Gender Classification Based on Fusing Shape and Texture Based Approaches

Mayibongwe H Bayana, Serestina Viriri, and Raphael Angulu

School of Mathematics,Statistics and Computer Science,
University of Kwazulu Natal, Durban, South Africa

Abstract. In this paper we seek to improve gender classification accuracy by fusing shape features, the Active Shape Model with the two appearance based methods, the Local Binary Pattern and the Local Directional Pattern. The Support Vector Machine with an Radial Basis Function kernel is used as the classifier of choice.To improve classification accuracies we have proposed the fusion of appearance and shape features as gender classification techniques have not achieved an ideal zero Mean Absolute Error. We use the whole facial area as input into the feature extractor on images of two datasets the FG-NET and the Pilots Parliaments Benchmark which contains a huge variation in images from light to extremely dark individuals. The findings of this paper show that fusion of the LBP and LDP with the Active Shape Model increased classification accuracy on the FG-NET fusion of LDP and ASM moved it from an accuracy of 94.5% from 92.8% before fusion. The Local Binary Pattern on the FG-NET increases from 85.33% to 89.53% after fusion.

Keywords: Active Shape Model, fusion, Local Binary Pattern, Local Directional Pattern

1 Introduction

The human face is very important when it comes to communication as it conveys a lot of information, such as gender [1] or age [2], which helps one to know how to then approach an individual to initiate a conversation. With Artificial Intelligence making inroads into every corner of society the human face has been used as the source of information from age [2], ethnicity [3], [4], emotion [5]. However this information may not be enough to uniquely identify an individual and in such cases it is referred to as a soft biometric [6]. Gender classification has been applied to the field of Human Computer Interaction which may improve interactions and services tailoring them to each gender [7], another important application of gender classification is with marketing electronic billboards which may be programmed to show male products when there are more males than females standing in front of it [8].

A number of methods have been put forward to obtain the gender of an individual given an image. Researchers have used the face as a whole, which may not always be possible as a result of environmental constraints and as a result fusion may be used put forward to enhance gender classification accuracies is fusion, which can occur at feature or classifier level. According to Yang et al [9] there have been a few papers which have discussed classification of gender by fusing global and local features. Of particular interest in this research is the Local Binary Pattern [10]. Hence since previous gender classification techniques have not achieved the Mean Absolute Error [11] and [12] obtained improved classification results after applying feature fusion.

In this paper we attempt to improve gender classification accuracies using the whole facial area as input by fusing the appearance (Local Binary Pattern, Local Directional Pattern) and the shape features (Active Shape Models). The Support Vector Machine is used as the classifier of choice inspired by the findings of [13] who obtained improved results from fusing the DCT, LBP and a novel feature extractor they termed the Geometric Distance Feature using a Support Vector Machine as the preferred classifier. We use the FG-Net dataset and the Pilots Parliament Benchmark project [14] which is made up of images of parliamentarians drawn from Africa and Europe. The database used in this paper uses images from different races however the AR [13] is made up of individuals from Iran.

1.1 Literature Review

Approaches to gender classification have been categorized into two which are the appearance and the geometric based approach [7]. The appearance based approach uses image pixels and performs statistical computations on them, on the other hand the geometric approach uses the distances between the facial features to determine an individuals gender. The appearance based approach hence makes use of an image as a high dimensional vector and extracts features from its statistical features making a decision based solely on it. Feature extraction attempts to develop a transformation of the input space onto a low-dimensional subspace whilst maintaining the most relevant information [15].

In the area of pattern recognition there are two categories when it comes to one is feature combination and the other being classifier combination [16]. A hybrid approach was put forward by Mozaffari et al [13], in which they utilized two appearance based features the Discrete Cosine Transform (DCT), Local Binary Pattern (LBP) and the novel Geometric Distance Feature (GDF) to enhance gender classification and compared the results on two datasets, Ethnic and AR using ellipse fitting during the localization process. On the Ethnic dataset a combination of the LBP and DCT yielded a 84.6% compared to 97.1% for LBP, DCT and GDF showing an enhancement of 12.5%. On the AR dataset combination of the LBP and DCT yields a 80.3% a combination of LBP, DCT and GDF yield a 96% accuracy having and enhancement of 15.7% using frontal images.

A hybrid approach is proposed by Xu et al [17], combining the Appearance based and geometric features. 83 landmarks are detected from the local features using the Active Appearance Model, leading to $3403(C_{83}^{2})$ features from which 10 are selected to form a feature vector and as a result can be further optimized to detect fewer features for classification to improve efficiency. Global features are extracted using the Adaboost algorithm and a Support Vector Machine is used as the classifier of choice. The results from the hybrid method proposed by Xu et al [17] led to a recognition rate of 92.38% compared to 82.97% and 88.55% for appearance and geometry features when used individually

Other researchers have attempted to carry out fusion at classifier level [16] by combining four different Support Vector Machines for the eyes, nose, mouth and hair using the fuzzy integral. Of particular interest is the use of hair, as proposed by Lapedriza et al [18] as previous researchers had shunned its use because of its large variations [16]. From their results [16] found that individually they eyes had the highest classification of 84.24% followed by the mouth and the hair had a better accuracy than the nose at 75.42% compared 70.94%. The fuzzy integral was used for classification after combining classifiers for the eyes, nose, mouth and hair recorded the highest accuracy, 90.61% compared to the weighted sum with 99.73%, the Maximum Margin Criterion using only facial features recorded 87.32% .

1.2 Local Binary Pattern

The Local Binary pattern is one the simplest texture based feature extraction techniques [1]. This technique makes use of pixel values to perform calculations on them. The LBP is a non parametric approach which summarizes local structures by comparing each pixel with its neighboring pixels(denoted by P) [19]. In the equation below radius is denoted by(R), the central pixel which is being thresholded is denoted by g_c) converting them into a decimal number as shown in equation 1.

$$\begin{aligned} LBP_{P,R(X_c,Y_c)} &= \sum_{P=0}^{P-1} s_{(g_p-g_c)} 2^{p} \\ where s(x) &= \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \end{aligned} \tag{1}$$

Although the LBP provides an invariant description in the presence of monotomic illumination it does not perform well under non-monotomic illumination. As a result various forms of the LBP have been introduced to make it more robust such as the uniform patterns, which contain at most two bitwise transitions from 0 to 1 or 1 to 0.

In previous research experiments Jabied et al [20], showed that the LDP using an SVM classifier outperformed the LBP using an Adaboost classifier and the LBP using the weighted Chi-square classifier, as facial components have been shown to have different gender discriminatory levels. Classification accuracies of 95,05% against 92.25% and 81.90% are obtained respectively. The findings of the paper also further revealed that increasing block size by doubling it from blocks of 5(5 by 5) to 10 only increased accuracy by 2% and further increasing it to 20 by 20 blocks the accuracy went back down by 1%.

Shape features can be extracted using the Active Shape Model [21] and Active Appearance Model [22]. The latter seeks to match the position of the model points against a representation of the texture of an object to the image whilst the ASM matches the model points to the image whilst being limited by a model of the shape. Lakshmiprabha [23] compared the performance of four major classifiers the AAM, Gabor wavelet, LBP and Wavelet Decomposition for gender classification of the FG-NET dataset with a Neural Network with three layers being used. The AAM obtained 92.523%, Gabor 90.03%, LBP 90.34% and the Wavelet Decomposition (WD) had the least with 89.72% accuracy.

2 Method

The gender classification model proposed is composed of a few steps the first being converting the image to grayscale then detecting the face and extracting the facial area, the region of interest. After facial components have been detected they are extracted using the LBP and the Active Shape Model after which dimensionality reduction is carried out using PCA and LDA. The final stage is classification by the Support Vector Machine as illustrated in Figure 1.

2.1 Face detection

Face detection is one of the most fundamental steps in gender classification and other facial recognition tasks. The findings of previous researchers have shown that the Viola-Jones algorithm [24] has outperformed all other facial detectors and hence it is used in this research. The goal of face detection is to find any face present in a given image returning the given face as an output [25]. There are three components which make the algorithm efficient which are the integral image which enables rapid computation of haar-like features, adaboost learning which combines weak classifier to find a accurate stronger classifier. The haar-like features, which are digital image features used in object detection in the form of rectangles of black and white. The integral image is calculated by summing the pixels in the white area which are then subtracted from those in the black area as shown in equation 2.

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \tag{2}$$

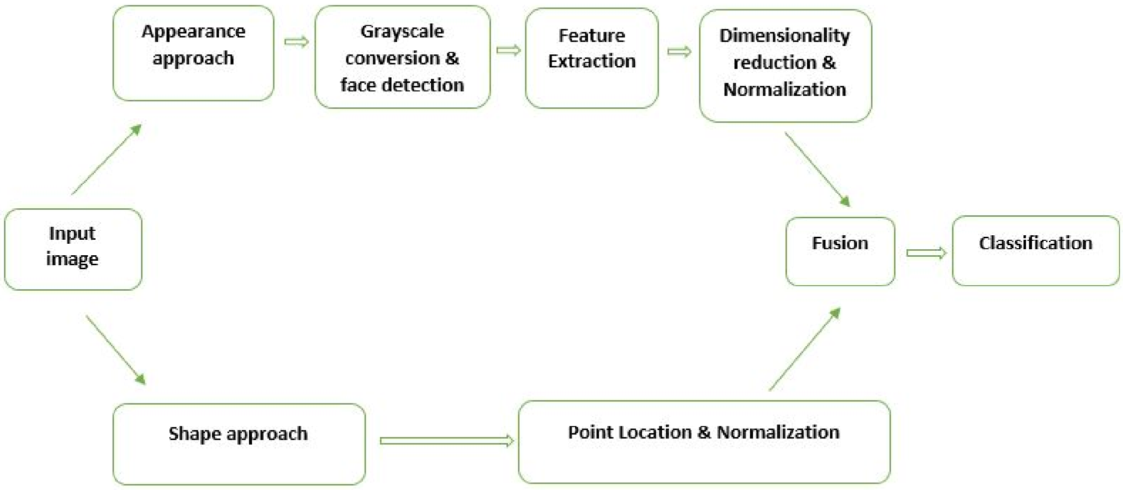

Fig. 1. Proposed Model

In equation 1, $ii(x, y)$ is the integral image at pixel location (x,y) which is used to find the sum of the rectangle where we subtract the white area from the black area as shown in the equation that follows .Given a rectangle WXYZ we then calculate the sum of pixels as shown by equation 3.

$$\sum_{(x,y)} \in WXYZ = ii(Z) + ii(W) - ii(X) - ii(Y) \tag{3}$$

The cascade structure increases efficiency by rejecting negative sub-windows before further processing.

2.2 Local Directional Pattern

The Local Directional Pattern [26] has been shown to have great resistance to noise and hence was chosen a the descriptor of choice. The LDP therefore computes the binary code for each pixel in the image by comparing edge responses of each pixel in the 8 different directions. The Kirsch edge detector is used in this paper. Hence when given a centre pixel in an image$P(i, j)$, 8-directional responses are computed by convolving the neighbouring pixels. The LDP determines the K significant directional responses setting them to a bit value of 1 and the rest to 0. The binary response bit for each bit is shown in equation 4.

$$
\begin{aligned}
LDP_K &= \sum_{i=0}^{i=7} b_i((m_i - m_k) \times 2^i) \\
b_i(a) &= \begin{cases} 1, \text{if a} \geq 0 \\ 0, \text{if a} < 0 \end{cases}
\end{aligned}
\tag{4}
$$

In equation 4, m_k is the K^{th} significant directional response. For $K = 3$ as used we generate $C_3^8(56)$ distinct patterns and a histogram $H(i)$ with C_k^8 bins can be used to represent an image $I_L(x, y)$, shown in equation 5.

$$
\begin{aligned}
H(i) &= \sum_{x,y} P(I_L(x, y) = C_i), \\
\text{where } C_i \text{ is the } i^{th} \text{ LDP Pattern}(0 \leq i < 56) \\
\text{where}P(A) &= \begin{cases} 1 \; if \; A \; is \; true \\ 0 \; if \; A \; is \; false \end{cases}
\end{aligned}
\tag{5}
$$

2.3 Active Shape Model

The Active Shape Model put proposed by Cootes [27], is used in medical segmentation and facial feature localization to mention a few areas, it fits a set of local feature detectors to an object taking into account global shape considerations. A manually labelled training set is used to come up with a linear shape model shown by the equation 6.

$$x = \bar{x} + P_s b_s \tag{6}$$

In equation 6 above, $\bar{x}$ is the mean shape, with P_s being the set of orthogonal modes of variation and b_s is the set of shape parameters, hence fitting an ASM is a non-linear optimization problem to minimize a squared error measure of the output that is desired and that of the model. This is shown in the equation below in which Y represents feature points in an image plane, the shape model parameters b_s are determined by minimizing the expression as shown in equation 7.

$$|Y - T_t(\bar{x} + P_s b_s)| \tag{7}$$

Constraints are placed on the allowable shape parameters b_s the shape model estimate of the current feature points $T_t(\bar{x}) + P_s b_s)$ are constrained to form a plausible shape, an example of an image with the detected facial points is shown in FIGURE 2.

2.4 Normalization and Feature Fusion

After feature extraction has been carried out normalization is carried out on both feature vectors of the Active Shape Models and the Local Binary Pattern, since

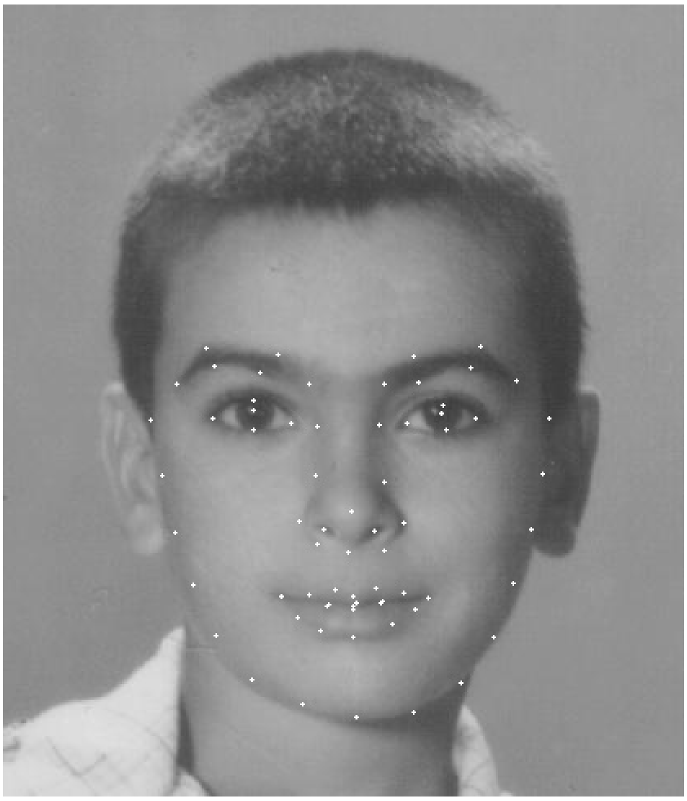

Fig. 2. Detection of 68 facial points on image

we will add the features. From the various methods studied for normalization the Min-Max method is used as it is simpler [28]. The Min-max equation is shown in equation 8.

$$n = \frac{x - min}{max - min} \tag{8}$$

The end-points of the range are specified by max and min and x represents the new matrice value to be determined . Min-max maps the values to within the range $[-1, 1]$. After which feature fusion is carried out by addition and dimensionality reduction is carried out.

2.5 Dimensionality Reduction

The fact that the facial image data is always high dimensional has meant that it requires considerable computing time for classification [29]. Hence gender classification like other image recognition practices has the stumbling block of high dimensionality and hence dimensionality reduction is carried out to improve the learning process and resulting in comprehensibility as irrelevant and redundant data are removed as we move the data from a high dimensional space to one with a low dimension.

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} dimensionality\ reduction- > \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \tag{9}$$

Dimensionality reduction, equation 9, shows dimensionality reduction where the number of variables is reduced from N to K as $K \leq n$ from a high dimensional data space [N], representing our normalized matrice to a lower dimensional dataspace [k]. In this research we have used the Principle Component Analysis and the Local Discriminant Analysis. The Principal Component analysis is a supervised learning method which is also referred as the Singular Value Decomposition.

The PCA reduces dimensionality by finding the direction of greatest variance in the given data. The first principal components show most of the variance, hence the rest can be thrown away as they contain little information [30]. The LDA is also performed in this experiment as to maximize between the classes scatter and minimizing within class scatter without losing data and hence for all sample classes the LDA defines two measures, the first being the within class scatter:

$$S_W = \sum_{j=1}^{c} \sum_{N_{i=1}}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{10}$$

In equation 10 c is the number of classes and$x(i)^j$is the i-th sample of the class j, μ_j and N_j the number of samples in class j. Below is the equation for the between class scatter.

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T \tag{11}$$

Above μ represents mean of all classes. After dimensionality reduction has been carried out the vector is then fed into a support Vector Machine [31], which finds the optimal linear hyperplane which will minimize the classification error for unseen test samples. The RBF kernel was chosen as previous researchers had found that it performed better than others [31], with the SVM type C-Support Vector classification being chosen as it applied on classes greater than or equal to two.

2.6 Datasets

We have used the FG-NET dataset to determine gender estimation using either LDP or LBP for feature extraction and SVM classification. The dataset is built up of 1002 images, coming from 82 different individuals drawn up from ages 1-69 years. Images have a wide variation in illumination, color amongst others.

The Pilots Parliament Benchmark has been used to provide a better intersectional representation on gender and skin type and consists of 1270 individuals drawn from three African and three European countries giving a good reflection on algorithm performance. Images from the FG-Net dataset are also used made up of 1002 images belonging to 82 different individuals with ages ranging from 0-69.

3 Results

The experiments carried out in this paper included feature using the LBP, LDP, ASM with the SVM as the classifier and the results are shown in Table 1.

Table 1. Gender classification using the SVM with various feature extractors

Feature Extractor	**FG-NET**	**PPB**
LBP	85.33%	83.13%
LDP	92.85%	89.26%
LDP + ASM	94.53%	81.56%
LBP + ASM	89.53%	85.43%

Table 1 shows the performance of the feature extractors before and after fusion. The LDP outperformed the LBP on both datasets as expected as it computes edge responses values in all directions [20] rather than using pixels as done by the LBP. However the LDP obtained an accuracy of 92.85% but it is reduced on the FG-NET to 89.26% which could be attributed to the presence of darker skin tones on the PPB dataset. When fusion is carried out the LDP is fused with ASM recording the highest accuracy with 94.53% on the FG-NET and 81.56% on the PPB. Gender accuracy is improved from 92.85% to 94.53% on the FG-NET through fusion of LDP and ASM . Fusion also improves LBP accuracy on the PPB from 83.13% to 85.43% on the PPB.

4 Conclusion and Future work

In this paper, a method for fusing shape and appearance based features is put forward we have used the LDP and LBP to extract the appearance based features and the Active Shape Model to extract the shape features. Of particular interest is the fact that for all tests, the fusion based approach outperforms the singular feature extractor. Also the PPB has a consistantly lower classification accuracy, which is also in line with the finding of [14] who found that images with darker complexion have the highest misclassification. However there are some aspects to consider to optimize the research as using all 68 points to create a feature vector may not be efficient as we may only need the points which contribute to gender as shown by [32] that around 85% of features show significant differences in male and female features. The research by Mozaffari [13] uses ellipse fitting, however research has shown that not all faces are ellipsical as some are round, square and oblong [33].

5 Acknowledgments.

The authors of this papers would like to thank the authors of the Pilot Parliaments Project (PPB) dataset and the Face and Gesture Recognition Network(FG-NET).

References

1. ming Cheung, Y., Deng, J.: Ultra local binary pattern for image texture analysis. Secutrity,Pattern Analysis and Cybernetics (2014)
2. R.Angulu, Tapamo, J., Adewum, A.: Human age estimation using multifrequency biologically inspired features (mf-bif). IEEE AFRICON, 2017 (2017)
3. Wang, W., Feixiang, Zhao, H.: Facial ethnicity classification with deep convolutional neural networks. Chinese Conference on Biometric Recognition (2016)
4. A, B., V. S, .: component-based ethnicity identification from facial images, computer vision and graphics. Lecture Notes in Computer Science,, 2016. ICCVG 2016. (2016)
5. L.Luoh, Huang, C.: International coriference on system science and engineering image processing based emotion recognition. 20IO International Coriference on System Science and Engineering (2010)
6. Anil K. Jain, S.C.D.N.: Soft biometric traits for personal recognition systems. Biometric Authentication. Lecture Notes in Computer Science ((2004))
7. Springer, ed.: Advances in Cross-National Comparison A European Working Book for Demographic and Socio-Economic Variables. Springer (2003)
8. Yu, S., Tan, T., , Huang, K., Jia, K., Wu, X.: A study on gait-based gender classification. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 8, AUGUST 2009 (2008)
9. Yang, W., Chen, C., Ricanek, K., Sun, C.: Biometric Recognition. Gender Classification via Global-Local Features Fusio (2011)
10. Ojala, T., Maenp, T.P.M.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971987 (2002)
11. A.Dehghan, E., G.Shu: Dager: Deep age, gender and emotion recognition using convolutional neural networks. IET Image Processing (2017)
12. Lian, H., Lu, B.: Multi-view gender classification using local binary patterns and support vector machines. International Symposium on Neural Networks (2006)
13. S.Mozaffari, H.Behravan, R.Akbari: Gender classification using single frontal image per person: Combination of appearance and geometric based features. Conference: Conference: 20th International Conference on Pattern Recognition, ICPR 2010, (2010)

14. J.Buolamwini, T.Gebru: Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research 81:1 -15 (2018)
15. Chumerin, N., Hulle, M.M.V.: Comparison of two feature extraction methods based on maximization of mutual information. 2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing (2006)
16. Chen, X., Lu, B.L.: Gender classfication by combining facial and hair information. Advances in Neuro-information Processing:15th International Conference (ICONIP)
17. Ziyi Xu, L.L., Shi, P.: A hybrid approach to gender classification from face images. 2008 IEEE (2008)
18. A.Lapedriza, D, M., J, V.: Are external face features useful for automatic face classification? CVPR 2005, Vol. 3, pp. 151157 (2005 (2005)
19. Cao, L., Dikmen, M., Fu, Y.: Gender recognition from body. MM 08 Proceedings of the 16th ACM international conference on Multimedi (2008)
20. Jabid, Kabir, M., O.Chae: Local directional pattern for face recognition. International Conference on Consumer Electronics (2010)
21. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: active shape modelstheir training and application,. Computer Vision and Image Understanding, vol. 61, pp. 3859, 1 1995. (1995)
22. T.Cootes, G.Edwards, C.Taylor: Active appearance models. Computer Vision ECCV'98 (1998)
23. Lakshmiprabha, N.: Face image analysis using aam, gabor , lbp and wd features for gender age ,expression and ethnicity classification. Computer Vision and Pattern Recognition (cs.CV)2016 (2016)
24. Viola, P., Jones, M.: Robust real-time object detection. Second International workshop on statistical and computational theories of vision - modelling,learning,computing and sampling (2001)
25. Ming-Hsuan, Kriegman, D.J., Ahuja, N.: Detecting faces in images: A survey. EEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 1, JANUARY 2002 (2002)
26. Jabid, T., Kabir, H., Chae, O.: Robust facial expression recognition based on local directional pattern. ETRI Journal (2010)
27. Cootes, T., G.Edwards, Taylor, C.: Comparing active shape models with active appearance models. In British Machine Vision (1999)
28. Xu, Z., Li Lu, P.S.: A hybrid approach to gender classification from face images. 2008 19th International Conference on Pattern Recognition (2008)
29. V.Jadhao, R.S.D.: Feature extraction and dimensionality reduction using radon and fourier transforms with application to face recognition. Conference on computational intelligence and multimedia applications (2007)
30. Y.pang, Y.Yuan, Li, X.: Effective feature extraction in a high dimensinal space. IEEE Trans.Syst (2008)
31. Yang, M.H., Moghaddam, B.: Support vector machines for visual gender classification. Proceedings 15th International Conference on Pattern Recognition. ICPR-2000 (2000)
32. A.Samal, V.Subramani, D.Marx: Analysis of sexul dimorphism in human face. J.Vis.Commun.Image R (2007)
33. A.Zafar, T.Popa: Face and eye-ware classification using geometric features for a data-driven eye-ware recommendation system. Graphics Interface Conference 2016 1-3 June, Victoria, British Columbia, Canada Copyright held by authors.