A Random Network Ensemble for Face Recognition

Kwontaeg Choi[1], Kar-Ann Toh[2], and Hyeran Byun[1]

[1] Dept. of Computer Science, Yonsei University, Seoul, South Korea, 120-749
[2] School of Electrical & Electronic Engineering, Yonsei University
Seoul, South Korea, 120-749

Abstract. In this paper, we propose a random network ensemble for face recognition problem, particularly for images with a large appearance variation and with a limited number of training set. In order to reduce the correlation within the network ensemble using a single type of feature extractor and classifier, localized random facial features have been constructed together with internally randomized networks. The ensemble classifier is finally constructed by combining these multiple networks via a sum rule. The proposed method is shown to have a better accuracy(31.5% and 15.3% improvements on AR and EYALEB databases respectively) and a better efficiency than that of the widely used PCA-SVM.

Keywords: Face recognition, random projection, neural network.

1 Introduction

Face recognition is among the most active areas of research because it is a difficult and representative problem of high dimensionality with high commercial application values. Many methods have significantly advanced the field over recent years. However, the robustness remains an important issue under an unconstrained environment because the limited number of available training data can not well represent all possible facial variations such as lighting, pose and expression. A single classifier using such a training set may be biased and possesses a large variance, resulting in a poor performance[1][2].

An ensemble of classifiers or a multiple classifiers system(MCS) is becoming relevant in face recognition due to the remarkable improvements over the single classifier system. Since different representation of the input face images have different sensitivity to various face variations, an ensemble of different face classifiers which integrates the complementary information leads to an improved classification accuracy[3]. There are two types of ensemble. Intermodal ensemble combines different types of modalities such as audio, fingerprint, 3D face and 2D face images, and intramodal ensemble combines multiple classifiers within a single modality such as a 2D image. We shall focus on intramodal ensemble in this paper.

In [4], the resampling techniques was utilized to generate several subsets of samples from the original training dataset for LDA based classifiers. In [5], an

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 92–101, 2009.

ensemble learning framework was developed based on a random sampling of three key components of a classification system namely the feature space, the training samples, and the subspace parameters. In [6], a random discriminant analysis was proposed to construct an optimal random subspace. In [7], a multi-classifier architecture was utilized using HMM, DCT, EigenFace and EigenObjects. In [3], two combination strategies namely, sum rule and RBF network, was adopted to integrate the outputes of three feature extraction methods, namely PCA, ICA, LDA. These works utilized various types of feature extractors and classifiers in order to increase the overall accuracy of the combined classifier.

These intramodal fusion networks appear somewhat complex due to a combination of heterogenous systems in terms of different kinds of feature extraction method and classifier. This is indeed a natural idea since a heterogeneous combination reduces the correlations among the classifiers. However, limitations of such a heterogenous system are the heavy computational cost and the tedious tuning of learning parameters. Moreover, they are not directly applicable in real-time problems such as video-based face tracking/recognition since most methods adopt the batch approach. An ensemble network with efficient online learning is thus an area worth exploring.

In this paper, we propose a homogeneous random network ensemble for face recognition particularly dealing with a large facial variation and with a limited number of training data. In order to reduce the correlation among classifiers in a homogeneous ensemble system, two ingredients namely, random projection and randomized network, have been utilized. The proposed method is differentiated from the above mentioned networks in the following ways: 1)the proposed network is a homogenous ensemble system whereas most other systems belong to the heterogeneous type, 2)an incremental(online) learning has been adopted whereas other ensemble networks adopt a batch learning.

The rest of this paper is organized as follows. Section 2 describes the concept of using localized random projection to extract facial feature by means of random basis. In Section 3, we present the proposed ensemble framework using randomized learning parameters selection. Section 4 describes an incremental learning formulation. In section 5, several experiments are presented to evaluate the proposed method in terms of its efficiency and generalization capabilities. Finally, our conclusion is given in Section 6.

2 Feature Extraction Using Random Basis

In this section, we present some backgrounds on random projection for immediate reference. This will be followed by an introduction of our proposed localized random projection method for face feature extraction.

2.1 Background of Random Projection(RP)

Unlike most training based feature extraction methods which require a lot of training samples and certain optimizing criteria, Random Projection(RP) has emerged as an efficient dimensional reduction method which does not require any training samples. In RP, the original high-dimensional data can be projected onto a low-dimensional subspace using a set of randomly generated basis where the orthonormal projection matrix consists of i.i.d. entries with zero mean and constant variance. According to [8], RP preserves approximately the pairwise distance of points in Euclidean space, volumes and affine distance and the structure of data without introducing significant distortion.

In order to significantly reduce the computational cost during the generation of projection matrix and projection onto a low-dimensional subspace, Achliopta[9] proposed a sparse random projection(SRP) with elements belonging to $\pm\sqrt{s}(s \in 1, 3)$ and 0. A projection matrix $R \in \mathbb{R}^{d \times D}$ with original face dimension D and reduced dimension d, consists of the following entities.

$$R_{ji} = \sqrt{s} \begin{cases} 1 \quad with\ prob\ \frac{1}{2s} \\ 0\ \ with\ prob\ 1 - \frac{1}{s} \\ -1 \quad with\ prob\ \frac{1}{2s} \end{cases}. \tag{1}$$

When sparsity s is 3, then $\frac{2}{3}$ of entities contain zero values. This leads to a threefold speedup. Moreover, the multiplications with $\sqrt{s}$ can be delayed and no floating point arithmetic is needed.

More recently, Li[10] proposed a very sparse random projection method. They showed that one can use $s >> 3$ (e.g. $s = \sqrt{D}$, or even $s = \frac{D}{logD}$) to significantly speedup the computation.

2.2 Proposed Localized Random Projection(LRP)

The relevance of SRP for feature extraction relies on how well the random distribution of non-zero entities (see Fig 1 (e) $\sim$ (f)) represents the distribution of facial features. While SRP focuses on efficient dimension reduction, several subspace methods such as PCA, LDA and ICA, consider both dimension reduction and effective feature representation. A uniform distribution of non-zero entities in SRP may not be representative for 2D face images. Hence, we cast a question, "How to change the distribution of the non-zero entities for face feature extraction while preserving the generation rule in Eq. (1)".

Inspired by localized feature extraction methods such as ICA, LFA and LNMF (see Fig. 1 (a) $\sim$ (c)), we propose a localized random projection in this work. Localized basis offers several advantages including stability to local deformations, lighting variations, and partial occlusion[11]. The proposed method modifies the uniform distribution of non-zero entities to localized distribution of non-zero entities as illustrated in Fig. 1 (g) $\sim$(i). The following provides more details.

Our main problem is about how to generate the sparse projection matrix with localized distribution of non-zero entities. In other words, we will need to determine the quantity, the position and the size(width, height) of localized blocks of non-zero entities. We will treat each basic vector as a 2D face image basis because we deal only with 2D images. Without loss of generality, we assume that each basis represents only one localized block. The position of a localized block can be set randomly within the size of the image template(we use 32 × 32 size).

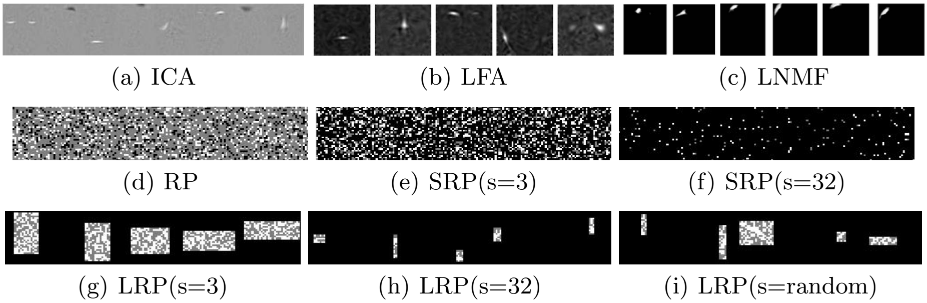

(a) ICA (b) LFA (c) LNMF
(d) RP (e) SRP(s=3) (f) SRP(s=32)
(g) LRP(s=3) (h) LRP(s=32) (i) LRP(s=random)

Fig. 1. Examples of basis obtained from various subspace methods. In (e) $\sim$ (i), black indicates 0, white indicates $sqrt(s)$ and gray indicates $-sqrt(s)$.

The size of each block is related to the amount of non-zero entities, given by sparsity s in Eq. (1). For example, when $s = 3$, $\frac{2}{3}$ of entities are non-zeros. In the case of a 32×32 template image, any size of rectangle block with $width \times height = 32 \times 32 \times \frac{2}{3}$ can be used. Fig. 1 (g) shows some possible localized basis and Fig. 1 (h) shows some possible localized basis when we use $s = 32 (= \sqrt{D} = \sqrt{32 \times 32})$.

One consideration here is regarding how to choose the best s for effective facial feature extraction. When s is small, many pixels are sampled for projection. When s is large, a small number of pixels are sampled for projection. Here we adopt the dynamic sparsity instead of fixed sparsity proposed by Achliopta and Li in order to capture small and large scale structures of face features as shown in Fig. 1 (i).

3 A Network Ensemble Using Randomized Parameters

In this section, we propose an ensemble framework of combining multiple networks with randomly selected learning parameters in order to reduce the correlation among classifiers.

3.1 Background of Single Layer Feedforward Network(SLFN)

Given n samples $x_i \in R^d$ and correspond target y_i, a standard SLFN classifier with h hidden nodes and activation function g can be mathematically modeled as

$$y_i = \sum_{j=1}^{h} \beta_j g(w_j x_i + b_j) \tag{2}$$

where $w_j = [u_{j1}, u_{j2}, ..., u_{jd}]^T$ is the weight vector connecting the j^{th} hidden node to the input nodes, b_j is the threshold of the j^{th} hidden node and $\beta_j \in R^h$ is the weight vector connecting the j^{th} hidden node to the output nodes.

For k-class problem, denote $\Theta = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_k] \in \mathbb{R}^{h \times k}$ as the weight matrix which is a collection of weight vector $\boldsymbol{\beta}$ and $Y \in \mathbb{R}^{n \times k}$ as the indicator matrix. The n equations above can be written more compactly as

$$H\Theta = Y \tag{3}$$

where

$$\underset{n\times h}{H} = \begin{bmatrix} g(w_1x_1+b_1) & \dots & g(w_hx_1+b_h) \\ \vdots & \ddots & \vdots \\ g(w_1x_n+b_1) & \cdots & g(w_hx_n+b_h) \end{bmatrix} \tag{4}$$

The weight parameters Θ can be estimated by minimizing the Least-Squares Error giving the solution[12]:

$$\Theta = (H^TH)^{-1}H^TY \tag{5}$$

3.2 A Random Network Ensemble Framework

Face images with high dimensionality and a limited number of training examples for each class easily lead to overfitting, overtraining, small-sample effect and singularity particularly when using an RBF network for face recognition[13]. In order to solve these problems, we combine multiple SLFN classifiers instead of using a single Multi-Layered Perception(MLP) by the following strategies.

1. Reducing the number of empirically fixed learning parameters : A neural network has various learning parameters which affect the recognition accuracy and computation. In the ensemble approach, it is particularly a time consuming task to determine the empirically fixed learning parameters of all networks. The classifier with such parameters may be not proper when a small number of training samples are used. Here, we adopt the Extreme Learning Machine(ELM)[12] classifier which comes with a small number of learning parameters. In ELM, two weight parameters w and b are arbitrarily chosen and need not be adjusted at all unlike the conventional SLFN classifier. We train an individual ELM classifier using random number of features and hidden units. Thus, the proposed method does not have empirically fixed learning parameters.
2. Reducing the computational cost : The many number of classifiers is combined to increase the overall accuracy. However this leads to an increase in computational cost. Unlike MLP which uses an iterative search based on gradient descent approach, the ELM estimates deterministically only β parameter in closed-form for fast training.
3. Reducing the correlations between classifiers : In order to improve the overall accuracy, each classifier should be trained differently to reduce the redundant classification. In ELM, two weight parameters w and b are arbitrarily chosen and need not be adjusted at all. As we train multiple number of ELM classifiers, this generates different decision boundaries even though the same training set is used. Moreover, the facial features have been extracted using random basis. This, again, generates different decision boundaries among ELMs which used the similar weight parameters w and b within the activation function.

In short, we extract the face features using the proposed LRP and train multiple ELMs using randomly selected learning parameters, finally we combine these multiple number of ELMs using a sum rule. We will call this method LRP-ELM. When C number of LRP-ELM classifiers are combined via a sum rule, the class label of unseen sample x can be predicted using

$$class(x) = \arg\max_{i} \sum_{c=1}^{C} \Theta_{c,i}^{T} g(W_c R_c x + B_c), i = 1, 2, ..., k \tag{6}$$

where R_c is c^{th} projection matrix by LRP, W_c and B_c are c^{th} parameters of activation function, and $\Theta_{c,i}$ is i^{th} column vector of c^{th} weight matrix Θ_c

4 Incremental Learning

Most classifiers adopt a batch based approach. This means that the classifiers need to be retrained when additional samples arrived. This could be a very time consuming task for ensemble approach.

The proposed LRP for facial feature extraction does not require the training set. Therefore an updating rule is not needed unlike the incremental PCA and incremental LDA. Moreover, the ELM can be retrained efficiently using only additional samples based on a recursive least square formulation. For this purpose, Huang[14] proposed an Online Sequential ELM(OSELM). When new feature block of data H_{t+1} and the corresponding indicator matrix Y_{t+1} are received, the parameter Θ_{t+1} can be estimated recursively as:

$$\begin{array}{l} P_{t+1} = P_t - P_t H_{t+1}^T (I + H_{t+1} P_t H_{t+1}^T)^{-1} H_{t+1} P_t \\ \Theta_{t+1} = \Theta_t + P_{t+1} H_{t+1}^T (Y_{t+1} - H_{t+1} \Theta_t) \end{array} \tag{7}$$

Therefore the proposed random network ensemble can be applied efficiently once a new training set is available.

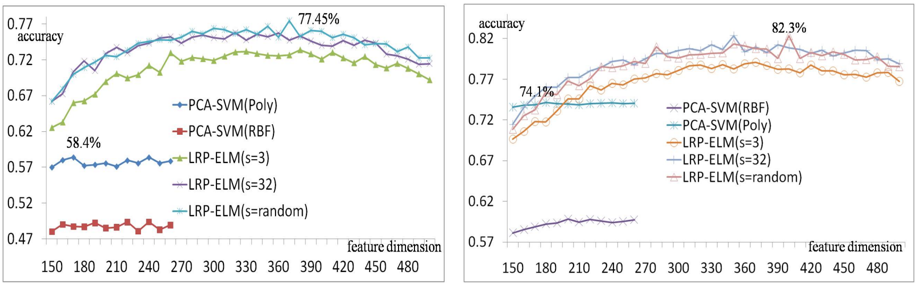
5 Experiment

Several experiments are conducted to evaluate the proposed method in terms of its generalization capability and efficiency. The experiments are repeated 10 times and the average test accuracy with the computational cost will be reported. The combination of snapshot based PCA and SVM[15] adopting polynomial and RBF kernels(PCA-SVM(Poly) and PCA-SVM(RBF), respectively) is used as the baseline classifier to compare the generalization capability and time complexity with the proposed method. We will use following three public face databases which contain many images per person and a large variation of imaging conditions.

1. AR database[16] (AR) : 126 identities/26 samples - different facial expressions, illumination conditions, and occlusions (sun glasses and scarf).
2. The extended Yale Face Database B[17] (EYALEB) : 28 identities / 1 frontal pose of 6 poses / 64 illumination conditions.
3. The Pose, Illumination and Expression Database [18](PIE) : 68 identities / 5 near frontal poses of 10 poses/ 5 illumination / 3 expression conditions.

We first evaluate the test accuracy of a single classifier using AR and EYALEB databases. Particularly, we evaluate the stability using a small number of training set, i.e. 30% of the dataset is used as training set with the remaining used as test set.

Fig. 2 (a) shows the test accuracy of compared methods at different number of feature dimensions using the AR database. In AR dataset with large face variation and with a limited number of training data in each class, the accuracy of PCA-SVM seems to be very low compared to the that of the proposed method because the training set is insufficient to predict the 70% test set which contains more face variations than that in the training set. The accuracy of PCA-SVM is seen to be 19.0% lower than the proposed method and seems to saturate at high dimensions. According to Fig. 2 (a), sparsity appears to affect the accuracy. For example, LRP-ELM(s=32) records a higher accuracy than that of LRP-ELM(s=3).



(a) AR database

(b) EYALEB database

Fig. 2. Average test accuracy comparison on the proposed method and the PCA-SVM

Fig. 2 (b) shows the test accuracy of compared methods on the EYALEB database. The proposed method records a much higher accuracy(an improvement of 8.2%) than that of PCA-SVM. The accuracy of LRP-ELM is seen to be lower than that of PCA-SVM at low feature dimensions(< 150). However, the accuracy of the proposed method increases when more features have been used.

In the second experiment, the accuracy of an ensemble classifier is evaluated. Since there are three controllable parameters (feature dimension, the number of hidden units and the number of classifiers) that can affect the accuracy, we construct 5 cases of ensembles of classifiers according to the feature dimension(220 $\sim$ 500) and the number of hidden units(200 $\sim$ 450). These values are set randomly because we do not seek an empirically determined values by experiments. As seen from Fig. 3, as we increase the number of classifiers, the accuracy of the ensemble classifier via sum rule increases. And the ensemble classifier with large number of features and hidden units records good accuracy. This result shows a significant improvement of 12.5% in terms of test accuracy over that before fusion. In EYALB database, the best accuracy of the proposed method is 89.4%(7.1% accuracy improvement over that before fusion).

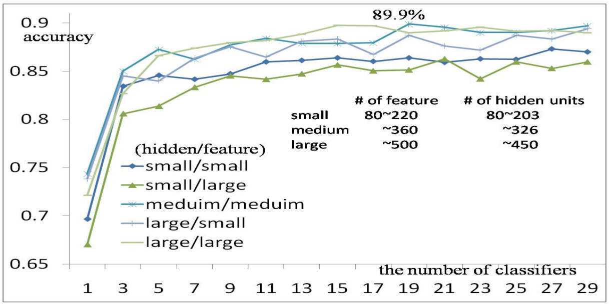
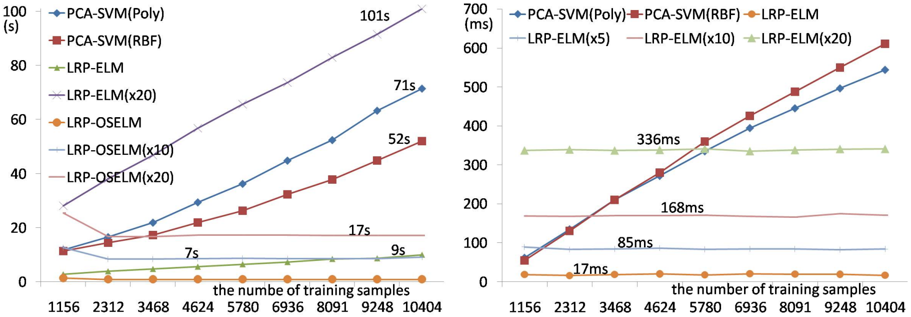


Fig. 3. Test accuracies of ensemble LRP-ELM classifiers(AR database)



(a) Training

(b) Test

Fig. 4. Execution time comparison on the proposed method and the PCA-SVM using PIE database. The number in parenthesis indicates the number of the combined classifiers.

To summarize, the accuracy of the combined classifier is 31.5% and 15.3% higher than that of a single PCA-SVM classifiers in AR and EYALEB databases respectively.

Finally, we compare the CPU execution times for both training and test using the PIE dataset which contains a large number of samples. Fig. 4 shows the CPU times for training and test. Here PCA-SVM used 100 eigen-faces and LRP-ELM used 400 localized random basis and 400 hidden units because PCA-SVM requires smaller dimension than that of LRP-ELM as shown in Fig. 2.

As seen from Fig. 4 (a), the proposed single LRP-ELM shows a much faster training speed than that of PCA-SVM(RBF). The ensemble classifier LRP-ELM(x20) using 20 classifiers is seen to be slower than single PCA-SVM(RBF). Since multiple LRP-ELM classifiers are combined via sum rule, LRP-ELM(x10) is 10 times slower than single LRP-ELM classifier. The computational cost of LRP-OSELM using Eq. (7) is evaluated. This approach records an almost constant execution time when the number of training samples increase, whereas the execution time of PCA-SVM and LRP-ELM increases due to retraining of the entire data set which includes the new and existing data.

Fig. 4 (b) shows the test execution time of the proposed method and PCA-SVM classifying one hundred samples. All methods show a low computational time within 610ms to classify one hundred samples. However, in mobile device or real-time applications, a fast test time is important. According to Fig. 4 (b), the proposed single LRP-ELM is seen to be 79 times faster than that of SVM when the number of training samples is 11,560. The ensemble classifier LRP-ELM(x20) using 20 classifiers is faster than single PCA-SVM when more than 5780 training samples have been used. Another benefit of the proposed LRP-ELM is that the execution time is almost independent of the total number of training samples because LRP-ELM is dependent on feature dimension and the number of hidden units, whereas PCA-SVM is dependent on the total number of training samples because the number of support vectors is proportional to the number of training samples.

6 Conclusion

In this paper, we proposed a random network ensemble for face recognition particularly for problems with a large appearance variation and with a limited number of training set. Unlike the conventional heterogeneous ensemble method, the proposed method is based on a homogenous system using random projection and randomized networks. In order to reduce the correlation between classifiers in the homogenous combination, we proposed localized random projections using sparse random basis. Next we train multiple number of single layer feedforward networks using randomly selected learning parameters. Finally we combined multiple number of randomized networks via a rum rule. The proposed ensemble network is subsequently extended to an incremental learning formulation. The proposed method is seen to improve the recognition accuracy of about 31.5% and 15.3% compared to that of SVM classifier using features extracted by PCA on AR and EYALEB databases, respectively. In terms of the computational cost, the proposed method has shown a better efficiency than that of PCA-SVM.

Acknowledgements

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center(BERC) at Yonsei University. (R112002105070030(2008)(2008)).

References

1. Raudys, S.J., Jain, A.K.: Small Sample Size Effects in Statistical Pattern Recognition Recommendations for Practitioners. IEEE Trans on Pattern Analysis and Machine Intelligence 13(3), 252–264 (1991)
2. Skurichina, M., Duin, R.: Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal. Appl., 121–135 (2002)

3. Lu, X., Wang, Y., Jain, A.K.: Combining classifiers for face recognition. In: ICME (2003)
4. Lu, X., Jain, A.K.: Resampling for Face Recognition. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688. Springer, Heidelberg (2003)
5. Wang, X., Tang, X.: Random sampling face recognition. International Journal of Computer Vision (IJCV) (2005)
6. Zhang, X., Jia, Y.: A linear discriminant analysis framework based on random subspace for face recognition. In: Proceedings of Pattern Recognition, pp. 2585–2591 (2007)
7. Lemieux, A., Parizeau, M.: Flexible multi-classifier architecture for face recognition systems. In: The 16th International Conference on Vision Interface (2003)
8. Goel, N., Bebis, G., Nefian, A.: Face recognition experiments with random projection. In: SPIE (2005)
9. Achlioptas, D.: Database-friendly random projections. In: ACM Symposium on the Principles of Database Systems, pp. 274–281 (2001)
10. Li, P., Hastie, T., Church, K.W.: Very sparse random projections. In: KDD, pp. 287–296 (2006)
11. Feng, Li, S.Z., Shum, H.Y., Zhang, H.J.: Local non-negative matrix factorization as a visual representation. In: The Second International Conference on Development and Learning, pp. 178–183 (2002)
12. Er, M.J., Wu, S., Lu, J., Toh, H.L.: Face recognition with radial basis function (RBF) neural networks. IEEE Trans. Neural Netw. 13, 697–710
13. Huang, G.-B., Siew, C.-K.: Extreme learning machine with randomly assigned RBF kernels. International Journal of Information Technology 11(1) (2005)
14. Liang, N.Y., Huang, G.B., Saratchandran, P., Sundararajan, N.: A Fast and Accurate On-Line Sequential Learning Algorithm for Feedforward Networks. IEEE Trans. Neural Networks 17, 1411–1423 (2006)
15. `http://www.ece.osu.edu/osusvm`
16. Martinez, A., Benavente, R.: The AR Face Database (1998)
17. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23(6), 643–660 (2001)
18. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression Database. IEEE Trans. Pattern Anal. Mach. Intelligence 25(12), 1615–1618 (2003)