Age and gender estimation using Region-SIFT and multi-layered SVM

Hyunduk Kim, Sang-Heon Lee, Myoung-Kyu Sohn, Byunghun Hwang

Hyunduk Kim, Sang-Heon Lee, Myoung-Kyu Sohn, Byunghun Hwang, "Age and gender estimation using Region-SIFT and multi-layered SVM," Proc. SPIE 10696, Tenth International Conference on Machine Vision (ICMV 2017), 106962J (13 April 2018); doi: 10.1117/12.2309441

Event: Tenth International Conference on Machine Vision, 2017, Vienna, Austria

Age and Gender Estimation Using Region-SIFT and Multi-Layered SVM

Hyunduk Kim, Sang-Heon Lee, Myoung-Kyu Sohn, Byunghun Hwang
Daegu Gyeongbuk Institute of Science & Technology (DGIST)

ABSTRACT

In this paper, we propose an age and gender estimation framework using the region-SIFT feature and multi-layered SVM classifier. The suggested framework entails three processes. The first step is landmark based face alignment. The second step is the feature extraction step. In this step, we introduce the region-SIFT feature extraction method based on facial landmarks. First, we define sub-regions of the face. We then extract SIFT features from each sub-region. In order to reduce the dimensions of features we employ a Principal Component Analysis (PCA) and a Linear Discriminant Analysis (LDA). Finally, we classify age and gender using a multi-layered Support Vector Machines (SVM) for efficient classification. Rather than performing gender estimation and age estimation independently, the use of the multi-layered SVM can improve the classification rate by constructing a classifier that estimate the age according to gender. Moreover, we collect a dataset of face images, called by DGIST_C, from the internet. A performance evaluation of proposed method was performed with the FERET database, CACD database, and DGIST_C database. The experimental results demonstrate that the proposed approach classifies age and performs gender estimation very efficiently and accurately.

Keywords: Age estimation, gender estimation, local descriptor, SIFT, Multi-layered approach.

1. INTRODUCTION

Recently, techniques for analyzing and understanding a user's face as a core technology for human and machine interaction have been actively researched. In particular, automatic demographic classification has found its way into industrial applications such as surveillance monitoring, security control, and targeted marketing systems [1]. In general, age and gender estimation systems consist of the following: a face detection and alignment stage, a feature extraction stage, a dimension reduction stage, and a classification stage. In the feature extraction stage, several algorithms are widely used: Gabor wavelet [2], Local Binary Pattern (LBP) [3], and Scale Invariant Feature Transform (SIFT) [4]. The Gabor feature based method generates a bunch graph as a feature vector using a Gabor wavelet filter and uses it as a pattern. This method is generally known to provide outstanding performance even in the case of changes of illumination, but it has disadvantages of high computational complexity and slow speed. Another commonly used local pattern is the LBP descriptor proposed by Ojala et. al. [3]. Due to its robustness to illumination change, LBP is widely used in the face analysis research area including for face recognition, facial expression recognition, and gender estimation. The SIFT algorithm was first proposed by David Lowe [4] in 2004 and is a robust feature extraction algorithm for image rotation, scale change, and illumination change. However, despite its high performance, it requires long time to perform.

A comprehensive survey of methods and data has previously been offered by [5, 6, 7]. Gao et. al. [8] used Gabor descriptors along with a Fuzzy-LDA classifier that considers a face image as belonging to more than one age class. Guo et. al. [9] used a combination of Biologically-Inspired Features (BIF) and various manifold-learning methods for age estimation. Choi. et. al, [10] used Gabor and LBP features along with a hierarchical age classifier composed of a Support Vector Machine (SVM) to classify the input image to an age class followed by the support vector regression (SVR) to estimate a precise age. Moghaddam et. al. [11] used a SVM with a Radial Basis Function (RBF) and Baluja et. al. [12] used an AdaBoost classifier. Fazl-Ersi et. al. [13] proposed a feature-based method using a SVM and a LBP operator. Beikos-Calfa et al. [14] proposed a holistic but resource intensive strategy that employed a Linear Discriminant Analysis (LDA) and a Principal Component Analysis (PCA) for gender recognition.

Our contributions are summarized as follows and delineated by section: (i) face alignment using face landmarks including both the central face region and the surrounding contextual information (e.g., facial shape, ears, and hair style) in Section 2-1, (ii) feature extraction using region-SIFT and dimension reduction in Sections 2-2 and 2-3, and (iii) age and gender estimation using multi-layered SVM classifier in Section 2-4. Finally, we present our experimental results and conclusions in Sections 3 and 4, respectively.

Tenth International Conference on Machine Vision, edited by Antanas Verikas, Petia Radeva, Dmitry Nikolaev, Jianhong Zhou, Proc. of SPIE Vol. 10696, 106962J · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2309441

2. METHODOLOGY

2.1 Face Alignment

Face alignment is one of the important stages of face analysis, such as face recognition, age and gender estimation, and facial expression. Previously, a popular approach for face alignment was the positioning of a frontal face image into an upright canonical pose based of the position of the eyes [1]. However, the previous face alignment method is unsuitable for age and gender expression estimation since the eye distance is various in an uncontrolled environment. Thus, we devise a landmark based face alignment method to acquire the detailed face region. To obtain the landmarks, we use the open-source OpenFace library [15] which is a real-time and accurate facial landmark detector. Fig. 1 shows some detected facial landmark points using the OpenFace library.

Figure 1. Example of detected facial landmark points.

The coordinates of both eyes are positioned and turned to align the face in such a way that both eyes are positioned in the proper location. The aligned face region is then resized to $L_0 \times L_0$ and converted to grayscale. First, the eyes can be aligned horizontally by an in-plane rotation of the face image into an upright pose using the angle $\theta = \arctan\left(\frac{p_{l,x} - p_{r,x}}{p_{l,y} - p_{r,y}}\right)$ where the points $(p_{l,x}, p_{l,y})$ and $(p_{r,x}, p_{r,y})$denote the center positions of the left and right eye. As shown in [1], we also use the horizontal positions of the nose $p_{n,x}$ and the center of the face $p_{c,y}$ to compensate for the over-scaling in face alignment. In (1), we apply the ratio of these points to find the scale factor S_0.

$$S_0 = \left(\frac{d_{eyes}}{L_0 - 2\Omega_0}\right) * \max\left(\frac{p_{n,x}}{p_{c,x}}, \frac{p_{c,x}}{p_{n,x}}\right) \tag{1}$$

Where d_{eyes} denote the distance between the eyes. The aligned face region is then defined by a $L_c \times L_c$square, and the cropped face is defined by a $L_0 \times L_0$ square image in which the left eye is fixed at the top-left offset Ω_0. From the scale actor S_0, we compute the dimensions $L_c \times L_c$ of the cropped area with $L_c = S_0 * L_0$, its horizontal offset $\Omega_x = p_{l,x} - S_0\Omega_0$, and its vertical offset $\Omega_y = p_{l,y} - S_0\Omega_0$. Fig. 2 shows some aligned face regions.

Many published methods utilize the central face region (Fig. 2(b)) to perform demographic estimation. However, in the age and gender estimation problem, the contextual region around a face may complement the central face region by providing additional information such as the face contour, ears, hair style, etc. Hence, we extract the face region with the context (Fig. 2(c)).

(a) Input 2D face image (b) Face region by tight cropping (c) Face region with context

Figure 2. Example of detected facial landmark points.

2.2 Feature Extraction

Local features are important to describe facial components in which wrinkles appear such as the eyes, mouth contour, cheeks, and nose. Recently, the SIFT feature extraction method [4] has been widely applied to many recognition problems. However, it requires long processing time. In order to solve this problem, we introduce the region-SIFT feature extraction method based on facial landmarks. For example, the forehead sub-region can be determined by following equations

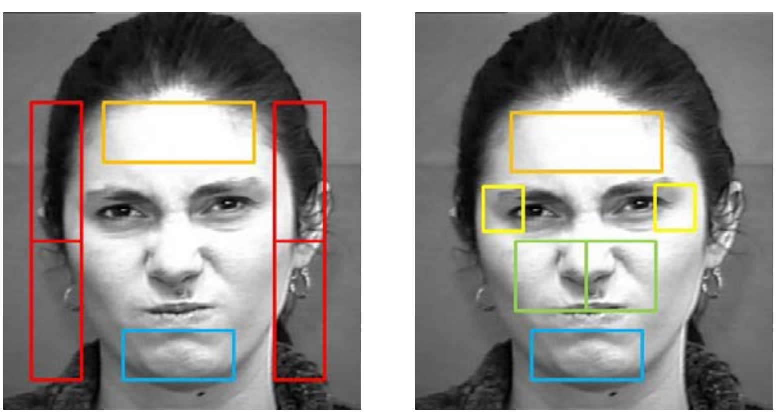
$$x = \frac{x_{face}+2*x_{leye}}{3}, y = y_{face}, w = \frac{x_{face}+w_{face}+2*x_{reye}}{3}, h = \frac{y_{leye}-y_{face}}{2} \quad (2)$$

Where (x, y, w, h) and $(x_{face}, y_{face}, w_{face}, h_{face})$ mean the rectangle correspond to the forehead sub-region and face region, respectively. (x_{leye}, y_{leye}) and (x_{reye}, y_{reye}) mean the position correspond to the left and right eye, respectively. In this process, we define sub-regions of the face related to age and gender. Table 1 presents detailed information on sub-regions that are manually defined using related landmarks. Once the sub-regions are defined, it is necessary to crop them. This is essential for normalization as it eliminates unnecessary areas and preserves the most relevant information. Hence, as shown in Table 1, each sub-region is resized to a predefined size.

Table 1. The Detailed Information on Sub-regions

Sub-region type	**Size**	**Dim. of Feature**	**Sub-region type**	**Size**	**Dim. of Feature**
Left-Top	52×16	7×128	Left-Eye	16×16	1×128
Right-Top	52×16	7×128	Right-Eye	16×16	1×128
Left-Bottom	52×16	7×128	Left-Cheek	28×28	9 ×128
Right-Bottom	52×16	7×128	Right-Cheek	28×28	9 ×128
Forehead	58×16	8×128	Chin	34×16	4 ×128

In particular, the following sub-regions were chosen to capture the signatures of the age and gender classification process: contextual regions, the forehead, and the chin, totaling six sub-regions that are highly correlated with aging. Moreover, the forehead, the corners of the eyes, the cheeks, and the chin total six sub-regions that are highly correlated with gender. Fig. 3 shows the sub-regions of face for age and gender estimation. We then extract SIFT features f_i by combining the histograms at each sub-region i. The raw feature vector $f = [f_1, f_2, \cdots, f_N]$ is the ensemble of SIFT descriptors, and is meant to feed the classifier's input.



(a) sub-regions for gender estimation (b) sub-regions for age estimation

Figure 3. Example of sub-region for age and gender estimation.

2.3 Dimension Reduction

Generally, high dimensionality is impractical due to large processing time and space complexity. Moreover, it also contributes to accuracy degradation due to data redundancy and noise. Hence, dimensionality reduction algorithms are used to reduce the number of features under consideration. In this work, the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used to reduce the dimensionality of the feature vectors derived from region-SIFT. First, we independently apply the PCA to each feature f_i extracted from the *i*-th sub-region. Because each feature f_i has a high dimensionality, only the principal components are selected with the PCA, and they are called f_i^{pca}. The feature $f^{pca} = \left[f_1^{pca}, f_2^{pca}, \cdots f_N^{pca}\right]$ is then composed by concatenation of the PCA feature vectors calculated from each sub-region. Finally, we apply a LDA to capture the most discriminant features from the face representation.

2.4 Age and Gender Estimation

In this work, we employ the multi-layered SVM classifier with an RBF kernel to guarantee accurate classification in the feature space. Typically, a soft margin SVM with a penalty cost C_p is used to compensate for misclassification due to asymmetric class sizes and over-proportional influence of larger classes. Similar to [1], we also obtain the optimal values for the RBF constants γ and C_p using *trainauto* in *OpenCV* with a 5-fold cross-validation method to avoid under or over-fitting in the training stage. Moreover, the classifier is balanced using a dedicated weight for each class. Here, a dedicated weight is determined by number of data contained in each class. For example, in age classification, the penalty cost of a smaller dataset should be decreased to counterbalance and diminish the influence of a larger dataset.

In this work, we propose a novel demography based multi-layered model. As shown in Fig. 4, we employ a novel multi-layered approach to improve the performance and tackle the complexity problem. In the gender classification layer, we classify gender from the input image using a two-class SVM classifier with an RBF kernel. We then calculate the probability of gender by applying a sigmoidal function to the signed margin distance. In the age classification layer, we classify age using two separate classifiers using a multi-class SVM classifier with an RBF kernel. We subsequently determine the final predicted age by the weight average of the predicted age. The rationale behind this method lies in the differences of facial structures among different genders. For example, middle-aged females and males generally do not show the same facial aging signs due to better skin care and cosmetic use by females.

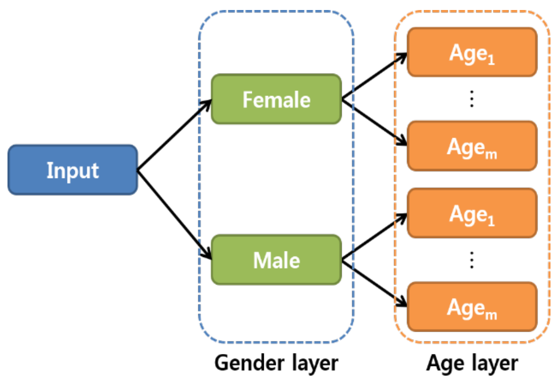


Figure 4. Multi-layered SVM classification model.

3. EXPERIMENTS

We perform age and gender classifications on the FERET database [16], Cross-Age Celebrity Dataset (CACD) [17], and DGIST_C database. On each database, the proposed approach is evaluated using a five-fold cross-validation protocol with a confusion matrix and the overall classification accuracy. Our age classifier categorizes four age groups: 0-19, 20-39, 40-59, and 60+ years old. Table 2 shows the composition of the data used in the experiment. Employing the same notations as used in previous sections, we begin by aligning the detected face and cropping it to size $L_0 = 100$ pixels with the left eye offset at $\Omega_0 = L_0/3$. Our system was implemented using c++ and tested it on a computer with an Intel i7-

2600K CPU @3.40GHz and 32GB RAM. Moreover, several standard routines from the OpenCV Library [18] have been integrated into our framework for face detection and SVM training.

Table 2. The Detailed Information of Database

Database	#Images	Controlled	Gender		Age Group (male + female)			
			Male	**Female**	**0-19**	**20-39**	**40-59**	**60+**
FERET(fa/fb)	2,662	Yes	1657	1005	23+14	942+733	621+250	71+8
CACD	17,620	No	8913	8707	116+821	3329+5498	5261+2319	207+69
DGIST_C	20,741	No	13324	7417	238+955	4456+4589	6525+1408	2105+465

3.1 Results on FERET Database

The FERET dataset [17] is a standard dataset used for facial recognition system evaluation. This dataset contains images of 1,196 different individuals, with up to five images of each individual captured under different lighting conditions, with non-neutral expressions and over a period of three years. From the entire FERET database, we extracted data with letter code 'Fa' and 'Fb'. We then removed data with missing labels. As a result, we finally construct 2,662 images.

As shown in Table 3, the proposed approach still achieves significantly better overall age group classification accuracy (91.17%) than the SIFT feature based method (89.70%). Moreover, the proposed approach also achieves significantly better overall gender classification accuracy (96.62%) than the SIFT (94.77%). As shown in Table 4, while the SIFT feature extraction step consumes about 50ms and classification time of 0.9 ms, the proposed method consumes about 16 ms in the region-SIFT feature extraction step and 0.3 ms in the classification step. This difference is because the dimension of region-SIFT is smaller than that of SIFT. The performance of the proposed method for age group and gender classification on the FERET database is outlined in Table 5. For age group classification, 0-19 and 60+ are found to be easily confused with subjects in their neighboring age groups. We attribute this to the following reason. The number of subjects in different age group in the FERET database is significantly unbalanced (14% and 3% of all subjects are in the age groups 0-19 and 60+, respectively). Regarding the per gender accuracy, the misclassification error of females (5.08%) is higher than that of males (2.35%). We also compare the processing time of the proposed method with that of the SIFT feature based method.

Table 3. The Comparison of the Proposed Method with SIFT

Database	SIFT + SVM		Region-SIFT + Multi-layered SVM	
	Gender	**Age**	**Gender**	**Age**
FERET(fa/fb)	94.77%	89.70%	**96.62%**	**91.17%**
CACD	96.16%	70.46%	**97.17%**	**74.80%**
DGIST_C	72.54%	65.13%	**76.08%**	**67.57%**

Table 4. Computational Analysis Per Cflassification Stage

Stage	SIFT + SVM with RBF	Region-SIFT + multi-layered SVM
Face Alignment	2 ms	**2 ms**
Feature Extraction	50 ms	**16 ms**
Classification	0.9 ms	**0.3 ms**
Total	52.9 ms	**18.3 ms**

Table 5. Computational Analysis Per Cflassification Stage

Overall Acc. 91.17±1.3	**0-19**	**20-39**	**40-59**	**60+**
0-19	**13.51%**	86.49%	0%	0%
20-39	0%	**95.58%**	4.42%	0%
40-59	0%	11.25%	**88.18%**	0.57%
60+	0%	3.80%	29.11%	**67.09%**

(a) Confusion matrix for age group classification

Overall Acc. 96.62±0.5	**Male**	**Female**
Male	**97.65%**	2.35%
Female	5.08%	**94.92%**

(b) Confusion matrix for gender classification

3.2 Results on CACD Database

CACD is short for Cross-Age Celebrity Dataset [18], a large scaled age database. It contains more than 163,446 images of 2,000 celebrities with age ranging from 16 to 62 years. It was originally designed for investigating the problem of age-invariant face recognition and retrieval. Nevertheless, annotation in this database cannot be guaranteed to be correct. Hence, from the entire CACD database, we extracted data with rank 1 to 5.

As shown in Table 3, the proposed approach still achieves significantly better overall age group classification accuracy (74.80%) than the SIFT feature based method (70.46%). Moreover, the proposed approach also achieves significantly better overall gender classification accuracy (97.17%) than the SIFT (96.16%). The performance of the proposed method for age group and gender classification on the CACD database is outlined in Table 6. For age group classification, the subjects in the age groups 0-19 and 60+ tended to be confused with subjects in the age groups 20-39 and 40-59, respectively. One possible explanation for this misclassification is that most of the subjects (93.12%) in CACD are public figures that are in the age range of 20-59, who tend to use facial makeup, resulting in a younger appearance than their real age. The proposed approach nevertheless achieves an overall age group classification accuracy of 74.80%. Moreover, the proposed approach also achieves an overall gender classification accuracy of 97.17%. Regarding the per gender accuracy, the misclassification error of females (2.91%) is higher than that of males (2.75%).

3.3Results on DGIST_C Database

The publicly available face image datasets are not sufficiently containing Asian faces. Hence, we collected 20K face images of Korean celebrities from internet, called DGIST_C. We then manually labeled age, gender, and 7 landmarks.

As shown in Table 3, the proposed approach still achieves significantly better overall age group classification accuracy (67.57%) than the SIFT feature based method (65.13%). Moreover, the proposed approach also achieves significantly better overall gender classification accuracy (76.08%) than the SIFT (72.54%). The performance of the proposed method for age group and gender classification on the DGIST_C database is outlined in Table 7. For age group classification, the subjects in the age group 0-19 tended to be confused with subjects in age group 20-39. We attribute his to the following reasons: (i) the number of subjects in different age groups in the database is significantly unbalanced (e.g., 5% of all the subjects are in the age group 0-19), and (ii) unlike FERET database, DGIST_C database are captured in uncontrolled condition. The proposed approach nevertheless achieves an overall age group classification accuracy of 67.57%. For gender classification, the proposed approach achieves an overall gender classification accuracy of 76.08%. %. According to our observation, the reason for lower gender classification rate can be attributed to the existence of numerous children under ten-years old who are very similar in appearance to females.

Table 6. Computational Analysis Per Cflassification Stage

Overall Acc. 74.80±0.6	**0-19**	**20-39**	**40-59**	**60+**
0-19	**19.21%**	91.72%	6.07%	0%
20-39	2.72%	**80.95%**	16.33%	0%
40-59	0.5%	22.24%	**76.58%**	0.68%
60+	0%	8.36%	75.27%	**16.36%**

(a) Confusion matrix for age group classification

Overall Acc. 97.17±0.3	**Male**	**Female**
Male	**97.25%**	2.75%
Female	2.91%	**97.09%**

(b) Confusion matrix for gender classification

Table 7. Computational Analysis Per Cflassification Stage

Overall Acc. 67.57±1.8	**0-19**	**20-39**	**40-59**	**60+**
0-19	**28.94%**	66.53%	3.61%	0.92%
20-39	3.69%	**82.22%**	12.41%	1.68%
40-59	0.80%	19.71%	**60.10%**	19.39%
60+	0.31%	5.38%	37.34%	**56.97%**

(a) Confusion matrix for age group classification

Overall Acc. 76.08±8.3	**Male**	**Female**
Male	**75.62%**	24.38%
Female	23.10%	**76.90%**

4. CONCLUSION

In this paper, we proposed age and gender estimation framework using region-SIFT and multi-layered SVM. First, we detect facial landmark points and perform face alignment using these landmarks. We then define sub-regions of the face that is related with age and gender estimation and extract region-SIFT feature. Finally, we classify age and gender using multi-layered SVM. By using the multi-layered SVM, we can improve the classification rate by constructing a classifier that estimate the age according to gender. The experimental results demonstrate that the proposed approach classifies age and performs gender estimation very efficiently and accurately.

In order to improve the performance of age and gender estimation, it is necessary to study how to combine various features extracted from specific areas. In the future, the proposed method will be applied to digital signage to provide customized service for advertising and marketing.

ACKNOWLEDGMENT

This research was supported by the DGIST R&D Program of the Ministry of Science and ICT (17-IT-01). It was also supported the technology transfer and commercialization Program through INNOPOLIS Foundation funded by the Ministry of Science, ICT and Future Planning (2017DG0002).

REFERENCES

[1] R. Azarmehr, R. Laganiere, W. S. Lee, C. Xu, and D. Laroche, “Real-time embedded age and gender classification in unconstrained video,” IEEE Conf. on Computer vision and Pattern Recognition Wrokshops, 57-65, (2015).

[2] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with gabor wavelets,” IEEE Conf. on Automatic Face and Gesture Recognition, 200-205, (1998).

[3] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognition, 29(1), 51-59, (1996).

[4] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Computer Vision, 60(2), 91-110, (2004).

[5] H. Han, C. Otto, and A. K. Jain, “Age estimation from face images: Human vs. machine performance,” IEEE Int. Conf. on Biometrics, (2013).

[6] D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross, “Soft biometrics for surveillance: an overview,” Machine learning: theory and applications. Elsevier, 327–352, (2013).

[7] G. Levi and T. Hassner, “Age and gender classification using convolutional neural netowrks,” IEEE Conf. on Computer Vision and Pattern Recognition Workshops, (2015).

[8] F. Gao and H. Ai, “ Face age classification on consumer images with gabor feature and fuzzy lda method,” In Advances in biometrics, 132–141, Springer, (2009).

[9] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang, “A study on automatic age estimation using a large database,” IEEE Int. Conf. on Comput. Vision, 1986–1991, (2009).

[10] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, “Age estimation using a hierarchical classifier based on global and local facial features,” Pattern Recognition, 44(6), 1262–1281, (2011).

[11] E. Makinen and R. Raisamo, “Evaluation of gender classification methods with automatically detected and aligned faces,” IEEE Trans. on Pattern Analysis and Machine Intelligence., 30(3), 541–547, (2008).

[12] S. Baluja and H. A. Rowley, “Boosting sex identification performance,” Int. J. Comput. Vision, 71(1), 111–119, (2007).

[13] E. Fazl-Ersi, M. E. Mousa-Pasandi, R. Laganiere, and M. Awad, “Age and gender recognitionusing informative features of various types,” IEEE Int. Conf. on Image Processing, (2014).

[14] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela, “Revisiting linear discriminant techniques in gender recognition,” IEEE Tran. on Pattern Analysis and Machine Intelligence, 33(4), (2011).

[15] T. Baltru, P, Robinson, and L. P. Morency, “OpenFace: an open source facial behavior analysis toolkitt,” IEEE Winter Conf. on Applications of Computer Vision, (2016).

[16] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, “ The FERET database and evaluation procedure for face recognition algorithm,” Image and Vision Computing, 16, 259-306, (1998).

[17] B. C. Chen, C. S. Chen, and W. H. Hsu, “Cross-Age Reference Coding for Age-Invariant Face Recognition and retrieval,” In European Conf. on Computer Vision, 786-783, (2014).

[18] G. Bradski, “The OpenCV Library,” Dr. Dobb’s Journal, 25(11), 120-126, (2000)