SURVEY ON FACE DETECTION METHODS

Sharmeena Naido[1], Rosalyn R. Porle[2]

[1,2]*Faculty of Engineering, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.*

Abstract - *Face detection has attracted attention from many researchers due to its wide range of applications such as video surveillance, face recognition, object tracking and expression analysis. It consists of three stages which are preprocessing, feature extraction and classification. Firstly, preprocessing is the process of extracting regions from images or real-time web camera, which then acts as a face or non-face candidate images. Secondly, feature extraction involves segmenting the desired features from preprocessed images. Lastly, classification is a process of clustering extracted features based on certain criteria. In this paper, 15 papers published from year 2013 to 2018 are reviewed. In general, there are seven face detection methods which are Skin Colour Segmentation, Viola and Jones, Haar features, 3D-mean shift, Cascaded Head and Shoulder detection (CHSD), and Libfacedetection. The findings show that skin colour segmentation is the most popular method used for feature extraction with 88% to 98% detection rate. Unlike skin colour segmentation method, Viola and Jones method mostly comprise of face regions and other parts of human body with 80% to 90% detection rate. OpenCV, Python or MATLAB can be used to develop real-life face detection system.*

Key Words: **Face Detection; Haar features; Skin Colour Segmentation; Viola and Jones**

1.INTRODUCTION

In this 21st century, advancement in computer technology has facilitated interest in human- computer interaction (HCI) system. Face detection is a fundamental procedure for any HCI system. It can be considered as one of the most challenging areas of research yet applicable in various areas such as security, gesture analysis and biometrics. Variations in pose, lighting condition, occlusion, facial expression, orientation, scale and location are some of the challenges associated with face detection.

Face refers to an individual's front part of the head, which includes areas of the forehead to the chin. It is a vital part of the human body as it represents information such as expression and identity [1]. Detection here refers to the process of identifying the existence of individuals.

Four major categories of face detection methods are knowledge-based, appearance-based, feature-based and template matching [2]. Knowledge-based method contains set of rules to detect face depict by human knowledge, for instance, facial features include nose at the centre, mouth under the nose and a pair of symmetric eyes. Appearance-based method uses a face model based on analysis perform on a set of training samples. Feature-based method aims to detect faces by extracting facial features. Although this method is flexible and invariant to changes in pose and orientation, it is easily affected by noise, illumination, occlusion and other environmental factors. Template matching method correlates input and template images for face detection. This method is simple, but rotation, direction and size are some of the factors that can affect the effectiveness of one's system.

Over the years, there are numerous research done on face detection to improvise an existing face detection system or to propose a new algorithm for face detection. The primary purpose of this paper is to provide insight on some of the well-known methods used for the feature extraction and classification process.

2. REVIEW OF FACE DETECTION METHODS

2.1 Skin Colour Segmentation

The skin colour segmentation method involves separation of skin and non-skin pixels. RGB, YCbCr and CbCr are some of the colour space used for skin colour modelling.

Face detection using CbCr colour space in video is presented in [3], which aimed to have a negligible misclassification rate. This colour space was used for human skin colour is less dependent to brightness. Experimentation on the performance of face detection was carried out on 4000 images of Ytcelebrity, YouTube and FJU database. Images in the video comprise of only face region. The detection accuracy achieved was 98% and 95% ,whereas false rejection and acceptance rate obtained were approximately 10%.

Face detection using YCbCr colour space is presented in [4]. The YCbCr colour space was used for skin colour detection and segmentation because this colour space has been encoded in most video media. Experimentation on the performance of face detection was carried out on 150 images which comprise of only face region. 95 images were correctly detected, and 55 images were falsely detected due to low image quality or face size below than 32×32.

In [5], face detection using a combination of motion and skin colour segmentation is presented. YCbCr colour space was used to perform skin colour segmentation. Experimentation on the performance of face detection was carried out on seven different ChokePoint video datasets, each consists of 200 frames. Images in the video comprise of face region and other parts of the human body. Frame size 600×800 with 3.16 frames per second processing speed recorded the highest detection rate which was 95.5% with one false positive.

2.2 Viola and Jones

Face detection, which aimed to detect human faces from video sequence using Viola and Jones method is presented in [6]. Experimentation on the performance of face detection was carried out on three different PETS video database which comprise of face region and other parts of the human body. The image resolution for the first and second video sequence was 160 ×120 pixels and 720×560 pixels for the third video sequence. The computation time for the first, second and third video were 0.637 seconds, 0.371 seconds and 9.258 seconds respectively. Apart from that, this method successfully detected frontal and slightly rotated-in-plane faces.

Viola and Jones method also adopted in [7]. The preprocessing stage was carried out using Viola and Jones method. First, the image was transformed from Red-Green-Blue (RGB) colour space to greyscale. Then, histogram equalization was performed to adjust the contrast. The output of this stage was used as the candidate image which may be face or non-face. Experimentation on the performance of face detection was carried out on 65 images in CMU database whose sizes ranged between 130×130 to 456×463 pixels. These images consist of face region and other parts of the human body. When tested, the number of incorrectly detected faces was reduced from 11 to three. Besides, detection rate increased from 86.23% to 90.31%. Apart from that, false positive error rate reduced from 5.61% to 1.53%, but false negative error rate remains at 8.16%.

2.3 Haar Features

Real time face detection based on Haar features is presented in [8]. The paper aimed to detect human faces rapidly besides achieving a high detection rate. The advantage was background regions can be quickly discarded while more computation was done on face-like regions. Experimentation on the performance of the face detection was carried out on 1390 video clips which comprise of only face regions. The detection rate for 24×24 video resolution which contained various position, scale and orientation of faces was 89%. For 18×18 and 40×40 video resolution, detection rate was 85% and 82% respectively.

The paper in [1] aimed to minimize the processing time of face detection. Improvised Viola and Jones method for face detection was based on Haar feature extraction. The disadvantage of this method was it cannot detect faces with small sizes and pose variation. Experimentation on the performance of face detection was carried out with and without background subtraction on video resolution 160 × 120 with 2587 frames. Tested images comprise of face region and other parts of the human body. Although this method can detect multiple faces, the processing time increased and the detection time decreased.

Size-based performance analysis of Haar-features for detection of face images is presented in [9]. Haar features presented was characterized by their orientation, size and type. This method was presented due to its simplicity and fast computation. Experimentation on the performance of face detection was carried out on 650 images in SCFace database of size 75×100 and 14×14. The detection rate was higher for image size ranged from 21×21 to 24×24 pixels. In addition, image up-scaling resulted in a higher number of false positives.

In [10], multi-view face detection system which aimed to minimize search space and boost detection rate is presented. Face detector presented was able to detect face with roll (-45 to +45) and yaw (-90 to +90) orientations based on Haar features. Experimentation on the performance of face detection was carried out on 7500 positive and negative samples of images which comprise of face region and other parts of the human body. The detection rate for non-frontal and frontal faces were 91.30% and 94.10%. True positive, false negative and false positive values obtained were 415, 25 and 47 for frontal faces and 210, 20 and 42 for non-frontal faces.

2.4 Other Face Detection Methods

In [11], depth sensor-based face detection for indoor surveillance is presented. The paper aimed to analyse complex scene features and integrate the colour and depth information. The method presented consists of RGBD+ViBe and 3D Mean-Shift. RGBD+ViBe was used to detect foreground. Firstly, the CIELAB colour space was applied to determine the light intensity change. Then, foreground pixel was calculated. The 3D mean-shift was done to detect facial features. This stage includes blob analysis, morphological opening and closing operation and hole filling. Blob analysis was used to examine the connectivity of similar foreground pixels. When the blob size was lower than 5, detected moving target was not identified as human. Bounding box width and height was set to values larger than 64 and 128 respectively. This box was used as a detection window. Morphological opening and closing operation was adopted to suppress noise data with a threshold value set to 5. Edge smoothing was done using a hole filling. This method has a promising outcome in moving object detection, but it has limitation in insufficient illumination condition. Experimentation on the performance of face detection was carried out on 380 frames for bright environment, 403 frames for dark environment and 250 frames for variable environment. Images in the video comprise of face region and other parts of the human body. When experimented in a bright environment, true positive rate (TPR) and false positive rate (FPR) were 89% and 3.8% respectively. As for dark environment, it recorded 88.62% (TPR) and 5.71% (FPR) while variable environment recorded 90.13% (TPR) and 5.83% (FPR).

The face detection method based on photoplethysmography is presented in [12]. The paper aimed to detect faces by estimating physiological measures. The first step involved region of interest (ROI) creation. In the paper, ROI size used was 40×30 pixels and 18×18 pixels. Next, heart rate was estimated by computing the mean pixel values for RGB colour channel of each ROI. Finally, index of heart rate was computed to ensure that detected peak matches the real heart rate. Index rate closer to one indicates a high possibility in the presence of living body parts in the ROI. Experimentation on the performance of face detection was carried out on video with 640 x 480 pixels stream size which comprise of only face region. The author claimed that 18×18 pixels detection rate were lower than 40×30 pixels and time window less than ten seconds also resulted in lower detection rate.

In [13], video-frame based face detection system which aimed to detect faces from the video is presented. The method presented was Libfacedetection which combined local binary pattern (LBP) feature and Adaboost algorithm. First, the training samples were labelled and divided into positive samples and negative samples. For each sample, average weight distribution was assigned. Then, the classifier was trained to select the best weak classifier from all features so that the average square error of the sample was minimized. Finally, the weight of training set was updated. Face angle of −40° to 40°, −60° to 60° and −90° to 90° can be detected using this method. Experimentation on the performance of face detection was carried out on 6100 images of CAS-PEAL, FMedia and MaskNet datasets. Images in the video comprise of only face region. Detection time obtained for frontal and multiview face were 2.92ms and 3.83ms at 342.5 and 140.4 frames per second respectively.

The face detection method based on multi-scale histograms is presented in [2]. The paper aimed to improve computational efficiency. The method presented consists of preprocessing, coarse-to-fine texture descriptor and descriptor construction. Preprocessing was done on input images to normalize its size to 128×128 pixels. Coarse-to-fine texture descriptor was used to extract feature in images. Support Vector Machine in the descriptor construction stage was used as a classifier to distinguish faces and non-faces. Experimentation on the performance of face detection was carried out on 4000 positive and negative image samples of Libor Spacek's database. Tested images comprise of only face region. When experimented, 4 × 4, 8 × 8, 16 × 16 and 32 × 32 block sizes achieved 95% precision for positive samples and 90% for negative samples. Detection time for 4 × 4, 8 × 8, 16 × 16 and 32 × 32 block size were 3.51799ms, 2.63614ms, 1.22268ms and 0.833912 respectively. This method achieved 10 times faster in detection rate when the size of the block was set accurately.

Contour-based procedure for face detection and tracking from video is presented in [14]. The paper aimed to detect the location of faces from the video. The method presented was based on moving face contour. First, frames extracted from video were converted into grayscale images. Then, the Robert edge detector was used to detect face edges. After that, Gaussian filtering method was used to remove non-desired edges and noise followed by computation of logical operation for edge detection. Finally, the moving face contour of each frame was determined. This method was used because the face can be easily detected in separate frame of two adjacent video frames. Experimentation on the performance of face detection was carried out on HONDA/UCSD video database with 640×480 resolution. Images in the video comprise of only face region. The author claimed that correct detection was achieved and moving faces were also efficiently tracked.

Human detection using Cascade Head and Shoulder Detection (CHSD) method is presented in [15]. The paper aimed to detect face region and filter out non head-shoulder regions. The method presented consists of initial feature rejecter, haar-like rejecter and HoG feature classifier. This method can handle side view detection of ±30° in pan and 0° to 60° tilt. Experimentation on the performance of face detection was carried out on 2000 images of the Pascal face dataset which comprise of face region and other parts of human body. The presented method achieved 83.9% detection rate.

3. FINDINGS

Generally, there are three stages which are preprocessing, feature extraction and classification stage. Based on the past researches, the most popular methods used for face detection are skin colour segmentation, Viola and Jones method, and Haar features. Skin colour segmentation method achieved 88% to 98% detection rate. However, detection on images comprise of only face region may result in this high detection rate. Viola and Jones method, and Haar features achieved 80% to 90% detection rate which was mostly carried out on images comprised of face region and other parts of the human body. Not only that, these methods can also detect slightly rotated faces. Overall, the methods presented was tested on 60 to 8000 images with 24×24 pixels as the most common resolution. Side view face detection has been carried out in only four out of 15 research papers.

4. CONCLUSION

In this paper, 15 papers from journals and were reviewed. Skin colour segmentation achieved 88% to 98% detection rate , whereas Viola and Jones method, Haar features achieved 80% to 90% detection rate. Classification is the final step in the face detection process which aims to distinguish faces from non-faces. The most popular classification method used for skin colour segmentation and Viola and Jones method is cascade classifier based on Adaboost algorithm. This classifier discards unnecessary features that is not part of a face. The other classification method includes back propagation neural network and support vector machine.

REFERENCES

[1] S. V. Tathe, A. S. Narote, and S. P. Narote, "Human Face Detection and Recognition in Videos," Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 2016, pp. 2200 – 2205.

[2] C. Y. Lin, J. T. Fu, S. H. Wang, and C. L. Huang, "New Face Detection Method Based on Multi-Scale Histograms," IEEE Second International Conference on Multimedia Big Data, 2016, p.p. 229 – 232.

[3] U. Priya, S. Vasuhi, and V. Vaidehi, "Face Detection Using CbCr Color Model in Video," 3rd International Conference on Signal Processing, Communication and Networking (ICSCN), 2015, p.p. 1 – 5.

[4] A. S. Dhavalikar, and R. K. Kulkarni, "Face Detection and Facial Expression Recognition System," International Conference on Electronics and Communication System (ICECS -2014), 2014, p.p. 1 - 7.

[5] V. Mutneja, and S. Singh, "GPU Accelerated Face Detection from Low Resolution Surveillance Videos using Motion and Skin Color Segmentation," Optik, vol. 157, 2018, p.p. 1155 – 1165.

[6] W. Zou, Y. Lu, M. Chen, and F. Lv, "Rapid Face Detection in Static Video using Background Subtraction," 10th

International Conference on Computational Intelligence and Security, 2014, p.p. 252 – 255.

[7] M. Da'san, A. Alqudah, and O. Debeir, "Face Detection using Viola and Jones Method and Neural Networks," International Conference on Information and Communication Technology Research (ICTRC2015), 2015, p.p. 40 – 43.

[8] P. I. Rani, and K. Muneeswaran, "Robust Real Time Face Detection Automatically from Video Sequence Based on Haar Features," International Conference on Communication and Network Technologies (ICCNT), 2014, p.p. 276 – 280.

[9] V. Mutneja, and S. Singh, "Size-based Performance Analysis of Haar-features for Detection of Facial Images from Low Resolution Surveillance Videos," International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 2017, p.p. 1191 – 1195.

[10] A. K. Gor, and M. S. Bhatt, "Fast Scale Invariant Multi-view Face Detection from Color Images using Skin Color Segmentation & Trained Cascaded Face Detectors," International Conference on Advances in Computer Engineering and Applications (ICACEA), 2015, p.p. 688 – 694.

[11] T. Hu, H. Zhang, X. Zhu, J. Clunis, and G. Yang, "Depth Sensor Based Human Detection for Indoor Surveillance," Future Generation Computer Systems, vol. 88, 2018, p.p. 540 – 551.

[12] G. Gilbert, D. D'. Alessandro, and F. Lance, "Face Detection Method Based on Photoplethysmography," Workshop on Low-Resolution Face Analysis (LRFA) in conjunction with 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2013, p.p. 449 – 453.

[13] G. Niu, and Q. Chen, "Learning An Video Frame-Based Face Detection System for Security Fields," Journal of Visual Communication and Image Representation, vol. 55, 2018, p.p.457 – 463.

[14] A. Dey, "A Contour Based Procedure for Face Detection and Tracking from Video," 3rd Int'I Conf. on Recent Advances in Information Technology, 2016, p.p. 83 – 488.

[15] Q. Liu, W. Zhang, H. Li. and K. N. Ngan, "Hybrid Human Detection and Recognition in Surveillance," Neurocomputing, vol. 194, 2016, p.p. 10 – 23.