MULTI SCALE MULTI DESCRIPTOR LOCAL BINARY FEATURES AND EXPONENTIAL DISCRIMINANT ANALYSIS FOR ROBUST FACE AUTHENTICATION

Abdelmalik Ouamane[1], Bengherabi Messaoud[1], Abderrezak Guessoum[2] Abdenour Hadid[3] and Mohamed Cheriet[4]

[1] Centre de Développement des Technologies Avancées, Algeria, [2]Blida University, [3]University of Oulu, [4] ETS, Québec, Canada

ABSTRACT

In this paper we present an efficient face verification system based on the fusion of multi-scale multi-descriptor local binary features. First, the face is divided into regions and each region is divided into several patches. For each patch and at every specific scale, the statistics of the baseline Local Binary Pattern (LBP), the Local Phase Quantization (LPQ) and the recently proposed Binarized Statistical Image Feature (BSIF) are summarized by histograms. The histograms of different patches belonging to the same region are concatenated to form a highly dimensional feature vector representing a specific descriptor at a specific scale. Second, we propose an efficient dimensionality reduction technique based on Exponential Linear Discriminant Analysis EDA coupled with Within-Class Covariance Normalization (WCCN) to downgrade the effect of the directions of high intravariability and to enhance the discrimination power of the EDA. The projected histograms for each region are scored using the cosine similarity metric. Lastly, the different region scores corresponding to different descriptors at different scales are fused using support vector machine classifier (SVM). Experimental verification results demonstrate that the proposed authentication pipeline outperforms all the existing systems on the XM2VTS controlled database and interestingly compete with the top performing systems on the challenging LFW database.

Index Terms— face verification, local binary descriptors, Exponential Discriminant Analysis, Within-Class Covariance Normalization.

1. INTRODUCTION

Face recognition is one of the most important research topics in pattern recognition and artificial intelligence. To achieve robust face verification systems, the first challenge focuses on constructing informative handcrafted or learned features from the image [1]. The second important stage is to learn a classification model and the third challenging research direction is to develop new distance metric learning methods [2]. Among the most successful local descriptors for face recognition are the local binary patterns and its variants [3]. Numerous local binary descriptors have been proposed to extract effective information, such as the handcrafted Local Binary Patterns (LBP) [3] and the Local Phase Quantization (LPQ) [4]. Recently, the learned Binarized Statistical Image Features (BSIF) is proposed where very promising results are obtained for both texture classification and face recognition [5]. The extension of LPQ to Multi-Scale LPQ (MLPQ) is proposed to describe the face image more efficiently. Multi-scale Local Phase Quantization histogram (MLPQH) have gained reputation as powerful and attractive texture descriptors showing excellent results in terms of accuracy and computational complexity in face recognition [6]. Furthermore, to bolster the performance, researchers have tried different methods by fusing multiple types of local features at different scales [6], [7]. In this work, we present a highly efficient face verification system based on the fusion of multi-scale LBP histograms (MLBPH), multi-scale LBQ histograms (MLPQH), and multi-scale BSIF histograms (MBSIFH). The face image is divided into several regions and each region is divided into small patches. For each patch and at every specific scale, the statistics of the baseline Local Binary Pattern (LBP), the Local Phase Quantization and the Binarized Statistical Image Features (BSIF) are summarized by the corresponding histograms. The histograms of different patches belonging to the same region are concatenated to form a highly dimensional feature vector representing a specific descriptor at a specific scale. For dimensionality reduction, we propose to use the Exponential Linear Discriminant Analysis (EDA) [8] coupled with Within-Class Covariance Normalization (WCCN) [9] to downgrade the effect of the directions of high intravariability and to enhance the discrimination of the cosine similarity scoring. Lastly, the different region scores corresponding to different descriptors at different scales are fused using SVM. The main contributions of this paper are summarized as follows: First, demonstrating via experiments that multi-scale BSIF is as efficient as multi-scale LPQ and even slightly better in the task of face authentication. Second, demonstrating the efficiency of exponential LDA for face verification and proposing a complete and highly efficient face verification system surpassing all the verification rates existing in literature on the XM2VTS database [10] and competing with the top performing systems on the LFW database [11].

2. DESCRIPTION OF THE PROPOSED RECOGNITION PIPELINE

Most state-of-the art face verification systems are component-based where the whole image is divided into smaller regions, and the similarity is measured component by component.The final decision is based on fusing these component scores. By doing this, more locality information can be preserved. In addition, compared to holistic representation, the region-based methods are much more robust to illumination and expression variations. The reason is that the illumination and expressions variation within the whole image is much greater than that within each component. In this work, the image is divided into regions after applying the corresponding local binary descriptor, we divide the image into 10 regions (5 rows by 2 columns) and each region is subdivided into 15 patches. For each patch and at every specific scale, the statistics of the baseline Local Binary Pattern (LBP), the Local Phase Quantization and the Binarized Statistical Image Features (BSIF) are summarized by histograms. The histograms of different patches belonging to the same region are concatenated to form a highly dimensional feature vector representing a specific descriptor at a specific scale.This strategy of taking the histogram of the whole region rather than the small patches as done in [6] makes the system more robust to occlusion and also results in low number of scores to be fused. However, the concatenation of 15 histograms still results in high dimensional vectors containing redundant information, for this we propose the use of Exponential LDA (EDA) to overcome the small-sample-size (SSS) problem without discarding the discriminant information contained in the LDA null space [8]. The EDA projection is followed by Within-Class Covariance Normalization (WCCN) to downgrade the effect of the direction of high intervariability and to enhance the discrimination of the cosine similarity scoring for each regions. At the end, the region scores are fused by SVM. Fig.1 gives an general diagram of the proposed system.

2.1. Features Extraction

2.1.1. Local Binary Patterns (LBP)

The idea behind LBP is that an image is composed into micro patterns and the statistics of these micro patterns represented by histogram contains information about the distribution of edges, spots and other micro textures in an image [3]. Given an image I and denoting q_c as the grey level of the pixel c of the image I, the LBP operator on this pixel is defined as follows: $LBP(P,R)=\sum_{p=0}^{p-1} s(q_p-q_c)2^p$, where P is the number of pixels in the neighborhood, R is the radius, and $s(x)=1$ if $x\geq 0$, otherwise $s(x)=0$.The LBP images for P=8 and R=2, 4, 6, 8 are shown in Fig.2.a.

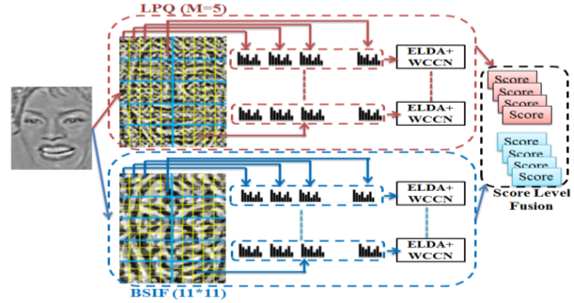

Fig. 1. General Diagram of our proposed framework showing the partition of the face and an example of fusing LBQ and BSIF)

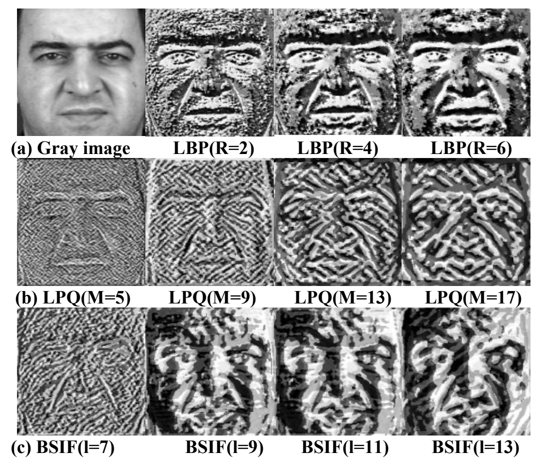

Fig. 2. Local binary features: **(a)** LBP, **(b)** LPQ, **(c)** BSIF

2.1.2. Local Phase Quantization (LPQ)

The principle of LPQ [4] is to extract the phase information in frequency domain which is robust to blur variation. Firstly, Short time Fourier Transform (STFT) is computed over $M\times M$ local rectangular window N_x at each pixel x of the image f(x), that is $F(u,x)=\sum_{y\in N_x} f(x-y)e^{-j2\pi u^T y}=w_u^T f_x$, where w_u is the basis vector of the STFT at frequency u, and f_x is another vector containing all M^2 image pixels from N_x. Only 2-D frequencies u_1= [a, 0]^T , u_2= [0, a]^T , u_3=[a, a]^T , u_4= [a,-a]^T are considered in LPQ.

Let: $F_x^c=[F(u_1,x),F(u_2,x),F(u_3,x),F(u_4,x)]$ and $F_u=[\text{real}(F_x^c), \text{imaginary}(F_x^c)]^T$, The corresponding 8 by M^2 transformation matrix is:

$W=[\text{real}(W_{u1},W_{u2},W_{u3},W_{u4}), \text{imaginary}(W_{u1},W_{u2},W_{u3},W_{u4})]$

So that: $F_x=Wf_x$. For robust phase feature, F_x is postprocessed via the whitening transformation [12] and singular value decomposition (*SVD*), resulting in the final feature G_x. The quantized coefficients are represented as integer values between 0-255 using binary coding:

$b=\sum_{j=1}^{8} g_j 2^{j-1}$, where g_j is the j^{th} component of G_x. Fig. 2.b, illustrates the LPQ images under different scales in which LPQ window size takes different values.

2.1.2. The Binarized Statistical Image Features (BSIF)

Unlike LBP and LPQ which can be seen as statistics of labels computed in the local pixel neighborhoods through predefined set of linear filters and then binarizing the filter responses. The recently proposed local descriptor BSIF(Binarized Statisitcal Image features) is based on the automatic learning of fixed set of filters from a small set of natural images [5], instead of using handcrafted filters such as in LBP and LPQ. To characterize the texture properties within each image sub-region, the histograms of pixels BSIF code values are then used. The value of each element (i.e. bit) in the BSIF binary code string is computed by binarizing the response of a linear filter with a threshold at zero. Each bit is associated with a different filter and the desired length of the bit string determines the number of filters used. The set of filters is learnt from a training set by maximizing the statistical independence of the filter responses. There are two parameters in BSIF descriptor: the filter size l and the length n of the bit string. The corresponding BSIF code images are shown in Figure 2.c, where the filter size l varies from 7 to 13 and the number of bits is 8.

2.2. Exponential Discriminant Analysis (EDA)

Let the training set contain L classes and each class X_i contains n_i samples. The within-class scatter matrix (S_w) and between-class scatter matrix (S_b) are defined as:
$S_w = \sum_{i=1}^{L} \frac{1}{n_i} \sum_{j=1}^{n_i} (x_j^i - m^i)(x_j^i - m^i)^T$ and,
$S_b = \sum_{i=1}^{L} (m^i - \bar{m})(m^i - \bar{m})^T$ where: $m^i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^i$ and $\bar{m}$ is the global mean. The objective of LDA is to find an optimal projection matrix A by maximizing the ratio of between-class scatter S_b to the within-class scatter S_w as follows:

$$A = \underset{A}{argmax} \left\{ tr \left[\left(A^T S_w A \right)^{-1} \left(A^T S_b A \right) \right] \right\} \tag{1}$$

Where V_w is the eigenvector matrix of S_w and $\Lambda_w = diag(\lambda_{w1}, \lambda_{w2}, \dots, \lambda_{wn})$ representing the corresponding eigenvalues of S_w, V_b is the eigenvector of S_b and $\Lambda_b = diag(\lambda_{b1}, \lambda_{b2}, \dots, \lambda_{bn})$ representing the eigenvalues of S_b and n is the sum of the $n_{i's}$.

In general, the matrix S_w is not a full-rank matrix due to the small-sample-size problem (SSSP). In fact, the null space of Sw contains the most discriminative information especially when the projection of S_b is not zero in that direction [8]. LDA cannot extract discriminant information contained in the null space of S_w. To extract this kind of discriminant information, we replace λ_{wi}, i.e., the eigenvalues of S_w, by $\exp(\lambda_{wi})$ and λ_{bi}, i.e., the eigenvalues of S_b, by $\exp(\lambda_{bi})$ and denote $exp(\Lambda_b) = diag(exp(\lambda_{b1}), exp(\lambda_{b2}), \dots, exp(\lambda_{bn}))$, $exp(\Lambda_w) = diag(exp(\lambda_{w1}), exp(\lambda_{w2}), \dots, exp(\lambda_{wn}))$. Then, equation (1) is transformed into:

$$\begin{aligned} \hat{A} &= \underset{A}{argmax} \left\{ tr \left[\left(A^T \left(V_w^T exp(\Lambda_w) V_w \right) A \right)^{-1} \left(A^T \left(V_b^T exp(\Lambda_b) V_b \right) A \right) \right] \right\} \\ &= argmax_A \left\{ tr \left[\left(A^T exp(S_w) A \right)^{-1} \left(A^T exp(S_b) A \right) \right] \right\}, \end{aligned} \tag{2}$$

According to matrix exponential property [13], [14]: If V are eigenvectors of B that correspond to the eigenvalues $[\lambda_{1b}, \lambda_2, \dots, \lambda_n]$, then V are also eigenvectors of matrix $exp(B)$ that correspond to eigenvalues $[exp(\lambda_1), exp(\lambda_2), \dots, exp(\lambda_n)]$.

The matrix $exp(S_w)$ is a full-rank matrix; therefore, the discriminant information contained in the null space of S_w can be extracted by using EDA. EDA leads to the projection matrix $\hat{A}$ that comprises the leading eigenvectors of $(exp(S_w))^{-1} exp(S_b)$. More details about exponential LDA can be found in [14], and the references therein.

2.3. Within-Class Covariance Normalization (WCCN):

Within-Class Covariance Normalization (WCCN) is a technique initially introduced for SVM-based speaker authentication [9]. It has since be successfully applied to i-vectors for speaker verification [15]. It is found in [15] that the best approach is to project the LDA reduced total variability I-vectors to a new subspace specified by the square-root of the inverse of the within-class covariance matrix:

$$W = \sum_{i=1}^{L} \frac{1}{n_i} \sum_{j=1}^{n_i} (\hat{A}^T x_j^i - \widetilde{m^i})(\hat{A}^T x_j^i - \widetilde{m^i})^T \tag{3}$$

Where: $\widetilde{m^i} = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{A} x_j^i$, and $\hat{A}$ is the EDA projection matrix found in Eq. 2. The WCCN projection matrix B can be obtained by Cholesky decomposition of the inverse of W ($W^{-1} = BB^T$). WCCN down-regulate the contribution of the directions in the vector space that account for much of the within-class variability yielding to an improvement in between-class separation [16].

2.4. Cosine similarity measure

The advantage of the cosine similarity measure after discriminant analysis comes from its connection to the Bayes decision rule, as the Bayes classifier is the optimal one for minimizing the classification error [17]. In this work, the cosine similarity score between the two local features ($\boldsymbol{x_1}$ and $\boldsymbol{x_2}$) in the EDA+WCCN projection space:

$$S_{cos}(x_1, x_2) = \frac{\left(B^T \hat{A}^T x_1 \right)^T \left(B^T \hat{A}^T x_2 \right)}{\left\| B^T \hat{A}^T x_1 \right\| \left\| B^T \hat{A}^T x_2 \right\|} \tag{4}$$

3. EXPERIMENTAL RESULTS

In order to ensure a reproducibility of the experiments and the comparability with other methods, we tested our approach on the well-known, XM2VTS and LFW image databases using the common protocols.

3.1 Experiment on the XM2VTS database

The XM2VTS database contains images of 295 subjects, captured over 4 sessions in a controlled environment.We carried out face verification experiments on the test set of Configuration LPI.A set of 200 training clients, 25 evaluation impostors and 70 test impostors constitute the database.Additional details on the XM2VTS database and the used protocol LPI can be found in [10]. The images were cropped and resized to a standard size of 190 × 160. The total error rates in evaluation and test sets (the threshold obtained from the evaluation set) using individual face descriptors, multiple descriptors and some state state-of the-art methods are presented in table 1.

Table.1 shows that the verification rates obtained by MLPQH using four scales and MBSIFH using 5 scales (windows) outperform all the state of the art results according to the best of our knowledge. Furthermore, by using multiple descriptors, the fusion results in a drastic decrease of both the evaluation and test total errors rates. The multiple descriptors (MLBPH+MLPQH+MBSIFH) followed by (EDA+WCCN) achieved a total error rate equal to 0.31%, which is far away from the best results reported on the XM2VTS database (0.96).

	Descriptor	Eva Set	Test Set
Single Face Descriptor	MLBPH^{u2}(R=2+4+6+8,P=8)	0.98	2.41
	MLPQH (M=5+9+13+17)	0.27	0.81
	MBSIFH(7+9+11+13+15)	**0.28**	**0.72**
	LBPH_MAP [18]	/	2.84
	LBPHMM [19]	/	2.74
	Gabor : ICB2006-Best [20]	1.63	0.96
Multiple Descriptor	MLBPH^{u2}+MLPQH+MBSIFH	**0.11**	**0.31**
	Gabor+ *S*component(*HSV*)[21]	1	2.66

Table. 1 Total Error Rates on the evaluation and test sets of the XM2VTS Lausanne Protocol I

3.2 Experiment on the LFW database

Labeled Faces in the Wild (LFW) [11] is a database collected from the web (yahoo images) for studying the problem of face recognition in constrained environment (non-frontal poses, low resolution and non-frontal illumination, varying expressions). There are 13,233 images from 5,749 different people with high variations in position, pose, lighting, background, camera and quality. In testing phase, researchers are suggested to report performance as 10-fold cross validation using splits that are randomly generated and provided by the organizers. In this experiment, we use the aligned images (LFW-a) [11] and we crop and resize the images to 130×90 standard size. The cropped faces are photometrically normalized by the preprocessing sequence approach (PS) [22]. In LFW, most subjects have only one image sample, only Thesample set of the subjects having more than two images ischosen for ELDA and WCCN training. For scores fusion we use 5-folds for SVM training, 4-folds for testing SVM scores fusion and last fold for performance evaluation.

A comparison of the classification accuracies of the proposed systems on the test data “View 2,” with the state-of the-art methods is presented in Table 2.

The first remark from table 2 is the proven efficiency of MBSIFH which is slightly better than MLPQH and outperforms the MLBPH^{u2}. This can be explained by the superiority of the learned descriptors over the handcrafted ones. It is also worthy to mention that our system based on multi-scale LPQ (MLPQH +EDA+WCCN) outperforms the system presented in [6] using the same descriptor

	Descriptor	$\mu \pm S_E$
Single Face Descriptor	MLBPH^{u2}(R=2+4+6+8,P=8)	88.60±0.87
	MLPQH (M=5+9+13+17)	91.27±0.71
	MBSIFH(7+9+11+13+15)	**91.40±0.94**
	MLPQH(EXT+INT)[6]	89.00±1.55
	CMDH [22]	91.70±1.1
Multiple Descriptor	MLBPH^{u2}+MLPQH+MBSIFH	**93.03±0.82**
	LBP+TPLBP+FPLBP+SIFT [23]	89.5±0.51
	MLPQH+MLBPH (EXT+INT) [6]	90.02 ±1.45
	face.com r2011b [24]	91.30±0.3
	CMD+SLBP [22]	92.58±1.36
	DM+PCA fusion [16]	92.05 ±0.45

Table. 2 Mean± Standard Error Score of Our Systems on the Aligned Versions of LFW under Unrestricted protocol without extra data (View 2)

(91.27±0.71% versus 90.02 ±1.45%) Our system is more resilient to misalignment because our histograms represent bigger regions than in [6]. and also due to the efficiency of EDA and WCCN. Comparing the proposed framework to the existing state-of-the-art methods in the same unrestricted configuration using multiple descriptors, the average accuracy of our system (MLBP+MLPQ+MSIFT +EDA+WCCN+ cosine similarity +SVM fusion) attains 93.03±0.82% which is one of the best results in literature.

4. CONCLUSION

In this work, we proposed a highly efficient face verification system based on the fusion of two handcrafted and one learned local binary descriptor using multi-scale representation. Dimensionality reduction is achieved by exponential and between class separation is enhanced using WCCN. Using single multi-scale descriptors, state of the art results on the controlled XM2VTS database are surpassed by both MLPQH and MBSIFH and the fusion of the local multistage multi-descriptor results in a total error rate of 0.31%, which far away from the best results exiting in literature. The results on the challenging LFW database demonstrate that the proposed system is among the top performing systems.

REFERENCES

[1] Z. Lei, M.Pietikäinen, and S. Z. Li, " Learning Discriminant Face Descriptor," *IEEE transactions on Pattern Analysis and Machine Intelligence,*vol. 36, no. 2,pp. 289-302,2014.

[2] Z. Cui, W. Li, D.Xu, S. Shan and X. Chen, "Fusing Robust Face Region Descriptors via Multiple Metric Learning for Face Recognition in the Wild," *Computer Vision and Pattern Recognition(CVPR)*, pp. 3554-3561, 2013.

[3] M. Pietikäinen, A.Hadid, G. Zhao, T.Ahonen, "Computer Vision Using Local Binary Patterns," *Computational Imaging and Vision*, vol. 40, 2011.

[4] V. Ojansivu and J. Heikkila, "Blur insensitive texture classification using local phase quantization,"*International Conference on Image and Signal Processing (ICISP08)*, pp. 236–243,2008.

[5] J. Kannala and E. Rahtu, "BSIF: Binarized Statistical Image Features," *21st International Conference on Pattern Recognition (ICPR)*, pp.1363-1366, 2012.

[6] C. Ho Chan, M. A.Tahir, J. Kittler, and M. Pietikäinen, Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors," *IEEE transactions on Pattern Analysis and Machine Intelligence,*vol. 35, no. 5, pp. 1164-1177,2013.

[7] X. Tan, B.Triggs, "Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition*Lecture Notes in Computer Science*, vol. 4778, 2007.

[8] T. Zhang, B. Fang, Y. Y. Tang, Z. W. Shang, and B. Xu, "Generalized discriminant analysis: A matrix exponential approach,"*IEEE transactions on Pattern Analysis and Machine Intelligence,*vol. 40, no. 1, pp. 186–197,.2010.

[9] A. O. Hatch, S. Kajarekar and A. Stolcke, "within-class covariance normalization for SVM-based speaker recognition," *ICSLP*, pp. 1471-1474, 2006.

[10] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database,"*2nd Int. Conf. AVBPA*, pp. 72–77, 1999.

[11] L. Wolf, T.Hassner, and Y.Taigman, "Effective Face Recognition by Combining Multiple Descriptors and Learned Background Statistics," *IEEE transactions on Pattern Analysis and Machine Intelligence,*vol33, no. 10,pp. 1978-1990, 2011.

[12] J. Heikkila, V. Ojansivu, and E. Rahtu, "Improved blur insensitivity for decorrelated local phase quantization," *International Conference on Pattern Recognition*, pp. 818–821, 2010.

[13] R. Sidje and W. Stewart, "A numerical study of large sparse matrix exponentials arising in Markov chains," *Comput.Statist.Data Anal.*, vol. 29, no. 3, pp. 345–368, 1999.

[14] S. J. Wang, S. Yan, J. Yang, C. G. Zhou, and X. Fu, "A General Exponential Framework for Dimensionality Reduction," *IEEE transactions on Pattern Analysis and Machine Intelligence,*vol. 23, no. 2, 2014.

[15] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[16] O. Barkan,J. Weill, L. Wolf, H.Aronowitz, Fast, "High Dimensional Vector Multiplication Face Recognition" *ICCV*,2013.

[17] C. Liu, "Discriminant analysis and similarity measure," *Pattern Recognition*, vol. 47, 359–367, 2014.

[18] Y. Rodriguez and S. Marcel, "Face authentication using adapted local binary pattern histograms," *ECCV*, pp. 321–332, 2006.

[19] G. Heusch, Y. Rodriguez, and S. Marcel, "Local binary patterns as an image preprocessing for face authentication," *7th International Conference on Automatic Face and Gesture Recognition*, pp. 9-14, 2006.

[20] K. Messer, J. Kittler, J. Short, G.Heusch, F.Cardinaux, S. Marcel, Y. Rodriguez, S.Shan, Y. Su, W.Gao, and X.Chen, "Performance characterisation of face recognition algorithms and their sensitivity to severe illumination changes," *ICB*, pp. 1-11, 2006.

[21] A. Ouamane, M. Belahcene, A. Benakcha, M. Boumehrez, and A. T. Ahmed, "The Classification of Scores from Multi-classifiers for Face Verification," *Sensors & Transducers Journal*, vol. 145, no. 10, pp. 116-118, 2012.

[22] C. Huang, S. Zhu, and K. Yu, "Large Scale Strongly Supervised Ensemble Metric Learning, with Applications to Face Verification and Retrieval," *Technical Report TR115, NEC Laboratories Am.*, 2011.

[23] Y. Taigman, L. Wolf, and T. Hassner, "Multiple One-Shots for Utilizing Class Label Information," *British Machine Vision Conference (BMVC)*, 2009.

[24] D. Chen, X. Cao, F. Wen and J. Sun, "Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification," *Computer Vision and Pattern Recognition (CVPR),*pp. 3025-3032, 2013.