Sukaran Golani et al. International Journal of Recent Research Aspects ISSN: 2349-7688,
Vol. 5, Issue 1, March 2018, pp. 169-172

Technical Survey of Face Detection Techniques

Sukaran Golani[1], Shubham Jain[2], Suraj Sahal[3], Tanvi Dhingra[4]

Department of Information Technology, Northern India Engineering College, New Delhi
[1]sukarang9@gmail.com
[2]sj100197@gmail.com
[3]sahalsuraj007@gmail.com
[4]tanvi.dhingra@niecdelhi.ac.in

***Abstract* - This paper provides a concise analysis of some of the well known and important techniques used for face detection. Face detection is a term which involves the computer system to detect faces in a digital image. A common issue which the face detection algorithms are the environmental and real time noise occurrence in the images clicked through the cameras or the mobile device. So various methodologies have been discussed like the Viola Jones method, Gabor Wavelet Transformation and the multi-task cascaded convolutional networks (MTCNN), etc. so as to deal with the above mentioned disadvantage. Also a concluding remark has been made about their efficiency.**

Keywords - Face detection, Neural network, Gabor Wavelet, Viola Jones.

I. INTRODUCTION

Face detection is referred as the technique of identifying faces in any given scene. The methodologies that have been mentioned deals with the frontal as well as side view of human faces. Nowadays, the necessity of keeping a vigilance in terms of security aspects is quite likely much applied in daily life. Right from the areas of banking, corporate, and other public amenities like airports, railway stations, etc. are in favour of face detection technology. Face detection has found various applications in areas such as capturing of image, coding of video and video conferencing [1]. It can also be utilized for locating or indexing face data from videos or images containing faces of various positions, sizes, and backgrounds. Machines are developed by humans and need the data to be fed so as to follow our instructions. There are various challenges while detecting faces like the presence of noise, occlusions and the illumination issues. If any scene or a two dimensional image is mentioned, the aim is to find faces in it. This paper is motivated from the various works of the famous authors. They have done a remarkable work by providing the upcoming detection techniques for finding faces. We have also tried to do the same by giving a survey on some important and recent techniques. Face detection techniques can be segregated into four categories as follows [10]:

1. Knowledge-based procedure: They are established on human understanding of the common human face configuration and arrangement of facial features. Taking benefit of natural face alignment and the unprocessed left-to-right and top-to-botttom order in which attributes appear in the human face, these methodologies find similarities to relate the size, shape, texture and other detailing of facial features (such as nose, eyes, chin, eyebrows) and association between them (relative distance and positions).
2. Feature invariant methodology: These focuses to find systematic features that prevail even when the lighting or viewpoint conditions differ and then involve these to locate faces. Various structural features are being used: facial local features, shape, skin color and texture. Local characteristics such as nose, eyebrows, eyes, and mouth are found using multifarious-resolution or derivative filters, detectors of edge and morphological behaviour.
3. Template-based approach: To find a face in a new photo, first the head is chalked out, which is uniformly roughly elliptical in shape and is detected using filters and edge detectors. Then the outline of local facial features are extricated in the similar way, utilizing knowledge of feature and face geometry. Lastly, the relationship between features derived from the source image and existing stored framework of face and facial features is calculated to determine whether there is face available in the image.
4. Appearance-based methods: While template-copying methods depend on an already defined model, appearance-based approach use huge numbers of examples (images of facial and \ or face features) highlighting many variations (face shape, eye color, skin color, closed\open mouth, etc). Face detection can be defined as a pattern categorization problem with two classes: “non-face” and “face”. The “non-face” class consists of images that may show anything that is not a face, whereas the “face” class has all face images.

II. ALGORITHMS

A. *Viola-Jones Method On CCTV Camera Using Embedded System*

© 2018 IJRAA All Rights Reserved

The Viola-Jones[4] face detection method is used to differentiate faces based on simplified values of features. There are several reasons so as to make use of the feature in comparison with the pixel directly. The most prevalent reason is that characteristics can be used to format the knowledge of temporary domains that are hard to learn about the finite training data. The second most useful reason for using features based on the operating system is that they are quite rapid than pixel-based systems. Figure 1(a),(b),(c) and (d) show the pixel value computation of an image. The procedure used in this research is to extract the face of an operational object that is captured by a camera which aims to improvise the efficiency of security related systems in a closed room. [3]

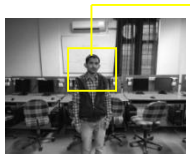

Fig 1a) Initial Picture

Fig1b) Magnified Picture

4	2	6	1
3	7	3	2
8	2	8	9
7	5	7	9

Fig1c) Feature's Pixel Values

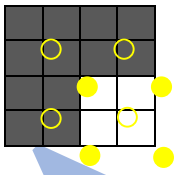

Fig 1d) Pixels in chosen areas

Figure 2 depicts the flow chart of the total system , the nascent stage starts with the collection of a video image sourced from the recording of the camera, which is then executed in the management of the image to change to be converted into a binary form i.e black and white. Next, the detection process is implemented using Viola–Jones is done. This process is repeated on loop till the full facial data with respect to it's position is made available. On the occurrence of a human face , the said system will quickly capture the detected image and store it in jpg format and immediately send it to the server for further steps to be performed by the mobile application.[3]

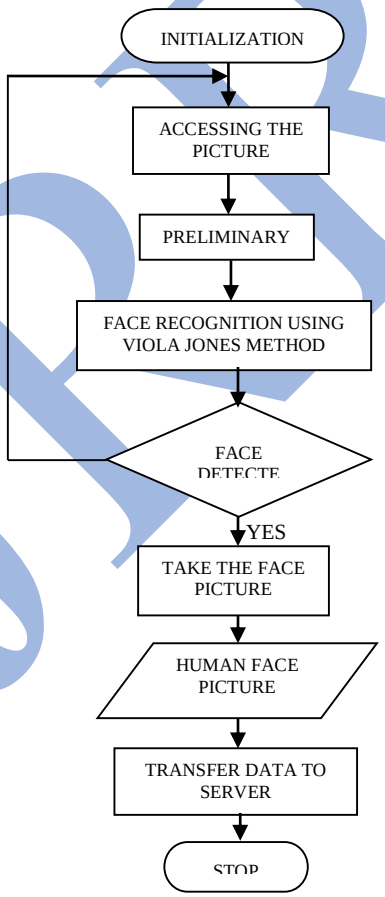

Fig.2 Face Detection flow chart

Fig. 3(a) Web Camera

Fig.3(b) Raspberry Pi 3B

Figure 3(a) and (b) shows a brief insight of the system to be made with required main equipments:

© 2018 IJRAA All Rights Reserved

1. An ordinary Camera 2. Raspberry Pi hardware 3. A Server

B. Multi –task Cascaded Convolution Networks

The MTCNN[5], has derived relationship between the face detection and it's alignment to increase up their performance. It mainly consists of following parts: (1) a Proposed Network (P-Net) for formulating a list of the candidate windows. Then, we use this information for distinguishing between face and non-face and find a tentative enclosed box for regression vectors to find location of face, and further merging is done with the non-maximum suppression (NMS) candidate. (2) A Refine Network (R-Net), in comparison with P-Net, does not only obtain the proposed regions but discards a lot of false candidates. (3)Third component which is similar to R-Net, is called O-Net. In this particular network, it considers the output of five facial landmarks' positions. As shown in figure 4, The training phase involves three tasks: classification of face, bounding of box regression, and classification of facial emotions. The calculative functions of face classification and regression of bounding box are the same as with MTCNN. By making use of the cross-entropy loss which is required to solve the face classification as well as the Euclidean loss for each of the individual sample.

$$Z_j^{rc} = -\left(a_j^{rc}\log\left(q_j\right) + \left(1 - a_j^{rc}\right)\left(1 - \log\left(q_j\right)\right)\right) \quad \ldots\ldots (1)$$

$$Z_j^{grid} = \| \hat{a}_j^{grid} - a_j^{grid} \| \quad \ldots\ldots (2)$$

Equation (1) mentioned above shows the cost function of face classification, where q_j symbol is the probability that is produced by the network that confirms the sample being a face. The notation $a_j^{rc} \epsilon \{0,1\}$further denotes the ground-truth image. [5] Equation (2) is constructed as a regression problem, where $\hat{a}_j^{grid}$ regression target is obtained from the network and $a_j^{grid} \epsilon$ R is the ground-truth value for the coordinate. There are four of the coordinates, including the left, top, right and bottom. Infact, it is similar to the face detection task, where face emotions are further classified as a seven-classification problem, we also make use of the cross-entropy loss:

$$Z_j^{lbl} = -(a_j^{lbl} \log(q_j) + (1 - a_j^{lbl})(1 - \log(q_j))) \quad \ldots(3)$$

where $a_j^{lbl} \epsilon \{0,1,2,3,4,5,6\}$denotes the label of the ground-truth. The label q_j is the probability that is produced by the network which indicates a sample consisting of one emotion in the facial expressions. Label $\boldsymbol{q_j}$ is the probability which is produced by the network that confirms it being a sample having only one emotion on the facial expressions. Equation (3) depicts the multi-source formulae. [5]

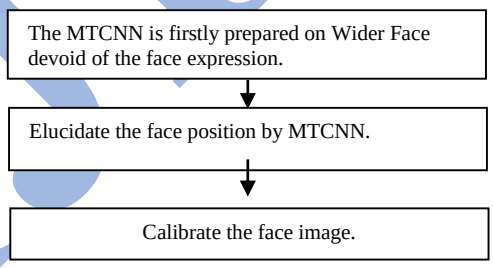

Fig 4 Flowchart of MTCNN training procedure

C. Gabor Wavelet Transformation

This paper involves a method which has been formed so as to detect human faces and the corresponding gray levels within a camera picture. The formulated method involves two parts. The first part consists of finding windows that includes the possibility of detecting face at a higher success rate. In order to find the middle coordinate of the window, which are also termed as face inaccurate centers, the histogram of an image captured by the camera is convolved with that of an image of a human face. The peak obtained is considered as the face centers which are incorrect such that the chance of a human face existence in such areas is much greater than the rest of other regions. On investigating the last part in this incorrect center and a bounded box that is inclined to the center and instance face dimensions will be studied. In the second part, by finding the product of Fourier transform of forty Gabor wavelet cells in the fast fourier transform[2] for each of the face inconsistent center, then matrix of Gabor features is found out and after calculating a normalized Gabor feature vector [11] it is fed back into the neural network, it is trained with a large number of face and non-face images. In this study, the neural

© 2018 IJRAA All Rights Reserved

network recognizes inconsistent center as the accurate center of the face or non-face image by mapping input pattern and trained patterns. It should be kept as a fact that the advantage of Gabor features involves resistance towards the intensity increase method and against ambient light variations.[6, 11]

D. Convolution-Net for Face detection:

Convolutional neural network[7] is an exceptional family of neural network which are designed so as to specially handle the image or two dimensional spatial data. The framework of convolutional network is motivated from the human visual and perception system. ConvNet[7] is formed mainly of three different kinds of layers like the most important 'Convolutional layers', an additional and modifiable 'Pooling layers' and a 'Fully Connected layer'. The fully connected layer can exceptionally be an entire multi-layer perceptron (MLP)[9]. But the base model described is not the only one to be used. As per different application requirements, several structures are used for face detection or for localization of object[7]. The network consisted only of the convolutional layers and additional pooling layers. The target of the network should be so as to produce the output shown in figure 5. Since the required output of the network has to be a binary mask, the last layer is to be comprises of sigmoid as a non-linearity function .In theory, if the described model is trained on data like in figure 5, several of the layers (mainly higher convolutional layers) will be learnt as face detectors.

$$Sigmoid(z) = \frac{1}{1+\exp(-z)} \in (0,1) \quad \text{.....(4)}$$

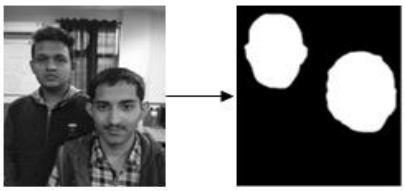

Fig. 5 Input and Output of Convolution Net Face Detection Technique

III. CONCLUSION

Face detection is the primary step of an automatic recognition system, since a face has to be present in the source image before it is identified. A denotation of face detection can be: given a snapshot, find all faces present in it (if any) and discover their precise size and positions. Generally, face detection sytem is a two-step method: first the entire image is scrutinized to detect regions that are located as "face". In this survey paper, we discussed both state-of-art as well as modern techniques for face detection. We analyzed the method of finding hand-crafted features and differentiated them along with a more new neural network based methodology. In future scope, we may perform some extensive working and a more laborious analysis of the work.

REFERENCES

[1] Hjelmas, Erik, Bon Kee Low:Face Detection: A Survey,Computer Vision and Image Understanding 83(3):236-274,2001

[2] Alejandro Ballado; Jose Lazaro; Glenn Avendano; Clarice An Rosette M. De Claro; Gabriel Kristofer Sandoval; Rez A. Viloria: Characterization using spectral signature of tropical plants through fast fourier transform, 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Pages:1-6,2017

[3] I Gusti Ngurah Made Kris Raya, Agung NugrohoJati, Randy Erfa Saputra, Analysis realization of Viola Jones method for face detection on CCTV camera based on embedded system, International Conference on Robotics, Biometrics and Intelligent Computational Systems (Robionetics), pages 1 -5, 2017

[4] Viola, Paul, M. Jones, "Robust object detection using a boosted cascade of simple features", Proc Cvpr1,Vol.1,pp.:I-511-I-518, 2001

[5] Jia Xiang; Gengming Zhu:Joint Face Detection and Facial Expression Recognition with MTCNN, 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Pages: 424 – 427, 2017

[6] Vahid Ghods, Mohammad Karim Sohrabi,Mehdi Joulaeian: Accurate detection of human face position in the environmental images using Gabor wavelet transformations, Computing Conference, Pages: 633 – 637, 2017

[7] SubhamMukherjee; SumalyaSaha; SounakLahiri; Ayan Das; Ayan Kumar Bhunia; Aishik Konwer; Arindam Chakraborty "Convolutional Neural Network based face detection", 1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech),Pages: 1 – 5,2017

[8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner: Gradient based learning applied to document recognition, Proceedings of the IEEE, 86(11):Pages:2278–2324, 1998

[9] Dabiah Ahmed Alboaneen, Huaglory Tianfield and Yan Zhang: Sentiment analysis via multi-layer perceptron trained by meta-heuristic optimisation, 2017 IEEE International Conference on Big Data (Big Data), Pages: 4630 – 4635,2017

[10] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," IEEE Trans. on PAMI,vol.24, pp. 34–58, Jan 2002.

[11] Gaurav Kumar, Pradeep Kumar Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems", IEEE 4th International Conference on Advanced Computing & Communication Technologies, pp. 5-12, Feb. 2014.

© 2018 IJRAA All Rights Reserved