

Non-Uniform Illumination Pre-Processing for Facial Recognition with Bidimensional Empirical Mode Decomposition

Lyndon Hin-Cheung Chan
Edward S. Rogers Sr. Department of Electrical & Computer Engineering
University of Toronto
Student Number: 999875817
Email: lyndon.chan@mail.utoronto.ca

Abstract—Non-uniform illumination is a significant obstacle to facial image recognition algorithms. Apart from developing illumination-invariant feature descriptors for the images so that facial images of different illuminations can still be classified correctly, non-uniform illumination can be corrected by applying a pre-processing step beforehand. However, non-uniform illumination pre-processing can be a challenging task, since it is difficult to eliminate illumination variations without eliminating meaningful facial feature information. Illumination variations on facial images are dependent on the facial geometry, the illumination characteristics of the facial surface, and both the direction and type of illumination, which are very difficult to estimate from merely a facial image. This report proposes using a novel method for non-uniform illumination pre-processing which, among other steps, is based on bidimensional empirical mode decomposition, which approximates well spatial frequency for non-linear and non-stationary data. Experimental results show that this method performs similarly to competing methods and better in the case of small training sets.

I. INTRODUCTION

Facial recognition is the task of finding descriptors for facial images which are both discriminative and invariant to viewing conditions, so that a new facial image can be correctly assigned a class label. This task is most hindered by variations in facial appearance, particularly in the camera and facial pose, facial expression, and partial occlusions. As [21] and [4] show, the biggest factors affecting facial appearance among these are illumination and pose variation. With non-uniform illumination of facial images, bright highlights and dark shadows can accentuate or diminish facial features to the detriment of facial recognition performance.

There are generally two ways to address this problem: (1) to design an illumination-invariant descriptor, or (2) to pre-process the images to normalize the illumination variations. Although many facial recognition algorithms have in-built measures to address illumination variations, extreme variations in lighting often still detrimentally affect recognition performance. For this reason, this report will focus on the pre-processing approach.

The challenge of removing illumination variations from specific facial images lies in the drastic local variations in shadows and highlights due to the illumination direction, type

of lighting, and facial geometry. Estimating all these factors is very difficult from a single two-dimensional greyscale image, so assumptions have to be made about the illumination variations. One approach is to assume that illumination variations are low spatial frequency components of the image, and filter all but mid-frequency components, which are assumed to contain facial structure information useful for recognition. Another approach is to assume that useful facial structure information lies in the mid-intensity levels of the image, and hence enhance the mid-intensity features at the expense of low- and high-intensity features.

Bidimensional Empirical Mode Decomposition offers an alternative approach to decomposing the spatial frequency components of a two-dimensional image and works well for nonlinear and nonstationary image information. In this paper, I propose a novel illumination pre-processing algorithm based on an hybrid exponential-logarithmic transform for intensity correction, BEMD for noise reduction and shadow removal, and DoG (Difference of Gaussian) filtering for edge enhancement.

The rest of the paper is organized as follows. In section II, I will review the prior art in facial recognition and illumination pre-processing methods. In section III, I will review bidimensional empirical mode decomposition and the prior art in applying it to image analysis. In section IV, I will describe in detail the proposed method and the rationale behind the algorithm. In section V, I will display the experimental results for classification accuracy of various facial recognition algorithms after applying the proposed method and other illumination pre-processing methods. And I will conclude in section VI, summarizing my findings and their significance.

II. REVIEW OF FACIAL RECOGNITION AND ILLUMINATION PRE-PROCESSING

A. Facial Recognition

In general, facial recognition methods can be separated into three categories: (1) appearance-based methods, (2) geometric feature-based methods, and (3) soft computing-based methods.

Appearance-based methods. Appearance-based methods of facial recognition characterize facial images by summariz-

ing the image intensities at either the global scale (i.e. appearance of the overall image), intermediate scale (i.e. appearance of sub-image objects), or the local scale (i.e. appearance of local neighbourhoods centred around specific pixels). Holistic approaches capture global information from an image, and these include the PCA (Principal Component Analysis) or Eigenfaces [26], KPCA (Kernel PCA) [31], and LDA (Linear Discriminant Analysis) [32] methods. Hybrid approaches, on the other hand, capture both global and local information, such as Gabor features [16] and LBP (Local Binary Patterns) [2].

Geometric Feature-based methods. Geometric feature-based methods use a-priori information to characterize facial images as collections of specific image features, which are certain geometric shapes or patterns assumed to compose a facial image. These methods include EBGM (Elastic Bunch Graph Matching) [28], the combined SIFT descriptor and Common Representation Matching Space projection of [12], and the Gabor-Kernel face recognition method [15].

Soft Computing-Based Face Recognition. Soft computing-based methods use fuzzy logic and Bayesian uncertainty analysis to classify facial images. Some methods use ANNs (Artificial Neural Networks), such as the hybrid supervised/unsupervised FFNN (Feed-Forward Neural Network) by Intrator et al. [8], the BPNN (Feed-Forward Back Propagation Neural Network) by Agarwal et al. [1], the Hybrid Gabor-Neural Network by Khatun and Bhuiyan [11] which uses Gabor filters as neural network features, and the learned LBP method by Jing and Zhang [9]. Fuzzy-based approaches use non-linear constraints to learn non-linear characteristics in facial images - these include Fisherface [30], Fuzzy LDA [13], Fuzzy SVM (FSVM) [24], DT-FSVM which combines FSVM with a decision tree [24], and Robust Kernel Fuzzy Discriminant Analysis (RKFDA) which is a nonlinear robust variant of LDA [6]. And Genetic Algorithm (GA)-based approaches use a stochastic search and optimization algorithm to design optimal facial recognition solutions - Sinha and Singh use a Breeder Genetic Algorithm (BGA) to design an optimal Composite Wavelet Matched Filter (CWMF) [23], Peretz et al. created Local Matching Gabor (LMG) to select Gabor jets with entropy and Genetic Algorithms [20], and Melin et al. created a Modular Neural Network (MNN) for facial recognition [17].

B. Illumination Pre-Processing

Generally, illumination pre-processing methods can be separated into two categories: (1) intensity-based methods, and (2) frequency-based methods.

Intensity-based Methods. Intensity-based methods of illumination pre-processing assume that image illumination variations consists of image pixels with either very high or very low image intensity. Hence, they transform the facial image into another number space where higher resolution is provided to mid-intensity levels which presumably hold useful information, and less resolution is provided to high and low intensities. Oftentimes, either a logarithmic function is applied to the image or else the image intensity values are raised to the

exponent of some value less than one to normalize the image intensities. Histogram equalization is one approach, used to redistribute the image intensities to cover the full available intensity range. Edge extraction enables filtering out all but edge features which presumably carry useful information.

Frequency-based Methods. Frequency-based methods of illumination pre-processing assume that the facial image consists of high-frequency edges and noise, mid-frequency facial structure components, and high-frequency illumination variations (e.g. shadows and highlights). Hence, these methods ignore the high- and low-frequency components of the image and enhance the mid-frequency components of the image. A prominent example of this approach is Jobson et al.'s Multiscale Retinex (MSR) [10], which divides images by smoothed versions of itself. Wang et al.'s Self Quotient Image (SQI) [27] takes a similar approach. Chen et al.'s Logarithmic Total Variation (LTV) factorizes the facial surface by retaining small intrinsic facial structures [3]. Gross and Brajovic's GB anisotropic smoothing relies on an iterative estimation of the blurred version of the original image [5]. A more thorough comparison of the various frequency-based illumination pre-processing methods available is provided by [22].

III. REVIEW OF BIDIMENSIONAL EMPIRICAL MODE DECOMPOSITION

Empirical mode decomposition (EMD) was first introduced as the first stage of the Hilbert-Huang transform (HHT) [7], which was originally used to decompose any one-dimensional data series into a finite sum of component "intrinsic mode functions" (IMFs). These intrinsic mode functions (1) have approximately the same number of zero-crossings and local extrema (their counts not differing by more than one), and (2) have symmetric envelopes defined by the local maxima and minima. The IMFs are extracted in a process called sifting; the sifting process for a discrete function $x(n), n = 1, 2, \dots, N$ in I stages consists of the procedure outlined in 1:

In the Hilbert-Huang transform, the next step after the EMD step is to apply the Hilbert transform to the IMFs, and thus yield instantaneous frequency spectra for each time-series value.

Nunes et al. extended Huang's original one-dimensional empirical mode decomposition only (and omitted the Hilbert transform) to two-dimensional image data as the so-called Bidimensional Empirical Mode Decomposition and successfully applied it to image and texture analysis [18] [19]. In this method, an image is decomposed two-dimensional IMF functions called Bidimensional IMFs (BIMFs) which are computed from two-dimensional envelopes on two-dimensional local extrema. Following Nunes' work, other researchers investigated other methods of applying BEMD to image recognition tasks to varying degrees of success. Linderherd used BEMD to obtain pixel-wise feature vectors for texture images [14], while Xie used BEMD to decompose face images into components of varying spatial frequency and then reduce shadows [29].

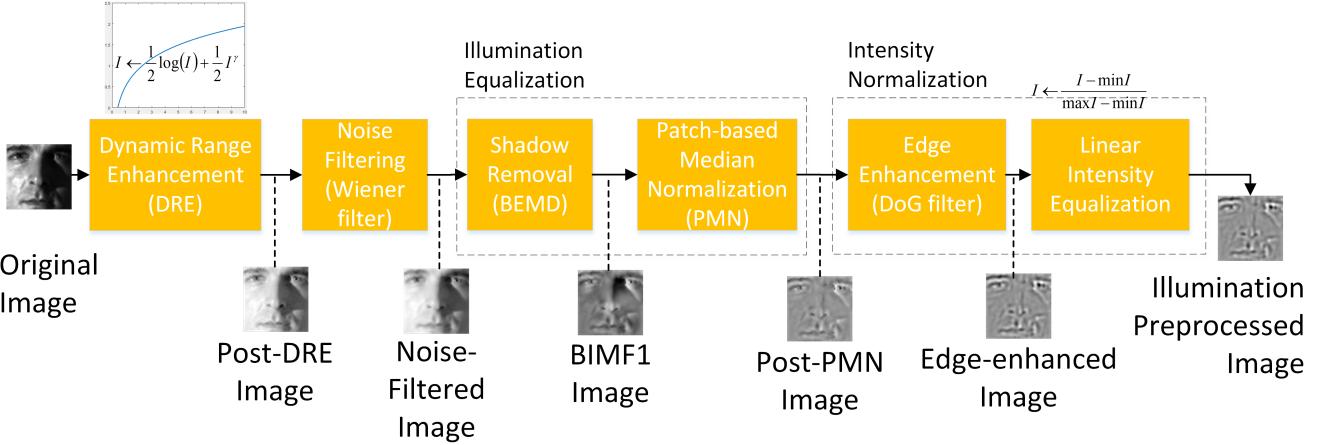


Fig. 1. A top-level overview of the proposed solution, with the intermediate images between steps displayed.

Result: I IMFs ($c_1(n), c_2(n), \dots, c_I(n)$) and a residue $r_I(n)$
Set first stage's initial proto-IMF to $h_{11}(n) = x(n)$ and the 0-th IMF $c_0(n)$ to $x(n)$;
while $i = 1, 2, \dots, I$ **do**
 Set $j = 1$, $SD_{i1} = 0.2$;
 while $SD_{ij} \geq 0.2$ **do**
 Increment j by 1;
 Identify local minima and maxima in $h_{i(j-1)}$;
 Find upper envelope by connecting local maxima by cubic spline, $u_{i(j-1)}(n)$;
 Find lower envelope by connecting local minima by cubic spline, $v_{i(j-1)}(n)$;
 Find the envelope mean,
 $m_{i(j-1)}(n) = \frac{1}{2}u_{i(j-1)}(n) + \frac{1}{2}v_{i(j-1)}(n)$;
 Find the proto-IMF
 $h_{ij}(n) = h_{i(j-1)}(n) - m_{i(j-1)}(n)$ Find the sum of difference between $h_{ij}(n)$ and $h_{i(j-1)}$,
 $SD_{ij} = \sum_{n=1}^N \frac{|h_{i(j-1)}(n) - h_{ij}(n)|^2}{h_{i(j-1)}^2(n)}$;
 end
 Set i -th IMF as current proto-IMF $c_i(n) = h_{ij}(n)$;
 Set i -th initial proto-IMF to the residue
 $h_{11}(n) = c_{i-1}(n) - h_{(i-1)1}(n)$;
end
Set I -th residue $r_I(n) = c_{I-1}(n) - h_{(I-1)1}(n)$;

Algorithm 1: One-dimensional empirical mode decomposition

IV. PROPOSED SOLUTION

In this section, I will describe in detail my proposed method for illumination pre-processing. It uses a series of steps chosen to enhance dynamic range, filter out image noise, equalize illumination, and normalize intensity, while still preserving the distinguishing elements of facial images necessary for recognition. These steps and the intermediate images between steps are visualized in IV.

A. Dynamic Range Enhancement

This is a nonlinear operation applied to the image grey levels in order to enhance the image's dynamic range. In a non-uniformly illuminated image, the dark and light regions are saturated, making it difficult for facial recognition algorithms to extract distinguishing details in either. By enhancing dynamic range, further steps in the illumination pre-processing pipeline are able to extract more meaningful information from formerly saturated regions. Xie proposes to use a natural logarithm in [29], while Tan et al. claim that gamma correction with $\gamma = 0.2$ works best. After some experimentation, I noticed that the logarithm transform tended to retain more noise in the dark saturated regions, while the gamma correction transform tended to over-correct the dark saturated regions. Hence, I use a uniformly weighted sum of the natural logarithm and the gamma correction (see IV-A for plots of possible dynamic range enhancement transform curves).

For a grey-level image I , the resultant dynamic-range-enhanced version I_d is derived as follows:

$$I_d = \frac{1}{2} \log I + \frac{1}{2} I^{\gamma}, \gamma = 0.2$$

B. Noise Filtering

After enhancing the dynamic range of the original image, there is typically a significant amount of image noise in the formerly-dark regions. Various types of noise can be observed and are consequences of capturing images in low-light conditions with external signal interference. Many images appear to have additive Gaussian noise, which is the result of capturing an image with excessive exposure in low-light settings. In some images, there is periodic noise, which comes from the image sensor or from electronic interference. In a few cases, it appears that certain images have been corrupted, and very little useful information can be extracted. Strangely, it appears that not many illumination pre-processing schemes consider filtering out the noise, perhaps in order to prevent

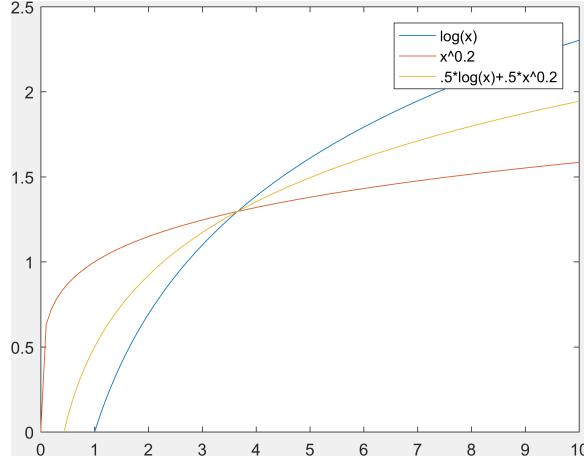


Fig. 2. Plots of various possible dynamic range enhancement transform curves.

useful details from being lost. However, noise filtering is important for my proposed solution, because subsequent steps can be very sensitive to noise artifacts.

After some experimenting with various commonly-used noise filtering algorithms (such as the Gaussian smoothing filter and edge-aware local contrast manipulation), I decided to use a 2×2 adaptive Wiener filter. Most of the image noise appears to have stationary spectra and also to be additive, and for these cases, the Wiener filter minimizes the mean square error between the noisy and denoised images. However, this does not get rid of the periodic image noise, and it seems that nothing can be done in the case of image corruption.

C. Illumination Equalization

This is the step where the non-uniform illumination of the image is equalized to approximately the same level. This operation consists of two stages: (1) Shadow Removal, and (2) Patch-based Median Normalization.

Shadow Removal. In this stage, bidimensional empirical mode decomposition (BEMD) is applied to remove the shadows from the image. Two stages of decomposition are applied, and the image is decomposed as the sum of two bidimensional implicit mode functions (BIMFs) and a single residue image (as shown in IV-C). It is possible to decompose the image into more than three BIMFs, but for the case of facial images, it seems that two levels of decomposition suffice for shadow removal: the first BIMF generally contains the edges of the facial features useful for facial recognition, the second BIMF generally contains the mid-scale shadows of the facial features, and the residue image generally contains the large-scale shadows of the entire face. However, this characterization of extracted features for each level of BEMD is not precise - since BEMD is an empirical technique which iteratively decomposes an image into the sum of envelope functions varying in spatial frequency, it often happens that smaller-scale shadows are extracted by the first BIMF. Nonetheless, mid-size and large-size shadows are always captured in the second BIMF and the residue image, so I remove most of the

shadows by discarding these two components and retaining only the first BIMF.

Patch-based Median Normalization. At this stage, the first BIMF tends to contain the important facial features for recognition for well-illuminated images. However, it tends to contain smaller-scale shadows and non-uniform illumination artifacts as well for poorly-illuminated images. In order to eliminate these residual shadows, I use a patch-based median normalization approach (visualized in IV-C). I sub-divide the image into overlapping square patches, and then set the middle pixel to the difference between the original pixel value and the patch's median intensity. This ensures that smooth patches (whether bright or dark) are normalized to the same normalized level, while patches with sudden illumination changes are smoothed out, since patches with many dark pixels and few bright pixels will take a smaller dark value, and vice versa for patches with many bright pixels and few dark pixels. Through experimentation, I have determined that a patch size of 7×7 gives optimal performance.

D. Intensity Normalization

This is the step where the normalized intensities are normalized so that different images are more comparable for the later classification stage. By this point, most of the illumination variations have been normalized, although the edges of the face are often smoothed out by the patch-based median normalization, and the dynamic range of the image is reduced. Hence, I first apply a Difference-of-Gaussians (DoG) filter to the image to enhance the edges (I found that this works best with a first Gaussian of $\sigma_1 = 1$ and a second Gaussian of $\sigma_2 = 2$). Then, I apply a simple linear image intensity equalization operation to ensure that the image's dynamic range is always normalized to the range $I_n \in [0, 1]$ as follows:

$$I_n = \frac{I - \min I}{\max I - \min I}$$



Fig. 3. A two-stage bidimensional empirical mode decomposition of an image decomposes it as the sum of the first BIMF, the second BIMF, and a final Residue image

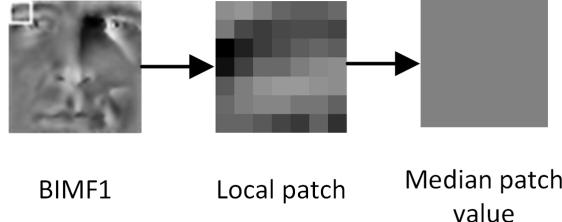


Fig. 4. A 7×7 patch is taken around each pixel and the median value extracted to normalize the intensities

V. EXPERIMENTS

In this section, I will describe the experiments conducted on evaluating the proposed illumination pre-processing algorithm against alternative methods, with the quality of the algorithm being determined by the overall accuracy of the facial recognition classification applied after the illumination pre-processing step. I will first describe the illumination pre-processing algorithms evaluated, the facial recognition algorithms used, the facial image database used, the experimental setup, the experimental results, and a discussion of the significance of the results.

A. Illumination pre-processing Algorithms

The following illumination pre-processing algorithms were evaluated:

Proposal A. This is an earlier iteration of my illumination pre-processing algorithm. It uses the same hybrid logarithmic-gamma dynamic range enhancement as detailed in section IV, extracts the first BIMF in the same way, then it uses a patch-based mean normalization (as opposed to median normalization) extracting on 8×8 patches outputting 2×2 patches instead of single pixels, and then performs linear intensity equalization as before. For Proposals A through D, the BEMD MathWorks File Exchange implementation by "Sasikanth"¹ was used with the bilinear interpolation option.

Proposal B. This is another earlier iteration of my illumination pre-processing algorithm, which is similar to Proposal A, but includes more considerations for noise reduction and patch boundary artifact reduction (caused by the patch-based normalization). First, the image undergoes gamma correction to enhance dynamic range, then noise is filtered out by local contrast reduction, the first BIMF is extracted as usual, then a patch-based mean normalization identical to that in Proposal A is carried out, followed by a second local contrast reduction

(to remove patch boundary artifacts), and linear intensity equalization as before.

Proposal C. This is the final earlier iteration of my illumination pre-processing algorithm, which uses a superior patch-based normalization method and compensates the smoothed-out edges by using local contrast enhancement. First, the hybrid logarithmic-gamma dynamic range enhancement is applied, followed by a Wiener filter of size 2×2 to filter out noise, then the first BIMF is extracted, and a patch-based median normalization extracting on 7×7 patches and outputting single pixels is applied, followed by local contrast enhancement to boost the edges, and then linear intensity equalization is applied.

Proposal D (recommended). This is the recommended iteration of my illumination pre-processing algorithm, and the algorithm detailed in section IV. The difference between this iteration and Proposal C is that the edge boosting by local contrast enhancement is replaced by a Difference-of-Gaussians (DoG) filter. This iteration is demonstrably superior in performance to the previous three iterations.

Difference-of-Gaussians. This method normalizes image illumination by filtering the image with the difference between two Gaussians of different scale, the first one having $\sigma_1 = 1$ and the second with $\sigma_2 = 2$. This extracts features of only a particular spatial frequency range and tends to boost edge features while filtering out low-scale noise and shadows.

Gamma Correction. This method normalizes image illumination by enhancing the dynamic range of the image in order to provide more grey-level resolution to the bright and dark areas of the image. This corresponds to the gamma correction used in image encoding, which adjusts image intensity to better fit the nonlinear human visual system's sensitivity to luminance. The gamma correction curve starts at $(0, 0)$ and increases steadily at a decreasing rate.

Natural Logarithm. Similar to the gamma correction method, applying the natural logarithm to an image also provides increased grey-level resolution to the bright and dark

¹<https://www.mathworks.com/matlabcentral/fileexchange/28761-bidimensional-emperical-mode-decomposition-bemd>

areas of an image. However, the natural logarithm starts lower than the gamma correction curve and increases at a faster rate - it starts at $(1, 0)$ (its value approaches $-\infty$ as the argument approaches 0 from the right) and increases steadily at a decreasing rate.

Gamma Correction and Difference-of-Gaussians. Tan and Triggs combine both a gamma correction (with $\gamma = 0.2$) and a Difference-of-Gaussians filter ($\sigma_1 = 1$ and $\sigma_2 = 2$) applied in sequence [25]. This enhances the dynamic range of the image and then boosts the edges.

Histogram Equalization. Equalizing the image histogram assumes that, in a poorly-illuminated image, the pixel intensities are not well distributed across the full range of possible pixel intensities. Hence, equalizing the histogram values redistributes the pixel intensities so that pixel values are more evenly distributed.

Sobel Edges. Assuming that the edge features of a face are the most characteristic for facial recognition, the Sobel edge method extracts the magnitude of the first-order image gradient (which largely correspond to edges) and discards all other information.

B. Facial Recognition Algorithms

The following facial recognition algorithms were evaluated:

Dense Scale-Invariant Feature Transform (SIFT). The Dense SIFT facial recognition algorithm extracts local scale-invariant SIFT features on a dense grid of interest points placed uniformly on the facial image. Each local SIFT feature consists of a normalized feature vector containing the histogram of oriented gradients pooled in space and orientation, relative to the peak orientation of the patch.

Dense Histogram of Gradients (HOG). The Dense HOG facial recognition algorithm extracts a histogram of oriented gradients (known as a cell) on a dense grid of interest points placed uniformly on the facial image. Each HOG cell is then grouped into a 2×2 "block", and block features are vectorized to form a feature vector.

Dense Local Binary Patterns (LBP). The Dense LBP facial recognition algorithm assigns each image pixel to one of 59 different local texture patterns, depending on the local binary characteristics of the image around that pixel. Then, the image is sub-divided into cells and a histogram of the texture identities in each cell is tabulated. The concatenation of the cell histograms forms the feature vector.

Dense Speeded-Up Robust Features (SURF). The Dense SURF facial recognition algorithm assigns a feature vector to each point in a dense grid of interest points placed uniformly on the facial image, which consists of the sums and absolute sums of first-order derivatives using Haar wavelets.

C. Databases

Yale B Extended Database. The Yale B Extended Database² was used for facial recognition evaluation. It contains a total of 640 grey-scale images divided into 10 classes of 64

images each. For each of the 10 classes, grey-scale images were captured of the same test subject facing toward the camera but with 64 different lighting conditions. Each of the images are sized 50×50 . Although many of the images contain at least some degree of noise, and some images appear to have been corrupted, these images are useful for illumination pre-processing evaluation because: (1) facial pose is approximately constant and (2) images captured under many different illumination conditions are provided.

D. Setup

The purpose of the experiment is to determine which illumination pre-processing method is the most effective. To that end, I created a software framework on MATLAB R2016b to apply the different illumination pre-processing algorithms on the images, train a Bag-Of-Visual-Words (BOVW) linear SVM classifier on the various facial recognition algorithms, and then test the trained classifier on unseen facial images.

Feature Encoding and Classifier. A Bag-Of-Visual-Words (BOVW) feature encoder was designed with a vocabulary of 600 words for all image recognition algorithms. It was trained with VLFeat's K-means function using the approximated nearest neighbours (ANN) algorithm and an iteration limit of 50. Then, spatial histograms were computed and a simple linear SVM trained on the BOVW spatial histograms.

Training and Test Set Split. The splitting of the images in each class into training and test sets was randomized, so the mean and standard deviation of the classification accuracy across ten random splits was recorded. Furthermore, since image classification algorithms are inevitably very sensitive to the selection of training and test images, I tested the facial recognition algorithms with different four different proportions of training and test images for each class to evaluate how well the facial recognition algorithms function when given inadequate training data.

E. Results

Classification Accuracy. In Table I, the facial classification results are listed for the different proposed iterations of the BEMD-based illumination pre-processing algorithm. In Table II, the facial classification results are listed for all the competing illumination pre-processing algorithms, and the recommended proposal algorithm (i.e. Proposal D). Note that the "Original" pre-processing setting merely refers to applying facial recognition on the original images without any pre-processing. C_{32}, C_{16}, C_8, C_4 refer to training and test set splits of (32, 32), (16, 48), (8, 56), (4, 60) respectively.

Subjective Quality. Aside from the objective numerical classification accuracies, I also provide some representative examples of the ten illumination pre-processing algorithms being applied on six facial images in V-E. "Person 10, Sample 21" represents images with horizontal illumination, "Person 3, Sample 31" represents images with diagonal illumination, "Person 1, Sample 49" and "Person 1, Sample 57" represent images with noisy information in dark areas, "Person 2, Sample 52" represents heavily-corrupted images,

²<http://www.cs.ucsd.edu/classes/sp05/cse152/faces.zip>

TABLE I
FACIAL CLASSIFICATION RESULTS FOR PROPOSED PRE-PROCESSING ALGORITHMS (HIGHEST MEAN ACCURACY **BOLDED**)

| | | Mean Accuracy (\pm Standard Deviation) | | | | |
|------|---------|---|------------------|----------------------------------|----------------------------------|--|
| | Setting | Proposal A | Proposal B | Proposal C | Proposal D (recommended) | |
| SIFT | C_32 | 98.44 \pm 0.7% | 95.09 \pm 0.9% | 98.87 \pm 0.7% | 98.97\pm0.5% | |
| | C_16 | 96.00 \pm 1.1% | 89.02 \pm 1.4% | 96.81 \pm 1.2% | 97.98\pm0.5% | |
| | C_8 | 90.95 \pm 1.7% | 80.46 \pm 2.9% | 93.07 \pm 2.1% | 95.77\pm1.1% | |
| | C_4 | 77.75 \pm 4.1% | 63.35 \pm 5.5% | 82.63 \pm 2.8% | 88.05\pm2.7% | |
| HOG | C_32 | 94.78 \pm 1.5% | 90.19 \pm 1.6% | 97.94 \pm 1.0% | 98.87\pm0.4% | |
| | C_16 | 91.73 \pm 1.5% | 83.13 \pm 2.3% | 95.81 \pm 0.7% | 97.37\pm0.6% | |
| | C_8 | 84.75 \pm 1.7% | 74.16 \pm 3.2% | 90.48 \pm 2.0% | 92.84\pm2.1% | |
| | C_4 | 74.03 \pm 3.6% | 64.62 \pm 2.7% | 83.03 \pm 3.3% | 86.45\pm3.5% | |
| LBP | C_32 | 73.78 \pm 2.0% | 68.19 \pm 1.9% | 89.94 \pm 2.0% | 93.22\pm1.7% | |
| | C_16 | 60.83 \pm 1.6% | 57.98 \pm 2.0% | 81.04 \pm 2.2% | 86.88\pm3.0% | |
| | C_8 | 47.66 \pm 2.1% | 44.29 \pm 4.0% | 64.89 \pm 2.4% | 76.46\pm2.3% | |
| | C_4 | 39.47 \pm 4.4% | 38.22 \pm 4.9% | 58.70\pm4.6% | 55.40 \pm 7.2% | |
| SURF | C_32 | 75.03 \pm 1.7% | 62.00 \pm 2.2% | 81.03 \pm 2.0% | 91.53\pm1.9% | |
| | C_16 | 65.42 \pm 2.8% | 51.31 \pm 2.7% | 69.79 \pm 3.3% | 84.31\pm2.8% | |
| | C_8 | 50.80 \pm 1.7% | 39.68 \pm 3.0% | 57.12 \pm 2.1% | 72.39\pm3.6% | |
| | C_4 | 43.08 \pm 3.2% | 34.87 \pm 3.3% | 46.77 \pm 2.2% | 59.00\pm2.4% | |

TABLE II
FACIAL CLASSIFICATION RESULTS FOR COMPETING PRE-PROCESSING ALGORITHMS (HIGHEST MEAN ACCURACY **BOLDED**)

| | | Mean Accuracy (\pm Standard Deviation) | | | | | | | |
|------|---------|---|------------------|----------------------------------|------------------|------------------|------------------|------------------|------------------|
| | Setting | Proposal D (recommended) | Original | Gamma+DoG | Log | Histeq | Sobel Edges | DoG | Gamma |
| SIFT | C_32 | 98.97 \pm 0.5% | 96.41 \pm 1.1% | 99.47\pm0.3% | 97.88 \pm 0.8% | 97.56 \pm 1.0% | 92.66 \pm 1.7% | 87.56 \pm 1.4% | 98.38 \pm 0.6% |
| | C_16 | 97.98 \pm 0.5% | 87.58 \pm 1.9% | 98.71\pm0.6% | 94.98 \pm 1.3% | 91.42 \pm 1.7% | 85.40 \pm 1.8% | 78.98 \pm 3.2% | 94.50 \pm 0.9% |
| | C_8 | 95.77\pm1.1% | 75.25 \pm 4.3% | 94.91 \pm 1.4% | 85.61 \pm 2.8% | 79.91 \pm 3.0% | 72.45 \pm 2.7% | 64.75 \pm 3.1% | 80.80 \pm 1.8% |
| | C_4 | 88.05\pm2.7% | 52.68 \pm 3.3% | 83.88 \pm 4.1% | 66.50 \pm 4.9% | 59.77 \pm 4.3% | 55.02 \pm 5.5% | 50.80 \pm 2.7% | 62.05 \pm 3.8% |
| HOG | C_32 | 98.87 \pm 0.4% | 81.19 \pm 2.8% | 98.94\pm0.6% | 94.06 \pm 1.4% | 91.47 \pm 1.9% | 76.16 \pm 2.0% | 77.72 \pm 1.6% | 92.03 \pm 1.1% |
| | C_16 | 97.37 \pm 0.6% | 76.54 \pm 2.2% | 97.92\pm0.5% | 89.87 \pm 1.8% | 84.48 \pm 3.0% | 71.88 \pm 2.2% | 69.69 \pm 2.6% | 87.79 \pm 1.6% |
| | C_8 | 92.84 \pm 2.1% | 68.20 \pm 2.4% | 94.18\pm1.2% | 81.23 \pm 3.0% | 76.62 \pm 2.0% | 62.82 \pm 2.7% | 60.54 \pm 3.0% | 78.54 \pm 3.4% |
| | C_4 | 86.45 \pm 3.5% | 55.35 \pm 3.1% | 87.93\pm2.6% | 67.08 \pm 3.3% | 64.00 \pm 3.8% | 52.55 \pm 4.2% | 51.18 \pm 5.2% | 66.63 \pm 4.9% |
| LBP | C_32 | 93.22 \pm 1.7% | 91.53 \pm 1.5% | 93.91\pm1.9% | 91.53 \pm 1.6% | 90.22 \pm 1.8% | 70.16 \pm 3.2% | 54.81 \pm 4.9% | 91.00 \pm 1.9% |
| | C_16 | 86.88\pm3.0% | 84.31 \pm 2.0% | 86.83 \pm 2.2% | 84.56 \pm 1.8% | 83.83 \pm 1.3% | 58.19 \pm 3.3% | 46.46 \pm 3.9% | 82.98 \pm 1.9% |
| | C_8 | 76.46\pm2.3% | 70.55 \pm 3.6% | 76.18 \pm 3.2% | 75.11 \pm 3.2% | 71.23 \pm 4.3% | 46.36 \pm 3.0% | 42.39 \pm 3.5% | 70.48 \pm 4.5% |
| | C_4 | 55.40 \pm 7.2% | 62.55 \pm 3.3% | 63.60\pm5.3% | 63.93 \pm 2.8% | 62.95 \pm 3.7% | 45.90 \pm 3.7% | 45.48 \pm 3.7% | 62.00 \pm 3.0% |
| SURF | C_32 | 91.53 \pm 1.9% | 66.09 \pm 2.4% | 92.47\pm1.3% | 80.28 \pm 2.0% | 77.72 \pm 2.0% | 59.59 \pm 2.4% | 67.25 \pm 1.6% | 77.72 \pm 1.5% |
| | C_16 | 84.31\pm2.8% | 58.35 \pm 2.4% | 83.29 \pm 2.3% | 72.54 \pm 3.0% | 71.10 \pm 2.5% | 53.79 \pm 2.7% | 59.56 \pm 1.8% | 70.12 \pm 1.7% |
| | C_8 | 72.39\pm3.6% | 49.20 \pm 3.1% | 71.64 \pm 3.3% | 61.79 \pm 2.3% | 57.13 \pm 2.1% | 46.36 \pm 4.2% | 49.89 \pm 1.5% | 60.30 \pm 3.2% |
| | C_4 | 59.00 \pm 2.4% | 39.35 \pm 4.2% | 61.75\pm3.1% | 49.40 \pm 3.4% | 46.40 \pm 3.5% | 45.90 \pm 2.8% | 42.20 \pm 4.5% | 49.42 \pm 3.2% |

and "Person 5, Sample 31" represents images with periodic noise artifacts. As can be seen, facial details are barely distinguishable in the heavily shadowed original images, and edge-based methods do not correct the illumination well by themselves for such images. Visually, the "Gamma", "Log", and "Histeq" methods work very well already in enhancing the images' dynamic range, but shadows remain to hinder facial recognition. The ideally illumination-preprocessed image only contains the useful facial edges and does not show any non-uniform illumination. It is clear by visual inspection that the recommended "Proposal D" and "Gamma+DoG" eliminate non-uniform illumination the best. It can also be seen that "Proposal D" is superior at equalizing the intensity of the

shadowed and unshadowed areas than "Gamma+DoG".

F. Discussion

From the results in Table I, it can be seen that the recommended configuration of the algorithm, Proposal D, is superior to all the other three proposed configurations, out-performing them in all four tested facial recognition algorithms. From the results in Table II, it can be seen that the recommended configuration of the algorithm and the "Gamma+DoG" out-perform the other illumination pre-processing algorithms. For dense SIFT, Gamma+DoG out-performs Proposal D for larger training sets, but Proposal D out-performs Gamma+DoG for smaller training sets. For dense HOG, Gamma+DoG out-

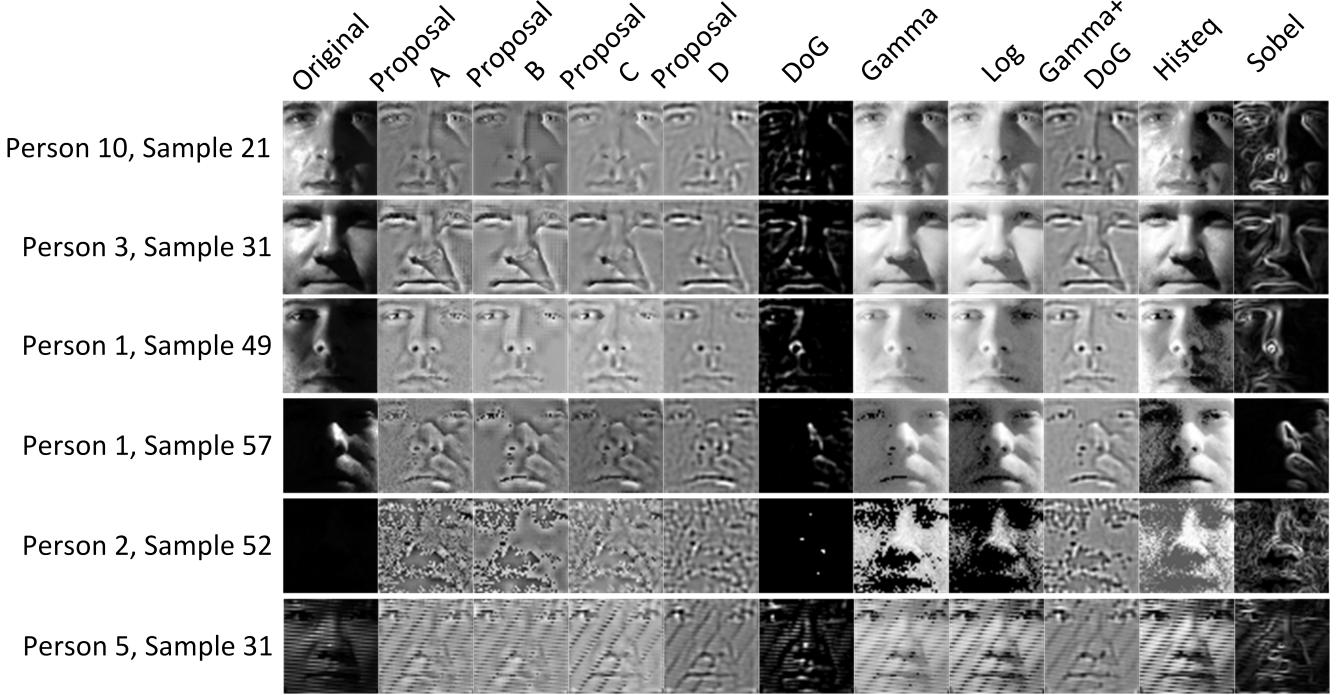


Fig. 5. Displays of six illumination pre-processed images: (left to right) Original (unprocessed), Proposals A-D, DoG, Gamma Correction, Natural Logarithm, Gamma+DoG, Histogram Equalization, and Sobel Edges.

performs Proposal D for all training set sizes, but the gap is particularly small for the larger training set sizes (only 0.69% and 0.55% for C_{32} and C_{16}). For dense LBP, Gamma+DoG out-performs Proposal D for the larger training sets, but Proposal D appears to perform better for smaller training sets (note that, for C_4 , Gamma+DoG does better, but the standard deviation in classification accuracy is abnormally high for both methods). For dense SURF, a similar pattern is observed, whereby Gamma+DoG performs better for larger training sets, but Proposal D does better for smaller training sets, with the exception of C_4 .

In general, it may be seen that the proposed method, Proposal D, either approaches the illumination pre-processing performance or does better than that of the next best method, Gamma+DoG. In the case of adequate training data (i.e. C_{32}), Gamma+DoG tends to do slightly better than Proposal D, but in case of inadequate training data (i.e. C_{16}, C_8, C_4), Proposal D tends to do better. Although Proposal D does not out-perform Gamma+DoG in the case of large training sets, its better performance for small training sets means that Proposal D generalizes better to the facial training data. Furthermore, a visual inspection indicates that Proposal D does much better than Gamma+DoG in equalizing the non-uniform illumination than a simple inspection of the classification accuracies would suggest.

VI. CONCLUSION

To conclude, this report shows that bidimensional empirical mode decomposition is effective for isolating shadow and

non-shadow features in facial images, and is thus a useful component of an overall non-uniform illumination normalization pre-processing algorithm, and experimental results indicate that it helps improve facial recognition accuracy for a variety of image recognition algorithms. When compared with other competing illumination normalization pre-processing algorithms, the proposed method out-performs all other methods for small training sets and performs with similar accuracy to the next best method for larger training sets. This indicates that the algorithm is able to preprocess facial images into a form that allows good generalization of training data for subsequent image feature description and classification. Furthermore, a subjective evaluation of the illumination normalization indicates that the proposed method out-performs all other methods in eliminating the sharp contrasts between dark and bright regions in a non-uniformly illuminated facial image. Although the various components of BEMD do not always directly correspond with clear-cut distinctions in illumination, the results of this report indicate that it is useful for illumination normalization.

REFERENCES

- [1] M. Agarwal, N. Jain, M. M. Kumar, and H. Agrawal. Face recognition using eigen faces and artificial neural network. *International Journal of Computer Theory and Engineering*, 2(4):624, 2010. [2](#)
- [2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006. [2](#)
- [3] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total variation models for variable lighting face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(9):1519–1524, 2006. [2](#)

- [4] R. Gross, S. Baker, I. Matthews, and T. Kanade. Face recognition across pose and illumination. In *Handbook of Face Recognition*, pages 197–221. Springer, 2011. 1
- [5] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *AVBPA*, volume 3, pages 10–18. Springer, 2003. 2
- [6] G. Heo and P. Gader. Robust kernel discriminant analysis using fuzzy memberships. *Pattern recognition*, 44(3):716–723, 2011. 2
- [7] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 454, pages 903–995. The Royal Society, 1998. 2
- [8] N. Intrator, D. Reisfeld, and Y. Yeshurun. Face recognition using a hybrid supervised/unsupervised neural network. *Pattern Recognition Letters*, 17(1):67–76, 1996. 2
- [9] X. Jing and D. Zhang. Face recognition based on linear classifiers combination. *Neurocomputing*, 50:485–488, 2003. 2
- [10] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 2
- [11] A. Khatun and M. Bhuiyan. Neural network based face recognition with gabor filters. *International Journal of Computer Science and Network Security*, 11(1):71–74, 2011. 2
- [12] B. Klare and A. K. Jain. Sketch-to-photo matching: a feature-based approach. In *Biometric Technology for Human Identification VII*, volume 7667, 2010. 2
- [13] K.-C. Kwak and W. Pedrycz. Face recognition using fuzzy integral and wavelet decomposition method. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(4):1666–1675, 2004. 2
- [14] A. Linderhed. Image empirical mode decomposition: A new tool for image processing. *Advances in Adaptive Data Analysis*, 1(02):265–294, 2009. 2
- [15] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image processing*, 11(4):467–476, 2002. 2
- [16] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 200–205. IEEE, 1998. 2
- [17] P. Melin, D. Sánchez, and O. Castillo. Genetic optimization of modular neural networks with fuzzy response integration for human recognition. *Information Sciences*, 197:1–19, 2012. 2
- [18] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel. Image analysis by bidimensional empirical mode decomposition. *Image and vision computing*, 21(12):1019–1026, 2003. 2
- [19] J. C. Nunes, S. Guyot, and E. Deléchelle. Texture analysis based on local analysis of the bidimensional empirical mode decomposition. *Machine Vision and applications*, 16(3):177–188, 2005. 2
- [20] C. A. Perez, L. A. Cament, and L. E. Castillo. Methodological improvement on local gabor face recognition based on feature selection and enhanced borda count. *Pattern Recognition*, 44(4):951–963, 2011. 2
- [21] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, volume 1, pages 947–954. IEEE, 2005. 1
- [22] J. Short, J. Kittler, and K. Messer. A comparison of photometric normalisation algorithms for face verification. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 254–259. IEEE, 2004. 2
- [23] A. Sinha and K. Singh. The design of a composite wavelet matched filter for face recognition using breeder genetic algorithm. *Optics and Lasers in engineering*, 43(12):1277–1291, 2005. 2
- [24] X.-n. Song, Y.-j. Zheng, X.-j. Wu, X.-b. Yang, and J.-y. Yang. A complete fuzzy discriminant analysis approach for face recognition. *Applied Soft Computing*, 10(1):208–214, 2010. 2
- [25] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Analysis and modeling of faces and gestures*, pages 168–182, 2007. 6
- [26] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991. 2
- [27] H. Wang, S. Z. Li, Y. Wang, and J. Zhang. Self quotient image for face recognition. In *Image Processing, 2004. ICIP’04. 2004 International Conference on*, volume 2, pages 1397–1400. IEEE, 2004. 2
- [28] L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):775–779, 1997. 2
- [29] X. Xie. Illumination preprocessing for face images based on empirical mode decomposition. *Signal Processing*, 103:250–257, 2014. 2, 3
- [30] J. Yang, J.-y. Yang, and A. F. Frangi. Combined fisherfaces framework. *Image and Vision computing*, 21(12):1037–1044, 2003. 2
- [31] M.-H. Yang. Face recognition using kernel methods. In *Advances in neural information processing systems*, pages 1457–1464, 2002. 2
- [32] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 336–341. IEEE, 1998. 2