PH-BRINT: Pooled Homomorphic Binary Rotation Invariant and Noise Tolerant Representation for Face Recognition under Illumination Variations

Raqinah Alrabiah[1,2,*] Muhammad Hussain[2,**], Hatim A. Aboalsamh[2], Mansour Zuair[3], George Bebis[4]
[1]Department of Information Technology, College of Computer, Qassim University, Qassim, Saudi Arabia
[2]Department of Computer Science, [3]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
[4]Department of Computer Science and Engineering, University of Nevada, Reno, USA
[*]raqinah@qu.edu.sa, [**]mhussain@ksu.edu.sa

Abstract. Face recognition under varying illumination conditions is a challenging problem. We propose a simple and effective multiresolution approach Pooled Homomorphic Binary Rotation Invariant and Noise Tolerant (PH-BRINT) for face recognition under varying illumination conditions. First, to reduce the effect of illumination, wavelet transform based homomorphic filter is used. Then Binary Rotation Invariant and Noise Tolerant (BRINT) operators with three different scales are employed to extract multiscale local rotation invariant and illumination insensitive texture features. Finally, the discriminative information from the three scales is pooled using MAX pooling operator and localized gradient information is computed by dividing the pooled image into blocks and calculating the gradient magnitude and direction of each block. The PH-BRINT technique has been tested on a challenging face database Extended Yale B, which was captured under varying illumination conditions. The system using minimum distance classifier with L1-norm achieved an average accuracy of 86.91 %, which is comparable with the state-of-the-art best illumination-invariant face recognition techniques.

Keywords: Face Recognition · Local Binary Pattern · Feature Extraction · Illumination Invariant

1 Introduction

Face recognition is a complicated and challenging problem for computer vision experts. Although, the level of performance achieved so far by a lot of existing approaches has increased to a point where face recognition is assumed to be suitable and reliable for many applications, there are still various difficulties that need to be overcome to build accurate and robust face recognition systems. Varying illumination conditions, differences of pose, facial expressions changes, age variations and occlusion offer challenges for face recognition. The motivation of this research is to handle illumination variations in face recognition. There has been research to overcome this aspect of face recognition, but it is still a challenge.

We propose a new approach to recognize faces under the variations of illumination. For the description of face images, we propose a new method called Pooled Homomorphic Binary Rotation Invariant and Noise Tolerant (PH-BRINT). This method uses wavelet transform based homomorphic filtering technique to mitigate the effects of illumination and uses multiscale BRINT operators, which were originally proposed for texture description [1] and are state-of-the-art best local texture feature extractors, and pooling to extract the discriminative information from face images. For classification, we used minimum distance classifier with L1-norm. The performance of the method has been evaluated using a benchmark database for illumination invariant face recognition - Extended Yale-B. The performance is comparable with the state-of-the-art methods dealing with face recognition under varying illumination.

The rest of the paper is structured as follows. Section 2 gives an overview of some state-of-the-art techniques, which have addressed the same problem. Section 3 illustrates in detail the proposed PH-BRINT approach. Section 4 discusses the implementation detail and interprets the results. Finally, Section 5 concludes the paper.

2 Related Work

In this section, first we give an overview of the recent work on illumination invariant face recognition and then give an overview of BRINT.

2.1 Literature Review

In the literature, there are many methods that focus on face recognition under illumination variations using texture descriptors. One such technique uses Local binary patterns (LBP) [2]. Another approach [3] uses Local Ternary Pattern (LTP), which is similar to LBP and is based on extracting local texture micro-patterns.

In the illumination-invariant approach [4], the histogram-equalized faces are divided into tiny overlapping local patches (LPs), then illumination-invariance for these LPs is accomplished by the difference of the vectors and local average illumination, and these vectors are logarithmically normalized to improve the local disparity.

Moreover, Local Directional Pattern (LDP) [5] and Local Directional Number Pattern (LDN) [6], which are based on Kirsch compass masks and generate eight directional edge images from a face image and encode the directional information to represent illumination-invariant face image. The LDP technique represents each pixel using an 8-bit binary code, assigning ones to the three bits corresponding to the three dominant numbers, and zeros to the other five bits. The second approach, LDN represents each pixel using 6-bit binary code; the first three bits encode the position of the top positive directional number and the next three bits encode the position of the top negative directional number. The generated codes from both the approaches are transformed into their corresponding decimal values to create LDP and LDN images.

Faraji et al. [7] proposed a method called adaptive homomorphic eight local directional pattern (*AH-ELDP*) for illumination invariant face recognition. This method adapts adaptive homomorphic filtering to decrease the influence of illumination from an input face image. They apply an interpolative enhancement function to stretch the filtered image, and using Kirsch compass masks, generate eight directional edge images, which are used to create an illumination-insensitive representation.

2.2 Binary Rotation Invariant and Noise Tolerant (BRINT) feature

The idea of face description in our approach is based on BRINT, which was proposed for texture classification. For this reason, we give an overview of BRINT in this section. LBP is widely used for texture description, but there are two main problems with LBP: (i) the number of LBP codes increases exponentially with the number of neighbors R in the circular neighborhood (P, R) of radius R of a pixel, and (ii) it is not robust against noise. To overcome these drawbacks of LBP, Liu et al. [1] proposed BRINT, which is computationally simple and very fast. It is robust against noise, and illumination and rotation variations. Also, it deals with a large number of scales and large circular neighborhoods efficiently.

For computing BRINT feature of a pixel, the $BRINT_{r,q}$ operator is applied on its circular neighborhood consisting of $(8q)$ points located on the circle of radius r. The parameter r defines the spatial scale of the BRINT and q controls the quantization of the angular space. The interesting aspect of BRINT is that a BRINT code always consists of 8 bits regardless of the scale r and the number of sampling points $(8q)$.

The construction of Binary Noise Tolerant (BNT) code is shown in Figure 1. The sampling scheme in BNT is similar to that in LBP, the points around a central pixel $\boldsymbol{x_c}$ are sampled on the circle of radius r. The number of sampled points is a multiple of eight, thus the neighbors of the central pixel $\boldsymbol{x_c}$ sampled on radius r are $\underline{\boldsymbol{x}}_{r,8q} = [\boldsymbol{x}_{r,8q,0}, \ldots, \boldsymbol{x}_{r,8q,8q-1}]^T$. Figure 1 shows the case with $q = 3$.

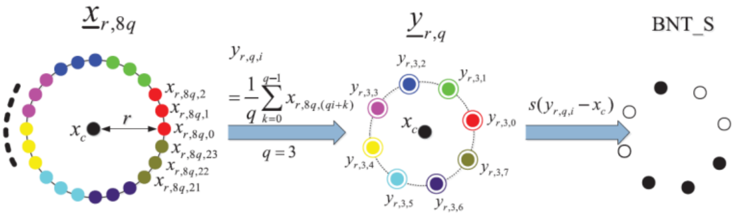

Fig. 1. The construction of BNT code [1]

BNT uses ABQ idea i.e. Average-Before-Quantization and converts the original grey-level values of the neighborhood into local average values $\{\boldsymbol{y}_{r,8q,0}, \ldots, \boldsymbol{y}_{r,8q,7}\}$, where

$$y_{r,q,i} = \frac{1}{q} \sum_{k=0}^{q-1} x_{r,8q,(qi+k)}, \quad i = 0, \ldots \ldots, 7 . \tag{1}$$

This averaging operation makes BNT robust against noise. After averaging, BNT applies thresholding function, which is similar to that used in LBP computation,

$$BNT\_S_{r,q} = \sum_{n=0}^{7} s\left(y_{r,q,n} - x_c\right) 2^n . \tag{2}$$

The thresholding function $s(x) = 1$ if $x \geq 0$, and $s(x) = 0$ otherwise. As result, there are $2^8 = 256$ BNT binary patterns for each (r, q). To make BNT rotation invariant i.e. BRINT, the following operation is applied, which is similar to the one used for rotation invariant LBP,

$$BRINT\_S_{r,q} = min\,\{ROR(BNT, i)|i = 0, .., 7\}\,. \tag{3}$$

Here ROR(x, i) is a rotation function that performs i times a circular bit-wise right shift on x and retains only rotationally-unique patterns; in this way the number of resulted bins for each scale is 36 [8].

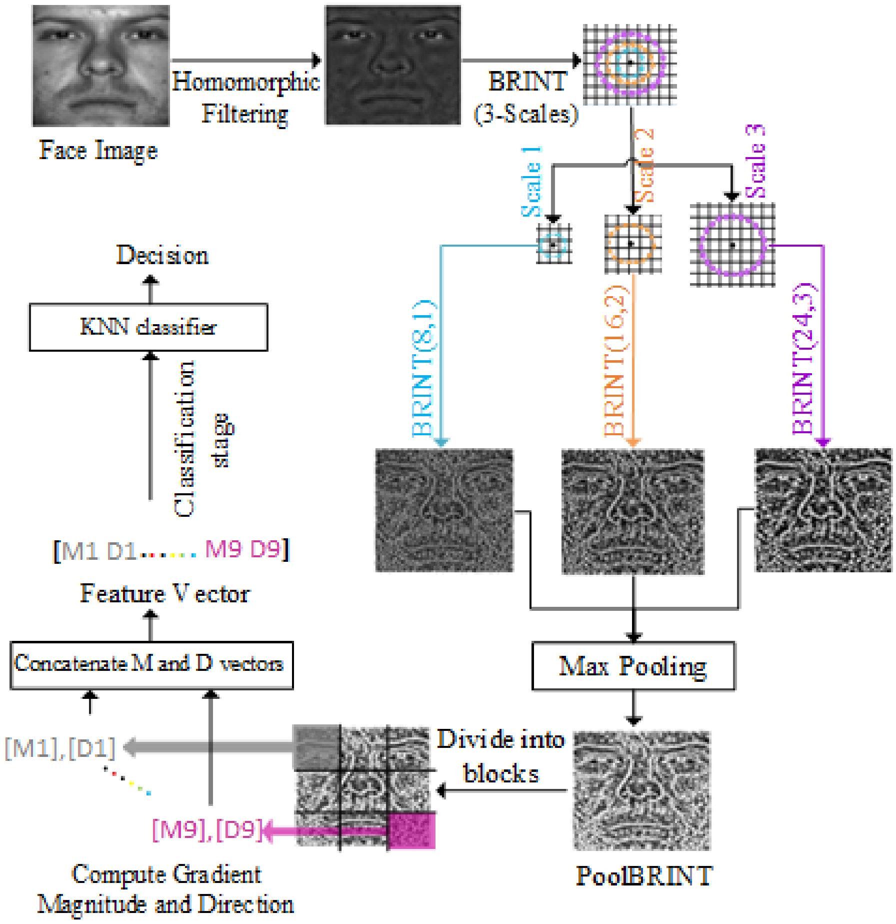

Fig. 2. The architecture of the PH-BRINT based face recognition

3 Pooled Homomorphic BRINT based Face Recognition

In this section, we describe in detail our proposed system for face recognition under varying illumination conditions, the structure of the system is shown in Figure 2. The two main components of the system are face description and classification. For classification, we used minimum distance classifier (KNN with K = 1) with L_1-norm [9] . For face description, we proposed a new descriptor, which is the main contribution of the paper. The construction of the descriptor involves three main operations: homomorphic filtering, generation of multiscale BRINT codes and pooling. In the following sections, give the detail of these operations.

3.1 Homomorphic filtering

According to the Lambertian-reflectance model, a face image I is represented as follows

$$\mathrm{I}(\mathrm{x},\mathrm{y}) = R(x,y)L(x,y) \tag{4}$$

where $R(x, y)$ and $L(x, y)$ are the reflectance and illuminance components. R is a high-frequency component and contains texture information of a face, also it embeds the detail of high frequency parts like skin, eyebrows, eyes and lips, and as such it plays the key role in recognition. On the other hand, L is a low-frequency component because the illumination values of neighboring pixels in face image are similar to each other [7],[10] For illumination invariant face recognition, it is important to mitigate the effect of illumination. The image I(x, y) is not separable, it can be separated using logarithmic operation as follows:

$$\ln\big(I(x,y)\big) = \ln\big(R(x,y)\big) + \ln\big(L(x,y)\big)\ . \tag{5}$$

The Discreet Wavelet Transform (DWT) of ln (I(x,y)) gives:

$$\mathrm{DWT}(\ln(\mathrm{I}(x,y))) = \textstyle\sum_{i=1}^{J}(LH^i\psi_v^i + HL^i\psi_h^i + HH^i\psi_d^i) + LL^J\phi^J\ . \tag{6}$$

The first term on the right hand side of Equation (6) is the high frequency component where the second term is the low frequency component. Keeping in view Equation (5), the first term on the right hand side of Equation (6) is the log of R i.e.

$$\ln(\mathrm{R}(x,y)) = \textstyle\sum_{i=1}^{J}(LH^i\psi_v^i + HL^i\psi_h^i + HH^i\psi_d^i)\ . \tag{7}$$

This analysis indicates that to eliminate the illuminance component, first take log of I, then discompose it using DWT, eliminate the low sub-band, reconstruct the image using only high frequency sub-bands, and undo the effect of log by applying exponential transformation. Figure 2 displays sample image from subset S0* (the detail of these subsets is given later), and the corresponding homomorphic filtered image, the illuminance component has been removed.

3.2 Multiscale BRINT Features

Texture micro-structures play key role in discrimination and these structures exist in a face image at different scales. For the reasons described in Section 2.2, we adopted BRINT for modeling multiscale texture patterns. Keeping in view the previous studies, we compute BRINT codes with three scales: $(r, 8q)$ ={(1, 8), (2, 16), (3, 24)}, as shown in Figure 1. The computed BRINT codes are robust against noise and the number of codes at each scale is same.

3.3 Pooling and PH-BRINT Descriptor

After computing BRINT codes (BRINT 1, BRINT 2, BRINT 3) at three scales, we apply pooling operation to extract the most discriminative information from the three scales. Different pooling operations are possible, but motivated by the success of MAX pooling operation in many applications, we apply MAX pooling operation i.e.

$$PoolBRINT = MAX(BRINT1, BRINT2, BRINT3)$$

Gradient information plays key role in discrimination. Finally, to capture the discrimination information keeping in view its spatial location, we divide PoolBRINT into a number of blocks, compute the gradient magnitude and direction vectors of each block and concatenate them.

4 Results and Discussion

In this section, we evaluate the performance of the proposed method, give comparison with the state-of-the-art methods. First, we give the detail of the Extended Yale B database as we used in our experiments, and the different parameters involved in the system and the evaluation protocol. Then we discuss the tuning of parameters and finally present the results and discuss them.

4.1 Evaluation Protocol

PH-BRINT approach was tested by conducting experiments on Extended Yale B face database. This database was chosen because the face images have considerable variation of illumination and is available on the Internet as a benchmark database for research purposes [11].

The Extended Yale B database contains greyscale face images of 38 subjects. The images were captured under 9 different poses and 64 illumination conditions. The frontal face images were used in the experiments for each subject. Following the protocol used in [7], the face images were classified into six subsets S0~S5 according to the angle between the light source direction and the central camera axis, as shown in Table 1.

The subject S0 contains 6 images per subject, out of these 6 images, one image is held out as training and the remaining 5 are used for testing. As such, the subject is partitioned into S0, consisting of one image per subject, and S0*, containing 5 images per subject. There are six possibilities of selecting one image, so there are six different versions of S0. Taking each version of S0, six experiments were performed for each of the six sub-sets S0*~S5. We computed accuracy as the percentage of correctly recognized faces, average accuracy over six different choices of S0 is used for performance evaluation.

Table 1. The detail of Extended Yale B face database used for the evaluation of PH-BRINT method. The total number of subjects is 38.

Set	S0	S0*	S1	S2	S3	S4	S5
Angle between light and camera axis	0°	0°	1°~12°	13°~25°	26°~50°	51°~77°	above78°
Images/subject	1	5	8	10	12	10	18
Total images	38	190	301	380	449	380	676

4.2 Parameter Tuning

The system involves several parameters, the tuning of these parameters is important for best performance. The parameters of BRINT were adjusted according to the recommendation given in [1] i.e. rotation invariant version was applied with three different scales $(r, 8q) \in \{(1, 8), (2, 16), (3, 24)\}$.

Homomorphic filtering with DWT involves the level of decomposition J. Several experiments were conducted to assess the effect of the level of decompositions (J), and it was concluded that four levels i.e. J= 4 gives the best performance.

Moreover, PoolBRINT face image is divided into blocks, so the number of blocks is another parameter to be tuned for best performance. We tested eleven different options: 1, 2×2, 3×3, 4×4, … until 11×11 with J = 4 and there is no mapping. The corresponding results are shown in Table 2. There is no significant difference between different choices, however the division 6×6 gives the best accuracy of 86.91%, so we fixed the number of blocks to 6x6 blocks in our next experiments to test the best type of mapping as shown in Table 3. The outcomes illustrate that the proposed simple PoolBRINT with 6x6 blocks gives the best accuracy.

Table 2. Recognition accuracy with different number of blocks: 1, 2x2, 3x3, 4x4, until 11x11, where is no mapping

Number of blocks	S0*	S1	S2	S3	S4	S5	Average
1x1	83.60	90.86	87.32	85.75	88.82	84.32	86.78
2x2	83.77	90.97	87.37	86.04	88.77	84.20	86.85
3x3	83.68	91.14	87.32	85.89	88.68	84.37	86.85
4x4	84.20	90.53	87.63	86.12	88.29	83.88	86.78
5x5	83.68	90.81	87.37	86.23	88.07	84.22	86.73
6x6	83.33	91.20	87.72	86.08	88.86	84.27	**86.91**
7x7	83.25	90.81	87.63	86.15	88.29	84	86.69
8x8	83.07	90.59	87.41	86.04	88.16	83.90	86.53
9x9	83.07	90.97	87.46	85.93	88.82	83.83	86.68
10x10	82.89	91.20	86.45	86.04	88.33	83.56	86.46
11x11	80.53	90.20	86.75	84.48	85.39	81.16	84.75

Table 3. Recognition accuracy with different mappings: uniform version (u2), rotation invariant (ri), rotation invariant uniform(riu2), and no mapping where blocks' number is 6x6

Mapping type	S0*	S1	S2	S3	S4	S5	Average
No mapping	83.33	91.20	87.72	86.08	88.86	84.27	**86.91**
u2	82.89	88.48	86.67	85.97	88.03	83.56	85.93
ri	80.09	84.33	84.39	85.23	86.32	79.44	83.30
riu2	73.42	81.71	79.82	80.55	86.32	71.15	78.83

4.3 Discussion of Results

In Extended Yale B database, there are six images per person with zero angle between light and camera axis, we selected one image per person at a time to form the training set S0, there were six choices for S0. For every choice of S0, taking S0*~S5 as testing sets, we performed experiments. The results are shown in Table 4. The PH-BRINT method achieved an average accuracy of 86.91%. The accuracy is lowest when image-6 is selected for training, whereas the accuracy is highest when image-2 is used as training image. With four levels of the variation of illumination (S1~S4), the average recognition accuracy is almost same, and better than those with the other two levels S0* and S5. The reason why the system does not give good result for S0* is that this set changes during each of the experiments. Overall, the system performs well when the variation of illumination is below 78^{o}.

Table 4. Recognition accuracy with differnt training images

Training image	S0*	S1	S2	S3	S4	S5	Average
Image1	86.84	87.38	95.79	79.29	88.95	76.33	85.76
Image2	86.84	95.02	92.89	86.41	91.05	84.62	89.47
Image3	83.16	88.37	82.37	88.86	84.21	88.76	85.96
Image4	79.47	85.38	73.95	92.65	93.42	98.22	87.18
Image5	83.68	92.69	98.42	82.85	94.21	79.88	88.62
Image6	80	98.34	82.89	86.41	81.32	77.81	84.46
Average	83.33	91.20	87.72	86.08	88.86	84.27	86.91

4.4 Comparison with exiting methods

The performance results of recognition accuracy of our system are compared with four recently proposed techniques, which we have discussed in Section 2.1. The state-of-the-art systems used for comparison are LBP[2], LDP[5], LDN[6], and AH-ELDP[7]. Table 5 and Figure 3 show the comparative results. It can be noted that PH-BRINT outperforms LBP, LDP, LDN, and AH-ELDP. It indicates that the performance of the proposed system is comparable that with the state-of-the-best algorithms.

Table 5. Recognition accuracy (%) of PH-BRINT and state-of-the-art methods

Method	S0*	S1	S2	S3	S4	S5	Average
LBP	63.07	82.45	76.32	67.52	72.54	75.17	72.29
LDP	67.89	84.60	77.10	73.31	65.96	64.32	70.06
LDN	61.58	83.17	74.30	65.07	66.27	64.08	67.39
AH-ELDP	77.46	88.26	84.82	81.89	86.32	89.79	84.42
PH-BRINT	**83.33**	**91.20**	**87.72**	**86.08**	**88.86**	**84.27**	**86.91**

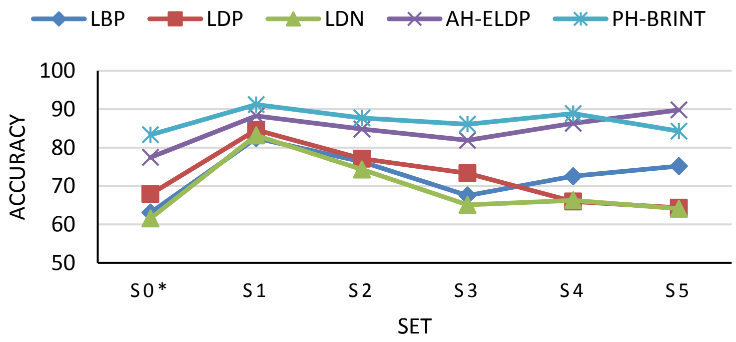

Fig. 3. Recognition accuracy comparison of PH-BRINT method while change training image

5 Conclusion

This paper proposes face-recognition system based on homomorphic filtering, multiscale BRINT and pooling to recognize face images which suffer from variations in illumination. The homomorphic filter is used to reduce the influence of illumination. To capture the multiscale texture micro-patterns, multiscale BRINT with three scales is used. In addition, max pooling is applied to extract the most discriminative information. The experiments on the Extended Yale B database show that the proposed PH-BRINT method achieved average accuracy of 86.91 %. The comparison reveals that the proposed method is comparable with the state-of-the-art techniques. The future work is to investigate alternative techniques for feature vector extraction from Pool-BRINT image and to further validate the performance of the method on other illumination invariant databases.

References

[1] L. Liu, Y. Long, P. W. Fieguth, S. Lao, and G. Zhao, "BRINT: Binary Rotation Invariant and Noise Tolerant Texture Classification," IEEE Trans. Image Process., vol. 23, no. 7, pp. 3071–3084, Jul. 2014.doi: 10.1109/TIP.2014.2325777

[2] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” Pattern Anal. Mach. Intell. IEEE Trans. On, vol. 28, no. 12, pp. 2037–2041, Dec. 2006. doi: 10.1109/TPAMI.2006.244

[3] Xiaoyang Tan and B. Triggs, “Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions,” Image Process. IEEE Trans. On, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.doi: 10.1109/TIP.2010.2042645

[4] A. A. Shafie, F. Hafiz, and Y. M. Mustafah, “Face recognition using illumination-invariant local patches,” in 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS), 2014, pp. 1–6.doi: 10.1109/ICIAS.2014.6869544

[5] T. Jabid, M. H. Kabir, and O. Chae, “Local Directional Pattern (LDP) for face recognition,” in 2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE), 2010, pp. 329–330.doi: 10.1109/ICCE.2010.5418801

[6] A. Ramirez Rivera, R. Castillo, and O. Chae, “Local Directional Number Pattern for Face Analysis: Face and Expression Recognition,” IEEE Trans. Image Process., vol. 22, no. 5, pp. 1740–1752, May 2013.doi: 10.1109/TIP.2012.2235848

[7] M. R. Faraji and X. Qi, “Face recognition under varying illumination based on adaptive homomorphic eight local directional patterns,” IET Comput. Vis., vol. 9, no. 3, pp. 390–399, 2015.doi: 10.1049/iet-cvi.2014.0200

[8] M. Pietikäinen, T. Ojala, and Z. Xu, “Rotation-invariant texture classification using feature distributions,” *Pattern Recognit.*, vol. 33, pp. 43–52, 2000.

[9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[10] Y.-S. Huang and C.-Y. Li, “An Effective Illumination Compensation Method for Face Recognition,” in *Advances in Multimedia Modeling: 17th International Multimedia Modeling Conference,* Springer Berlin Heidelberg, 2011, pp. 525–535.

[11] 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, Jun. 2001.