A Photometric-based Illumination Compensation Approach for Face Recognition

Guilherme F. Plichoski[1], Chidambaram Chidambaram[2], Rafael Stubs Parpinelli[1]

[1]Graduate Program in Applied Computing – State University of Santa Catarina
Joinville – SC – Brazil

[2]Department of Production Engineering and Systems – State University of Santa Catarina
Joinville – SC – Brazil

guilherme.plichoski@edu.udesc.br, {chidambaram, rafael.parpinelli}@udesc.br

Abstract. *Face recognition (FR) systems can perform well under uncontrolled illumination, but there is no general and robust technique with total immunity to all conditions. Hence, investigating it still is an open research area. In this work, we investigate methods that can be applied to illumination changes and, with this, it is proposed an illumination invariant FR approach that employs unlighting photometric-based techniques together with a Local Binary Patterns (LBP) texture descriptor. To conduct the experiments, we used the leave-one-out accuracy estimation model using the well known Yale B face dataset with different degrees of illumination. The results show that our methodology is robust since it achieves 100% FR rate in the most challenging subsets.*

1. Introduction

In the last decades, biometrics identification has brought attention due the exponential growth of technology expanding the wide range of possibilities for its application. Nowadays, we have biometric identification systems through fingerprint, iris and retina, voice, face among others. In these, face recognition (FR) systems strongly highlights because it does not need full interaction with the subject. However, some challenges arise such as occlusion, scale, pose variants and uncontrolled illumination. Among these conditions that affect the FR process, illumination variation can be considered as one of the most critical. It affects the classification rate in a more significant manner than even the effect caused by having different subjects in a database [Adini et al. 1997]. According to the extensive review [Ochoa-Villegas et al. 2015], we can identify many illumination compensation techniques that perform well on images obtained under uncontrolled conditions, but there is no generic technique with total immunity to overall conditions. This context strongly implies that there is more space for improvement and investigation to develop new approaches. The use of unlighting methods, which aim to suppress the variation, have been demonstrated to be a practical solution to address illumination challenges [Ochoa-Villegas et al. 2015]. As example, we can mention [Hu 2011] in which the authors proposed to use Discrete Wavelet Transform (DWT) and multi-scale reflectance model which preserves the illumination discontinuities by reducing the halo artifacts and finally, increases recognition performance. In the same way, the authors [Yuan et al. 2013] used the wavelet transform and mapped the image by applying a method that normalizes the illumination based on its energy. Another complex method was proposed by [Baradarani et al. 2013] using double-density dual-tree complex wavelet transform (DD-DTCWT) which demonstrated to be effective in compensating illumination variation. According to our survey, it can be found many photometric based works in the literature. Hence, some of them are mentioned in the following. [Lai et al. 2012] assigned different weights to the sub-bands of wavelet decomposition according to the contribution to the recognition task through a regularization criterion function. [Cheng et al. 2012] introduced an unlighting method based on the retinal illumination model meanwhile [Manikantan et al. 2012] proposed a methodology using 2D-DWT multi-level illumination suppression which proved to be effective in practical situations where the images are captured under uncontrolled and unknown surroundings. Based on the above discussions, the present work aims to investigate the cascade application of different unlighting photometric techniques in order to achieve illumination invariance. To validate the results, we employed the *leave-one-out* accuracy estimation model in the well known Yale B face dataset [Georghiades et al. 2001, Lee et al. 2005] which has a wide range of illumination degrees. The contributions of this work rely on the FR model and on the reliability of the results based on our experimental design.

This paper is presented as follows: in Section 2 we detail the preprocessing techniques employed in our FR model; in Section 3, the feature extraction techniques are discussed and; in Section 4 the proposed FR approach is exposed. Finally, in Sections 5 and 6, the results and conclusions are presented, respectively.

2. Preprocessing Techniques

The preprocessing is the initial step in a FR system where different techniques are usually employed, mostly, to deal with irregularities in images that might obfuscate discriminative features. In the following sub-sections, we detail the techniques that are used in the present work.

2.1. Bi-cubic Interpolation

According to [Han 2013], the most common interpolation techniques that are used to reduce image resolution are the Nearest Neighbor, Bilinear, Bi-cubic and Cubic-B-Spline. The Nearest Neighbor Interpolation is simple and fast, but it can bring significant distortion and it might produce mosaic and saw tooth phenomenon. The Bilinear and the Cubic-B-Spline Interpolation methods are more complex, but in this application the high frequency component will appear faded and the image contour can have some degree of fuzzy. On the other hand, Bi-cubic Interpolation keep the image information by maintaining its quality, in addition to this, it speeds up the computations without losing much information.

2.2. Wavelet-based Illumination Normalization (WBIN)

Wavelet decomposition has been successfully applied in the illumination normalization problem due its ability to provide a multi-resolution analysis of the image and the capability of decomposing the image into sub-bands in both time and frequency domains. In 2D-DWT, the image is represented in terms of translations and dilations of a scaling function and a wavelet function using a 2D filter bank consisting of low-pass and high-pass filters [Du and Ward 2005]. High pass filtering produces detail information (such as edges) and low pass filtering with scaling produces coarse approximations. At one level decomposition, the 2D-DWT produces four components, which are approximation (LL), horizontal (HL), vertical (LH), and diagonal (HH), as shown in Figure 1.

Figure 1. DWT decomposition levels. Left: original image. Middle: Decomposed components. Right: Components specification.

As pointed out in [Du and Ward 2005], if the original image does not use most of the available dynamic range, which is the distribution of light, the transformed coefficients will not use most of the dynamic range either. Therefore, histogram equalization is performed in the approximation coefficients (LL) to achieve contrast enhancement. The fine details in the image are emphasized by enlarging the amplitude of the detail coefficient matrices (HL, LH and HH), also multiplying each element with a scale factor (*Sf*) greater than 1. Then, the enhanced image is reconstructed using inverse wavelet transform. This procedure may be repeated *n* times.

2.3. Gamma Intensity Correction (GIC)

Depending on the illumination intensity level, the image may need improvement to reveal all discriminative power. According to [Shan et al. 2003], gamma correction, also known as power law transformation, can control overall brightness of an image by changing the parameter γ. It's formula is shown in Equation 1.

$$O(u, v) = c * I(x, y)^{\gamma} \tag{1}$$

where *O(u,v)* is the gamma corrected image, *c* is a constant, *I(x,y)* is the input image and γ is the correction factor [Manikantan et al. 2012].

2.4. Laplacian Edge Detection

An edge is a set of pixels in the image where intensity abrupt changes, representing the shape of face, eyes, nose and other discriminant features. In [Kushwah et al. 2017], a review of edge detection procedures was performed covering 1^{st} order derivative-gradient methods such as Sobel, Prewitt and Robert operators, and 2^{nd} order derivatives or zero crossing which includes Laplacian of Gaussian (LoG) and Difference of Gaussian, and Canny operator as optimal method. Among these, LoG presented better results for both, visual perception and edge quantity.

3. Feature Extraction

Feature extraction step consists in retrieving informative and non-redundant features leading to a better representation of interesting parts of an image as a compact feature vector. In fact, this step performs as a dimensionality reduction of features. In this section, we introduce the techniques that are used in our methodology.

3.1. Discrete Wavelet Transform

As mentioned in Section 2.2, the 2D-DWT produces four sub-bands (Figure 1) containing the approximation component (LL) and three detail components (HL, LH and HH). According to [Manikantan et al. 2012], the information in the approximation component have the most discriminant features. Hence, the required facial features for recognition remains in this sub-band, therefore, only these features are extracted. This procedure may have n levels of decomposition applied recursively.

3.2. Local Binary Patterns (LBP)

Texture is a fundamental characteristic of the appearance of surfaces, which is an important component of many FR systems. When choosing a texture descriptor, two competing goals should be considered, namely, low computational complexity, and capturing the most representative texture information. Thus, different texture classes can be distinguished despite the presence of various imaging distortions. LBP emerged as one of the most prominent and widely studied local texture descriptors [Liu et al. 2017]. Its canonical version is used in this work in which the features are extracted directly from the input image. Figure 2 illustrates LBP's procedure with a kernel size (*ksize*) of 3x3.

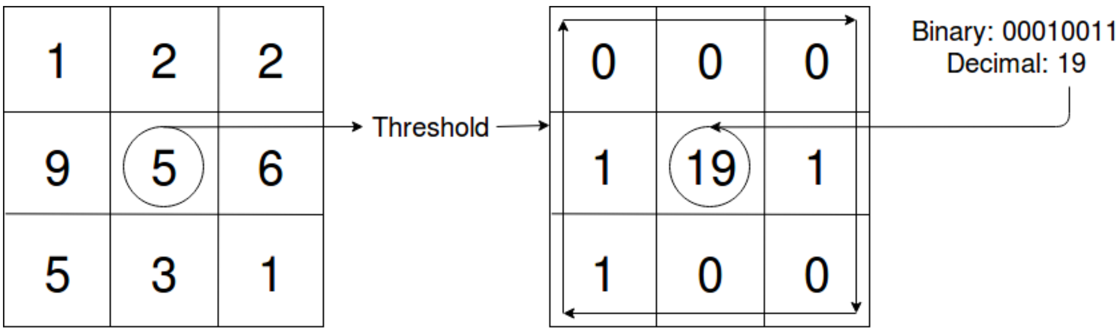

Figure 2. LBP texture descriptor procedure.

The neighborhood structure is a set of pixels taken from a square neighborhood, usually of 3x3 pixels, which are compared against the value of the central pixel (thresholded) resulting in a 8 bits binary vector which is converted in decimal. Then, the resultant value is used to represent the pixel.

4. Face Recognition Approach

The proposed FR system starts with a resolution reduction in the input images by using the Bi-cubic Interpolation, in which the output pixel values are calculated from a weighted average of pixels in the nearest 4-by-4 neighborhood [Gonzalez and Woods 2008]. Then, the WBIN technique is used to enhance the contrast and emphasize the details in the image through the multilevel haar wavelet transform since it is a powerful technique for multi-resolution decomposition of time series [Manikantan et al. 2012]. Figure 3 shows the resultant image using as input the image from Figure 1.

Figure 3. The resultant image after applying the WBIN technique.

Depending on the illumination intensity level, the image might still need improvement to reveal all discriminative power, thus, GIC is employed to adjust brightness. Figure 4 shows the gamma corrected image.

Figure 4. Gamma corrected image.

Next, we apply the LoG edge detector to simplify the analysis of images reducing the amount of data while preserving the structural information. Figure 5 shows the edge detected image with kernel size (*ksize*) of 3x3.

Figure 5. LoG edge detected image.

Now, the image is prepared for feature extraction using the LoG edge detected image where the 2D-DWT is employed and the approximation component (LL) is extracted. This is used for feature extraction since the required features are present in this sub-band. Finally, the texture descriptor LBP is performed resulting in the feature vector used for classification. The k-nearest neighbor (k-NN) algorithm is used for classification, which consists in returning the k closest training images given a test image. If k = 1, then the single nearest neighbor is the output. This classification method has been applied successfully to a wide range of problems, such as FR [García-Pedrajas et al. 2017]. The Euclidean distance was employed to measure the similarity between test and trained images using Equation 2.

$$D = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{2}$$

where x_i (or y_i) is the coordinate of $\vec{x}$ (or $\vec{y}$) in dimension i.

Figure 6 shows the flowchart representing the proposed FR approach. Both training and testing stages encompass the same routines.

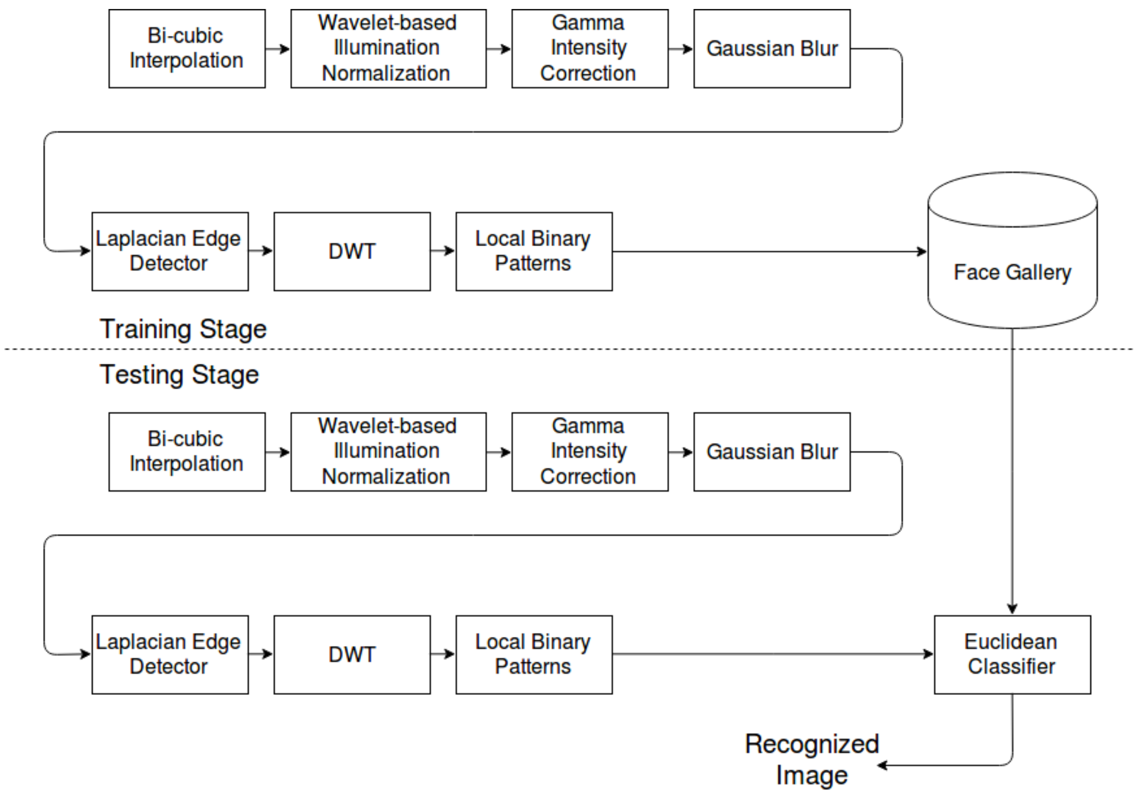

Figure 6. Flowchart of the FR approach.

5. Experiments and Results

The experiments were conducted in the Yale B face dataset [Georghiades et al. 2001, Lee et al. 2005] with the *leave-one-out* accuracy estimation model, which are detailed next. Table 1 shows the parameters used in the experiments. The values for the parameters were chosen based on an empirical experimentation. All experiments were performed in an AMD Phenom II X4 B93 Processor with 4 cores and 4 Gigabytes of memory running Ubuntu 16.04 LTS operating system. The algorithm was designed in C++ programming language with multi-threading OpenMP [Dagum and Menon 1998] library using 4 threads for classification. To be as fair as possible, parallelism was not employed in the experiments when comparing to the results found in literature.

Table 1. Parameters used in the experiments.

Technique	Parameter	Value
Bi-cubic Interpolation	Percentage of reduction	50%
WBIN	n	1
	Sf	2 to 4
GIC	γ	2.4
LoG	$ksize$	3x3
2D-DWT	n	1 to 3
LBP	$ksize$	3x3
k-NN	k	1

Yale B face dataset is a well known database which consists of 38 subjects under different conditions varying pose and illumination (9 poses x 64 illuminations). As our purpose is to compensate illumination variation, only the frontal face images were selected. This database is commonly divided into five subsets according to the angle between the light source direction and camera axis [Ochoa-Villegas et al. 2015]. Table 2 shows these variations and Figure 7 illustrates the images used in the subsets.

Table 2. Five subsets according to the light angle source directions.

	S1	S2	S3	S4	S5
angle	0^o-12^o	13^o-25^o	26^o-50^o	51^o-77^o	$>78^o$
number of images	6	12	11	12	16

Figure 7. Yale face database B subsets example from S1 to S5.

To ensure accuracy in the results, k-fold cross-validation model was employed. In this model, sometimes called rotation estimation, the dataset (D) is randomly split into k mutually exclusive subsets D_1, D_2,...,D_k of equal size, which are trained and tested k times. Each subset D_i is tested on the remained subsets (D-D_i). According to

[Kohavi et al. 1995], the cross-validation is an estimate that depends on the division into folds. A complete factorial cross-validation is known as *leave-one-out* where every image is tested with all the remaining images. Based on this definition, we chose to use the *leave-one-out* model.

Table 3 shows the recognition rates for each subset followed, between parenthesis, by the execution time of the whole system in seconds. The entire experiment was carried out varying the DWT level and the scale factor (*Sf*). As the images are resized applying DWT, its size is shown at each level of decomposition, e.g. L1(42x48). Last row indicates the recognition rates without any preprocessing techniques (RAW DB). Bold cells indicate overall best results.

Table 3. *Leave-one-out* recognition rates (%) and execution time(*s*) between parenthesis, respectively.

DWT Level	Sf	S1	S2	S3	S4	S5
L1 (42x48)	2	95.61 (0.515)	**100 (1.044)**	**100 (0.97)**	99.12 (1.049)	**100 (1.413)**
	3	95.61 (0.514)	**100 (1.01)**	**100 (0.949)**	99.56 (1.047)	**100 (1.381)**
	4	95.61 (0.511)	**100 (1.016)**	**100 (0.817)**	**99.78 (1.032)**	**100 (1.403)**
L2 (21x24)	2	98.24 (0.438)	**100 (0.886)**	99.52 (0.853)	99.56 (0.893)	**100 (1.176)**
	3	**98.68 (0.446)**	**100 (0.874)**	**100 (0.831)**	98.9 (0.897)	**100 (1.185)**
	4	**98.68 (0.438)**	**100 (0.872)**	99.52 (0.847)	99.12 (0.9)	99.83 (1.184)
L3 (10x12)	2	94.73 (0.43)	98.68 (0.84)	83.01 (0.797)	76.75 (0.87)	88.81 (1.144)
	3	96.05 (0.424)	99.34 (0.846)	79.9 (0.801)	75.21 (0.87)	88.32 (1.146)
	4	95.61 (0.425)	99.56 (0.856)	77.99 (0.805)	76.31 (0.858)	97.33 (1.149)
RAW DB	-	90.78 (0.416)	97.14 (1.696)	64.35 (1.390)	36.62 (1.650)	33.88 (2.913)

According to the experiments, it is worth pointing out that the best results were achieved with 2D-DWT levels 1 and 2. Also, the changes in the WBIN *Sf* parameter does not affect much the results. We also notice that the proposed methodology achieved lower accuracy in S1 subset regarding other subsets. The subset is composed of brighter images. In addition to this, the low rate might be due to the parameters we chose. But, overall, from the results in this dataset, we can conclude that the proposed methodology is robust enough to different lighting conditions. Among all results, the very low rate is obtained in 2D-DWT Level 3. It probably happened because the image lost discriminant information as it has low resolution (10x12). The low execution time is another thing to be considered, being ideal in real-world applications. Comparing the recognition rates with no preprocessing techniques (RAW DB), we can realize that the proposed method demonstrates strong illumination compensation effect.

In [Manikantan et al. 2012], the authors proposed a similar architecture, but they do not expose some parameters used in their work, and did not employed any feature extraction technique besides 2D-DWT. Table 4 shows the best results of both works on S5 subset. To promote a fair comparison, parallelism was not employed and the same validation model was used in this experiment: the dataset is divided equally with half images for training and other half for testing, randomly.

Table 4. Comparison with literature on S5 subset.

	DWT Level	Sf	Average Recognition Rate (%)	Average Execution Time (s)
Our method	**1**	**2**	**100**	**1.41**
[Manikantan et al. 2012]	1	3	99.84	26.42

Hereupon, we can conclude that considering some parameters effect (i.e. γ correction factor) and employing LBP, it can provide greater accuracy in the results. Also, [Manikantan et al. 2012] included a feature selection step in their system, which does not provide a better accuracy nor a better efficiency, actually making the system slower. In terms of execution time, our method showed to be at least 18 times more efficient.

6. Conclusions and Future Work

From the early studies to the most recent ones have shown that there is no generic technique that is immunity to all illumination conditions, pointing that the illumination compensation is still an open research topic. Hence, in this paper, we proposed an illumination invariant FR approach using an unlighting photometric-based methodology. The well known Yale B face dataset was used to test our approach. In addition to this , we have also applied the trustworthy *leave-one-out* validation accuracy estimation model due its experimentation completeness.

In this work, it has been shown that the present photometric-based methodology is suitable for the FR task achieving 100% in subsets S2, S3 and S5, which has the most dark images, following by 99.78% in S4 and 98.68% in S1. This last subset, S1, has the most brighter images. Hence, we believe that one alternative way to achieve the highest recognition rates would be the tuning of parameters that might lead to optimal accuracy. Regarding the execution time, our method reduced the execution time significantly showing high efficiency compared to [Manikantan et al. 2012].

We believe that future directions would be to employ an optimization algorithm to tune all the aforementioned parameters. Also, include some different feature extraction techniques combined with different classifiers in the optimization process could improve the accuracy. Another direction for future research will be the use face images with higher degrees of illumination variation to reveal the system's illumination compensation power.

References

Adini, Y., Moses, Y., and Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):721–732.

Baradarani, A., Wu, Q. J., and Ahmadi, M. (2013). An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering. *Pattern Recognition*, 46(1):57–72.

Cheng, Y., Li, Z., and Jiao, L. (2012). Illumination normalization for face recognition under extreme lighting conditions. In *International Conference on Intelligent Science and Intelligent Data Engineering*, pages 491–497. Springer.

Dagum, L. and Menon, R. (1998). OpenMP: an industry standard API for shared-memory programming. *IEEE computational science and engineering*, 5(1):46–55.

Du, S. and Ward, R. (2005). Wavelet-based illumination normalization for face recognition. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–954. IEEE.

García-Pedrajas, N., del Castillo, J. A. R., and Cerruela-García, G. (2017). A proposal for local k values for k-nearest neighbor rule. *IEEE transactions on neural networks and learning systems*, 28(2):470–475.

Georghiades, A. S., Belhumeur, P. N., and Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence*, 23(6):643–660.

Gonzalez, R. C. and Woods, R. E. (2008). Digital image processing, pretice hall, 2008. *ISBN-13*.

Han, D. (2013). Comparison of commonly used image interpolation methods. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineerings (ICCSEE)*, pages 1556–1559.

Hu, H. (2011). Variable lighting face recognition using discrete wavelet transform. *Pattern Recognition Letters*, 32(13):1526–1534.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA.

Kushwah, A., Gupta, K., Agrawal, A., Jain, G., and Agrawal, G. (2017). A review: Comparative study of edge detection techniques. *International Journal*, 8(5).

Lai, Z.-R., Dai, D.-Q., and Ren, C.-X. (2012). Face recognition under variable illumination by weighted-subband edge enhancement. In *Chinese Conference on Pattern Recognition*, pages 324–331. Springer.

Lee, K.-C., Ho, J., and Kriegman, D. J. (2005). Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on pattern analysis and machine intelligence*, 27(5):684–698.

Liu, L., Fieguth, P., Guo, Y., Wang, X., and Pietikäinen, M. (2017). Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition*, 62:135–160.

Manikantan, K., Shet, M. S., Patel, M., and Ramachandran, S. (2012). DWT-based illumination normalization and feature extraction for enhanced face recognition. *International Journal of Engineering & Technology*, 1(4):483–504.

Ochoa-Villegas, M. A., Nolazco-Flores, J. A., Barron-Cano, O., and Kakadiaris, I. A. (2015). Addressing the illumination challenge in two-dimensional face recognition: a survey. *IET Computer Vision*, 9(6):978–992.

Shan, S., Gao, W., Cao, B., and Zhao, D. (2003). Illumination normalization for robust face recognition against varying lighting conditions. In *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, pages 157–164. IEEE.

Yuan, X., Meng, Y., and Wei, X. (2013). Illumination normalization based on homomorphic wavelet filtering for face recognition. *Journal of Information Science and Engineering*, 29(3):579–594.