

Recent Trends in Face Detection Algorithm

Nivedha N

Department of CS&E

P. A. C. E., Mangaluru, India

Prof. Ganesh Pai

Department of CS&E

P. A. C. E., Mangaluru, India

Dr. M Sharmila Kumari

Department of CS&E

P. A. C. E., Mangaluru, India

Abstract - Face Detection has been a major field of study in facial analysis as it is being the preliminary phase in face processing. Face detection concerns finding whether or not there are any faces in a given image and if present, get the location and contents of each face. This paper reviews on some of the well-known face detection algorithms highlighting their merits and weaknesses.

Keywords - Face detection, Face Analysis, Face verification, Face Recognition

I. INTRODUCTION

Face detection is the first phase in face analysis. Face detection concerns finding whether or not there are any faces in a given image. If present, get the location and contents of each face. The detected face goes as input to various facial analysis processes for further processing whose outcome can be used in machine learning and computer vision.

Face analysis is concerned with processing of a human face in a given image. Face analysis comprises of face detection [1], face recognition and verification [2], face tracking for surveillance [3], facial behavior analysis [4], facial attribute recognition [5] (i.e., gender/age recognition [6] and assessment of beauty [7]), face relighting and morphing [8], facial shape reconstruction [9], face alignment and many more. It may be a trivial task for humans to detect a face in a given image that humans do effortlessly, but it is an extremely tough and tedious task for a computer.

Computer is given with a digital image as a data for processing and it has to determine the existence of a face and if exists, determine its location and retrieve it. But computer on its own cannot recognize human faces. It needs an algorithm that searches for a human face through the facial features or it needs to be trained through some set of training data to recognize a face.

But the problem here is that the image taken into consideration might have come from any source and the picture may be captured in unconstrained conditions. The human face has a high degree of variability in its appearance like, the face(s) in the image may vary in size, location, pose, quality, distance, alignment, lighting conditions, overlapping with other faces or objects, facial expression etc. Such unconstrained conditions make the accurate detection of face a challenging and tedious task. Several algorithms have been developed till date that tries to detect face like Viola-Jones, Local Binary Pattern, AdaBoost algorithm, SMQT Features and SNOW classifier, Neural Network based face detection etc. Each can detect faces but with certain constraints or drawbacks.

Automatic face detection is the key element in face analysis that has attracted tremendous applications. Most of the commercial digital cameras and mobile phones are embedded with face detector used to help auto-focusing. Social networking websites, such as Facebook, Instagram, use face detection mechanisms to tag a person. Image processing and its related subject of face analysis or in general object analysis overlaps with computer vision which fundamentally aims to duplicate human vision. It is a field that includes methods for acquiring, processing, analyzing and understanding images and in general high dimensional data from the real world in order to produce numerical or symbolic information. The result can be used in decision making. Face detection and recognition has found tremendous application in security systems, robotics, forensics, defense to name a few.

II. LITERATURE SURVEY

The essential first-step in face recognition systems is Face Detection, and it is used to detect faces present in an image or video, with the purpose of localizing and extracting the face region from the background. In the literature we find a wide variety of techniques proposed, ranging from simple edge-based algorithms to composite high-level methodology utilizing advanced pattern recognition methods. Certain algorithms use new innovative techniques and some as extensions with improvements to the existing techniques by addressing the issues. Innovative techniques to solve existing problems have opened the doors for more scope in the solution domain. Innovative solutions are necessary because of the upper bound that the researchers have reached on the solution that can be produced by the existing algorithms. Here we highlight on some of the major works done in the past on face detection.

Techniques that were used earlier for face detection were based on certain assumptions such as plain background, frontal face. To these systems, any change of image conditions would mean fine-tuning, if not a complete redesign. But image source available today are taken in unconstrained environments. Face detection algorithms developed in all these years are categorized into different categories based on how the problem is addressed. Time to time various survey papers classify these algorithms based on the techniques and utilization of face knowledge.

Face detection algorithms are classified as *feature-based approach* and *image-based approach* [10]. Feature-based approach explicitly uses face knowledge and follow the classic detection methodology where low level features are

derived. Properties of the face such as skin color and face geometry are exploited at different system levels. Here, face detection task is accomplished by manipulating distance, angles, and area measurements of the visual features derived from the scene. As face features are the main ingredients in this approach, these techniques are termed as feature-based approach. Feature-based approach can be further classified as *Low-level analysis*, *Feature analysis* and *Active shape models*. With cluttered scenes, low-level analysis first deals with segmentation of visual features using pixel properties such as gray-scale, color and edges. Features generated from this analysis are ambiguous due to the low-level nature. In feature analysis, using information of face geometry, visual features are organized into a more global concept of face and facial features. Through feature analysis, feature ambiguities are brought down and locations of the face and facial features are extracted. Active shape models were developed for the purpose of complex and non-rigid feature extraction such as eye pupil and lip tracking. In spite of the above techniques, the determination has been uncertain by the unpredictability of face appearance and environmental conditions. Even though certain feature-based attempts have improved ability to cope with unpredictability, many are still limited to head, shoulder and frontal faces. Therefore there is a necessity of techniques that can perform detection in more hostile environment such as detecting multiple faces with clutter-intensive backgrounds.

Image-based representations of faces, such as 2D intensity arrays, are classified into a face group through training algorithms without feature derivation and analysis. These techniques take into account the face knowledge implicitly into the system through mapping and training schemes. Most of the image-based approaches apply a window scanning technique for detecting faces. The window scanning algorithm is basically an exhaustive search of the input image for possible face locations at all scales, but variations do exist in the implementation of these algorithms for almost all the image-based systems. Normally, the sub-sampling rate, size of the scanning window, the number of iterations and the step size vary depending on the method proposed and the need for a computationally efficient system. Image-based representations are further classified as *linear subspace method*, *neural network* and *statistical approaches*. Linear subspace method processes with the fact that human face fall in a subspace of the overall image space. Some of the techniques developed are principal component analysis (PCA), linear discriminant analysis (LDA) and factor analysis (FA). Neural network based approaches are efficient in detection and recognition of two-dimensional images of the object or face for matching. It allows the detection of upright, tilted, non-frontal faces or faces that vary considerably with lighting, occlusion and facial expression in a cluttered image. Statistical approaches of face detection make use of probability theory and decision theory to develop an algorithm. Some of the standard techniques used to develop algorithm are information theory, support vector machine (SVM) and Bayes' decision rule.

Principal component analysis is a statistical procedure and uses orthogonal transformation to convert a set of correlated

variables into a set of values of linearly uncorrelated variables called *principal component*. The number of principal component is not greater than the number of original variables. The transformation is such that, the first principal component has the largest possible variance and each succeeding component has the highest variance possible with the constraint that it is orthogonal to the preceding component. The resulting vectors are an uncorrelated orthogonal basis set. The principal components are orthogonal as they are the eigenvector of the covariance matrix. Using PCA, a technique was developed to efficiently represent human faces by Sirovich and Kirby [27]. Given a subset of different face images, the technique first finds the principal components of the distribution of faces, expressed in terms of eigenvectors. Each individual face in the face set can then be approximated by a linear combination of the largest eigenvectors, more commonly referred to as eigenfaces, using appropriate weights. In [12], new technique for Face Detection based on Viola and Jones algorithm and principal component analysis has been proposed. They have shown simulation results for the proposed technique and established that proposed technique is performing better than the existing one. [13] explains Robust Principal Component Analysis (RPCA). It describes a robust M-estimation algorithm for learning linear multivariate representations of high dimensional data such as images. Quantitative comparisons with traditional PCA and previous robust algorithms illustrate the benefits of RPCA when outliers are present.

Linear Discriminant Analysis (LDA) is used to perform dimensionality reduction by projecting the input data to a linear subspace consisting of the directions which maximize the separation between the classes. Studies show that high classification accuracies have been achieved by LDA regularized by PCA i.e. above 90% on average for single fMRI volumes acquired 2 s apart during a 300 s movie (chance level $0.7\% = 2 \text{ s}/300 \text{ s}$). The largest source of classification errors were autocorrelations in the BOLD signal compounded by the similarity of consecutive stimuli [14].

Factor analysis is a multivariate statistical technique equivalent to PCA, but FA assumes that the observed data samples come from a well-defined model [15]. $\mathbf{x} = \mathbf{A}\mathbf{f} + \mathbf{u} + \mu$, where \mathbf{A} is a matrix of constants, \mathbf{f} and \mathbf{u} are random vectors, and μ is the mean. Factor analysis seeks to find \mathbf{A} and the covariance matrix of \mathbf{u} which best models the covariance structure of \mathbf{x} . If the specific variances \mathbf{u} are assumed to be 0, the procedure for FA can be equivalent to PCA.

Neural Network Approach: The first neural approaches to face detection were based on MLPs [17, 18, 19], where promising results were reported on fairly simple datasets. The first advanced neural approach which reported results on a large, difficult dataset was by Rowley *et al.* [20]. It incorporates face knowledge in a retinally connected neural network. The neural network is designed to look at windows of size 20×20 pixels. There is one hidden layer with 26 units, where 4 units look at 10×10 pixel sub-regions, 16 look at 5×5 sub-regions, and 6 look at 20×5 pixels overlapping horizontal stripes. The input window is pre-processed through lighting correction and histogram

equalization. A problem with window scanning techniques is overlapping detections. Rowley *et al.* handles the problem through two heuristic methods:

1. Thresholding: the number of detections in a small neighborhood surrounding the current location is counted, and if it is above a certain threshold, a face is present at this location.
2. Overlap elimination: when a region is classified as a face according to thresholds, then overlapping detections are likely to be false positives and thus are rejected.

In [21], an approach based on a neural network model, called constrained generative model (CGM), was proposed. CGM is an auto-associative, fully connected multilayer perceptron (MLP) with three large layers of weights, trained to perform nonlinear dimensionality reduction in order to build a generative model for faces. Multi-view face detection was achieved by measuring the reconstruction errors of multiple CGMs, combined via a conditional mixture and an MLP gate network. [22] proposed a face detection scheme based on a convolutional neural architecture. Compared to the traditional feature-based approaches, CNN automatically derives problem-specific feature extractors from the training examples without making any assumptions about the features to extract or areas of the face patterns to analyze. Deep convolutional neural networks (DCNN) showed remarkable performance for object categorization [23]. [30] proposed fast face detection method based on discriminative complete features (DCFs) extracted by an elaborately designed convolutional neural network, where face detection is directly performed on the complete feature maps. DCFs have shown the ability of scale invariance, which is beneficial for face detection with high speed and promising performance.

Statistical approach: It makes use of probability theory and decision theory to develop an algorithm. Some of the standard techniques used to develop algorithm are information theory, support vector machine (SVM) and Bayes' decision rule. Information theory approach uses the histogram of the face components and probability density function for the class of face and non-face structures. As mentioned in [10], SVM with 2nd degree polynomial as a kernel function is trained with a decomposition algorithm which guarantees global optimality. Training is performed with the boot-strap learning algorithm, and the images are pre-processed. SVM algorithms are used in real-time tracking and analysis of faces. Here, SVM algorithm was applied on segmented skin regions in the input images to avoid exhaustive scanning. Recently, face detection using Gabor feature extraction and support vector machines were developed.

Local Binary Pattern (LBP): This technique is very effective to describe the image texture features. LBP has advantages such as high-speed computation and rotation invariance, which facilitates the broad usage in the fields of image retrieval, texture examination, face recognition, image segmentation, etc. In [26], LBP was successfully applied to detection of moving objects via background subtraction. In

LBP, every pixel is assigned a texture value, which can be naturally combined with target for tracking thermo graphic and monochromatic video. The major uniform LBP patterns are used to recognize the key points in the target region and then form a mask for joint color-texture feature selection. LBP is tolerant against monotonic illumination changes and is computationally simple but using larger local regions increases the errors and it is inefficient for non-monotonic illumination changes.

AdaBoost (Adaptive Boost) is a machine learning approach based on the idea of creating highly accurate prediction rule by combining many relatively weak and incorrect rules. The AdaBoost algorithm in [25] was the first practical boosting algorithm, and one of the most widely used and studied. Boosting algorithm is used to train a classifier capable of processing images rapidly while having high detection rates. It is a learning algorithm which produces a strong classifier by choosing visual features in a family of weak classifiers and combining them linearly. It is sensitive to noisy data and outliers. It's called adaptive because it uses multiple iterations to generate a single composite strong learner. It creates the strong learner, i.e. a classifier that is well-correlated to the true classifier, by iteratively adding weak learners, i.e. a classifier that is only slightly correlated to the true classifier. Throughout each round of training, a new weak learner is added to the group and a weighting vector is adjusted to focus on examples that were misclassified in preceding rounds. The outcome is a classifier that has higher accuracy than the weak learner's classifiers.

The advantage of this algorithm is that it needs no prior knowledge. It only needs two inputs: a training dataset and a set of features. The training errors theoretically converge exponentially towards 0. It is very simple to implement and does feature selection resulting in comparatively simple classifier. But the result depends on the data and weak classifiers. The quality of the final detection depends highly on the consistence of the training set. It is sensitive to noisy data and outlier, quite slow training. Weak classifiers of higher complexity leads to over-fitting and weak complexity can lead to low margins.

Viola-Jones: is an object detection framework [24] and is the first object detection framework to provide competitive object detection rates in real-time. In 2001, Paul Viola and Michael Jones used this framework for face detection. This face detection framework is capable of processing images extremely rapidly while achieving high detection rates.

The three key phases:

1. *Integral Image* allows the features used by the detector to compute a summed area table necessary for quick calculations.
2. *AdaBoost learning algorithm* is a classifier creates a small set of only the best features to create more efficient classifiers.
3. *Cascaded Classifiers* allows discarding negative window early to focus more computational time on possible positive window and reduces the overall number of computations.

Its advantage is its high detection speed at real time with high detection accuracy with a very low false positive rate but it is limited in head poses, takes extremely long training time and does not detect black faces.

SMQT Features and SNOW classifier: SMQT stands for Successive Mean Quantization Transform and SNOW stands for Sparse Network of Winnows. This technique consists of two phase. The primary phase is face luminance. The operation of this phase is being performed to get pixel information of an image and further implemented to detection purpose. The second phase is detection. In this phase, local SMQT features are used as feature extraction for object detection. The features were found to be able to cope with illumination and sensor variation in object detection. The split up SNOW is proposed to speed up the standard SNOW classifier. It requires just training of one classifier network which can be arbitrarily divided into several weaker classifiers in cascade. All weak classifier uses the result from previous weaker classifiers which makes it computationally efficient [13].

FPGA based implementation: Peter Irgenset. al., [27] have developed an efficient and cost effective FPGA based implementation of the Viola-Jones face detection algorithm. The proposed system can be applied in a video surveillance and tracking application and was tested for the data received from a camera. This implementation is a complete system level hardware design described in a hardware description language and validated on the affordable DE2-115 evaluation board. The primary objective was to study the achievable performance with a low-end FPGA chip based implementation. Their proposed algorithm achieves a performance of 4.4fps for image sizes of 320×240 pixels. For larger image sizes and larger number of faces, the FPGA design potentially offers better performance since the FPGA design can execute in parallel more computations by exploiting the hardware parallelism offered by the FPGAs.

Fast Face Detection proposed in [28] aim to detect faces in violence scenes, in order to help the security control. The authors have used Violent Flow (ViF) descriptor with Horn-Schunck for violence scenes detection at first stage. Then non-adaptive interpolation super resolution algorithm is applied to improve the video quality and finalize with a Kanade-Lucas-Tomasi (KLT) face detector. In order to get a very low time processing, super resolution and face detector algorithms with CUDA are parallelized. Experiments are done using Boss Dataset and also a new violence dataset has been built, taking scenes from surveillance cameras. Success rate of up to 97% was achieved in certain input videos.

In [29] [1. [28] , a new type of feature, called Normalized Pixel Difference (NPD) is proposed, which is efficient to compute and has several desirable properties, including scale invariance, boundedness and enabling reconstruction of the original image. Deep quadratic tree learner is used to learn and combine an optimal subset of NPD features to boost their ability to discriminate. Compared to the absolute difference approach, NPD is invariant to scale change of the pixel

intensities. It uses AdaBoost algorithm to select the most discriminative features and construct single strong classifiers. NDP is advantageous as there is no need to label the pose of each face image manually or cluster the poses before training the detector. In the learning process, the algorithm automatically divides the whole face manifold into several sub-manifolds by the deep quadratic trees. NPD detector is also efficient and faster when it is compared with Viola-Jones face detector. On the other hand, NPD face detection algorithms performance is less when it is used on low image resolutions. NPD2 uses same features of NPD but pixel intensities are slightly modified. Performance NPD2 is considerably low when it is compared with NPD.

III. CONCLUSION

Face detection has been today applied and used in diverse application. Various techniques have been developed and implemented for various applications in diverse fields. The fundamental limits are unconstrained environment and processing power. Algorithms developed tries to address these two major factor that affects the overall performance of the detection algorithm. The aim of having high detection rates with low processing power has been the ultimate challenge. There has been many hardware based designs, software based designs, neural network based designs and many more to obtain optimized results. Some have proved to be optimal compared to some of the existing systems but none could be called as the best till date. The effort will continues in the research community, awaiting to get a complete solution to the problem under consideration.

REFERENCES

- [1] Stefanos Zafeiriou, Cha Zhang and Zhengyou Zhang, A Survey on Face Detection in the wild: past, present and future, Elsevier, 2015
- [2] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, Face recognition: A literature survey, *Acm Computing Surveys (CSUR)* 35 (4) (2003) 399–458.
- [3] Z. Kalal, K. Mikolajczyk, J. Matas, Face-tld: Tracking-learning-detection applied to faces, in: *Image Processing (ICIP)*, 2010 17th IEEE International Conference on, IEEE, 2010, pp. 3789–3792.
- [4] M. Pantic, L. J. M. Rothkrantz, Automatic analysis of facial expressions: The state of the art, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* 22 (12) (2000) 1424–1445.
- [5] N. Kumar, A. C. Berg, P. N. Belhumeur, S. K. Nayar, Attribute and simile classifiers for face verification, *IEEE 12th International Conference on Computer Vision*, 2009, pp. 365–372.
- [6] Y. Fu, G. Guo, T. S. Huang, Age synthesis and estimation via faces: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 (11) (2010) 1955–1976.
- [7] A. Laurentini, A. Bottino, Computer analysis of face beauty: A survey, *Computer Vision and Image Understanding*.
- [8] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, D. Samaras, Face relighting from a single image under arbitrary unknown lighting conditions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31 (11) (2009) 1968–1984.
- [9] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces, *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.
- [10] Erik Hjelm and Boon Kee Low, Face Detection: A Survey, *Computer Vision and Image Understanding* 83, 236–274 (2001)
- [11] Samiksha Agrawal, Pallavi Khatri, Facial Expression Detection Techniques: Based on Viola and Jones Algorithm and Principal Component Analysis, *Fifth International Conference on Advanced Computing & Communication Technologies*, 21-22 Feb 2015, pp. 108-112, ISSN 2327-0632

- [12] Fernando De la Torre, Michael J. Blacky, Robust Principal Component Analysis for Computer Vision, Int. Conf. on Computer Vision (ICCV'2001), Vancouver, Canada, July 2001. IEEE 2001.
- [13] Hendrik Mandelkow, Jacco A. de Zwart, and Jeff H. Duyn, "Linear Discriminant Analysis Achieves High Classification Accuracy for the BOLD fMRI Response to Naturalistic Movie Stimuli", Journal of Frontiers in Human Neuroscience, v 10, March 2016, Article 128.
- [14] K. V. Mardia, J. T. Kent, and J. M. Bibby, Multivariate Analysis, Academic Press, San Diego, 1979.
- [15] Varsha Gupta, Dipesh Sharma, A Study of Various Face Detection Methods, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 5, May 2014, pp6694-6697
- [16] G. Burel and D. Carel, Detection and localization of faces on digital images, Pattern Recog. Lett. 15, 1994, 963-967.
- [17] P. Juell and R. Marsh, A hierarchical neural network for human face detection, Pattern Recog. 29, 1996, 781-787.
- [18] M. Propp and A. Samal, Artificial neural network architecture for human face detection, Intell. Eng. Systems Artificial Neural Networks 2, 1992, 535-540.
- [19] H. A. Rowley, S. Baluja, and T. Kanade, Neural network-based face detection, IEEE Trans. Pattern Analysis and Machine Intelligence. 20, January 1998, 23-38.
- [20] R. F'eraud, O. J. Bernier, J.-E. Viallet, M. Collobert, A fast and accurate face detector based on neural networks, IEEE Trans. on PAMI 23 (1) (2001) 42-53.
- [21] C. Garcia, M. Delakis, Convolutional face finder: A neural architecture for fast and robust face detection, IEEE Trans. on PAMI 26 (11) (2004) 1408-1423.
- [22] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097-1105.
- [23] P. Viola and M. Jones, —Rapid object detection using a boosted cascade of simple features, in Proc. of CVPR, 2001.
- [24] R. Meir and G. R'atsch. —An introduction to boosting and leveraging. S. Mendelson and A. J. Smola Ed., Advanced Lectures on Machine Learning, Springer-Verlag Berlin Heidelberg, pages 118-183, 2003.
- [25] Ahonen T, Hadid A, Pietika'inen M, 'Face recognition with local binary patterns. In Proceeding of European conference on computer vision (ECCV2004), LNCS 3021, pp 469-481, 2004
- [26] L. Sirovich and M. Kirby, Low-dimensional procedure for the characterization of human faces, J. Opt. Soc. Amer. 4, 1987, 519-524.
- [27] Peter Irgenset.al., An efficient and cost effective FPGA based implementation of the Viola-Jones face detection algorithm, HardwareX 1 (2017) 68-75, Elsevier, 2017.
- [28] V. E. MachacaArceda et.al., Fast Face Detection in Violent Video Scenes, Electronic Notes in Theoretical Computer Science 329 (2016) 5-26, Elsevier, Science Direct.
- [29] Shengcai Liao, Anil K. Jain, and Stan Z. Li. "A Fast and Accurate Unconstrained Face Detector: IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. pp.723,727, February 2016
- [30] Guanjun Guo, Hanzi Wang, Yan Yan, Jin Zheng, Bo Li, A Fast Face Detection Method via Convolutional Neural Network, arXiv:1803.10103v1, 2018