The Utilization of Feature based Viola-Jones Method for Face Detection in Invariant Rotation

Tioh Keat Soon¹, Abd Samad Hasan Basari², Nuzulha Khilwani Ibrahim³, Burairah Hussin⁴, Ariff Idris⁵, Noorayisahbe Mohd Yaacob⁶

¹,²,³,⁵,⁶ BIOCORE and ⁴OptiMAS Research Group, Center for Advanced Computing Technology (C-ACT) Department of Intelligent Computing and Analytics Faculty of Information and Communication Technology Universiti Teknikal Malaysia Melaka Melaka, Malaysia

Mustafa Almahdi Algaet⁷
Department of Computer, Faculty of Education, Elmergib University, Alkhums, Libya

Norazira A. Jalil⁸
Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman (Perak Campus), Jalan Universiti Bandar Barat, 31900 Kampar, Perak

Abstract—Faces in an image consists of complex structures in object detection. The components of a face, which includes the eyes, nose and mouth of a person differs from that of ordinary objects, thus making face detecting a complex process. Some of the challenges encounter posed in face detection of unconstrained images includes background variation, pose variation, facial expression, occlusion and noise. Current research of Viola-Jones (V-J) face detection is limited to only 45 degrees in-plane rotation. This paper proposes only one technique for the V-J detection face in unconstrained images, which V-J face detection with invariant rotation. The technique begins by rotating the given image file with each step 30 degrees until 360 degrees. Each step of adding 30 degrees from origin, V-J face detection is applied, which covers more angles of a rotated face in unconstrained images. Robust detection in rotation invariant used in the above techniques will aid in the detecting of rotated faces in images. The images that have been utilized for testing and evaluation in this paper are from CMU dataset with 12 rotations on each image. Therefore, there are 12 test patterns generated. These images have been measured through the correct detection rate, true positive and false positive. This paper shows that the proposed V-J face detection technique in unconstrained images have the ability to detect rotated faces with high accuracy in correct detection rate. To summarize, V-J face detection in unconstrained images with proposed variation of rotation is the method utilized in this paper. This proposed enhancement improves the current V-J face detection method and further increase the accuracy of face detection in unconstrained images.

Keywords—Face detection; V-J face detection; unconstrained images; bicubic interpolation; SIFT

I. INTRODUCTION

There are mainly two methods that can be utilized for face detection. It is either a feature-based method or image-based method [1]. In this paper, feature based method is selected for face detection. Feature based method include skin colour, facial features and blob features. The advantages of the feature-based method are due to its rotation independence, scale independence, and quick execution time compare to image-based method [2]. Face detection is widely used in a multitude of preliminary applications. Face detection is utilized to locate a face or faces in an image. Face recognition, on the other hand, is utilized to find out who the person in the image is after face detection has been performed. Therefore, the preliminary accuracy of face detection is crucial to support face recognition. It is similar to the application of CCTV surveillance with built-in face recognition for the security purposes. Improving face detection for CCTV surveillance cameras will ensure that the faces of people can be easily identified compared to modern CCTV footage that produces blurry images. Modern cameras come with built-in functions to auto-focus on the face region. By being able to detect the face accurately, only then can unwanted red-eye effects be corrected. Another utilization of face detection can be seen in marketing methods in order to gather information on the types of customers that frequently pass by certain areas. The proposed V-J face detection method allows the detection of the faces of customers in different angles on the same plane. By utilizing customer classification, businesses can predict what type of customers is interested in certain product for advertisement purposes.

II. RELATED WORK

According to [3], there are four categories under face detection methods. They are the feature invariant approach, knowledge-based method, template-based method and appearance-based method.

Knowledge based method is known as the Rule based method. This method translates human knowledge of face features into a set of rules. These rules include the relationship of facial features. For instance, the intensity of the eye is darker than the forehead of face. Another example of frontal face in images is often with 2 symmetrical eyes, a nose and a mouth. The features then are represented as the distance and positions.

The limitation of Rule based method is that it may lead to high false positive if it is too general whereas false negative may increase if it is too detailed. Hierarchical knowledge based is
introduced to overcome the problems. However, it has the limited solution to find multiple faces in a complex image with the solution alone.

In contrast to knowledge based, feature invariant method aims to find structures of face features regardless of lighting conditions, different scaling and angles in complex images. Numerous feature invariants have been proposed such as human skin color, blob detection and moment to detect face features which then moves forward to classify face region. Merits of face detection based on human skin colour have a faster execution time in face detection despite different scaling and angles. Usually, it is utilized in preliminary process for dimension reduction to improve the speed detection. There are several types of colour space. Usually, there are 6 colour spaces for skin colour face detection. Face detection that based on colour space are YCbCr, RGB, HSI, nRG, HSV, and CIE. However, skin colour based method has skin colour-like background challenge. The author in [4] proposed to use skin colour with edges. YCbCr colour space was selected and classified skin or non-skin types by using Gaussian Mixture Model. Sobel edge detection was utilized after binary process. At least 3 ‘holes’ was created with Euler formula if a face has been detected. However, the researcher found out that fault detection was due to over-bounding if other regions are similar to skin colour. In order to resolve the challenge, skin modeling coefficient matrix technique and improved Gaussian distribution are proposed. The author in [5] explicit defined algorithm is chosen for the development due to simplicity and speed performance. Skin modeling coefficient matrix is used for segmentation process skin pixels or non-skin pixels. The Robert edge detection method was performed before post-processing. Connected component analysis is performed after post-processing. Finally, 2 conditions of aspect ratio must be fulfilled to classify face or non-face. The author in [6] proposed to use skin-colour model (RGB colour space), facial features (labial feature and holes feature) and improved Gaussian distribution model to detect multiple faces with good performance and remove skin colour-like background. Another challenge of skin colour is to resolve illumination problem. The author in [7] proposed to use the skin colour (HSV) in different range for indoor and outdoor environment. Erosion method was used to remove small non-face objects after it was converted to binary black and white. Active snake contour method was selected to detect maximum contour area. However, the researcher suggests changing to automatic threshold for better detection rate especially outdoor environment if the illumination is too bright. Noise challenge is further removed. The author in [4] proposed low pass filter was used to eliminate noises. Threshold value was determined via average sum of median and maximum values column scanning. Blob is also considered as an interest point in face detection. It has been widely used for face detection. For instance, for the blobs are Haar features, corner detection, Laplacian of Gaussian, Difference of Gaussian and component labeling. Later, the blobs are further analyzed by extracting the information of shapes of objects that are present in the image. This technique is also referring to image segmentation. Result of feature extraction is to identify the number of different objects, region information and other salient features. At an early stage, [8] was one of the first corner detection being carried out for interest point. It was improved by [9] to remove the noise. The author in [9] applied Gaussian to autocorrelation matrix for corner detection. However, it is limited to scaling invariant. The author in [10] proposed SIFT to overcome the scaling variant problem. It was dependent on the sigma or standard deviation. The author in [11] showed that the Laplacian response is decayed when the standard deviation or scale getting bigger. Superposition of two ripples results in the maximum response becoming blob-like. To keep the Laplacian response the same across the scale, second order of Gaussian must multiply by σ^2. In [10], the author proposed to use Difference of Gaussian (DoG), which is approximate for LoG. Optimization is improved by using DoG. Corner detection results in rotation invariant but not in scale variant [9]. Scale space theory was introduced and there are two important steps. These two steps are known as i) feature detection and ii) finding maxima and minima extrema. The author in [11] introduced automatic scale selection. There are many blob detections based on LoG or DoG in scale space. For instance, Determinant of Hessian (DoH), SIFT [10], Harris Laplacian [12], Hessian Laplacian [11], and Harris Affine Region [13]. The author in [14] proposed in-plane angle estimation for face images from multi-poses by applying SIFT to 2 reference points, which are midpoint of eyes and nose. The appearance descriptors consist of SIFT descriptors such as location, scale and orientation of reference points. 2 hypotheses were used to determine face or non-face via Bayesinan classifier. The proposed result outperformed in terms of low false face detection rate, low in-plane rotation error and speed performance. The author in [15] proposed an improved Haar-like feature so called Haar Contrast Feature, which efficiently for object detection under various illuminations with the Haar Wavelet based. The LoG can be represented by Haar Wavelet which proposed by [16]. Computation of Haar Wavelet can be done by utilizing integral imaging method. This method has speed up the process. The author in [17] proposed heterogeneous feature descriptors and feature selection for efficient and accurate face detection. To address the issue of distinctive representation for face patterns, the researcher proposed complementary feature descriptors Generalized Haar-like descriptor, Multi-Block Local Binary Patterns descriptor and Speeded-Up Robust Features (SURF) descriptor. Particle Swarm Optimization (PSO) algorithm was integrated into the Adaboost framework as feature selection and classifier learning. A three-stage hierarchical classifier structure and nonlinear support vector machine (SVM) classifier were used to rapidly remove non-face patterns. The experiment was tested on CMU+MIT data set. The proposed solution also worked well for faces with Yaw rotation between ±22.5°. The results show robustness and efficiency of the proposed solution with other state-of-the-art algorithms. The author in [18] proposed to use normalized RGB colour space to determine skin. Blob detection (Connected Component) was used later. The researcher [19] made a time consumption and accuracy comparative study of SIFT and its variants such as GSIFT, PCA-SIFT, SURF, ASIFT, and CSIFT in 4 situations. The results showed that, in scale and rotation situations, SIFT and CSIFT performed better compared to other variants. In affine image, ASIFT performed better compared to others. SURF gain the fastest speed performance compared to others. In blur
or illumination image, GSIFT performed better compared to others.

There are two techniques of face detection based on template matching. Template image includes either the face as a whole or face features separately. The stored predefined face features with eye, mouth, and nose are known as deformable template matching. It is then the stored predefined face features are correlated with the input face. For instance, template is matched with the input image through slide windows. However, it is limited to achieve better result with the variant of scale and pose. Deformable template is introduced to overcome the problems. In [20] enhanced the winner-update algorithm (WUA) with winner-update and integral image (WUI) for fast and full search algorithm. These algorithms were used for reducing the computational complexity. By exploiting the integral image, the method gained the speed performance. The author in [21] adopted template matching method for design pattern detection. This method was not only utilized to detect exact pattern, but on variation of patterns as well, based on normalized cross correlation.

Appearance based method is similar to template matching but learning from a set of stored example face images. This method depends on statistical analysis and machine learning. For instance, statistical analysis based on probability to determine a face or not. There are a lot of machine learning in this method such as logistic regression, discrimination analysis and other binary classification. Based on appearance method, statistical analysis is also known as feature representation. Example of feature representation methods are Haar feature, skin colour and shape. Usually machine learning used in face detection is mainly for feature selection. Most feature selection methods are based on machine learning such as Adaboost, neural-network and SVM. Pattern classification is a method to classify pattern vectors into several classes. It is often referred to machine learning in artificial intelligence field. The methods could be classified into supervised learning, unsupervised learning, reinforcement learning, evolutionary learning, and ensemble learning. Supervised learning is usually having past historical data and class of the subset of the data. Unsupervised learning is same as supervised learning but without knowing the class of the subset of the data. One of the famous ensembles learning method is Adaboost. It was utilized for the face detection classification. The method that helped to improve the training time during performing Adaboost is the Cascaded method. Most recent research is focusing on machine learning. They are multilayer Neural Network, Support Vector Machine, Adaboost, Hybrid Adaboost and Support Vector Machine, Model Based, Discriminant Analysis method and Deep Learning.

A recent progress of face detection is rotation invariant, fast speed detection, quality of the image which includes illumination, noise and blur. According to knowledge-based method, [22-23] proposed morphological technique to detect face. It is limited to accuracy of edge detection and multi face. The authors in [24-27] focus on deep convolutional Neural Network. However, appearance based requires more data to do the training and it is time consuming. The authors in [28-34] focus on V-J face detection. Survey study [35] on several techniques regarding the extraction and learning algorithms including Local Binary Pattern (LBP), Adaboost algorithm, SNOW classifier, SMQT features and Neural Network-Based face detection. It shows that V-J face detection is faster and accurate for frontal face detection.

Related works of V-J face detection. The author in [36] proposed real-time face detection. The author further proposed pose estimation during the Haar features training which covers (±15°), 30° covers (15° - 45°), 60° covers (45° - 75°), 90° covers (75° - 105°) until 360° with 12 detectors. However, training on rotation of Haar features require longer training time. Viola et al., 2004 continued to propose more robust real-time face detection but limited to ±15° only. The author in [37] proposed rotate input sub-windows with ±30° which could cover up to ±45° from 0°. The author in [14] proposed feature transform which covers ±10° only. Based on the previous authors, most of them focused on mainly speed and only a minor contribution to accuracy. Many of them are based on training method. In 2004, Viola and Jones [38] took about 2 weeks to complete the training, which was only limited to ±15° in-plane rotation. This thesis extends the rotation method from Li and Yang [37] which covers 360°. The method gained significantly better accuracy on in-plane rotation with low false positive without a longer training required.

Studies in [24-25, 39-41] tested the face detection from Face Detection Dataset and Benchmark (FDBB). It contains more than 5000 unconstrained faces such as large appearance variation in pose, occlusion, expression, illumination, and imaging conditions. Study [24] did the training dataset from Feret, PIE database which contains face with different poses, frontal, left/right half profile, and 0 till 30 degree in-plane rotations. Study [42] tested face detection from Feret database. Study [22] tested FEI database contains facial images including facial expressions, occlusion, lighting conditions, and background complexities. Studies [23, 40, 43] tested on IMM frontal face database contains variance of lighting conditions which was recorded in 2005 by Fagertun and Stegman at Technical University of Denmark. Study [23] tested face detection from FEI database. Studies [24, 44, 46] tested BioID database which consists of 1521 grey images with 384x286 pixels dimensions. Studies [4, 40, 32] tested Bao database that contains family images. Study [32] tested LFW database. Study [40, 43] tested Caltech database. Study [47] tested with XM2VTS contains occlusion faces. Study [5, 14, 44, 45, 48] tested and training from MIT+CMU database. Studies [14, 40] evaluated the testing from CMU database which contains total 50 face images with in-plane rotation and some with multiple faces. Study [14] tested with good quality images where some with poor quality of images were removed, left 40 images and 65 faces.

III. PROPOSED V-J FACE DETECTION IN UNCONSTRAINED IMAGES

This paper follows pattern recognition methodology by [49], which refer to the Figure 1. According to [49], there are 2 ways to do the recognition, either classification phase or training phase. Training phase can be incorporated into classification method. The included experiment has been based on our previous publication in [50].

www.ijacsa.thesai.org
The methodology starts with pre-processing the enhanced in-plane rotation image file so that faces in different angles could be detected. Then, the evaluation of V-J face detection is evaluated in established databases to proof it is more accurate in V-J face detection in unconstrained images. There is only one main part research design of V-J face detection in unconstrained images in this thesis. It involves the enhanced rotation in V-J face detection only. The data type utilized in this paper is grey colour images. The rotated face images are from established database CMU. CMU consists of grey colour with 50 rotated face images only. The CMU images have different size of image files with grey format. Study [37] utilized 50 CMU input images as image dataset. The enhanced rotation of V-J face detection consists of two parts 1) Rotation process, 2) Face detection process.

A. Rotation Process

The parameter of $\theta$ is represented as radian of the rotation. A negative angle represents clockwise rotation whereas positive value represents anticlockwise rotation. The enhancement of V-J face detection is carried out with rotated 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330° and 0°. In mathematics, rotation [51] is formulated as (1).

$$[x', y'] = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} [x, y]^T$$  \hspace{1cm} (1)

B. Face Detection Process Units

Equation (2) shows the grey colour conversion.

$$\text{grayColour} = (R + G + B)/3$$  \hspace{1cm} (2)

In Viola Jones’s face detection, Figure 2 shows Haar features are represented in two-rectangle features, three-rectangle features and four-rectangle features. The value of the rectangle feature is the sum of difference between black region and white region.

In order to speed up computation of rectangle-feature, Viola Jones proposed integral image. There are four regions (Figure 4) represent A, B, C, and D. To compute D region, $D = (X4,Y4) + (X1,Y1) - (X2,Y2) - (X3,Y3)$. Figure 3 shows integral images calculation.

$$\sigma^2 = M^2 - \frac{1}{N}\sum X^2$$  \hspace{1cm} (3)

Where $\sigma$ is the standard deviation, $M$ is the mean, $N$ is the region size and $X$ is the pixel value within the region.

AdaBoost is the machine learning algorithm. It will form a strong classifier. The AdaBoost performed the feature selection for Haar features in face detection. Figure 4 shows sample of rotation from CMU.
The proposed V-J face detection method has several better achievements to meet the objectives. The accuracy performance by rotating image file with 30° each step within 360° before performing V-J face detection which meets the pattern recognition methodology. V-J face detection provides standard couple of solution with rotation. The accuracy performance is increased by providing flexibility and prior knowledge to any face detection as pre-processing. The proposed V-J face detection in unconstrained images is significant better accuracy than previous method, which are demonstrated the results by the number of unconstrained images. For future works, the interpolation can be combined with rotated face to enhance the rotation accuracy. Besides that, the SIFT can be combined with convolutional neural network to find the eye region for accuracy of face detection.

ACKNOWLEDGMENT

This paper is part of PhD work in the Faculty of Information and Communication Technology, UTeM and funded by MyPhD scholarship under the Ministry of Higher Education, Malaysia.

REFERENCES


