

Face Recognition under Ageing Effect: A Comparative Analysis

Zahid Akhtar¹, Ajita Rattani², Abdenour Hadid³, and Massimo Tistarelli⁴

¹ Bahcesehir University, Istanbul, Turkey
zahid.eltc@gmail.com

² Michigan State University, East Lansing, MI, U.S.A.
ajita@msu.edu

³ University of Oulu, Oulu, Finland
hadid@ee.oulu.fi

⁴ University of Sassari, Sassari, Italy
tista@uniss.it

Abstract. Previous studies indicate that performance of the face recognition system severely degrades under the ageing effect. Despite the rising attention to facial ageing, there exist no comparative evaluation of the existing systems under the impact of ageing. Moreover, the compound effect of ageing and other variate such as glasses, gender etc, that are known to influence the performance, remain overlooked till date. To this aim, the contribution of this work are as follows: 1) evaluation of *static* baseline facial representations, based on local features, under the *ageing effect*, and 2) analysis of the compound effect of ageing with other variates, i.e., race, gender, glasses, facial hair etc.

Keywords: Biometrics, covariate analysis, facial ageing, facial features.

1 Introduction

Recognition difficulty based on facial biometric is linked to various factors (variate) such as illumination, expression, glasses, gender, race etc [11,9,4,10]. However another variate (factor) gaining increasing attention is the *ageing effect*. Existing studies on different databases and for different algorithms suggest degradation in the performance of face recognition system as a result of time lapse between the pair of facial images [8,15].

Advances have been made in the form of ageing-invariant solutions [13] that seek to learn an ageing model that is utilized for the age transformation of the input operational image to that of enrolled facial image. These solutions are usually integrated with existing face recognition engines to obtain invariance to the ageing effect.

Despite the increasing interest, there exist no study for the comparative performance evaluation of the existing face recognition systems under the ageing effect (temporal variance). This evaluation is important because it will a) allow to gauge the performance gain of the face recognition systems on employing ageing invariant solutions, and b) facilitate the designer/researcher in choosing the

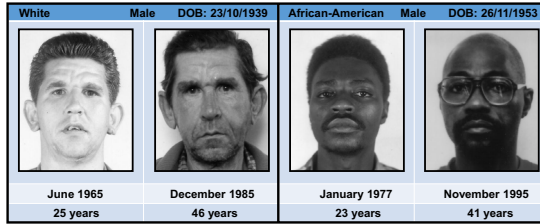


Fig. 1. Example of facial images from MORPH [1] data set containing variations like facial hair, glasses and illumination apart from the ageing effect

most robust face recognizer to be integrated with ageing invariant solutions for optimal performance.

However, none of the available facial ageing databases are specialized only on ageing but other variates, such as illumination, glasses, gender etc., known to influence the performance of the face recognition systems are also present. For instance, Figure 1 shows the example of facial images with variates such as pose, illumination changes and facial hairs apart from the ageing effect. As a consequence, sole impact of ageing to the performance degradation of the face recognition systems can not be evaluated. Therefore, it becomes necessary to study the compound effect of ageing with other variate present in the ageing database (covariate analysis). This covariate analysis will allow to analyze the contribution of other variate in the performance degradation of the facial system under the ageing effect. For instance, study in [4] indicate males are easier to be recognized than females under normal condition, it will be interesting to investigate whether this observation also holds under the impact of ageing.

To these aims, the two-fold contributions of this paper are as follows:

1. To evaluate and compare existing face recognition systems based on six different facial representations under the ageing effect.
2. To evaluate the compound effect (covariate analysis) of ageing with other variates such as gender, race, glasses and facial hairs, present in the MORPH database.

Six automated baseline facial representations based on local features namely LBP (local binary patterns) [2] and MLBP (Multi-scale local binary patterns) [6], LPQ (local phase quantization) [3], LTP (Local ternary patterns) [12], EBGM (elastic bunch graph matching) [14], SIFT (Scale invariant feature transform) [5] and SURF (speeded up robust features) [7] have been evaluated under this study. Facial features obtained using these local descriptors acquire more detailed facial observation that is critical to the recognition process. Moreover, face recognition systems based on these local features do not require training images for learning and efficient performance. Because of these reasons local feature based face recognition systems have been considered in this study. Experiments have been conducted on 631 subjects from MORPH facial ageing database.

The paper is organized as follows: Section 2 briefly explains six facial representations evaluated under this study. Section 3 explains the protocol of

performance evaluation and covariate analysis. Experimental setup and validations are described in sections 4 and 5. Conclusions are drawn in section 6.

2 Facial Representations under Study

In this section, we describe six facial representations considered in this study.

- **Local Binary Pattern (LBP) [2] and Multi-scale Local Binary Patterns (MLBP)[6]**: Ahonen et al. introduced a facial representation based on LBP texture descriptor. It is an efficient descriptor that assigns a label to every pixel of an image by thresholding the neighborhood of each pixel with the center pixel value and considering the result as a binary number. Then the histogram of the labels is obtained and used as a descriptor. Dissimilarity measure between the pair of facial images is obtained using χ^2 histogram distance. MLBP is an extended version of LBP by using multiple radii and offering the advantage of scale invariance.
- **Local Phase Quantization (LPQ)[3]**: LPQ is based on quantizing the Fourier transform phase in local neighborhoods. The phase has been shown to have a blur invariant property under certain commonly fulfilled conditions. These descriptors are obtained for facial images in the manner similar to LPB and matched using χ^2 histogram distance [3].
- **Local Ternary Patterns (LTP)[12]**: These descriptors utilize the idea that many facial regions are relatively uniform, it is potentially useful to improve the robustness of the underlying descriptors in these areas. This is obtained through, local ternary patterns, 3-valued codes, in which gray levels in a zone of width $\pm t$ around center pixel i_c are quantized to zero, ones above this are quantized to +1 and one below are quantized to -1. Thus the binary code in LBP are replaced by the ternary code. Application of LTP for face recognition has been evaluated in [12].
- **Elastic Bunch Graph Matching (EBGM)[14]**: Elastic Bunch Graph Matching localizes a set of landmark features and extracts Gabor jets at landmark positions. Similarity between pair of facial images is expressed as function of similarity between localized Gabor jets corresponding to facial landmarks.
- **Scale Invariant Feature Transform (SIFT)[5]**: Scale Invariant Feature Transform (SIFT) are invariant to image scaling, translation, rotation, and partially invariant to 3D projection. These features are efficiently detected through a staged filtering approach and are highly distinctive. SIFT has emerged as a very powerful image descriptor and its employment for face recognition was systematically investigated in [5]. Euclidean distance is used for pairing SIFT key-points obtained from the facial images to be matched. Number of matched key-points is used as a similarity measure.
- **Speeded Up Robust Features (SURF)[7]**: SURF (Speeded Up Robust Features) are known for repeatability, distinctiveness, and robustness, yet can be computed and compared much faster. This is achieved by relying

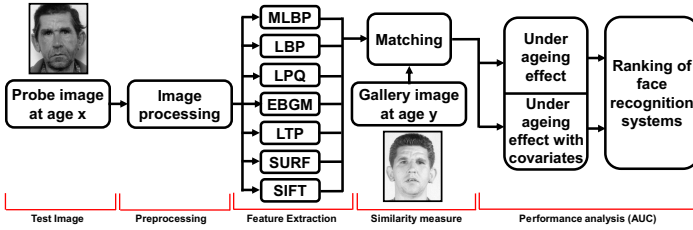


Fig. 2. Schematic of the procedure of performance evaluation and covariate analysis for the face recognition systems under ageing effect

on integral images for image convolutions (using a Hessian matrix-based measure for the detector, and a distribution-based descriptor). SURF has been successfully investigated as a descriptor for face recognition [7]. SURF descriptors are matched in a manner similar to SIFT.

3 Performance Evaluation and Covariate Analysis

In this section, we detail the protocol of performance evaluation and covariate analysis of the face recognition systems adopted in this study.

First of all, facial features are extracted from all images in the database using the representations mentioned in section 2. Then the following steps are performed for performance evaluation:

- Similarity(dissimilarity) matrix is computed using all-pair matching of facial features. Matching scores are divided into an authentic (genuine) and an impostor score distribution.
- The dataset (matching scores) is bootstrapped at the user level, i.e., subset of users are selected with replacement for performance evaluation.
- Performance has been evaluated on calculating area under curve (AUC) statistic on the bootstrapped dataset. AUC is computed as a function of true accept rate (TAR) and false accept rate (FAR) as :

$$AUC = \int_0^1 TAR(FAR)dFAR \quad (1)$$

The AUC value ranges from 0 to 1. An area (AUC) of 1 represents a perfect classifier; an area of 0.5 represents a random classifier (i.e., not able to distinguish between genuine and impostor samples). In order to provide comprehensive information regarding the performance of the face recognition systems, we have used AUC instead of equal error rate (EER) or false accept rate (FAR) and false reject rate (FRR) statistics at the selected operational points.

- Variation in AUC on the bootstrapped dataset is recorded as mean \pm std.
- Finally, face recognition systems are ranked in the descending order on the basis of their average AUC on the bootstrapped dataset.

Next, the contribution of other variates (such as gender, race, glasses) to the performance degradation of the face recognition system under the ageing effect has been evaluated. Previous studies [11,9,4,10] evaluated the recognition difficulty for facial biometrics to the imaging conditions, such as glasses, illumination etc., and subject covariates, such as gender, race etc., when the ageing effect is either absent or not very profound. None of these studies are conducted on the facial ageing databases. In contrary, the aim of this study is to gauge the impact of these covariates under the ageing affect.

To this aim, the following covariates are extracted from the MORPH meta-data file. These extracted covariates are explained as follows:

- Age {Young and Old}. Old age is assigned to subjects above 40 years.
- Race {White , Non-white}. Self-explanatory.
- Gender {Male , Female}. Self-explanatory.
- Glasses {Yes, No}. Self-explanatory.
- Facial Hair {Yes, No}. There were many subjects who had thin hairs, beards or were not clean shaven.

Using these extracted covariates, single and joint factor analysis is performed. In single factor analysis, only one of the available covariates is kept constant and others are allowed to vary. For joint factor analysis, values of multiple covariates are kept fixed and the performance of the system is gauged under the ageing impact. These analysis aimed at determining the favorable covariate values under the influence of ageing. Figure 2 illustrates the overall schematic of the procedure for performance evaluation and covariate analysis of face recognition systems under ageing effect.

4 Experimental Setup

4.1 Dataset

MORPH: MORPH [1] data corpus comprises of thousand of facial images of individuals across time, collected in real-world conditions (not a controlled collection). This dataset also include essential meta-data, such as age, sex, race,

Table 1. AUC values obtained on the performance evaluation of six facial representations under the ageing effect on the MORPH database

Reference	Face representation	mean \pm std	Percentiles
			[25%, 50%, 75%]
[6]	MLBP	0.66 \pm 0.02	[0.58, 0.60, 0.64]
[2]	LBP	0.64 \pm 0.08	[0.56, 0.59, 0.63]
[3]	LPQ	0.62 \pm 0.01	[0.59, 0.60, 0.61]
[14]	EBGM	0.60 \pm 0.02	[0.56, 0.57, 0.59]
[12]	LTP	0.55 \pm 0.00	[0.50, 0.51, 0.54]
[7]	SURF	0.52 \pm 0.01	[0.51, 0.52, 0.53]
[5]	SIFT	0.51 \pm 0.08	[0.50, 0.51, 0.51]

glasses, facial hair etc. A subset of 631 subjects from MORPH (1700 images) with about 3 images per subject are used in this study. Age range of the subjects are [16, 99].

4.2 Facial Image Preprocessing

First of all, face images are cropped to extract the face portion using the eye locations, retrieved through Meta-data file. These cropped images are first resized to 250×200 pixels. Then, the cropped images are illumination normalized by incorporating a series of steps [12] as follows 1) gamma correction, 2) difference of gaussian (DoG) filtering, 3) masking, and 4) contrast equalization. This preprocessing chain combat the effect of illumination, local shadowing and highlights. While still preserving the essential elements of visual appearance of use in recognition. Readers are referred to [12] for the detailed information on the preprocessing steps for illumination normalization.

5 Results

Facial features are extracted from all the preprocessed facial images using the six facial representations. The facial features based on LBP, MLBP (using 2,3 and 4 radii), LPQ and LTP (see section 2) are obtained from the 24×24 windows extracted from the preprocessed facial images. The similarity (dissimilarity) matrix is obtained from the all-pair matching using their respective distance (similarity) classifiers. The dissimilarity matrices are converted into similarity by normalizing the scores in the range $[0, 1]$ and subtracting from 1. Then the bootstrapped version of the score set is obtained at the user level. A subset of 50 % users are selected with replacement and the process of bootstrap is repeated at most 30 times. Then, the performance evaluation and covariate analysis for the six facial representations is performed using the procedure mentioned in section 3 and AUC values are recorded.

5.1 Performance Evaluation

Table 1 quotes the AUC values as *mean \pm std* and percentile statistics on the bootstrapped dataset for all the facial representations. These facial representations are mentioned in the descending order on the basis of their AUC values.

It can be seen that all the systems resulted in low performance on the MORPH facial ageing database. However, MLBP outperforms all the other facial representations. LBP and LPQ resulted in 3% and 6% relative performance degradation over MLBP. LPQ, shown to be invariant also to the blur affect [3] in the facial images, performed lower than MLBP and LBP. Further, SURF usually indicated to outperform SIFT under the absence of ageing [7], performed slightly better than SIFT in the presence of ageing as well.

EBGM resulted in 9% performance degradation over MLBP. The reason could be the failure to extract fiducial points or landmarks in the facial features due to the shape and texture changes on account of ageing.

LTP performed lower than MLBP by 17%. The reason could be that LTP based feature descriptors look for relatively uniform facial regions. However, these regions may become non-uniform under the ageing effect.

To sum up, MLBP based facial representation outperformed other facial representations under the ageing effect. These results indicate that MLBP is able to, to some extent, locate discriminative information even under the presence of profound facial ageing between the pair of images. Nevertheless, it could be interesting to integrate MLBP based facial recognition system with age invariant solutions and gauge the improvement over the baseline performance, evaluated to be 0.66 (recorded as AUC) in Table 1.

However, the low performance of these systems can not be solely attributed to the ageing effect, but also to other variates present in the database. Thus, we perform covariate analysis evaluating the compound affect of ageing with other variates (such as facial hair, glasses etc) available in the MORPH database.

5.2 Covariate Analysis

Covariate extracted from meta-file (explained in section 3) are used to perform single and joint factor analysis as follows: For the single factor analysis (only one of the available variate is kept fixed and the rest others are allowed to vary), the database is broken down into different subsets as follows:

Age {Young (550 subjects) and Old (130 subjects) }
Race {White (171 subjects) , Non-white (460 subjects)}
Gender {Male (515 subjects), Female (116 subjects)}
Glasses {Yes (36 subjects), No (612 subjects)}
Facial Hair {Yes (400 subjects), No (342 subjects)}

The bootstrapped version (similar to performance evaluation) of each subset of the database is evaluated using AUC for all the six facial representations. Results of the single factor analysis are shown in Table 2.

It can be seen from the table 2 that MLBP continues to outperform other face representations even for the single factor analysis. Findings of the single factor analysis, as supported by majority of the facial representations, is enumerated as follows:

- **Age:** Older people (above 40 years) performed better than younger. MLBP indicated a 7% higher performance for old people.
- **Race:** White subjects performed better than non-white under the ageing effect. MLBP indicated a 3% higher performance for white people.
- **Gender:** Males performed better than females under the influence of ageing. MLBP indicated 5% higher performance for males.
- **Glasses:** Presence of glasses degrades the performance of the system. MLBP indicated 5% performance degradation.
- **Facial hair:** Facial images with the presence of hair have resulted in better performance (5% higher performance for MLBP). This result is contrary to

Table 2. AUC values for the single factor analysis on the bootstrapped version of MORPH dataset

Covariates	Values	MLBP [6]	LBP [2]	LPQ [3]	LTP [12]	EBGM [14]	SURF [7]	SIFT [5]
Age	Young	0.64 ± 0.04	0.62 ± 0.1	0.55 ± 0.03	0.54 ± 0.01	0.54 ± 0.04	0.54 ± 0.02	0.53 ± 0.05
	Old	0.69 ± 0.03	0.67 ± 0.05	0.58 ± 0.03	0.56 ± 0.02	0.56 ± 0.05	0.56 ± 0.03	0.55 ± 0.06
Race	Non-white	0.73 ± 0.01	0.61 ± 0.02	0.54 ± 0.03	0.55 ± 0.01	0.55 ± 0.02	0.55 ± 0.02	0.53 ± 0.05
	White	0.75 ± 0.03	0.62 ± 0.08	0.55 ± 0.03	0.56 ± 0.02	0.56 ± 0.01	0.57 ± 0.03	0.55 ± 0.06
Gender	Male	0.75 ± 0.06	0.62 ± 0.04	0.57 ± 0.04	0.56 ± 0.01	0.55 ± 0.04	0.55 ± 0.01	0.53 ± 0.03
	Female	0.71 ± 0.08	0.61 ± 0.03	0.55 ± 0.03	0.54 ± 0.07	0.54 ± 0.03	0.55 ± 0.04	0.54 ± 0.04
Glasses	Yes	0.72 ± 0.03	0.60 ± 0.03	0.55 ± 0.02	0.55 ± 0.04	0.54 ± 0.03	0.54 ± 0.01	0.51 ± 0.01
	No	0.76 ± 0.06	0.62 ± 0.03	0.56 ± 0.01	0.56 ± 0.03	0.55 ± 0.02	0.55 ± 0.01	0.52 ± 0.02
Facial Hair	Yes	0.77 ± 0.05	0.63 ± 0.03	0.57 ± 0.02	0.55 ± 0.01	0.53 ± 0.01	0.53 ± 0.03	0.52 ± 0.02
	No	0.73 ± 0.04	0.61 ± 0.04	0.55 ± 0.03	0.54 ± 0.01	0.52 ± 0.02	0.52 ± 0.05	0.51 ± 0.04

the common belief as presence of facial hairs cause occlusion and result in loss of information.

For the joint factor analysis, the database is broken down into subsets as mentioned below (compound effect of only those covariates could be analyzed for which sufficient number of subjects were obtained)

Non-white and Male (B+M) {373 subjects}

Non-white and Female (B+F) {87 subjects}

White and Male (W+M) {142 subjects}

White and Female (W+F) {29 subjects}

Facial Hair and Glasses (Fh+Gl) {23 subjects}

Facial Hair and No Glasses (Fh+NGl){420 subjects}

Table 3 shows the results of the joint factor analysis for the top three good performing facial representations. Most of the results are in accordance with the results of single factor analysis. For instance, non-white males are easier to recognize than non-white females under the ageing affect (by 8% for MLBP). Similar observation holds for white males and white females (by about 2% for MLBP). Presence of facial hair and no glasses results in better performance than the presence of facial hair and glasses, in most of the cases (7% for MLBP) . MLBP usually performed better than other descriptors except for the case of facial hair

Table 3. AUC values for the Joint factor analysis on the bootstrapped version of MORPH data set

Multi-variates	MLBP[6]	LBP [2]	LPQ [3]
B+M	0.65 ± 0.01	0.62 ± 0.03	0.62 ± 0.04
B+F	0.60 ± 0.05	0.60 ± 0.02	0.61 ± 0.02
W+M	0.65 ± 0.07	0.64 ± 0.08	0.63 ± 0.06
W+F	0.64 ± 0.03	0.63 ± 0.04	0.62 ± 0.06
Fh+Gl	0.65 ± 0.01	0.63 ± 0.02	0.62 ± 0.04
Fh+NGl	0.70 ± 0.06	0.65 ± 0.05	0.74 ± 0.05

and no glasses, where LPQ performed better than MLBP by about 5%. The compound effect of facial hairs and presence/absence of glasses yielded different results for different representations. This could also be due to the sampling imbalances between the subset of users.

5.3 Comparison with Previous Studies on Covariate Analysis

Givens et al.'s study [9] on 2144 images from FERET database for three algorithms indicate that non-white people are easier to recognize than white people. Presence of glasses and facial hairs facilitate the recognition process. Further, their study did not indicate any gender preference. In [10], Lui et al. summarized existing results, indicating also that older people are easier to recognize than younger. Another study by Beveridge et al. [4] on 351 subjects from FRGC for covariates such as gender, age, glasses etc., indicates that males, older people and subjects without glasses are easier to recognize than their counterparts. However their study related to ageing variate (old and young) was not performed on the facial ageing database. Recent study by Paone et al. [11] on 659 subjects from FRGC dataset for commercial FaceVACS indicated smiling expression, controlled illumination as favorable covariates without any gender preference.

In accordance with the existing studies, our results suggest that covariate influence do exist even under the ageing effect. Our study suggests that older subjects and males are easier to recognize (consistent with [4]) under ageing effect. Presence of facial hair (consistent with [9]) and absence of glasses ([4]) are the favorable imaging conditions under the ageing effect. Results related to imaging conditions are quite obvious though. However, in contrary to previous studies, white people are indicated to be easily recognized than non-white under the ageing affect.

6 Conclusion

This study contributes to evaluate the six baseline facial representations based on local features under facial ageing impact. Further the compound effect of ageing with other variate are systematically analyzed. Covariate analysis indicate that white, older, males, no glasses and facial hair are favorable conditions under the ageing effect. In comparison to prior studies, most of the favorable covariate values remain same even under the ageing effect. Experimental results indicate MLBP to outperform other facial representations for majority of the experiments. However, definite conclusions may be drawn after thorough evaluation on different facial ageing databases, as a part of future work. Such an evaluation will further expand on the considered covariates. Joint factor analysis will be further explored and sampling imbalances between different subset of users will be accounted for.

Acknowledgments. Abdenour Hadid was partially supported by the TABULA RASA project, 7th Framework Research Programme of the European Union (EU), grant agreement number: 257289.

References

1. MORPH database, <http://www.faceaginggroup.com/projects-morph.html/>
2. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 28(12), 2037–2041 (2006)
3. Ahonen, T., Rahtu, E., Ojansivu, V., Heikkil, J.: Recognition of blurred faces using local phase quantization. In: *Proc. Int. Conf. on Patt. Reco.*, pp. 8–11 (2008)
4. Beveridge, J.R., Givens, G.H., Phillips, P.J., Draper, B.A.: Factors that influence algorithm performance in the Face Recognition Grand Challenge. *Computer Vision and Image Understanding* 113(6), 750–762 (2009)
5. Kisku, D.R., Rattani, A., Grosso, E., Tistarelli, M.: Face Identification by SIFT-based Complete Graph Topology. In: *Proc. of 5th IEEE Int'l Workshop on Automatic Identification Advanced Technologies*, pp. 63–68 (2007)
6. Chan, C.-H., Kittler, J., Messer, K.: Multi-scale Local Binary Pattern Histograms for Face Recognition. In: Lee, S.-W., Li, S.Z. (eds.) *ICB 2007*. LNCS, vol. 4642, pp. 809–818. Springer, Heidelberg (2007)
7. Dreuw, P., Steingrube, P., Hanselmann, H., Ney, H.: SURF-Face: Face Recognition Under Viewpoint Consistency Constraints. In: *Proc. BMVC*, pp. 1–11 (2009)
8. Flynn, P., Bowyer, K.W., Phillips, P.J.: Assessment of Time Dependency in Face Recognition: An Initial Study. In: Kittler, J., Nixon, M.S. (eds.) *AVBPA 2003*. LNCS, vol. 2688, pp. 44–51. Springer, Heidelberg (2003)
9. Givens, G., Beveridge, J.R., Draper, B.A., Grother, P., Phillips, P.J.: How Features of the Human Face Affect Recognition: A Statistical Comparison of Three Face Recognition Algorithms. In: *Proc. IEEE Conf. on CVPR*, pp. 381–388 (2004)
10. Lui, Y.M., Bolme, D., Draper, B.A., Beveridge, J.R., Givens, G., Phillips, P.J.: A meta-analysis of face recognition covariates. In: *Proc. 3rd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pp. 139–146 (2009)
11. Paone, J., Biswas, S., Aggarwal, G., Flynn, P.: Difficult imaging covariates or difficult subjects? - An empirical investigation. In: *Proc. IJCB*, pp. 1–8 (2011)
12. Tan, X., Triggs, B.: Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. *IEEE Trans. on Image Processing* 19(6), 1635–1650 (2010)
13. Park, U., Tong, Y., Jain, A.K.: Age-Invariant Face Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 32(5), 947–954 (2010)
14. Wiskott, L., Fellous, J.M., Kruger, N., Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Trans. on PAMI* 19(7), 775–780 (1997)
15. Yamaguchi, O., Fukui, K., Maeda, K.: Face Recognition Using Temporal Image Sequence. In: *Proc. Third IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 318–323 (1998)