

An Optimized Illumination Normalization Method for Face Recognition

Jukka Holappa, Timo Ahonen and Matti Pietikäinen

Abstract—Differences in illumination conditions cause significant challenges for any 2-D face recognition algorithm. One of the methods to counter these effects is image preprocessing before feature extraction. In this paper we present a new preprocessing approach that uses custom filters obtained through an optimization procedure striving for most suitable preprocessing filters for the selected feature extractor and distance measure. We experiment with it using Local Binary Pattern texture features and χ^2 histogram distance metric. Results are provided for Face Recognition Grand Challenge (FRGC) 1.0.4 dataset.

I. INTRODUCTION

In the recent years, different problems in face image analysis, such as face detection, face recognition and facial expression recognition have received very much attention in computer vision research. These problems are interesting from the viewpoints of basic research aiming to efficient descriptors for facial images and of applications such as biometrics, surveillance, and human-computer interaction [10].

A key issue in face analysis is finding such representation for face appearance that is robust to different perturbations and changes such as illumination and pose changes, aging of the subjects, etc, but still very discriminative to be able to handle the low inter-person variation in face image.

Different illumination conditions by uncontrolled conditions cause significant challenges in face recognition tasks. Self-shadowing by nose and other facial features can cause large variations in the image depending on the light directions. Also multiple lights may be present in the images.

Different approaches to handling the problem of illumination changes have been discussed in the literature. The most straightforward method is to have a large enough number of training images of each subject under different illuminations. As acquiring such training data in real applications is laborious or even impossible, Huang *et al.* proposed creating 3D models of the gallery faces and using them to get synthetic training images of the faces in varying illuminations [9].

Another possibility is to deal with illumination variations in the feature extraction phase. Aggarwal and Chellappa proposed illumination invariant features derived directly from face images under unknown lighting conditions [2]. However, clearly the most widely studied approach is to apply some kind of preprocessing to the face image, aiming to

transform it into a *canonical face image* free of illumination artefacts prior to feature extraction. Shashua and Rilkina-Raviv proposed creating a quotient image using a class model of faces. In their work the surface shape was assumed to be identical for all the faces and only the texture changes. This allows creation of illumination invariant signature image while assuming Lambertian reflections and a single directional light source [17]. Among the best known face image preprocessing methods are, e.g., the self-quotient image [20], total variation models [6], and anisotropic smoothing [7]. Experimental comparisons of preprocessing algorithms have been done by Short *et al.* [18] and Tan and Triggs [19].

Some useful texture features that can be used for face recognition are already invariant to some of the illumination differences. The good performance of Gabor filters in face representation has been attributed in part to their robustness to illumination changes [12]. On the other hand, the well known Local Binary Pattern (LBP) texture descriptors are invariant to monotonic gray level changes. The use of LBP for face description was proposed by Ahonen *et al.* [3], [4] in a study where such description was shown to perform well on the FERET face image dataset containing different challenges such as different facial expressions, aging of the subjects, and illumination changes.

Despite their invariance to monotonic gray scale mappings, the LBP representations have been shown not to perform well under heavy lighting changes. Strong lines caused by self-shadowing and changes in dominant gradient orientations due to changes in illumination direction still cause variations in LBP codes as they change the local texture patterns. To alleviate this problem, Tan and Triggs proposed a preprocessing chain for illumination normalization. Furthermore, they introduced a three-level version of local patterns and a Hausdorff-like distance between query and gallery label images to replace histogram based approach. [19]

The problem with preprocessing is always about selecting the correct normalization method. It is very easy to accidentally remove information from the image that should have been preserved for improved recognition rate. What needs to be filtered away and what to preserve depends on the features extracted and the type of variations there are in the images.

In this work we propose a preprocessing chain and optimization method for setting its parameters so that the preprocessing chain explicitly tailors for the specific feature extractor. This is done by stochastic optimization of the preprocessing parameters using a simple probability value derived from intra- and inter-class differences of the extracted features as the cost function. Moreover, due to the general

This work has been performed with partial support from the MOBIO project (contract number IST-214324), which belongs to the 7th Framework Research Programme of the European Union.

The authors are with Machine Vision Group, University of Oulu, PL 4500, FI-90014 Oulun yliopisto, FINLAND {jukkahol, tahonen, mkip}@ee.oulu.fi

3D structure of faces, illumination changes tend to cause different effects at different parts of the face image (e.g., strong shadows on either side of the nose, etc.). This is taken into account in the preprocessing chain by making the parameters spatially variant.

We use the local binary pattern histograms to describe the preprocessed faces and χ^2 distance to compare the histograms. It should be noted that the optimization method aims to find the optimal preprocessing for the currently selected representation and distance measure. This preprocessing chain and related optimization method is however not specifically designed to LBPs and it can be used for finding a preprocessing method for other face descriptors as well.

II. RELATED WORK

The use of local binary pattern histograms computed in local image regions was first proposed for face image description by Ahonen *et al.* [3], [4]. This representation has successfully been applied to face recognition [4], face detection [8], facial expression recognition [16], demographic classification [22] and also other tasks such as general object recognition [25].

Some of the limitations of that method have also been discussed in the literature and different extensions and modifications have been proposed. In [24], Zhang *et al.* used AdaBoost learning algorithm for selecting a set of local blocks and their weights. Then, the LBP methodology was applied to the obtained blocks yielding in smaller feature vector length. Rodriguez and Marcel noticed that the method as such does not suit well for the face verification task and proposed an approach based on adapted, client-specific LBP histograms [15].

Other proposed enhancements include multi-scale block LBP which considers mean gray values from larger pixel blocks than original LBP [11], using patterns at multiple scales for representation [5] and combining LBP representation with other information such as Gabor filter based recognition [21], [26].

In the research of illumination effects on the face appearance it has been concluded that illumination induces larger changes on the unprocessed grey-level image than differences between individuals [1]. To compensate for these changes, several different image processing algorithms have been introduced.

Many preprocessing algorithms proposed in the literature for illumination invariant try to extract the reflectance component free of effects caused by illumination. The Self-Quotient Image (SQI) [20] aims for solving the reflectance component of a face image by dividing the perceived image by an approximation of the lighting component. In the SQI model the lighting component is approximated by a smoothed version of the input image.

In the local total variation model [6] the input image is decomposed into large- and small-scale components, where the latter is assumed to contain mostly illumination invariant information. Anisotropic smoothing [7], on the other hand,

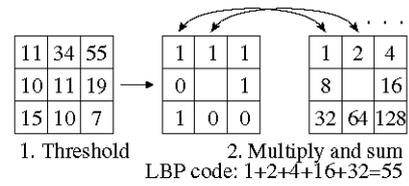


Fig. 1. The basic LBP operator.

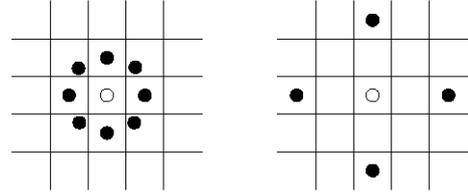


Fig. 2. Two different circular neighborhoods: 8 sampling points at radius 1 and 4 sampling points at radius 2. If the sampling point is not in the center of a pixel, the pixel values are bilinearly interpolated.

aims to find a lighting component from the input image through a constrained optimization procedure.

In first papers discussing LBP for face recognition, no preprocessing was used for the images except scaling and rotating to normalize eye coordinates and image sizes. The first work applying and comparing illumination normalization preprocessing for LBP representation was done by Tan and Triggs [19]. In that work they compared numerous different preprocessing methods using FRGC 1.0.4, CMU PIE and Extended Yale-B datasets and introduced an effective preprocessing chain for face images [19]. They implemented a four step preprocessing chain for the image: gamma correction, difference of gaussians filtering, masking and contrast equalization. In addition to preprocessing, they demonstrated an improvement with recognition rate by using Local Ternary Patterns (LTP) instead of LBP and Distance Transform (DT) instead of χ^2 distance of LBP histograms.

III. LOCAL BINARY PATTERNS

The local binary pattern operator [13] was originally designed for texture description. It has nevertheless shown very good performance also in many other tasks, and recently one of the most important application areas has been in facial image description. The basic LBP operator labels the pixels of an image by thresholding each 3×3 pixel neighborhood of the input image with the center value, multiplying the thresholded values by powers of two and summing them. Usually the histogram of the resulting labels is then used as an image descriptor. The operation of the basic LBP operator is illustrated in Fig. 1.

The operator is extended to use neighborhoods of different sizes [13] by using a circular neighborhood around the center pixel and bilinearly interpolating the pixel values. This allows any radius and number of pixels in the neighborhood. See Fig. 2 for an example of two different circular neighborhoods.

Another extension to the original operator is using *uniform* binary patterns [13]. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010011 (6 transitions) are not. In the computation of the LBP histogram, uniform patterns are used so that the histogram has a separate bin for every uniform pattern and all non-uniform patterns are assigned to a single bin.

IV. FACE IMAGE PREPROCESSING AND DESCRIPTION

To achieve good performance under illumination changes, methods based either on normalization or illumination invariant descriptors have been proposed. Deriving from [19], we propose a combined chain of operations for face image preprocessing prior to feature extraction that is based on local binary pattern histograms. As discussed in [4], the local binary pattern descriptor is relatively robust to different illumination conditions but severe changes in lighting still pose a problem.

To combat those changes, we strive for a preprocessing method that *explicitly reduces such intra-class variations* that the LBP description is sensitive to. This is done through an optimization procedure which is explained in more detail in the following section.

Many existing preprocessing methods use iterative methods and are therefore relatively slow. The preprocessing and description chain presented in this paper uses only logarithmic transformation of pixel values and convolution of the input image region with small sized filter kernels, which makes the method very fast.

The complete preprocessing and feature extraction chain is presented in Figure 3. First the logarithm of the input image $I(x, y)$ is taken

$$I^L(x, y) = \log(I(x, y) + 1) \quad (1)$$

To avoid taking logarithm of zero, 1 is added to the input image. Then, following the procedure of [4], the image is divided into N local rectangular regions $I_1^L(x, y), \dots, I_N^L(x, y)$. Each of these regions is convolved with a preprocessing filter F_n .

$$I_n^F(x, y) = I_n^L(x, y) \star F_n \quad (2)$$

Different filter kernels are applied to different regions because the effects caused by illumination changes to different parts of the facial image are not alike. Finally, the LBP operator is applied to all processed local windows $I_n^F(x, y)$, the histogram $h_n(i)$ of obtained labels is computed and the histograms are concatenated to build a global description of the face.

When this method is compared to Tan and Triggs' preprocessing chain, the main distinction is that different filter kernels are applied to different part of the face. Instead of using a fixed difference-of-gaussians filter, the kernels for

local regions are obtained through an optimization procedure explained in the following section. Moreover, the last step of the Tan and Triggs preprocessing chain, a gray scale transformation, does not affect LBP features at all because the transformation is monotonically increasing so it is not included in this system.

V. FILTER OPTIMIZATION

Filter optimization is done individually for each separate area that is processed for feature extraction.

While the goal is maximizing the recognition rate with the actual face data, this cannot be used as a criteria for optimization. It would create an optimization problem with huge number of variables that most likely would not generalize from the training data.

A better approach is maximizing the probability that the features calculated from an image region, that the filter to be optimized is applied to, are closer to each other in the intra class case than in the extra class case. To compare two histograms $h_n^A(i)$ and $h_n^B(i)$ computed from region n of images A and B , we use the Chi square distance measure

$$\chi^2(h_n^A(i), h_n^B(i)) = \sum_i \frac{(h_n^A(i) - h_n^B(i))^2}{h_n^A(i) + h_n^B(i)}. \quad (3)$$

Now the function to be maximized is the probability that, for randomly selected $A1$, $A2$, and B , the histogram $h_n^{A1}(i)$ lies closer to $h_n^{A2}(i)$ than to $h_n^B(i)$, where images $A1$ and $A2$ represent the same subject and image B represents another subject. In other words, we try to find the filter F_n that maximizes

$$P(\chi^2(h_n^{A1}(i), h_n^{A2}(i)) < \chi^2(h_n^{A1}(i), h_n^B(i))).$$

This optimization problem is not an easy one. The cost function is not monotonic or smooth and it takes discrete steps when distances between image pairs change. Also the dimensionality of the optimization space gets very large very quickly as the filter size is increased.

Our implementation solves above optimization problem by using a variation of Improving Hit and Run (IHR) method [23]. In the original Improving Hit and Run, iterative steps are taken in random direction from current state using uniform distribution and the new state is used if it improves the result. In our work, two separate states are tracked: Current best and an improving state. Both states are optimized by taking random steps using IHR algorithm.

Initially both states start from random values. If improving state ever gets better results than current best state, current best is replaced with improving state and new random state is assigned to improving state. If improving result does not improve in a limited amount of steps, a new random state is generated as the new starting point. The limit of 100 non-improving steps was used in this experiment to prevent improving state from getting stuck in another local maximum. Each filter is normalized to length of 1 using L_2 -distance after random step is taken.

This allows constant small improvements to be made from the known best solution for locating local maximum. At

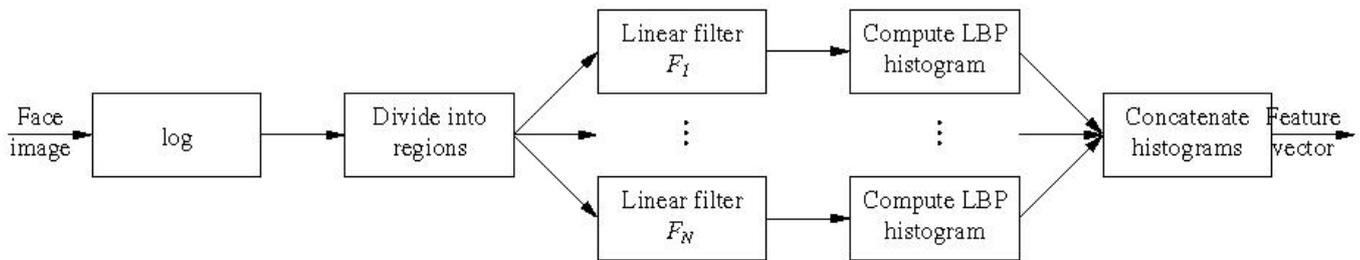


Fig. 3. The face image preprocessing and description chain



Fig. 4. Optimized 7×7 filters for logarithm case and LBP histogram features. Each filter is displayed in matching histogram grid position.

the same time alternative solutions are tried so that global maximum recognition rate can be found.

States that are optimized are actual filter matrix coefficients. Example filters are shown in Fig 4. Filters look very noisy because of the nature of the problem. As only the sign of the LBP operator bits are significant, small differences in filter values do not change the result in most images but may just be enough to do that in a single image. This causes very fast overfitting when there are too many coefficients to update.

In our experiments, each histogram window has its own optimized filter that is convolved with image contents. Optimization tries to maximize the probability that two histograms taken from the same person have smaller distance when compared to a histogram from another person. Each person was assumed to have equal probability for appearing and each image of each person was also equally likely.

Various filter sizes were tried from 3×3 to 11×11 . While larger filter sizes can always contain smaller filters and their theoretical performance is therefore always better, in practice larger filter sizes suffer from overfitting much more. What size works best for each problem depends on the amount of



Fig. 5. Example images from the FRGC 1.0.4 dataset. From left: gallery image and two probe images.

training data and can be determined with testing.

We experimented with and without taking logarithm of the image beforehand. Optimization process was done twice to train best possible filters for both cases.

VI. EXPERIMENTS

The efficiency of the proposed face preprocessing scheme in the face recognition scenario was tested and compared to that of Tan and Trigg's method using the Face Recognition Grand Challenge (FRGC) experiment 1.0.4 [14] dataset. That dataset is divided into training, probe and gallery sets which have no overlap. The probe and gallery images represent 152 subjects, and there are 1 gallery image and 2–7 probes per subject, totaling 152 images in the gallery set and 608 images in the probe set. The gallery images were taken with good quality camera under controlled conditions whereas the probe images were taken with a pocket digital camera under uncontrolled conditions. Examples of the gallery and probe images are shown in Fig. 5.

Image is divided into 14×14 grid of 8×8 pixel histogram windows. Performance improves with smaller histogram window sizes but feature vector length also becomes significantly longer. 8×8 size gives reasonably good results with short enough feature vector length and is therefore a good compromise.

Each histogram window is preprocessed, LBP features extracted and histogram calculated. Large enough area around histogram windows is processed so that filter and LBP calculations can produce results for full histogram window area. Process is illustrated in Fig. 6. Processing only required parts of the image gives a significant performance boost and should be used when possible. At the borders of the image, border line is duplicated to extend the image size until enough pixels are available for processing.

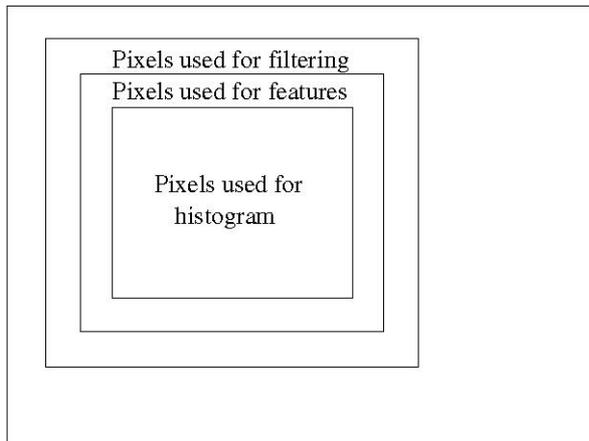


Fig. 6. Illustration of image areas used for processing each step.

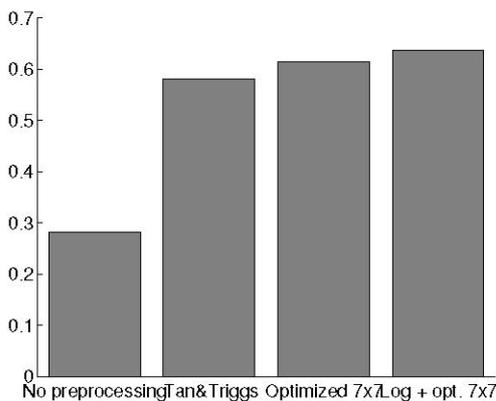


Fig. 7. Best results with different processing methods.

Faces are classified with nearest neighbour classifier by using the sum of χ^2 histogram distances (3) for all histogram windows. Weights for local regions are not used, i.e. each histogram is weighted equally.

For the optimization procedure in which the filter kernel values were determined, a set of FRGC training images was used. This set did not contain any of the images or the persons that are in the FRGC 1.0.4 dataset used for reporting the performance.

VII. RESULTS

Recognition rates for different preprocessing methods with FRGC 1.0.4 dataset are provided in Fig. 7. and in Table I. Note that these figures are not comparable to those reported in [19]. This is due to the different experimental setup that was utilized in [19]. In that work several gallery images per subject were used, whereas we followed the original FRGC protocol using only one gallery image per subject.

Performance for pure LBP histogram distance is 28.1% recognition rate. With Tan and Triggs' preprocessing chain, performance improves to 58.1%. Using optimized filters directly to image data gives 61.5% recognition rate when

TABLE I
RECOGNITION RATES ON FRGC 1.0.4 USING LBP AND DIFFERENT
PREPROCESSING METHODS

Filter	Without log	With log
Optimized 3×3	48.5%	51.0%
Optimized 5×5	61.2%	62.3%
Optimized 7×7	61.5%	63.7%
Optimized 9×9	61.3%	63.2%
Optimized 11×11	54.9%	57.9%
No preprocessing	28.1%	
Tan & Triggs [19]	58.1%	

using filters with size 7×7 . Taking logarithm of the image before optimized filters improves recognition rate to 63.7%.

Original LBP performance with the data is not good. Varying focus settings cause low-pass filtering for the face areas in some of the images and that changes local texture patterns significantly which can be easily seen from resulting LBP code images.

Using Tan and Triggs' preprocessing chain improves performance from 28.1% to 58.1%. The most significant part of this preprocessing chain is Difference of Gaussians filtering that acts as bandpass filter for the images [19]. It preserves most of the important details as well as larger scale features and gives a very nice boost in the performance.

When using optimized filters directly to the image, the result improves to 61.5% with 7×7 filters. The actual recognition rate varied between 61% and 64% depending on the exact filters that were tested while the optimization process was still ongoing. This shows that while small changes in training data can have significant effect on the actual result, the overall recognition rate was still better than single filter solution all the time.

Smaller filters like 3×3 show clearly worse performance than larger filters. Smaller filters are unable to capture essential larger features of the image and mostly provide local derivatives in various dimensions and provide some blurring. The differences between 5×5 , 7×7 and 9×9 filters are small but consistent according to our testing. Performance starts to decrease significantly at 11×11 filter size that has clearly overfitted for the training data.

Taking a logarithm of the image data before filtering shows 1–2.5% recognition rate increase with the test data. The idea is to convert the product of incoming illumination and surface reflectance into a sum. That way a given reflectance change produces a constant step in the logarithm image in different illumination intensities [19]. This seems to make it easier to generate good filters for the images as the results are clearly better.

VIII. DISCUSSION AND CONCLUSIONS

In this paper we proposed a filtering based method for illumination normalization for face recognition. A stochastic optimization procedure for setting the filter kernel values was introduced. Using the challenging FRGC 1.0.4 dataset it was shown that the optimized preprocessing filters can significantly improve the performance of local binary pattern

based face recognition in comparison to no preprocessing or manually selected bandpass filters.

While the representativeness and amount of training data for filter optimization has significant influence on the generated filters, it was shown that even with small amount of training data, a significant improvement can be achieved in the recognition rate. Filter size selection is also important because of problems with overfitting. In our experiments, 7×7 filter size gives the best results with FRGC 1.0.4 data.

When used with simple histogram features, filtering needs to process only little extra image area per histogram and is almost as fast to compute as a single pass using a filter kernel of equal size, assuming that windows do not overlap.

The proposed method could be used with any classification algorithm that extracts features from various image areas as long as suitable training material is available. Also classifiers that extract multiple features from the image would most likely benefit from custom preprocessing before extraction of each feature.

REFERENCES

- [1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):721–732, July 1997.
- [2] G. Aggarwal and R. Chellappa. Face recognition in the presence of multiple illumination sources. In *ICCV 2005*, pages 1169–1176, 2005.
- [3] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *ECCV 2004*, volume 3021 of *Lecture Notes in Computer Science*, pages 469–481. Springer, 2004.
- [4] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.
- [5] C.-H. Chan, J. Kittler, and K. Messer. Multispectral local binary pattern histogram for component-based color face verification. In *BTAS 2007*, pages 1–7, 27–29 Sept. 2007.
- [6] 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.
- [7] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *AVBPA 2003*, pages 10–18, 2003.
- [8] A. Hadid, M. Pietikäinen, and T. Ahonen. A discriminative feature space for detecting and recognizing faces. In *CVPR 2004*, volume 2, pages 797–804, 2004.
- [9] J. Huang, B. Heisele, and V. Blanz. Component-based face recognition with 3d morphable models. In *AVBPA 2003*, pages 27–34, 2003.
- [10] S. Z. Li and A. K. Jain, editors. *Handbook of Face Recognition*. Springer, 2005.
- [11] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *ICB 2007*, volume 4642 of *Lecture Notes in Computer Science*, pages 828–837, 2007.
- [12] C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):725–737, May 2006.
- [13] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, Jul 2002.
- [14] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. M. ues, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR 2005*, pages 1: 947–954, Washington, DC, USA, 2005. IEEE Computer Society.
- [15] Y. Rodriguez and S. Marcel. Face authentication using adapted local binary pattern histograms. In *ECCV 2006*, volume 4, pages 321–332, 2006.
- [16] C. Shan, S. Gong, and P. W. McOwan. Robust facial expression recognition using local binary patterns. In *ICIP 2005*, pages II: 914–917, 2005.
- [17] A. Shashua and T. Riklin-raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23:129–139, 2001.
- [18] J. Short, J. Kittler, and K. Messer. A comparison of photometric normalisation algorithms for face verification. In *FGR 2004*, pages 254–259, 2004.
- [19] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *AMFG 2007*, 2007.
- [20] H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. In *FGR 2004*, pages 819–824, 2004.
- [21] S. Yan, H. Wang, X. Tang, and T. S. Huang. Exploring feature descriptors for face recognition. In *ICASSP 2007*. IEEE Signal Processing Society, 2007.
- [22] Z. Yang and H. Ai. Demographic classification with local binary patterns. In *ICB 2007*, volume 4642 of *Lecture Notes in Computer Science*, pages 464–473, 2007.
- [23] Z. B. Zabinsky, R. L. Smith, J. F. McDonald, H. E. Romeijn, and D. E. Kaufman. Improving hit-and-run for global optimization. *Journal of Global Optimization*, Volume 3(2):171–192, June 1993.
- [24] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu. Boosting local binary pattern (LBP)-based face recognition. In *Proc. Advances in Biometric Person Authentication*, pages 179–186, 2004.
- [25] H. Zhang, W. Gao, X. Chen, and D. Zhao. Object detection using spatial histogram features. *Image and Vision Computing*, 24(4):327–341, 2006.
- [26] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In *ICCV 05*, pages 1:786–791, 2005.