An Extensive Analysis of Human Gender Prophecy using Computer Vision

Devjyoti Saha[1], Diptangshu De[2], Pratick Ghosh[3], Sourish Sengupta[4] and Tripti Majumdar[5]

[1]*Department of Computer Science and Engineering, Bengal Institute of Technology, Kolkata, India*
[2]*Department of Computer Science and Engineering, Bengal Institute of Technology, Kolkata, India*
[3]*Department of Computer Science and Engineering, Bengal Institute of Technology, Kolkata, India*
[4]*Department of Computer Science and Engineering, Bengal Institute of Technology, Kolkata, India*
[5]*Asst. Professor, Department of Computer Science and Engineering, Bengal Institute of Technology, Kolkata, India*

---***---

Abstract - *One of the most important demographic features of human beings is gender. One can easily distinguish between male and female human faces by looking at their images, but enabling computers to identify a person's gender is a tough job. Human gender recognition with computer vision has found its relevance in a number of fields, ranging from day to day software applications to forensic science, particularly due to the certain surge in usage of social media and social networking websites. This paper, in particular, provides a review of human gender recognition in computer vision from still images and video recording datasets. A survey of different approaches using information ranging from faces to the full body (either from still images or video sequences) have been taken into account and carefully presented. Through this paper, we have reviewed the methods incorporated by various authors in their approaches. Various performances and accuracies have been achieved with the different approaches using different datasets under controlled and uncontrolled environments, though there is much more yet to be done on this topic.*

Key words: Machine Learning, Gender Recognition, Pattern Recognition, Image Processing, Computer Vision, Convolutional Neural Network.

1. INTRODUCTION

Human gender recognition has gained much popularity in recent times. The main reason behind this popularity is due to the usage of gender recognition in various domains such as surveillance, security, Image forgery, and identity verification. Gender is the most important aspect of a human being. A lot of research work has been done on this topic. It has been observed from studies that a human being can easily identify the gender of another human being and classify them as male or female, but it is a tedious job using Computer Vision. Gender classification can be seen as a binary classification approach for classifying into male and female. Various computerized methods are already implemented in order to recognize gender. When classifying gender, some distinguishable features exist between males and females which can be of much use when classifying genders. Most researchers have concentrated on recognizing the gender using facial features such as eyes, nose, mouth, etc. for classifying the gender using machine learning, techniques. In this paper, we have focused on the approaches which use facial features in the field of 2D (still images) and 3D (motion pictures).

A concise report of all the related research works direct us to conclude that the research works, the process of classifying human gender can be divided into the following sequence of steps:

1.1 Face detection and Pre-processing

This step involves the detection of a human face from an image and cropping out the exact facial region and reducing the dimension of the larger image. An example of such an algorithm is Bilateral Histogram Equalization with pre-processing (BHEP) [1].

1.2 Extraction of facial features

After performing the above step, the facial features are extracted from the "pre-processed" image. One such algorithm is Cascaded Deformable Shape Model (CDSM) [2] which masks out different face regions by single-colour rectangles and the process is repeated until no more faces are left undetected. Another such algorithm is tCENsus TRansform hISTogram (tCENTRIST) [3], which produces an eight-bit string for a pixel, to capture the image structure, by comparing its intensity value with eight neighbouring pixels.

1.3 Classification using various classifiers

Classification is a technique in which a computer program learns from data provided to it as an input and then uses this learning to classify new observation. Classification can be done by means of various algorithms like Support Vector Machines (SVM) and Convolutional Neural Networks (CNN), etc. to name some.

2. LITERATURE SURVEY

Azzopardi et al. [4] applied the Viola-Jones algorithm [5] on the input image to detect faces. A face alignment algorithm is used to normalize the pose and to crop the detected faces to 128x128 pixels. The Combination of Shifted FIlter REsponses (COSFIRE) based classifier [43] is then applied to the obtained face image. Another Speeded up Robust Features (SURF) based classifier [44] is used where 51 facial landmark points are detected which belong to the eyes, nose, and mouth, SURF descriptors at the key points which indicate the facial landmarks. The final result is obtained by fusing the outputs of the SURF based classifier and the COSFIRE based classifier [43] through an SVM classifier to obtain the final result. Figure 1 represents the flow diagram as approached by Azzopardi et al. [4].

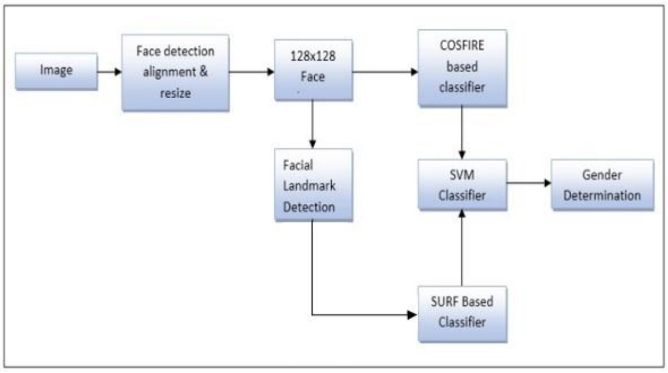

Fig-1: The work flow diagram of the approach

According to Afifi et al. [6], the face images are modulated based on different facial components such as eyes, nose, eyes, etc. These components are used to train a CNN followed by an Adaboost-based score fusion as used by Afifi et al. [6] to infer the final gender class. Initially, a retinex image [7] is obtained by applying the lighting-invariant enhancement techniques. After that, the retinex image is supplied to the CDSM [2]. This model masks out different facial regions by single-colored rectangles and the process is repeated until no more faces are left undetected. After all the facial components are obtained, the face is left with nothing but the outline and this leads to the following step of generating the foggy face. The foggy face is generated using the PIE (Pose, Illumination, and Expression) equation [8]. Now, all of the above features are fed to the CNNs that are pre-trained to classify the biometric traits; i.e., separate CNNs are allotted for separate components (Precisely, 4 CNNs are used for a pair of eyes, mouth, nose and the foggy face) and five classification scores are obtained. These scores are applied to an Adaboost-based score fusion mechanism as stated by Afifi et al. [6]. The algorithm was trained using four different datasets namely: Labeled Faces in the Wild (LFW), Adience benchmark for age and gender classification and Facial Recognition Technology (FERET). But, all the above datasets suffer from a problem of face occlusions and illumination changes. To overcome that, a new dataset has been introduced named Specs on Faces (SoF). On analysis, the accuracy increased from 94.90% to 95.98% using the LFW dataset, 89.7% to 92.71% in the FERET dataset all when compared to SDL-GC [9]. Figure 2 represents the flow diagram as approached by Afifi et al. [6].

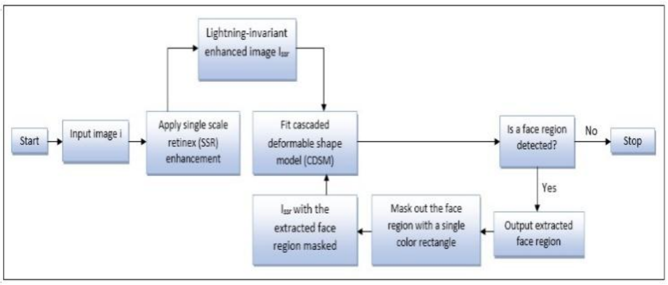

Fig-2: An outline of the method proposed

Nurul et al. [10] in their paper focuses on recognizing the gender in uncontrolled situations where there is a lack of illumination, high rate of noise, etc. To overcome these situations, the proposed framework demands the pre-processing of the whole image using Bilateral Histogram Equalization (BHEP) [1] and then extract facial features from the image using CENTRIST [3] algorithm and then reducing the dimension applying the Principal Component Analysis. The next step includes the use of Support Vector Machines (SVM) to classify the features from the image. CENTRIST [3] is based on the concept of Census transform (CT) [11]. It produces an 8-bit string for a pixel by comparing its intensity value to the neighboring 8 pixels. All the above steps are performed to capture the image structure. The image is enhanced in order for it to be classified on the basis of distinctive properties such as line, corners, spots, edges, etc. For this purpose, BHEP [1] is used as it has shown better results on low contrast images. For extracting the facial features, tCENTRIST [3] is used and PCA is used to reduce the dimension of the image. After extraction of the features, SVM is used for classifying the features of input images. For evaluation purposes, LFW was used as the dataset to train the algorithm and 1/5th of it was used for testing. The proposed method showed an improved accuracy of 94.29% when compared with state-of-art methods without BHEP [1] which showed an accuracy of 93.75%. Figure 3 represents the flow diagram as approached by Nurul et al. [10].

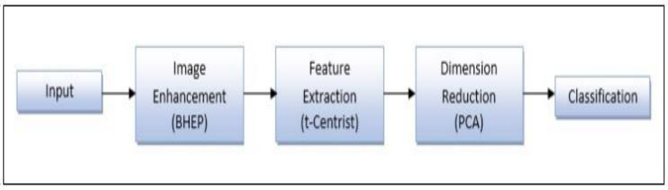

Fig-3: The flow of work of the proposed method

The main objective of Alexandre [12] was solving the problem of non-replicable datasets as well as improving the accuracy defined on the FERET dataset defined by

Makinen et al. [45]. The main idea in this method was to extract features at different image resolutions and classify using different shape and texture classifiers, obtain a classification and combine the decisions obtained. The only classifier used in this process was the Library for Support Vector Machine (LIBSVM) [13]. The SVM was trained with different intensity values for shape and texture features. The results from the shape feature showed an accuracy of 91.53% in the case of FERET database and 86.34% for UND database. The results from the texture feature show an accuracy of 93.46% (36X36 with 6X6 Overlay) in the case of FERET database and 80.18 (128X128 with 16X16 overlay) in case of UND database. The results from these two descriptors were fused and for the decision process, the majority rule was used. After the fusion, the results show an accuracy of 99.07% [12] in the case of FERET database and 91.19% [12] in the case of UND database. Figure 4 represents the flow diagram as approached by Alexandre [12].

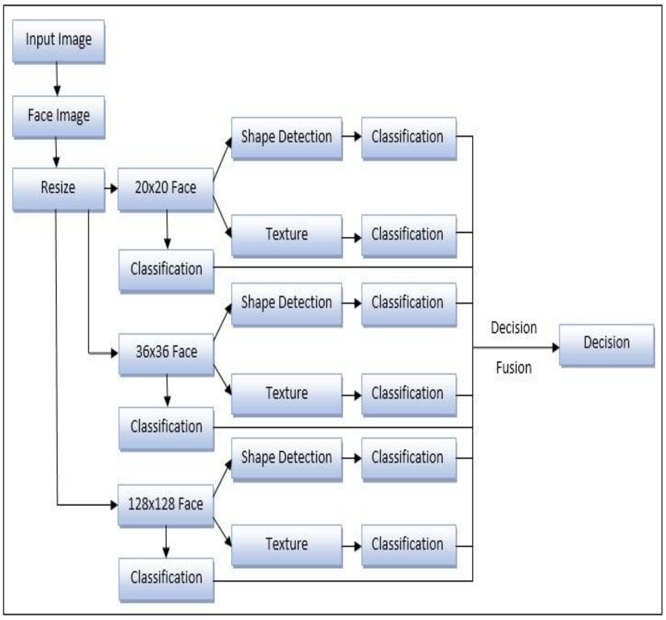

Fig-4: The proposed multi-scale decision fusion approach

In Multilabel Streaming Feature Selection – Conditional Random Fields (MSFS-CRFs) [14] first the face detection is done using a CNN based detector and position of various facial regions are estimated using a CRF [16] based model that was trained using manually labeled face images and then segmented into superpixels using SEEDs algorithm [17] and for shape information, Histogram of Oriented Gradients (HOG) algorithm [18] was used. A new face segmentation algorithm called Multi-class Face classification using conditional random fields [15] was formed. A new automatic gender classification algorithm, called GC-MSFS-CRFs [14] was used for classification. Features such as skin, nose, hair and eyes were considered. The Probability Classification Strategy (PCS) was used to generate probability maps for all classes. The probability maps are used as gender descriptors to train a Random Decision Forest (RDF) classifier and an automatic gender classification algorithm is developed. This method showed an accuracy of 91.4% for the Adience database, 100% for FERET dataset, 94.4% for LFW dataset was and 93.7% for FEI dataset.

Rössler et al. [19] in their work, made use of some modern facial forgery methods and examined various methods of detecting facial forgeries. First, a benchmark was created using four facial manipulation techniques, namely DeepFakes [20], Face2Face [21], FaceSwap [22] and NeuralTextures [23], containing over 1.8 million forged images. A survey with 204 participants was conducted to compare the accuracy of the human observers with the accuracy of the forgery detectors. The detection process involved the tracking and cropping of the face present in the input image. Six different architectures were used for the purpose of automatic forgery detection from which, one architecture was based on Steganalysis features and SVM. The other five were learned feature-based architectures - Cozzolino et al. [24], Bayar and Stamm [25], Rahmouni et al. [26], MesoNet [27] and XceptionNet [28]. Analysis of the resulting accuracies showed very high performances on raw input data, whereas, a drop in the performance for compressed videos. The highest accuracy was obtained with XceptionNet (Raw image: 99.26%, High-Quality image: 95.73%, Low-Quality image: 81%), which surpassed even human observers. The accuracies only improved further with the increase in the number of training images, further demonstrating that trained forgery detectors can be highly capable of detecting forged images. Figure 5 represents the flow diagram as approached in the paper by Rössler et al. [19].

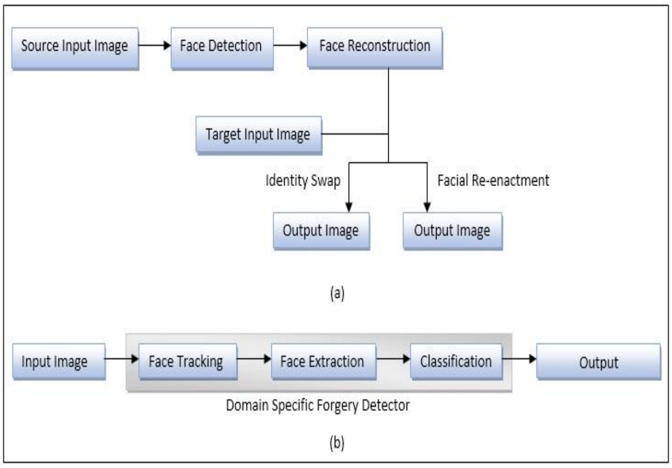

Fig-5: (a) Creation of the forged images dataset using the Identity Swap and Facial Re-enactment algorithms. (b) Detection of facial forgery using Domain Specific Forgery Detectors.

Ramachandran et al. [29] used a Convolutional Neural Network (CNN) as a classifier/algorithm. CNN was used as it considers maximum accuracy and classification speed with equal weightage. A 3 Convolutional layer CNN was used in the paper. Precisely, 16000 training images were used for training (8000 per class) and 2000 images were used for testing (1000 per class) which belong to a part of the UTKFace dataset. An accuracy of 90% has been achieved with the test dataset after 10 epochs. Whereas, Dwivedi et al. [30] use another deep Convolutional Neural Network (CNN) with 5 Convolutional layers. The Gender-FERET dataset was used which achieved an accuracy of 90.33%. Figure 6 represents the basic outline of the approach taken by Ramachandran et al. [29].

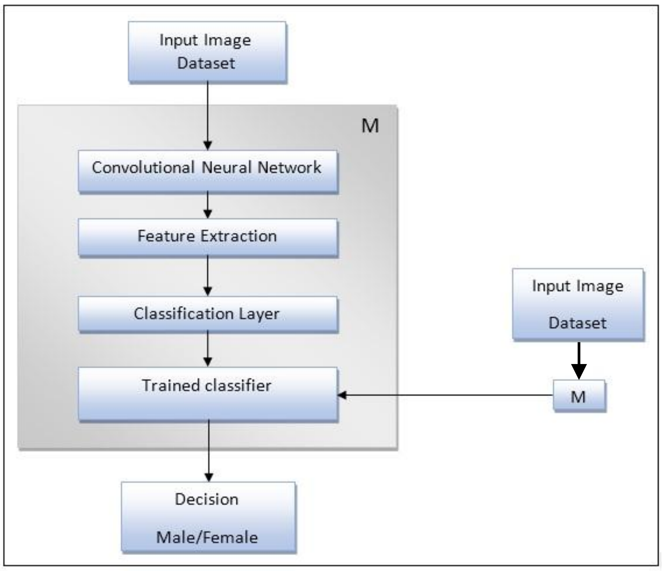

Fig-6: The basic outline of the proposed method

The main objective of El-Alfy et al. [31] was to detect the gender of a person from a video. The algorithm mainly focuses on recognizing a person from their dynamic features such as the way they walk. The gait sequence of a walking person was converted into a Gait Energy Image (GEI) [32] by subtracting the background and focus on the moving person and then making a low dimensional representative set of features relevant for identifying genders using Fuzzy Local Binary Pattern (FLBP) [33]. Different combination of parameters was used for getting the best pair of parameters for feature extraction while training the classification model. The extracted feature vectors were stored in a database and were used to build a linear kernel SVM classifier for gender identification using the LIBSVM package [34]. The relevant features from the GEI image are fed to the SVM classifier which classifies the Person as Male or Female. The CASIA-B multi-view gait database [35] was used for the evaluation of the system. The highest performance to be recorded was for a normal walking person with 96%. The FLBP*, a modified version of FLBP, designed by El-Alfy et al. [31] outperforms the other processes with normal walking at 96.44% and is a more robust variation in the camera. Figure 7 represents the flow diagram as approached by El-Alfy et al. [31].

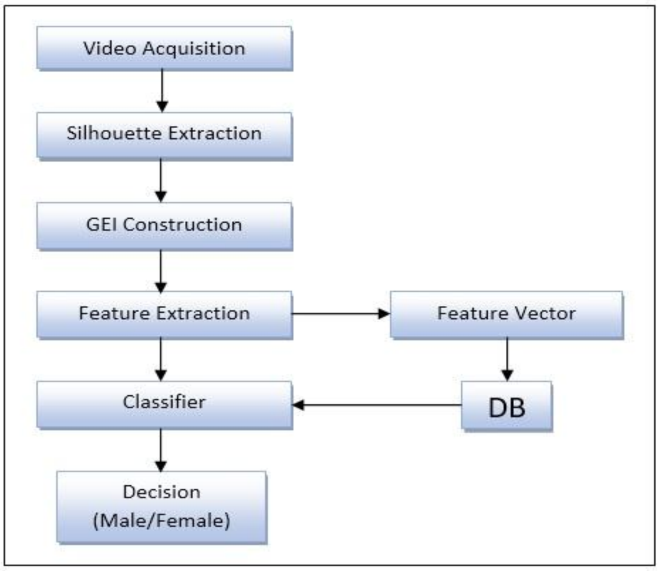

Fig-7: Workflow for gender identification from gait sequence

Zhang S et al. [36] in their paper, analyses the various methods of determining the age and gender of a person using their ear images. The paper employs four geometric based classifiers and four appearance-based classifiers for the process. The geometric features are based on eight landmark points on the ear and derived 16 features from them. The classifiers used were logistic regression, random forest, support vector machine (SVM) and a neural network using 3 hidden layers. For appearance-based methods, a large-scale ear dataset derived from the Multi-PIE face dataset [37] was used. Four popular CNN based architectures, namely AlexNet [38], VGG-16[39], GoogLeNet [40] and SqueezeNet [41] were used. The models were pre-trained on the ImageNet database [42] and then further fine-tuned through a two-stage training procedure using the Multi-PIE dataset and a pre-constructed target dataset. The resulting accuracies of gender classification show that the appearance-based classifiers proved to be superior to the geometrically based classifiers. The highest accuracy for the geometric based architectures was of SVM (65%) and that of the appearance-based classifiers was of GoogLeNet [40] (94%). Figure 8 represents the flow diagram as approached by Zhang S et al. [36].

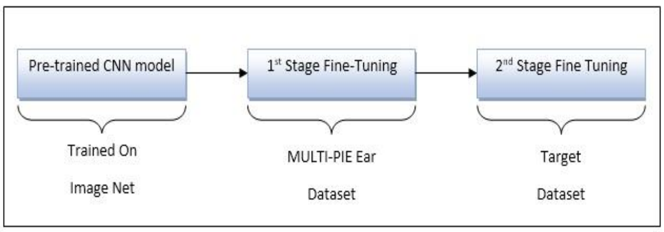

Fig-8: Visualization of the two step fine tuning approach.

While surveying all the approaches undertaken in the various research papers, we have found out that, amidst all the datasets used, there are two datasets used in the majority of the approaches. We have arranged the approaches according to their years of publication and their accuracy percentage in broadly two tables TABLE-I and TABLE-II on basis of the LFW and the Gender-FERET datasets. Figure 9 and Figure 10 graphically represent both the tables.

Year	Accuracy	Approach
2010	99.07%	Gender recognition: A multi-scale decision fusion approach [12]
2018	90.33%	Review of Deep Learning Techniques for Gender Classification in Images [30]
2018	94.7%	Fusion of Domain-Specific and Trainable Features for Gender Recognition from Face Images [4]
2018	98.3%	A Gender Recognition System from Facial Image [10]
2019	99.49%	AFIF4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces [6]
2019	100%	Automatic Gender Classification through Face Segmentation [14]

Table-1: The arrangement of approaches which used the FERET dataset in terms of their years of publication. The accuracies achieved by the approaches have also been listed

Table-2: The arrangement of approaches which used the LFW dataset in terms of their years of publication. The accuracies achieved by the approaches have also been listed

Year	Accuracy	Approach
2018	94.29%	A Gender Recognition System from Facial Image [10]
2018	95.40%	AFIF4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces [6]
2019	95.68%	Automatic Gender Classification through Face Segmentation [14]

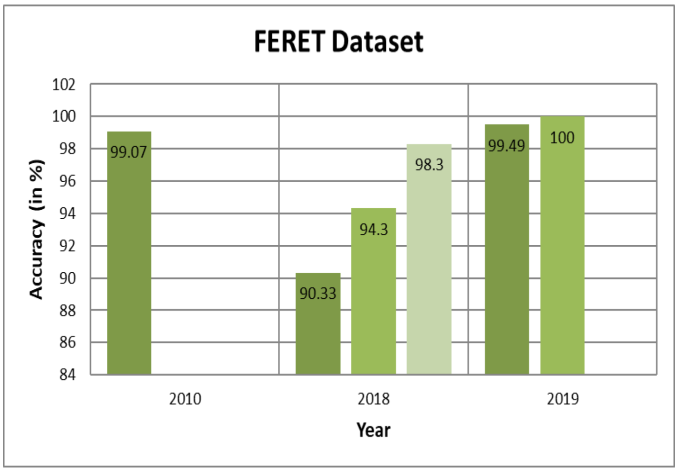

Chart-1: Graphical representation of Table-1

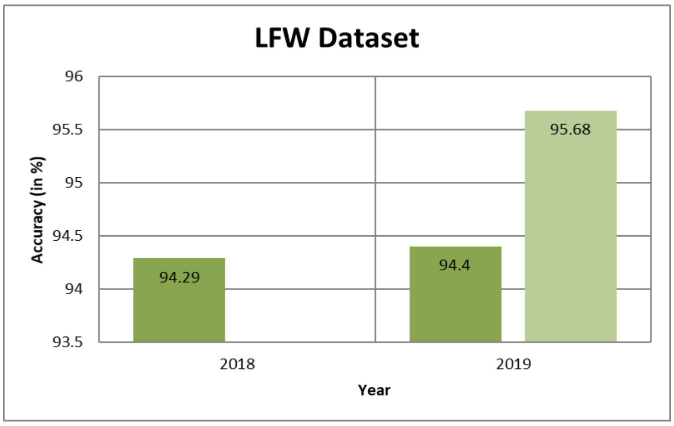

Chart-2: Graphical representation of Table-2

3. CONCLUSION

This paper revolves around various approaches that have been undertaken by different researchers in recent years. The paper also presents other various factors alongside the main approach like accuracies, datasets used and performances. While surveying various research works we found out that, in gender recognition with face, facial landmarks and distance between them played an important role. We also found out that, how genders can be classified from video sequences, which can be a very useful output in the fields of surveillance, security and for organized/customized manner of marketing for a particular product. Furthermore, we have also understood how gender recognition can be a crucial element in the field of forensics where forgery can be detected and eliminated completely through proper technological usage. In many cases, gender classification can be cumbersome from facial images, for those cases gender recognition can be done from analyzing the ears of human beings. Various results have been obtained while surveying different approaches, though a lot of research work is left in the field.

4. FUTURE PROPOSAL

After an extensive analysis of all the approaches, our analysis stands that the future approaches should concentrate on the improvement of three of the most important parameters: **accuracy**, **efficiency** and **reliability** from the existing approaches. The key to reaching ideal results is to improve pre-processing methods and extract more specific facial features to ease the process of detection and classification. A fusion of two or more approaches can be undertaken to improve accuracy and efficiency. The fusion should be effective and well-functioning to yield better results in terms of all the above-mentioned parameters than the already implemented approaches. It is also observed that most of the above approaches work fine where the images are free from non-uniform illuminations and occlusions but not many approaches discuss how to deal with these obstacles. For our future proposal, we expect to work on this domain and deal with the above-stated obstacles as well. If we are successful in overcoming these obstacles, then it will guarantee much better results in terms of all the above-mentioned parameters and would also be considered for frail affairs such as criminal identification and security.

REFERENCES

[1] Feroz Mahmud Amil, Md Mostafijur Rahman, Shanto Rahman, Emon Kumar Dey, and Mohammad Shoyaib. Bilateral histogram equalization with pre-processing for contrast enhancement. In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2016 17th IEEE/ACIS International Conference on, pages 231–236. IEEE, 2016.

[2] X. Yu, J. Huang, S. Zhang, W. Yan, D.N. Metaxas, Pose-free facial landmark fitting via part mixtures and cascaded deformable shape model, in: Proceedings of the IEEE International Computer Vision, 2013, pp. 1944-1951.

[3] Jianxin Wu and Jim M Rehg. Centrist: A visual descriptor for scene categorization. IEEE transaction on pattern analysis and machine intelligence, 33:1489–1501, 2011.

[4] Azzopardi, G., Greco, A., Saggese, A. and Vento, M. (2018). Fusion of Domain-Specific and Trainable Features for Gender Recognition from Face Images. *IEEE Access*, 6, pp.24171-24183.

[5] P. Viola and M. J. Jones, ''Robust real-time face detection,'' Int. J. Comput.Vis., vol. 57, no. 2, pp. 137–154, 2004.

[6] Afifi, M. and Abdelhamed, A. (2019). AFIF4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces. Journal of Visual Communication and Image Representation, 62, pp.77-86.

[7] H. Han, S. Shan, X. Chen, W. Gao, A comparative study on illumination pre-processing in face recognition, Pattern Recognit. 46 (6) (2013) 1691–1699.

[8] P. Pérez, M. Gangnet, A. Blake, Poisson image editing, ACM Transactions on Graphics (TOG), vol. 22, ACM, 2003, pp. 313–318.

[9] H. Moeini, S. Mozaffari, Gender dictionary learning for gender classification, J.Vis. Commun. Image Represent. 42 (2017) 1–13, https://doi.org/10.1016/j.jvcir.2016.11.002.

[10] Nurul, M. and Kumar, E. (2018). A Gender Recognition System from Facial Image. International Journal of Computer Applications, 180(23), pp.5-14.

[11] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In European conference on computer vision, pages 151–158. Springer, 1994.

[12] Alexandre, L. (2010). Gender recognition: A multiscale decision fusion approach. Pattern Recognition Letters, 31(11), pp.1422-1427.

[13] Chang, C.-C., Lin, C.-J., 2001. LIBSVM: A Library for Support Vector Machines. (http://www.csie.ntu.edu.tw/cjlin/libsvm).

[14] Khan, K., Attique, M., Syed, I. and Gul, A. (2019). Automatic Gender Classification through Face Segmentation. Symmetry, 11(6), p.770.

[15] Khan, K.; Mauro, M.; Leonardi, R. Multi-class semantic segmentation of faces. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Quebec, QC, Canada, 27–30 September2015; pp. 827–831.

[16] Lafferty, J.; McCallum, A.; Pereira, F.C. Conditional random fields: Probabilistic models for segmenting and labelling sequence data. In Proceedings of the 18th International Conference on Machine Learning 2001(ICML 2001), Williamstown, MA, USA, 28 June–1 July 2001.

[17] Van den Bergh, M.; Boix, X.; Roig, G.; Van Gool, L. Seeds: Superpixels extracted via energy-driven sampling. Int. J. Comput. Vis. 2015, 111, 298–314. [CrossRef]

[18] Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. Int. Conf. Comput. Vis. Pattern Recognit. 2005, 1, 886–893.

[19] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M. (2019). *FaceForensics++: Learning to Detect Manipulated Facial Images*. [Online] arXiv.org. Available at: https://arxiv.org/abs/1901.08971 [Accessed 12 Nov. 2019].

[20] DeepFakes github. https://github.com/deepfakes/faceswap. Accessed: 2018-10-29.

[21] Justus Thies, Michael Zollhofer, Marc Stamminger, Chris-tian Theobalt, and Matthias Nießner. Face2Face: Real-TimeFace Capture and Reenactment of RGB Videos. In IEEEConference on Computer Vision and Pattern Recognition, pages 2387–2395, June 2016.

[22] FaceSwap. https://github.com/MarekKowalski/FaceSwap/. Accessed: 2018-10-29.
[23] Justus Thies, Michael Zollhofer, and Matthias Nießner. De-ferred neural rendering: Image synthesis using neural textures. ACM Transactions on Graphics 2019 (TOG), 2019.
[24] Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection. In ACM Workshop on Information Hiding and Multimedia Security, pages 1–6, 2017.
[25] Belhassen Bayar and Matthew C. Stamm. A deep learning approach to universal image manipulation detection using anew convolutional layer. In ACM Workshop on Information Hiding and Multimedia Security, pages 5–10, 2016.
[26] Nicolas Rahmouni, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Distinguishing computer graphics from natural images using convolution neural networks. In IEEE Workshop on Information Forensics and Security, pages 1–6, 2017.
[27] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Mesonet: a compact facial video forgery detection network. arXiv preprint arXiv:1809.00888, 2018.
[28] Francois Chollet. Xception: Deep Learning with Depth wise Separable Convolutions. In IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[29] Ramachandran Nair, R. (2019). Automated Detection of Gender from Face Images. *International Research Journal of Engineering and Technology (IRJET)*, : 06.
[30] Dwivedi N., Singh D.K. (2019) Review of Deep Learning Techniques for Gender Classification in Images. In: Yadav N., Yadav A., Bansal J., Deep K., Kim J. (eds) Harmony Search and Nature Inspired Optimization Algorithms. Advances in Intelligent Systems and Computing, vol 741. Springer, Singapore.
[31] El-Alfy, E. and Binsaadoon, A. (2018). Automated gait-based gender identification using fuzzy local binary patterns with tuned parameters. Journal of Ambient Intelligence and Humanized Computing, 10(7), pp.2495-2504.
[32] Han J, Bhanu B (2006) Individual recognition using gait energy image IEEE Trans Pattern Anal Mach Intell 28(2):316–322.
[33] Iakovidis DK, Keramidas EG, Maroulis D (2008) Fuzzy local binary patterns for ultrasound texture characterization. In: Campilho A, Kamel M (eds) Image analysis and recognition, ICIAR 2008, Lecture notes in computer science (LNCS), vol 5112. Springer, Berlin, Heidelberg, pp 750–759.
[34] Chang CC, CJ Lin (2011) LIBSVM: a library for support vector machines. ACM Trans IntellSystTechnol 2:27:1–27:27.
[35] Zheng S, Zhang J, Huang K, He R, Tan T (2011) Robust view transformation model for gait recognition. In: International conference on image processing (ICIP), Brussels, Belgium.
[36] Yaman, D., Eyiokur, F., Sezgin, N. and Ekenel, H. (2019). Age and Gender Classification From Ear Images. [online] arXiv.org. Available at: https://arxiv.org/abs/1806.05742v1 [Accessed 12 Nov. 2019].
[37] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multipie," Image and Vision Computing, vol. 28, no. 5, pp. 807–813, 2010.
[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.
[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprintarXiv:1409.1556, 2014.
[40] [40] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
[41] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with50x fewer parameters and¡ 0.5 mb model size," arXiv preprintarXiv:1602.07360, 2016.
[42] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma,Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," arXiv preprintarXiv:1409.0575, 2014.
[43] G. Azzopardi and N. Azzopardi, ''Trainable COSFIRE filters for key-point detection and pattern recognition,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 2, pp. 490–503, Feb. 2013.
[44] H. Bay, T. Tuytelaars, and L. Van Gool, ''SURF: Speeded up robust features,'' in Proc. Eur. Conf. Comput. Vis., 2006, pp. 404–417.
[45] Mäkinen, Erno & Raisamo, Roope. (2008). Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces. IEEE transactions on pattern analysis and machine intelligence. 30. 541-7. 10.1109/TPAMI.2007.70800.