

## Brief review on gender classification techniques



Muhammad Humair Noor<sup>1</sup>, Sajid Ali Khan<sup>1,2,\*</sup>, Anmol Haider<sup>1</sup>, Ahmed Faraz<sup>1</sup>, Osama Khan<sup>1</sup>, Arsalan Aamir<sup>1</sup>, Nazish Noor<sup>3</sup>

<sup>1</sup>Department of Software Engineering, Foundation University, Rawalpindi, Pakistan

<sup>2</sup>Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan

<sup>3</sup>Harvest American, Inc. New York, USA

### ARTICLE INFO

#### Article history:

Received 6 February 2017

Received in revised form

21 April 2017

Accepted 18 May 2017

#### Keywords:

Gender classification

Feature selection

Feature extraction

### ABSTRACT

We carried a study of comparison for the gender classification methods for finding their pros and cons. The main primary contributions are comparable and comprehensive results for the classification of gender methods and combined with real-time automatic detection of face. Our research is focus on highlighting the limitations and strengths of different gender classification techniques by taking an overview of some major problems. Several areas of future research have been presented in this paper.

© 2017 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Gender classification has received great interest from researchers, because of its growing use in the current Human Computer Interaction (HCI) system. Its application area includes security, sales, access limitations, demographic studies, etc. Both the traditional and modern pattern classifiers such as Nearest Neighbor, Linear and Quadratic Discriminant, Support Vector Machine (SVM), and Neural Networks have been substantially used for gender classification in literature. Classification is one of the many techniques that are used for data mining, and serves as a very useful tool. The detection of distinctive patterns or hidden affinities throughout large datasets helps in making decisions or drawing predictions. One of the main uses of face recognition is the identification of a person's gender from either a captured image, or from a continuous stream. The visual information that is extracted from the human faces serves as the fundamental source of information for gender classification (Graf and Wichmann, 2002). A large number of psychophysical studies have investigated gender classification from facial perception in the humans (O'Toole et al., 1998). So far there has not been a single technique that provides the robust solution for application whose functionality involves the face recognition (Chen et al., 2005).

For gender classification from facial images, feature extraction and the selection of proper classifier are of major significance. Feature extraction techniques for captured images or continuous stream are usually categorized into two methods, geometric and template matching. Feature extraction techniques deal with measurement of distinct facial parts like mouth, eyes, chin, nose and other salient elements of the face.

Every feature extraction technique includes the following attributes (Khan et al., 2013):

- Geometry-based features are normally extracted by the use of geometric characteristics such as size of the facial parts and relative position. This said approach results in a large number of extracted features that can be used by the selected classifier.
- Template-based approach compares the elements of the face to templates that have been previously designed. This approach can sometimes be too complex due to the extensive and extremely detailed computation that is involved and works only when they entered and existing model images both have the same orientation, illumination properties and scale.
- Appearance-based technique deals with any of the extracted image characteristics, known as a feature. However, good quality images are needed to extract features properly.

## 2. Literature review

Lu et al. (2008) the author proposed gender recognition for real time by using a pixel pattern based texture feature (PPBTF). He uses (PCA) principal function for obtaining the image as the

\* Corresponding Author.

Email Address: [sajidalibn@gmail.com](mailto:sajidalibn@gmail.com) (S. A. Khan)

<https://doi.org/10.21833/ijaas.2017.07.013>

2313-626X/© 2017 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

pattern matching. support vector machine (SVMs) was used for classification and to select the most discriminative feature subset he used Adaboost. Both of these applied for the successfully face processing computational. Receiver operating characteristics (ROC) used to assess the accuracy. The proposed algorithm first tested by FERET database and trained. Gobar filter bank was used to evaluate and effectiveness of PPBTF by using of multiple channel and it is also used to make classification of Adaboost and SVM.

Carcagni et al. (2015) proposed Histogram of oriented gradients descriptor (HOG) to handle facial expression recognition problem. From the generic images Facial expression recognition requires pipeline algorithms that have many operating blocks. HOG may be used for the coherent spatial references which is position of eyes. It was also used for the facial emotions with the help of SVM technique. In face registration and detection input was the human face image then operation of registration was done. Face detection was done by the mean of Viola and Jones which became more complex by using of classifier in cascade. Feature vectors of HOG were input of Support Vector Machines (SVMs). Separating hyper plane was used to define the SVM discriminative classifier. For video sequence analysis and a decision strategy was based on temporal consistency of FER. In experimental setup of data the session was used to carry out the two available of dataset images that sequence acquires by FER issue. The pipeline was tested to prove that achieved HOG was general and it was not depending of data input. From this experiment it was prove that it can used for real world application context of video sequences.

Ng et al. (2012) proposed Vision based Human Gender Recognition. His approaches based on 2-D still images and videos rather than 3-D. SVM was widely used for gender classification and it is followed by Adaboost boosting approaches. SVM also used by value of pixel intensity. Principal component analysis (PCA) method used to reduce the dimension of images spaces. For same purpose he used (2DPCA) which is two dimensional, nonlinear data reduce by (CCA) Curvilinear Component Analysis, Rapid face detection rectangular feature introduced by Viola and Jones. Local binary pattern (LBP) used for texture classification of invariant rotation and gray scale. Scale invariant feature transform (SIFT) used to remove invariant images scaling, rotation and translation. Unconstrained face recognition compiled by (Labeled Faces in the Wild) LFW. Face database used in it is CAS PEAL

Castrillón-Santana et al. (2017) purposed the Demographic classification of people appearing in photos as well as in (real time) videos was attracting increasing interest in the scientific community, and especially among biometric system. His approach used to solve a problem of face images by GC automatic using multi expert. Local Binary Patterns (LBP) is used to applied facial analysis for the collection of descriptors. For the same purpose he

used Local Gradient Patterns (LGP), Local Derivative Patterns (LDP)), Weber Local Descriptor, Local Phase Quantization (LPQ). After the learning of class statistics Likelihood Ratio (LR) was used to evaluate the membership of specific class. Information is retained by feature level (FL) and information is loses by decision level (DL) before their final result. Therefore, particularly when the number of experts to combine increases, score level (SL) fusion achieves the best compromise among, speed, preserved information. (Ethnicity, Gender and Age face database) EGA is used to support experiments on face demographics. Images captured in no controlled and controlled condition by FRGC. He detecting the region of interest (ROI) containing the face using the algorithm by Viola-Jones. He offers a better trade-off between accuracy and cost with respect to fusion of features

Levi and Hassner (2015) proposed a Convolution Neural Networks for Age and Gender Classification. He used a network architecture for both gender and age classification. For human face parsing and pose estimation key point CNN was used. Framework Caffe open source was used for implemented of his method. CNN also used to improve the classification of gender result and also for the unconstrained image sets which labeled for gender and age. This system which was used to improving result of reported it is the simplicity of this system and it is elaborating the training data.

Lee et al. (2015) proposed a (cGPRT) cascade Gaussian process regression trees for the face alignment method. Cascade stage-wise manner and Gaussian process regression trees (GPRT) both were combined for this process. For the Face alignment is a task he use a cascade regression trees (CRT). Initial and output of the tree was sum of final shape. Current shape was recognized by each tree. Gaussian (DoG) features which were develop by local retinal pattern used to design an input feature of cGPRT. From Viola and Jones face detector he gather the images of face and data training which was first cropped by used of bounding boxes. Then from randomly sampled initialized a shape estimation. HELEN and LFPW used for the dataset comparison result. LBF and ERT used for feature comparison with two baseline index shape.

Han and Jain (2014) proposed an Automatic estimation of demographic attributes (e.g. Gender, age, and race) from an image face was a topic of growing interest with many potential applications. Given an input face image, he first normalizes it by performing pose and photometric corrections. Biologically inspired features (BIF). In the first layer of BIF, he applies Gabor filter to a normalized face image. He used three different SVM classifiers with RBF kernel to perform age group, gender, and race classifications to predict the age group (or exact age), gender, and race of a subject. For SVM implementation, he uses a publicly available LIBSVM library. His Experimental results shows on two large public-domain unconstrained face databases (Images of Groups and LFW) used to outperform by

state of art methods. No demographic estimation results have been reported on the (LFW) Labeled Faces in the Wild database. He was use (LFW+) of the public-domain LFW database. He applied 2D affine transformation based on the two eye locations to correct pose variations in unconstrained face images. In his experimental result Generalization ability of the proposed approach is evaluated through cross-database age group and gender classification between the Images of Groups and LFW+ databases, and the exact age estimation on the FG-NET database. For the exact age estimation, he report the cumulative score (CS) curve and the mean absolute error (MAE); his proposed approach on the public-domain Images of Groups and FG-NET databases and an extended version (LFW+) of the LFW database. Unconstrained face images of subjects.

Stewart et al. (2013) proposed that when real world images were extracted from a video stream, parts of the face may be occluded and hence classification via lips (static or dynamic) can become applicable. A Discrete Cosine Transform (DCT) is followed throughout and GMM approach is taken.

Bai et al. (2013) in their paper, proposed a new way of classifying gender via extracted images by using Local Directional Pattern (LDP). A set of pixels are focused on and a feature vector is formed which represent a coded image. The results are then taken to Google Images and may even be applied FERET and FG-NET face database. Results show that such an approach is about 95% accurate.

Khryashchev et al. (2012) proposed an automatic gender recognition algorithm that is based on machine learning methods. It is performed in two parts: adaptive feature extraction and support vector machine classification. A large image dataset is needed to ensure efficient and various results. The classifier works upon Adaptive Features and SVM (AF-SVM). The algorithm follows scaling of images, calculation of adaptive feature set, color space transform and Support Vector Machine classification. Input images are converted from RGB color space to HSV color space, and then are scaled accordingly. The results of the new classifier are shown to be at 79.6% which is 1.9% higher from average SVM. 65 faces can be processed per second, thus having much more improved time complexity.

Khan et al. (2013), in their paper, proposed the usage of decision trees to check the accuracy of their results regarding facial recognition and gender classification. One of the main classification technique used is Support Vector Machine (SVM). At least six face features are required to identify and extract a frontal face image. Eye glasses or any other accessory can hinder the process. The new classifier formed has an accuracy of 98.5%, out of 200 facial images studied from a dataset. This is a new approach into gender classification process and can be used to form a new rule set for a decision support system involving a large dataset with several decision parameters.

Han et al. (2013) propose the application of Age Specific Human Computer Interaction (ASHCI) in the practical work, by forming a novel technique for Age Estimation. Facial features were extracted by Active Appearance Models (AAMs). For the purpose of Age Regression, the Traditional Quadratic Model (QM) that was used based on Least Square Estimation (LSE), but as it is not robust, it has not been used and a new method Locally Adjusted Robust Regressor (LARR) has been designed. The LARR method clearly out-performs many of the state-of-the-art approaches that have been used for age estimation. This method's Mean Absolute Error (MAE) and Cumulative Score (CS) are both better than the previous approaches. The newly designed method has MAEs of 5.25 years for females, and 5.30 years for males, which are explicitly smaller than the previous results obtained under same experimentation. This system about has brought about 24% deduction of MAEs, as compared to the best results of most of the previous approaches.

Xu et al. (2011) claimed that it was very challenging to track all the individuals in a crowded scene. As a result, it is impractical to extract object trajectories as the feature to represent events in crowded scenes. This research investigates the suitability of using low level features such as dynamic textures and particle trajectories for event representation, which are more robust to an alternative method for local unusual even the detection is proposed. In this method, LBPTOPs are also extracted from regular grids of the video. Then a dimension reduction and whitening process is performed using Principal Component. The Gaussian Mixture Model defines the data is used for a mixture of Gaussian distributions. Once it is trained or built, it will be fast in computing the statistics of a sample input from the given data. The low probability patterns are detected as the unusual events besides the efficiency in the detection step especially when there are dependencies among the channels of feature inputs as these dependencies are captured. GMMs sometimes were used together with Hidden Markov Models. Andrade The purpose of feature extraction is to represent the surveillance events in to a feature vector, which is effective enough to separate the unusual events from usual events.

Savadi and Patil (2014) said that facial expression recognition is an important research problem working on the different fields and disciplines. This because Facial recognition expression , in additional to have a same work system of detecting these features, such as bankcard identification, access control of the system, security monitoring, and surveillance activity system, Facial expressions refer to movements of the mimetic musculature of the face. The main problem of automatic features of facial extraction from a still image is a challenge. Upside posed image and classified of facial expressions and also includes emotion and mood of a person. Facial Action Coding System (FACS) is a system to taxonomies humans Face expression by their appearance on the face. It is a general standard

to automatically categorize the physical expression of Emotion and it has proven useful to psychologist and to animator. FACS is working for automated computed system that detects the division of faces in images and videos, extracts the defined features of the faces through given techniques, and then

produces temporal profiles of each facial movement. The grid and tones tracking and deformed system used, based on deformable models, tracks the sides edges in continuous videos, as the facial expression evolves the greatest facial expression intensity. Comparison table was shown in Table 1.

**Table 1:** Comparison table

Ref.	Technique used	Pros	Cons
Lu et al. (2008)	ppbtf pixel-pattern-based texture feature 2) roc receiver operating characteristics 3) Gobar filter	1) it comprised the number of pixel in gray scale image belonging each pattern 2) used to assess the accuracy 3) used to evaluate and effectiveness of	1) crucial for improving the accuracy Only choose male under curve 3) Method just reached 66.3774%. accuracy
Carcagni et al. (2015)	1) histogram of oriented gradients 2) Fer	1) Efficiently exported facial expression 2) Accurate	not use of small cell 2) Loss of information
Ng et al. (2012)	1) principal component analysis 2) curvilinear component analysis	1) used for dimension reduction increases Used to reduce the dimensionality of nonlinear data.	hard to model Not applying any optimization technique, It is enable and optimize performance. Its decision level is limited possibility representation of
Castrillón-Santana et al. (2017)	local binary patterns feature level decision level	aim is to efficiently summarize the local structures of images information is retained information is loses	it forms very long codes which make the Lbp codes statistically weak. very complex and time consuming It is enable and optimize performance. Its decision level is limited possibility
Levi and Hassner (2015)	Cnn	more sensible way of images	High computational cost.
Lee et al. (2015)	(cgprt)	optimized exactly, given the values	the predictive mean of Cgprt can be computed in the crt framework with better generalization.
Han and Jain (2014)	biologically inspired features (bif)	achieve high accuracy for face age estimation	we modified a model of visual object recognition processing
Stewart et al. (2013)	discrete cosine transform (dct) and Gaussian mixture model (gmm) approach	High gender identification rate.	Only focuses on gender classification via lips and cannot function if lips are covered.
Bai et al. (2013)	local directional pattern (ldp)	Highly accurate and robust to variations.	Complex background may affect calculation time.
Khryashchev et al. (2012)	adaptive feature and support vector machine (af-svm)	More accurate than existing technique.	Mostly frontal face images used.
Khan et al. (2013)	decision trees and support vector machine (svm)	Novel technique for gender classification.	Multiple facial features required and can fail if some occlusion occurs.
Han et al. (2013)	locally adjusted robust Regressor (larr)	Novel technique for age estimation, high and improved accuracy.	Only frontal face images have been used.
Xu et al. (2011)	gaussian mixture model (gmm)	This is the fastest algorithm for learning mixture models.	This model can't work accurately when the many points are given on the faces. It may diverge and cannot find solution.
Savadi and Patil (2014)	facial action coding system (facs)	This system helps to detect the emotional response the human face.	High dimensional facial features cannot detect properly and it is difficult to use.

### 3. Conclusion

Gender recognition techniques are up-to-date in this paper which is represented by this survey. Face-based gender classification techniques are performed by critical analysis in this paper. The major problem which researchers are still facing is variations of facial like expression changes, occlusions, illumination effects, high data dimensions and computational time. Working on the pixels to classify gender is more accurate but computationally expensive so a researcher prefers to Feature extraction of face rather than to work directly on pixels. Feature-based methods are divided into two i.e. local features and global feature. We can conclude that in the step of classification, support vector machine (SVM) is performing excellent as compared to other classifiers.

### References

Bai MR, Prashanth, and Naveen (2013). A new approach for gender classification. International Conference on Innovations in Computer Science and Engineering (ICICSE-2013), Guru Nanak Engg College, Hyderabad, India.

Carcagni P, Del Coco M, Leo M, and Distanto C (2015). Facial expression recognition and histograms of oriented gradients: A comprehensive study. Springer Plus, 4(1): 1-25.

Castrillón-Santana M, De Marsico M, Nappi M, and Riccio D (2017). MEG: Texture operators for multi-expert gender classification. Computer Vision and Image Understanding, 156: 4-18.

Chen Y, Jiang S and Abraham A (2005). Face recognition using DCT and hybrid flexible neural tree. In the IEEE International Conference on Neural Networks and Brain, IEEE, Beijing, China, 3: 1459-1463. <https://doi.org/10.1109/ICNNB.2005.1614906>

Graf AB and Wichmann FA (2002). Gender classification of human faces. In: Bülthoff HH, Wallraven C, Lee SV, and Poggio TA

- (Eds.), *Biologically Motivated Computer Vision*: 491-500. Springer Berlin Heidelberg, Heidelberg, Germany.
- Han H and Jain AK (2014). Age, gender and race estimation from unconstrained face images. MSU Technical Report MSU-CSE-14-5. Michigan State University, East Lansing, USA.
- Han H, Otto C, and Jain AK (2013). Age estimation from face images: Human vs. machine performance. In the International Conference on Biometrics, IEEE, Madrid, Spain: 1-8. <https://doi.org/10.1109/ICB.2013.6613022>
- Khan MNA, Qureshi SA, and Riaz N (2013). Gender classification with decision trees. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(1): 165-176.
- Khryashchev V, Priorov A, Shmaglit L, and Golubev M (2012). Gender recognition via face area analysis. In the World Congress on Engineering and Computer Science, San Francisco, USA: 645-649.
- Lee D, Park H, and Yoo CD (2015). Face alignment using cascade Gaussian process regression trees. In the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Boston, USA: 4204-4212. <https://doi.org/10.1109/CVPR.2015.7299048>
- Levi G and Hassner T (2015). Age and gender classification using convolutional neural networks. In the IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE: 34-42. <https://doi.org/10.1109/CVPRW.2015.7301352>
- Lu H, Huang Y, Chen Y, and Yang D (2008). Automatic gender recognition based on pixel-pattern-based texture feature. *Journal of Real-Time Image Processing*, 3(1-2): 109-116.
- Ng CB, Tay YH, and Goi BM (2012). Vision-based human gender recognition: A survey. In the 12<sup>th</sup> Pacific Rim International Conference on Artificial Intelligence, 7458: 335-346. [https://doi.org/31\\_0-32695-642-3-978/10.1007](https://doi.org/31_0-32695-642-3-978/10.1007)
- O'Toole AJ, Deffenbacher KA, Valentin D, McKee K, Huff D, and Abdi H (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory and Cognition*, 26(1): 146-160.
- Savadi A and Patil CV (2014). Face based automatic human emotion recognition. *International Journal of Computer Science and Network Security (IJCSNS)*, 14(7): 79-81.
- Stewart D, Pass A, and Zhang J (2013). Gender classification via lips: Static and dynamic features. *IET Biometrics*, 2(1): 28-34.
- Xu J, Denman S, Fookes C, and Sridharan S (2011). Unusual event detection in crowded scenes using bag of LBPs in spatio-temporal patches. In the International Conference on Digital Image Computing Techniques and Applications, IEEE, Australia: 549-554. <https://doi.org/10.1109/DICTA.2011.98>