feature $F_{(t)}$ according to 3;
  $T \leftarrow F_{(t)}$, $t \leftarrow t + 1$;
  Delete $F_{(t)}$ and features in the same region of $F_{(t)}$ from F_C;
end while
return T;

Given a candidate feature set, we begin with an empty template corresponding to the negative examples distribution $p_0(\mathbf{I}) = q(\mathbf{I})$ at step 0. Then at each step t, we choose the max-gain feature

$$\begin{aligned} F_{(t)} &= \arg\max_{F_k} \mathrm{KL}(p_t(r_k(\mathbf{I}))||p_{t-1}(r_k(\mathbf{I}))) \\ &\approx \arg\max_{F_k} \mathrm{KL}(p_t(r_k(\mathbf{I}))||q(r_k(\mathbf{I}))) \\ &\approx \arg\max_{F_k} \mathrm{KL}(h(R_k^P)||h(R_k^N)). \end{aligned} \tag{3}$$

On our dataset, $\mathrm{KL}(\cdot||\cdot)$ is the Kullback-Leibler divergence, and $h(\cdot)$ denotes the histograms over a set of local lighting contrast feature responses, R_k^P and R_k^N denote the responses of lighting contrast feature k of positive and negative samples respectively (Fig. 4). The second approximation $\arg\max_{F_k} \mathrm{KL}(p_t(r_k(\mathbf{I}))||q(r_k(\mathbf{I})))$ applies empirical estimates of marginal probabilities with instances in the dataset. For the first approximation, where we assume $p_{t-1}(r_k(\mathbf{I})) \approx q(r_k(\mathbf{I}))$, to be feasible, we apply local inhibition in the template learning process to reduce the correlations among the selected features in the pursuit steps. Noticing the fact that the rectangular regions have no or very little overlap with each other, we simply assume independence of features in different rectangular regions, and use the inhibition strategy that selects only one feature from each rectangular part of the face. Meanwhile, for step t, the parameters $\lambda_{(t)}$ and $z_{(t)}$ can be computed by solving the system

$$\begin{aligned} \mathrm{E}_q\left[\frac{1}{z_{(t)}} \exp\{\lambda_{(t)} r_{(t)}(\mathbf{I})\} r_{(t)}(\mathbf{I})\right] &= \mathrm{E}_{p_t}[r_{(t)}(\mathbf{I})] \\ z_{(t)} &= \mathrm{E}_q[\exp\{\lambda_{(t)} r_{(t)}(\mathbf{I})\}] \end{aligned} , \tag{4}$$

with $\mathrm{E}_q[\cdot] \approx \mathrm{Mean}_{\Omega_N}(\cdot)$ and $\mathrm{E}_{p_t}[\cdot] \approx \mathrm{Mean}_{\Omega_P}(\cdot)$ as empirical estimates according to our dataset. Algorithm 1 describes

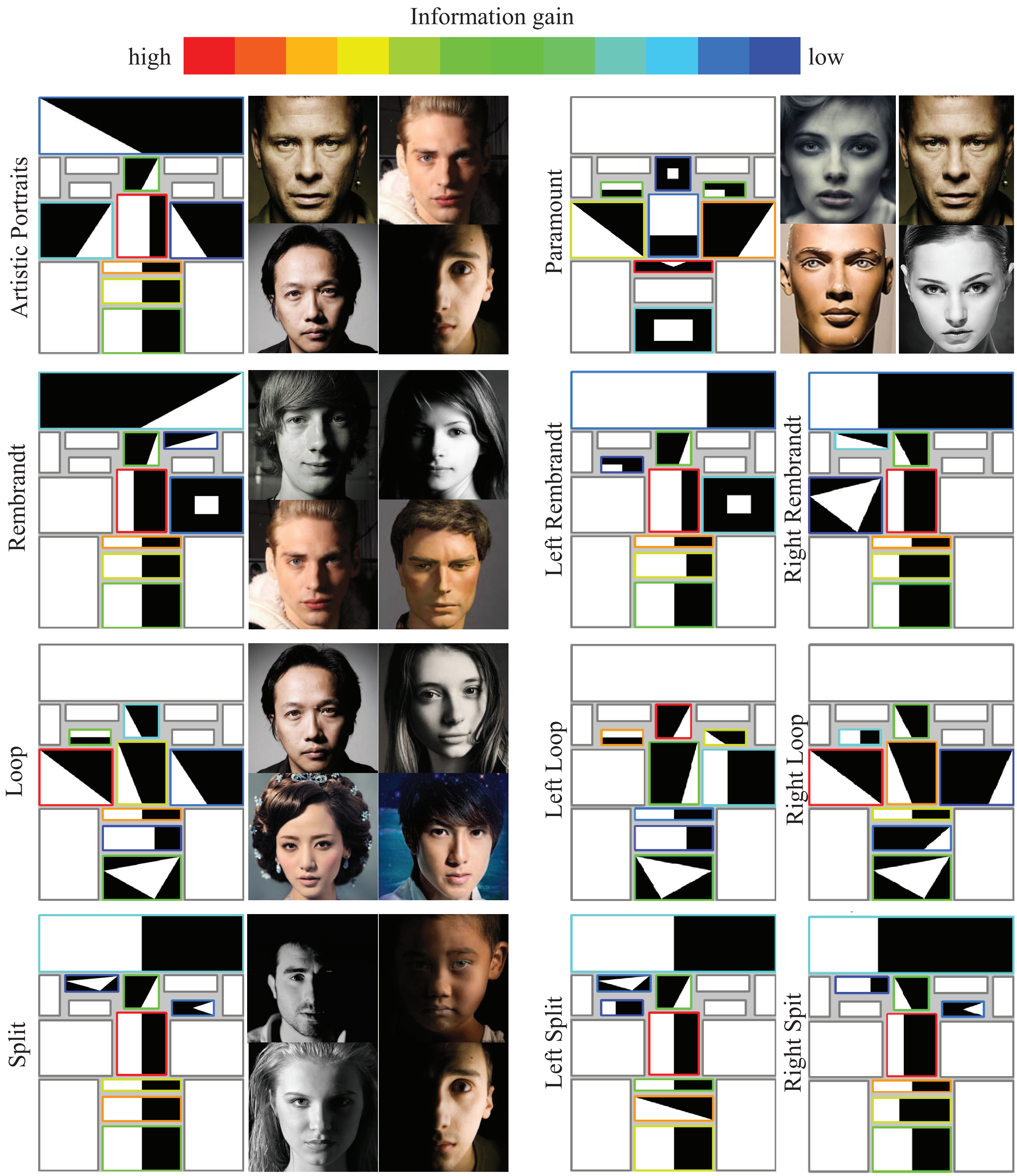


Fig. 5. The learned artistic lighting templates. The first row is the common template of the whole artistic lighting style and the template of *Paramount* style. *Paramount* style has no sub-classes according to key lighting directions. The second to the fourth rows show the learned templates for the other three typical artistic lighting styles. In each row, from the left to the right are the common template of this style, the samples of this style, and the two templates of the two sub-classes. All templates are shown with ranks marked by colored boundaries corresponding to the legend.

the stepwise pursuit algorithm for lighting contrast features. In order to keep the independence between the local contrast features of the template, we find a most informative feature for every face region, so that the $Thresh_t$ is equal to the amount of face region.

Fig. 5 displays the learned lighting templates (For clarity, the top 8 local contrast features of the template are shown). More discriminative local contrast types are selected than just left-vs-right, top-vs-bottom, and center-vs-periphery in [27]. For example, the triangle contrast type learned in the *Right Rembrandt* style captures the triangle light well. The *Paramount* template has the top triangle contrast type between the nose and the mouth. This is the most informative feature the *Paramount* style has.

The results show that the left and the right templates of each style are not strictly symmetrical. For example the left-split template is not a mirrored version of right-split template. Because the training images of the left and the right sub-classes are not symmetrical due to the asymmetries of human faces and the key light direction. Besides, individual differences and artistic creations of different artists also generate the non-symmetrical sub-classes.

The learned templates correspond to the studio lighting well. Most of the selected features are of the generalized left-vs-right spatial type, which matches the common lighting strategies for portrait photography in studios as photographers usually change the directions of light in the horizontal dimension. The parts for the nose and the area between the mouth and the nose are the top significant areas, which make sense because these two parts are relatively complex in geometry and have various appearances under different illumination conditions. It is worth noticing that all the color features are ignored by the template learning process. Our dataset contain only 63 monochrome portraits, and the monochrome portraits make up only a small proportion (850 portraits). This proves that color is not a distinctive lighting contrast feature.

V. Portrait Lighting Analysis

With the learned templates of the artistic lighting styles, we make analysis portrait lighting in this section, including lighting styles classification, quantitative assessment, and face illumination matching.

A. Classification

The template T_A for artistic lighting style A is learned by all the portraits with artistic lighting style A against portraits with other lighting style. It can be used for lighting style classification. Given an input portrait photo $\mathbf{I}$, we can calculate a template matching score with T_A by

$$\begin{aligned}\text{MatchingScore}(\mathbf{I}) &= \log \frac{p(\mathbf{I})}{q(\mathbf{I})} \\ &= \sum_{k=1}^{K} (\lambda_k r_k(\mathbf{I}) - \log z_k).\end{aligned} \tag{5}$$

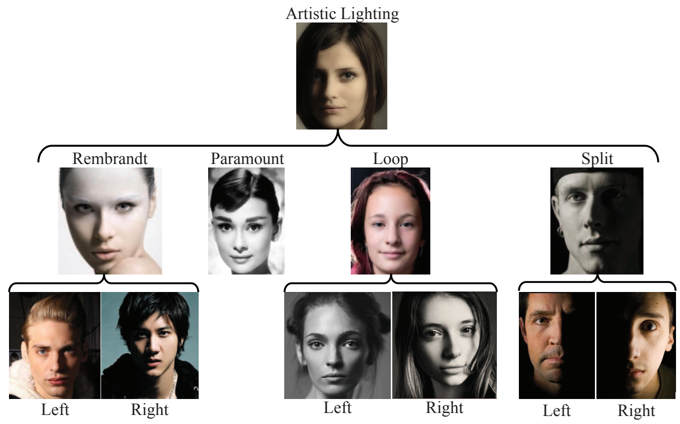


Fig. 6. Weakly supervised clustering. All the artistic lighting styles are divided into four typical artistic lighting styles. We make weakly supervised clustering within each style. Each style of *Rembrandt*, *Loop* and *Split* can be automatically divided into two sub-classes: Left and Right. The *paramount* has no such sub-classes because the key lighting direction is always above its subject.

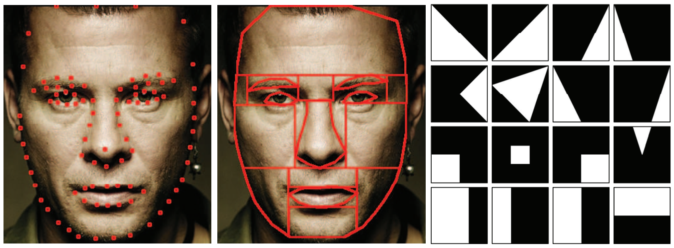
This follows a probability ratio test formulation. If the matching score is greater than a learned threshold corresponding to a certain significance level (e.g., the equal error rate (EER) threshold), the test photo is classified as lighting style P; otherwise, it is classified as other lighting style.

The template learned by portraits with artistic lighting against that without artistic lighting can be used to classify a test portrait as artist lighting or not. We can also learn the template by the portraits with one lighting style against the portraits with other lighting styles and the portraits without artistic lighting. The template learned for each lighting style can be used to classify a test portrait as that lighting style or not.

We focus on 4 popular artistic lighting styles: *paramount*, *Rembrandt*, *Loop* and *Split*. We invite the artists to divide the portraits with artist lighting in our dataset into those 4 styles.

Our dataset contains between-class and within-class variations. The within-class variation is mainly caused by the key lighting direction in each artistic lighting style. We make weakly supervised clustering [32] and the cluster adopted in our paper is APCluster [33], and the face illumination distance is defined in Section V. We empirically set the preference parameter of APCluster as 4 in our experiments.

We learn two artistic lighting templates per style to account for more within-class variation through weakly supervised clustering. Each style of *Rembrandt*, *Loop* and *Split* can be divided into two sub-classes: Left and Right, which means the key light comes from left and right respectively. The *paramount* has no such sub-classes because the key lighting direction is always above the subject. Then we obtain $3 \times 2 + 1 = 7$ templates for $3 \times 2 + 1 = 7$ styles (Fig. 6). After that we make one-to-all classification in our experiments as what we have done in our multi-style classification. Fig. 9 (the third row) shows that the classification performance is



(a) (b) (c)

Fig. 7. Definition of FID. (a) The portrait photo with facial landmarks, (b) The facial mask according to (a), (c) The 16 selected contrast types.

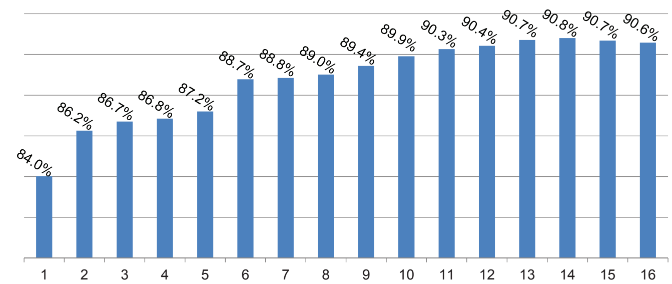


Fig. 8. Classification accuracy with the number of local contrast features. The accuracy of the classification of all the artistic lighting styles is tested using the local contrast features designed in this paper.

improved by the weakly supervised clustering.

B. Quantitative Assessment

In addition to classification, it is usually more useful to have a reasonable quantitative assessment of a portrait lighting usage. This can be achieved by extending our learning algorithm to a regression framework.

In an empirical manner, we define the lighting quality of a portrait photograph $\mathbf{I}$ as the probability p that it is better than another randomly chosen portrait photograph $\mathbf{J}$, namely,

$$p = \mathrm{E}_{f(\mathbf{J})}\left[\mathbf{1}(\mathbf{I} \text{ wins against } \mathbf{J})\right], \tag{6}$$

where, f is the distribution of all portrait images, and $\mathbf{1}(\cdot)$ is the indicator function. To predict the score p, we randomly choose a fixed number n of photographs to compare with $\mathbf{I}$, then winning number of $\mathbf{I}$ should follow a binomial distribution $\mathrm{binom}(n, p)$. We do such comparison experiments on example portraits and obtain their scores. We can then estimate the effects of the features on the score using logistic regression [34] by fitting the model

$$\begin{aligned} \log \frac{p}{1-p} &= \sum_{k=1}^{K} (\lambda_k r_k(\mathbf{I}) - \log z_k) \\ &= \lambda_0 + \sum_{k=1}^{K} \lambda_k r_k(\mathbf{I}), \end{aligned} \tag{7}$$

and this model is able to output a score $p \in (0, 1)$ for the quality of test photographs.

C. Face Illumination Matching

In order to quantify the illumination difference between two portrait images, we define a new **Face Illumination Descriptor (FID)** based on the local lighting contrast features. We extend the rectangle region to non-regular region with a mask of each parts such as forehead, nose, eyebrows, mouths, etc. (Fig. 7 (b)). In our previous work [2], the FID contains only 3 contrast types (left-vs-right, top-vs-bottom, and center-vs-periphery). In this paper, we use the more informative and a new set of lighting contrast features.

In the templates learning process, we accumulate the KL divergence between $h(R_k(P))$ and $h(R_k(N))$ for every local contrast type, target channel and target statistics. 5 contrast types with nearly zero accumulated KL divergence are ignored, and 16 contrast types are selected (Fig. 7 (c)). The L channel and the mean statistic have the highest accumulated KL divergence, meaning that the selected 16 contrast types, L channel and mean statistic can describe the characteristics of lighting contrast .

According to the 16 selected local contrast types, the non-regular region is divided into two non-regular subregions A^* and B^*. In order to model the lighting direction, we do not use the absolute value of the difference,

$$r^*(\mathbf{I}) = \mu_{A^*} - \mu_{B^*}. \tag{8}$$

The FID of portrait photo $\mathbf{I}$ is defined as a set of the local contrast feature responses $r_i^*(\mathbf{I})$ between the two subregions of the 16 non-regular regions on L channel using the mean statistic, which is a 16×16 d vector.

Face Illumination Distance. We define the illumination distance between two portrait images $\mathbf{I}_1$ and $\mathbf{I}_2$ by **FID** as

$$\mathrm{D}(\mathbf{I}_1, \mathbf{I}_2) = \sqrt{\Sigma_{i=1}^{256} (r_i^*(\mathbf{I}_1) - r_i^*(\mathbf{I}_2))^2}. \tag{9}$$

As an application of our lighting contrast features and the **FID**, we use the face illumination distance to match the portrait with the most similar illumination effects as those of the query one.

VI. Experiments

We validate our template and lighting contrast features in the aspect of classification, quantitative assessment and face illumination matching. To the best of our knowledge, there are no open data sets for our experiment. Therefore, we set up our own dataset. We collect 350 example portraits with artistic lighting styles from 3 sources: (1) masterpieces of portrait photography from famous photographers (e.g., Yousuf Karsh, Arnold Newman), (2) collections from professional photography websites (e.g., photo.net, portrait-photos.org), and (3) scanned copies from professional portrait photography books focusing on lighting [11, 12, 13, 14].

The 500 portraits with no obvious lighting effect for comparison are obtained from two main sources (Most of them are daily portraits): (1) popular photo hosting websites (e.g., flickr.com), and (2) image search engines' results for the keywords "face" and "daily life" (e.g., images.google.com).

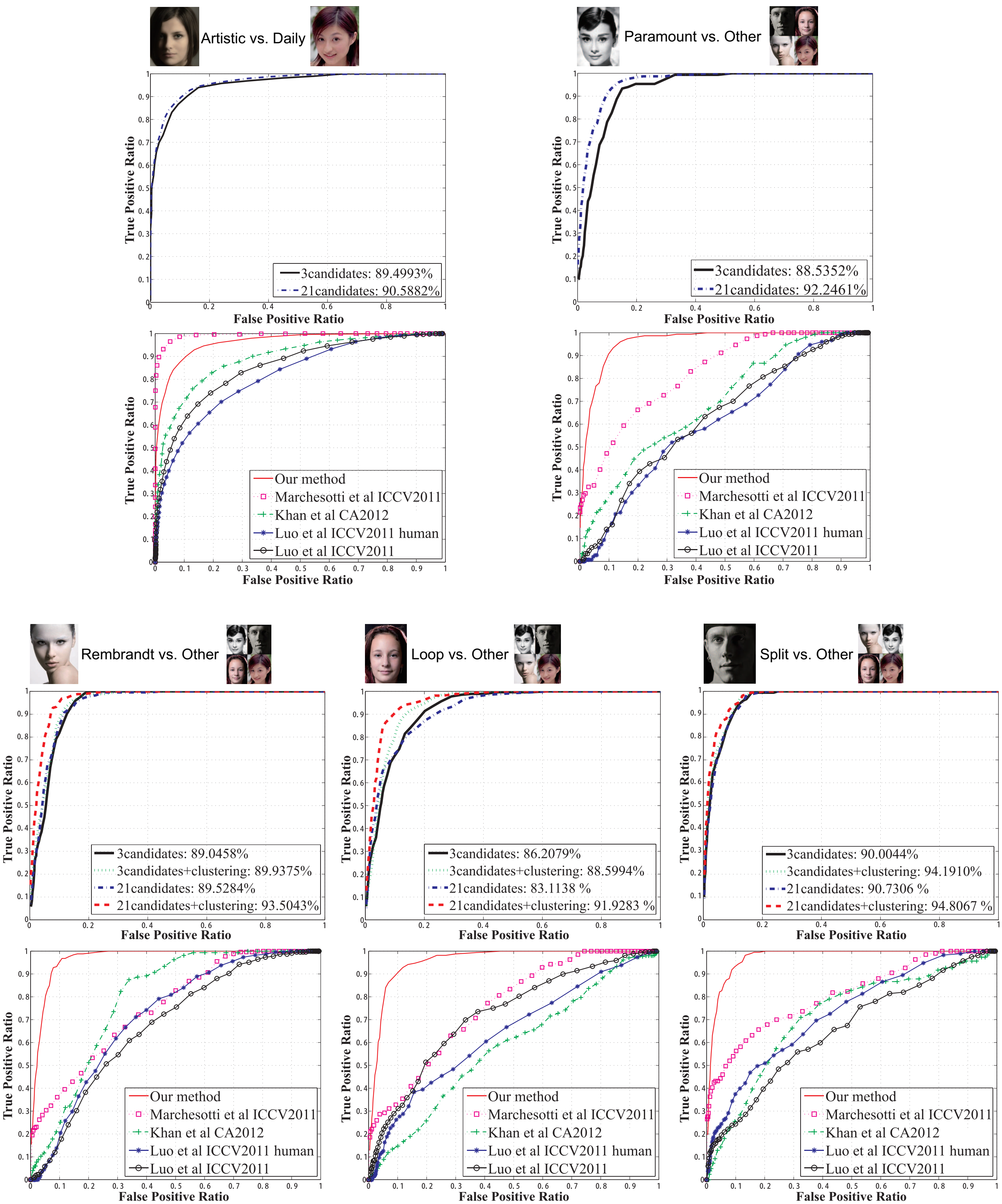


Fig. 9. In the first and third rows, we show the average accuracy rates and ROC curves of 5 times 5 folds cross validation (one-to-all classification). The *3 candidates* and *21 candidates* mean the classification results using 3 basic contrast types of [27] and the proposed 21 new contrast types. The *3 candidates+clustering* and *21 candidates+clustering* mean the classification results after weakly supervised clustering, which are all higher than those without clustering. All the highest accuracy rates are in *21 candidates+clustering*. In the second and fourth rows, we show the ROC curves of 5 times 5 folds cross validation in the classification of artistic lighting, compared with the methods in Marchesotti et al. [9], Khan et al. [10], Luo et al. human [8] (human face combined features) and Luo et al. [8] (all features).

We try to randomize irrelevant factors by spanning over various poses, ethnic groups, etc. All images in our dataset are aligned using AAM [35] to a standard frontal face before running our other tasks. Due to the shadow effects, the AAM sometimes cannot get the accurate position of facial feature points and some manual corrections are thus needed. The face image with inaccurate feature points can be detected by [36] in the future work.

The lighting template consists of lighting contrast features. Fig. 8 shows the lighting classification accuracy with different $Thresh_t$. The classification accuracy can reach 84.0% with only one lighting contrast feature, demonstrating good correspondence between our template and the experience of artists, who usually judge the lighting style by a small number of face regions.

Which shows that our template corresponds to the artist experience well: the artists usually judge the lighting style by small number of face regions.

A. Classification Results

Fig. 9 shows our classification performance. We compare the performance of using candidate features with 3 basic contrast types (left-vs-right, top-vs-bottom, and center-vs-periphery) in [27] and with 21 contrast types designed in this work. As shown by the ROC curves in the first row of Fig. 9, the performance with 21 contrast types is better than that with the 3 basic contrast types. Using the weakly supervised clustering, we obtain $3 \times 2 + 1 = 7$ templates for $3 \times 2 + 1 = 7$ styles (*Left Rembrandt, Right Rembrandt, Paramount, Left Loop, Right Loop, Left Split and Right Split*). If the test portrait belongs to the sub-class, its label is the same as that of the father node of this sub-class note in Fig. 6. The third row of Fig. 9 shows that the multi-style classification accuracy is improved by the weakly supervised clustering. The classification performance (the areas under ROC curves) is quite good using the relatively small training sample size.

Comparison with Previous Works. As shown in the second and fourth rows of Fig. 9, we compare our method (21 contrast types) with previous methods on our dataset (also using clustering for Rembrandt, Loop and Split). The low level generic features proposed by Marchesotti et al. [9] are designed for all kinds of images. However, the portrait lighting is implicitly encoded in the generic features. Other features proposed by Luo et al. [8] (human face combined features) and Khan et al. [10] are designed for the portrait aesthetic assessment, but are not designed specifically for the lighting usage. Our method outperforms others (including using all features proposed by Luo et al. [8]) in the classification experiments of four artistic lighting styles (considering the areas under ROC curves).

B. Quantitative Assessment Results

For the quantitative assessment task, training data with the quality of photographs are necessary for fitting the regression model (Section V). We obtain the consensus quality scores (i.e., winning probabilities against randomly chosen images) with human experiments.

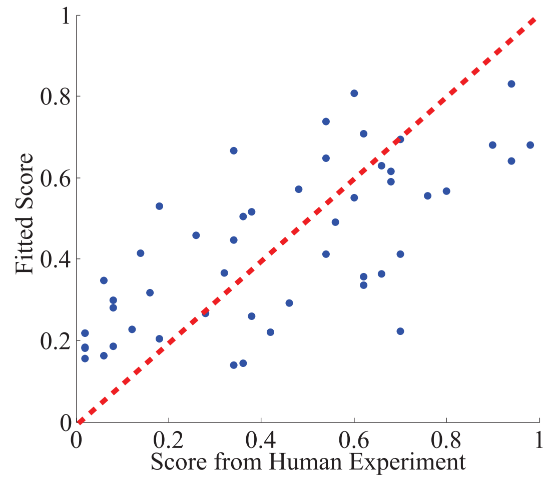


Fig. 10. Goodness-of-fit visualization of the logistic regression for quality assessment.

TABLE I
LOGISTIC REGRESSION COEFFICIENTS.

Feature	Est.	Std.Err.	z-score	p-value
(I)	-2.15	0.13	-17.14	$< 2e-16$
F_1	0.50	0.74	0.68	0.50
F_2	0.42	0.67	0.64	0.53
F_3	4.73	1.23	3.86	< 0.01
F_4	-0.35	0.79	-0.45	0.65
F_5	0.66	0.62	1.05	0.30
F_6	0.42	0.74	0.57	0.57
F_7	1.50	0.51	2.95	< 0.01
F_8	8.20	0.88	9.34	$< 2e-16$

Human Experiments. We randomly choose 100 portraits (either with or without artistic lighting style) from our dataset as the training examples. We also ask 15 professional artists from professional art studios and the department of fine arts as test subjects to do the comparisons between photographs.

For each $\mathbf{I}$ in the 100 training images, another 100 random images $\mathbf{J}_1, \cdots, \mathbf{J}_{100}$ from the rest of the dataset are sampled with replacement. Each of the 100 pairs $(\mathbf{I}, \mathbf{J}_1), \cdots, (\mathbf{I}, \mathbf{J}_{100})$ is displayed to a random test subject, who then compare the two images and report their relative rank order in the quality of lighting. In this way, a total of $10,000$ comparisons are performed, and the numbers of wins and losses of the 100 training images are obtained for fitting the prediction model in Section V. Here the replacement ensures the constant probability condition for the assumption of a binomial distribution [34].

Regression. Most features have favorable small p-values in the logistic regression (Table I). Fig. 10 plots the fitted scores vs. the scores obtained from human experiments (50 samples). Due to the small size of the training set and the small number of Bernoulli trials, there is still a considerable residual deviance in the fitting (null deviance: 936.85 on 49 degrees of freedom, residual deviance: 522.14 on 41 degrees of freedom). Since no heteroskedasticity or non-normal effects are noticed [34], we believe that our empirical definition for the aesthetic quality and the experimental design should make good sense.

Fig. 11. Quantitative assessment results. The portraits are displayed with predicted quality scores, and higher scores indicate higher lighting quality.

Fig. 11 shows the portrait lighting quality predicted by our regression model. The test portraits are not included in the training portraits. Experimental result shows that our regression model can assess the portraits lighting quality as human beings do. We use lighting contrast characteristic to predict the portrait lighting quality in this paper. In the future work, we will try to access portrait lighting quality by incorporating the statistical prior obtained from millions of unlabeled images (as in [37]).

C. Face Illumination Matching Results

Similar to [2], we test the face illumination matching in the Yale Face database B and the Extension [38, 39], which contain frontal face photos of 38 subjects under the same 64 lighting conditions (64 point light sources in 64 directions). Due to the dense light directions of Yale database, even humans cannot distinguish the illumination effects of two light directions with small intersection angles. We thus also relax the correct matching criterion, and set that if the same lighting condition as that of query portrait is in the top 3 matched results, we consider it as correct matching (Top 3 hit).

The accuracy rates are shown in Fig. 12. We outperform the FID defined in [2] with only 3 basic contrast types (about 16 per cent). And we substitute the 16 selected contrast types with the 21 contrast types. The average accuracy rate is higher using only the 16 selected contrast types (78.0%) than using the whole 21 contrast types (77.6%). This can validates the representativeness of the selected features.

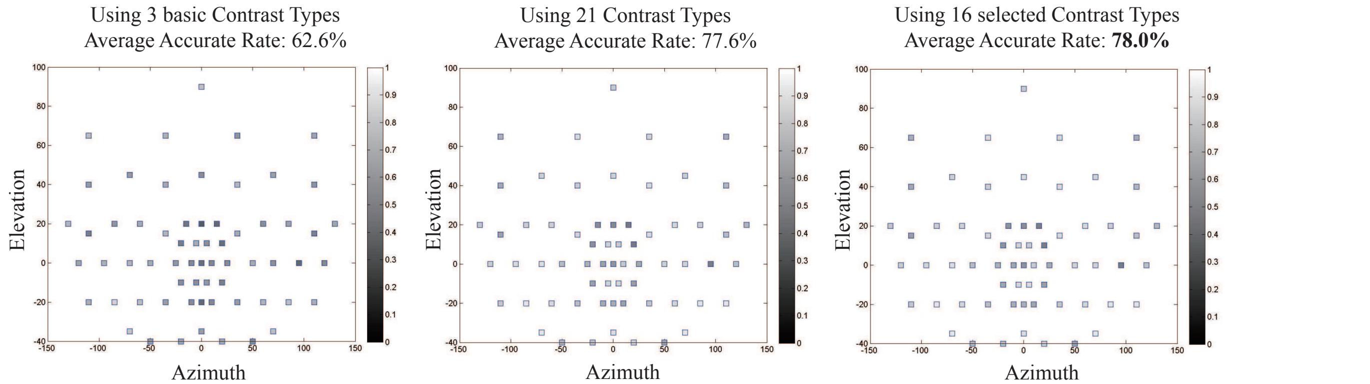


Fig. 12. Accuracy of face illumination matching on Yale Face database. The distribution of all the 64 light sources is shown with each direction as a square. The average accuracy of each direction is shown as the graylevel of the small squares. The whiter the squares are, the more accurate they are considered to be. The light source directions are with respect to the camera axis, which is perpendicular to human faces for frontal view photos. The azimuth is the horizontal angle, while the elevation is the vertical angle. We compare the performances of using 3 basic contrast types used in [27], 21 and 16 selected contrast types, and the average accuracy rates are 62.6%, 77.6% and 78.0%, respectively.

VII. Conclusion and Future Work

In this paper, we design a novel set of local lighting contrast features (21 contrast types) to model various local lighting patterns. Artistic lighting templates consisting of the local lighting contrast features are learned from example portraits. Based on the local lighting contrast features and the learned templates, we make various analyses of artistic portrait lighting including: portrait lighting classification, quantitative assessment and face illumination matching.

The method in this paper is more suitable for front view portraits. In our future work, we will collaborate with psychologists to find the psychological explanation of our lighting contrast features. We will extend the local contrast features to side view portraits. Besides portrait images, portrait lighting in videos are also significant for aesthetic feelings. The aesthetic assessment of the dynamic lighting can be our future work.

Acknowledgment

We would like to thank the anonymous reviewers for their suggestions in improving this paper. We would also like to thank Song-Chun Zhu and Mingtian Zhao for their insightful ideas. Parts of the results and figures presented in this paper have previously appeared in our previous work [27].

References

[1] X. Chen, M. Chen, X. Jin, and Q. Zhao, “Face illumination transfer through edge-preserving filters,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011.

[2] X. Chen, X. Jin, Q. Zhao, and H. Wu, “Artistic illumination transfer for portraits,” *Comput. Graph. Forum*, vol. 31, no. 4, pp. 1425–1434, 2012.

[3] X. Chen, H. Wu, X. Jin, and Q. Zhao, “Face illumination manipulation using a single reference image by adaptive layer decomposition,” *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4249–4259, 2013.

[4] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, “Estimating the natural illumination conditions from a single outdoor image,” *Int. J. Comput. Vision*, vol. 98, no. 2, pp. 123–145, 2012.

[5] M. K. Johnson, D. G. Stork, S. Biswas, and Y. Furuichi, “Inferring illumination direction estimated from disparate sources in paintings: an investigation into jan vermeer’s girl with a pearl earring,” in *Proc. Computer Image Analysis in the Study of Art*, 2008, pp. 1–12.

[6] K. Nishino and S. K. Nayar, “Using eye reflections for face recognition under varying illumination.” *ACM Trans. on Graphics*, vol. 23, no. 3, pp. 704–711, 2004.

[7] Z. Wen, Z. Liu, and T. S. Huang, “Face relighting with radiance environment maps.” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. 158–165.

[8] W. Luo, X. Wang, and X. Tang, “Content-based photo quality assessment,” in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2206–2213.

[9] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, “Assessing the aesthetic quality of photographs using generic image descriptors,” in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 1784–1791.

[10] S. S. Khan and D. Vogel, “Evaluating visual aesthetics in photographic portraiture,” in *Proc. Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging*, 2012, pp. 55–62.

[11] B. Hurter, *The best of photographic lighting — techniques and images for digital photographers*, 2nd ed. Amherst Media, 2007.

[12] F. Hunter, S. Biver, and P. Fuqua, *Light: science and magic: an introduction to photographic lighting*, 3rd ed. Focal Press, 2007.

[13] C. Grey, *Master lighting guide for portrait photographers*. Amherst Media, 2004.

[14] D. Prakel, *Basics photography: lighting*. AVA Publishing, 2007.

[15] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. I–511 – I–518.

[16] C. P. Papageorgiou, M. Oren, and T. Poggio, “A general

framework for object detection," in *Proc. Int. Conf. Comput. Vis.*, 1998, pp. 555–562.

[17] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, 2002.

[18] Z. Si, H. Gong, Y. N. Wu, and S.-C. Zhu, "Learning mixed templates for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 272–279.

[19] G. Xing, X. Zhou, Y. Liu, X. Qin, and Q. Peng, "Online illumination estimation of outdoor scenes based on videos containing no shadow area," *Sci China Ser F-Inf Sci*, vol. 56, no. 3, pp. 1–11, 2013.

[20] X. Chen, X. Jin, and k. Wang, "Lighting virtual objects in a single image via coarse scene understanding," *Sci China Ser F-Inf Sci*, vol. 57, no. 9, pp. 1–14, 2014.

[21] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios, "Illumination estimation and cast shadow detection through a higher-order graphical model." in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 673–680.

[22] J. Zhu, K. G. G. Samuel, S. Z. Masood, and M. F. Tappen, "Learning to recognize shadows in monochromatic natural images." in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 223–230.

[23] K. Nishino, P. N. Belhumeur, and S. K. Nayar, "Using eye reflections for face recognition under varying illumination." in *Proc. Int. Conf. Comput. Vis.*, 2005, pp. 519–526.

[24] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps," in *SIGGRAPH*, 2001, pp. 497–500.

[25] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: Automatically replacing faces in photographs," *ACM Trans. on Graphics*, vol. 27, no. 3, pp. 39:1–39:8, 2008.

[26] H. Han, S. Shan, L. Qing, X. Chen, and W. Gao, "Lighting aware preprocessing for face recognition across varying illumination." in *Proc. Eur. Conf. Computer Vision*, 2010, pp. 308–321.

[27] X. Jin, M. Zhao, X. Chen, Q. Zhao, and S.-C. Zhu, "Learning artistic lighting template from portrait photographs," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 101–114.

[28] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 679–714, 1986.

[29] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Trans. Info. Theory*, vol. 37, no. 1, 1991.

[30] S. Della Pietra, V. Della Pietra, and J. Lafferty, "Inducing features of random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 380–393, 1997.

[31] J. H. Friedman, "Exploratory projection pursuit," *J. Am. Stat. Assoc.*, vol. 82, no. 397, pp. 249–266, 1987.

[32] Z. Si and S.-C. Zhu, "Learning hybrid image templates (hit) by information projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1354–1367, 2012.

[33] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972–976, 2007.

[34] J. J. Faraway, *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Taylor & Francis Group, 2006.

[35] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. Eur. Conf. Computer Vision*, 1998, pp. 484–498.

[36] J. Li, D. Xu, and W. Gao, "Removing label ambiguity in learning-based visual saliency estimation," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1513–1525, 2012.

[37] J. Li, Y. Tian, and T. Huang, "Visual saliency with statistical priors," *Int. J. Comput. Vision*, vol. 107, no. 3, pp. 239–253, 2014.

[38] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, 2001.

[39] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, 2005.

[40] K. Nishino and S. K. Nayar, "Corneal imaging system: Environment from eyes," *Int. J. Comput. Vision*, vol. 70, no. 1, pp. 23–40, 2006.

[41] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.