A Review on Face Detection

Md Omor Faruk
Electrical and Computer Engineering Dept.
University of Waterloo
Waterloo, Canada
mofaruk@uwaterloo.ca

Miaoshu Wangxu
Mechanical and Mechatronics Engineering Dept.
University of Waterloo
Waterloo, Canada
mwangxu@uwaterloo.ca

***Abstract* - Face Detection has its wide range of applications in Face Recognitions, Facial Motion Capture, Photography, Lip Reading, and Emotional Inference. It can be regarded as a special type of objective detection. In today's most applications, frontal human face images are used. Different detection techniques have been set up over the years including knowledge-based algorithms, feature based algorithms, and appearance-based algorithms. The first two algorithms use the rules set up by human and does not reply on the data set, while appearance-based algorithms depend on the # of training sets that are fed in and are related to statistical learning and machine learning approaches. The difference between knowledge-based and featured based techniques is that feature-based technique extracts features out of the image for processing by using classifiers to differentiate between non-facial and facial regions. The improvement of feature-based algorithm is to overcome the limits of human knowledge.**

Keywords - knowledge-based algorithms, feature-based algorithms, appearance-based algorithms, colour segmentation, RGB, YCbCr, Morphology Filter, Haar feature, AdaBoost, Integral image, Cascade Classifier.

I. Introduction

In this report, two different face detection techniques will be introduced and discussed in detail. First algorithm is known as color segmentation where images are represented by RGB, YCbCr, or HSV and pixels that are determined to be human skin colors are marked for processing. It is a typical knowledge-based detection technique and is based on some general rules that people have set up over the years which can be applied to fit all different human skin tones. Second algorithm is called Viola-Jones, a feature-based detecting technique, where key features on a grey scale image such as eyes and ears are needed to be extracted to match the given patterns in order to draw the conclusion that the objective region is human face. So far, both techniques are well-developed, and have their own advantages and disadvantages.

II. Color Segmentation - Knowledge-Based Algorithm

For face detection by color segmentation, many general algorithms have been developed over the years to detect all human skin tones based on either RGB, YCbCr, HSV, or the combination of any of these two or three. It is a typical knowledge-based method using the rules set up based on human knowledge. In this report, RGB+YCbCr approach will be our focus.

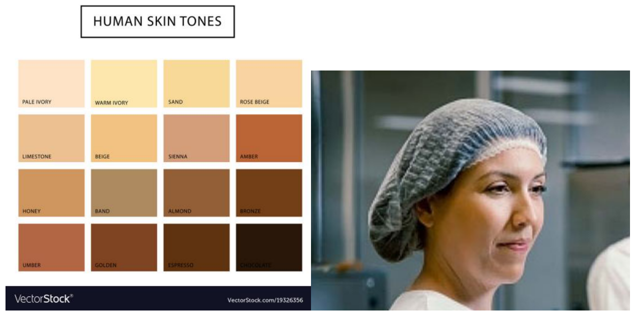

Fig.1: Different human skin tones

First of all, as this approach is quite sensitive to noise, denoising is always needed [8]. Second, RGB 3 channels will be processed independently when dealing with any input image of size m x n. Similar to grey scale image, each pixel in RGB and YCbCr is in the range of 0-255. YCbCr is another form of color representation, where Y stands for illumination, Cb is blue chrominance, an Cr is red chrominance. The conversion from RGB to YCbCr is concluded as follows by B. Iyer, S. Nalbalwar and R. Pawade [3]:

$Y = 0.257*R+0.564*G+0.098*B+16$

$Cb = -0.148*R-0.291*G+0.439*B+128$

$Cr = 0.439*R-0.368*G-0.071*B+128$

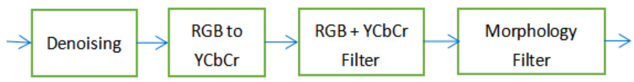

Fig.2: Steps for CSKB algorithm

So far, we have 6 input matrix of size m x n (R, G, B, Y, Cb, and Cr respectively). The third step is to apply the general skin color rules to the 6-input matrix; the corresponding matrix operations are [3]:

$R>95, G>40, B>20,$

$R>G, R>B, R-G>15,$

$Cr>135, Cb>85, Y>80,$

$Cr<(1.5862*Cb+20), Cr>(0.3448*Cb+76.2069),$

$Cr<(-1.15*Cb+301.75), Cr<(-2.2857*Cb+432.85).$

The resulting matrix after applying each matrix operation is actually the same size as the original input matrix which is m x n, and each pixel is marked to be either 0 or 1 depending on each matrix operation.

Take the below input matrix of size 3 x 3 as example, and assume it is red channel. If the detecting rule of R>95 is applied to the matrix, then the resulting matrix will only consist of 0s or 1s which corresponds to black and white pixels.

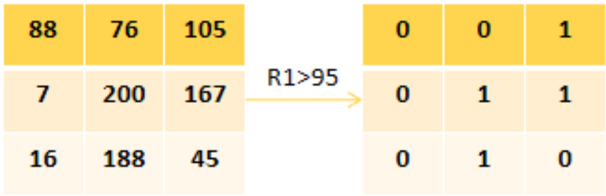

Fig.3: 3 x 3 matrix operation for CSKB

After all matrix operations are done, 'and' operation between all these outputs will be performed. Pixels that satisfy all the above conditions will be marked and considered to be the face region. As a result, there will only be one output matrix of m x n which is exactly the same size of the input matrix.

The last step is to use a morphology filter to eliminate any misclassified regions such as human body (arms, hands, etc.) which has similar skin color as human face. The general idea is to determine whether or not the detected regions match the shape of human face; this is usually done by comparing the similarities between the marked region and the given shape of human face. In this case, mathematic approach such as mean squared error could be applied here.

The above color segmentation technique has been tested in Octave/MATLAB, below is the output images generated from output matrix which contains black/white pixels and also RGB. Notice there is a piece of tiny region on the bottom left of the below images, which has similar color as human face, is misclassified. Therefore, morphology filter does play an important rule here for taking out the non-face regions.

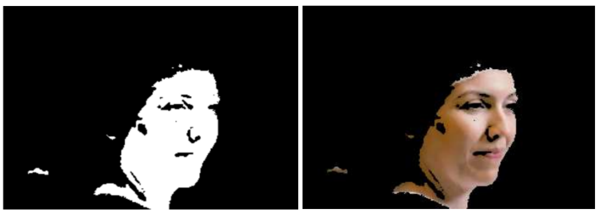

Fig.4: Applying Morphology Filter

III. VIOLA-JONES FACE DETECTION APPROACH

An exceptionally quick and direct way for object detection was contrived by viola and Jones [11] in the year 2001. They built up a general object detection structure that had the option to give competitive object detection rates progressively. It tends to be utilized to take care of number of detection problems. However, the principle inspiration originates from face detection based on human face features such as nose, eyes, mouth etc.

The Viola-Jones algorithm has 4 principal steps [11].

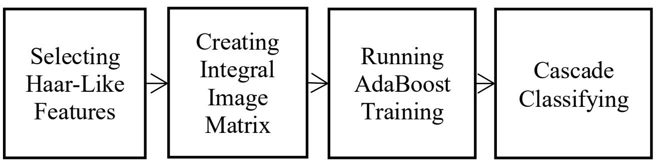

Fig.5: Steps of Viola-Jones algorithm

Similar to other feature-based face detection technique, Viola-Jones algorithms converts RGB images to GrayScale first which results in an m x n matrix with pixel range from 0 to 255.

With an image, the calculation technique takes an eye at various little subregions and attempts to discover a face by searching for explicit features in each subregion. It needs to check a wide range of positions and scales on the grounds that an image can contain numerous numbers of faces of different sizes. Viola and Jones utilized Haar-like features to recognize faces.

A. *Haar-Like Features*

Every single human face shares a few similitudes. For example, the eye district is darker than the scaffold of the nose. The cheeks are likewise lighter than the eye area. We can utilize these properties if a picture contains a human face or not.

A basic method to discover which area is lighter or darker is to summarize the pixel estimations of the two districts and contrasting them. The total of pixel esteems in the darker district will be less than the total of pixels in the lighter locale. This can be achieved utilizing Haar-like features. Some ordinarily utilized Haar features are:

a. Two rectangle features.

b. Three rectangle features.

c. Four rectangle features.

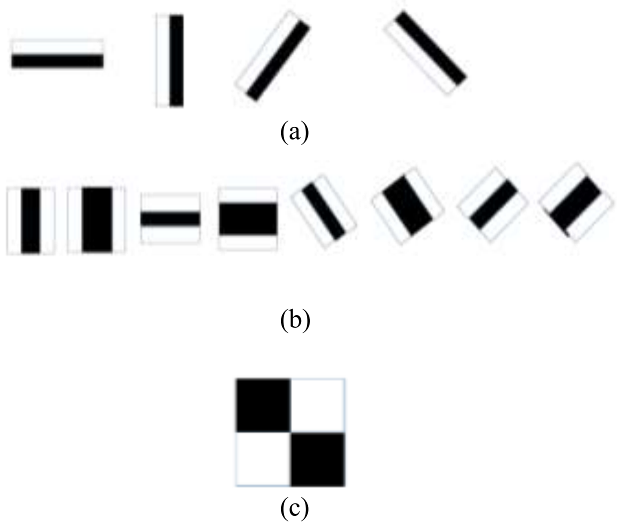

Fig.6: Common Haar- Features (a – edge feature, b – line feature, c – diagonal feature) [11]

The estimation of edge rectangle feature is the contrast between the total pixels inside edge rectangle districts as planted in Fig 6. While diagonal rectangle features compute the difference between the diagonal pairs of the rectangles, the value is center rectangle subtracted by the addition of the two surrounding rectangles in line feature rectangle [9] [10].

B. *Integral Images*

In an integral image, the value of each point is the sum of all pixels above and to the left, including the target pixel. The integral image can be calculated in a single pass over the original image [11]. It requires only three operation with four numbers, regardless of rectangle size containing N numbers of pixels where N ∈ $\mathbf{Z}^+$. Following example will demonstrate how it works for a rectangle region.

10	13	11	32
7	22	21	19
8	0	14	8
9	5	8	23
29	6	7	10

10	23 A	34	66 B
17	52	83	134
25	60	105	164
34	74	127	210
63	109 C	169	262 D

Fig.7: Demonstration of the concept of Integral Image

Let us assume that Fig. 7 (a) is rectangular subregion of an image. Fig. 7 (b) is the integral image matrix of fig. 7 (a). For instance, 74 is the summation of all pixels beginning from top-left corner i.e. 10 + 13 + 7 + 22 + 8 + 22 + 8 + 0 + 9 + 5 = 74.

Now, we want to compute the sum of the pixels of the rectangle area denoted by red area. The four points of rectangle A, B, C and D are determined as shown in fig. 7 (b). The sum of pixels in the rectangle ABCD can be derived from the values of points A, B, C and D using the formula D - B - C + A = 262 – 66 – 109 + 23 = 110 which is exact match of summation of the pixel inside red area. With this manner the algorithm compares a region with its neighbor regions to detect whether it is darker or lighter.

The purpose of using integral image is to detect the face in an image in a more efficient way [12]. But deciding correct features and size of the features is a huge dilemma which can be solved by a machine learning Algorithm named Adaptive Boosting.

C. Adaptive Boosting (AdaBoost)

AdaBoost training algorithm chooses little features from the face that encourages quick and simple calculation [11]. Unlike other strategies, AdaBoost calculation finds desired area of the object by omitting unnecessary area. The working model is based on neural network and can be found at [11].

In the Viola-Jones algorithm, each Haar-like feature represents a weak learner [11]. It utilizes a significant idea of joining various classifiers developed utilizing similar information set. As the technique progresses, it gives more emphasis on weakly classified image subregion by changing weights. To compute the performance of a classifier, all subregions of the considerable number of images utilized for preparing will be assessed. Some subregions will deliver strong response in the classifier. Those will be named positives, which means the classifier thinks it contains a human face. To decide the type and size of a feature that passed into the final classifier, AdaBoost checks the performance of all classifiers delivered during training and finally build a strong classifier that contains the best performing weak classifiers [12] [13].

D. Cascade Classifiers

Since Adaptive Boosting process builds a thousands of classifiers each candidate feature requires to pass every classifier in order to form a face. Therefore, it becomes computationally expensive to detect a face in an image [11].

To avoid this, Viola and Jones transformed their strong classifier into a cascade where each weak classifier denotes one single stage. The cascade disqualifies candidates by making stricter requirements in each stage because later stages are being considerably harder for a possibility to pass. Finally, if candidate passes all stages, then it exits the cascade with positive flag; otherwise it exits with a negative flag after failure in any of the stages.

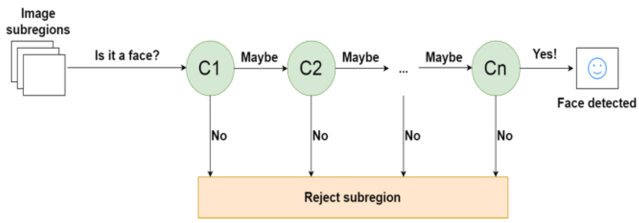

Fig.8: Filter with cascade classifier [13]

Fig.9: Face detection with Viola-Jones Algorithm [12]

IV. Advantages & Disadvantages

Notice the key advantage of using colour segmentation (RGB + YCbCr) algorithm is that this approach can be implemented using hardware such as Field Programmable Logic Array (FPGA) due to its less computation complexity. FPGA/ASIC allows parallel computations which would significantly speed up the detection by processing independent matrix operations all at a time. In comparison, feature-based algorithms are generally too much floating-point calculation involved and many iterations to go through therefore not possible to implement on hardware due to the cost [17].

Although the color segmentation approach is must faster and cheaper to implement, its accuracy is relatively low due to the capability of fitting all human skin tones. For feature-based face detection algorithms, although it is a relatively slow approach, researchers have found formulas for reducing the complexity of matrix operation such as calculating the sum of pixel values in a sub-region as well as set up the AdaBoost approach which can quickly and automatically fix any misclassified pixels that contain noise.

Unlike today's most face detection techniques, colour segmentation algorithm is not limited to frontal face detection, images with orientation are supported as well. However, the corresponding shape of none-frontal human face has not been designed yet for the morphology filter to process.

Besides running speed, computation complexity, cost and accuracy, the below limitations have also been found:

For colour segmentation method, morphology filter is always required to eliminate none-face region such as human body; however, people tend to use simple morphology filters for a fast detection. These simply morphology filters, such as only keep the largest detected regions or take out the lower half part of the image assuming face should be on the upper half [4], are usually limited to certain applications and could cause the whole algorithm to just corrupt due to the wrong assumptions.

For Viola-Jones method, it is sometimes difficult to detect features due to complex background and different orientations. Another issue found is that it can be hard to locate facial features because of several corruption such as illumination [16].

Since Viola-Jones uses the adaptive boosting technique, it has the highest accuracy among all the face detection techniques which is nearly 94% [11]. AdaBoost is an algorithm which only needs two inputs: a training dataset and a set of features (classification functions). There is no need to have any a priori knowledge about face structure. Meanwhile, there are some trade-offs using this technique. The result depends on the data and weak classifiers. The quality of the final detection depends highly on the consistence of the training set. Both the size of the sets and the interclass variability are important factors to be taken into account. The training process is slow because at each iteration step, the algorithm tests all the features on all examples which requires computation time and the computation time is directly proportional to the size of the features and example sets [14]. As it is a combination of complex weak classifiers, sometimes it leads to overfitting. Additionally, if the image contains too much noise, it would fail to detect the face due to no denoising process.

V. Conclusion and Future Work

In general, identifying face from image depends on several factors such as pose, facial expression, occlusion, orientation, illumination etc. For feature-based algorithm, like Viola-Jones the possibility primarily depends on how many face features it can identify. In comparison for knowledge-based techniques, the possibility does depend on features [15].

After implementing the codes in Octave and Python for these two face detection techniques, possible improvements are found. For color segmentation (RGB+YCbCr) algorithm, few papers were found regarding the hardware acceleration of face detection system on FPGA. Further implementation on FPGA shall be done for comparing with the existing results of running time and hardware cost from published papers. For feature based (Viola-Jones) algorithm, more testing could be done on the orientation of the images to find its limitations.

Besides these knowledge-based and feature-based algorithms introduced, appearance-based algorithms which relates to machine learning such as Bayes Learning, Neutral Network, and CNN are also quite popular for detecting features. Appearance-based algorithms have their own advantages and are very comparable with the introduced two face detection techniques in this paper.

Acknowledgment

Of the many who provided us with constructive comments, we are especially grateful to Prof. Dr. Zhou Wang and Mr. Shahrukh Athar for their kind feedback and clear direction. We would like to pay our sincere gratitude to all the authors of the main paper on which we based our work.

References

[1] Patil, Prajakta M., and Y. M. Patil, "Robust Skin Colour Detection and Tracking Algorithm", International Journal of Engineering Research and Technology Vol. 1. No.8 (October-2012), ISSN: 2278-0181 (2012).

[2] KB Shaik, P Ganesan, V Kalist, BS Sathish, JMM Jenitha:"Comparative study of skin color detection and segmentation in HSV and YCbCr color space", Procedia Computer Science57 ,41-48(2015).

[3] B. Iyer, S. Nalbalwar and R. Pawade (Eds.) ICCASP/ICMMD-2016. Advances in Intelligent Systems Research.Vol. 137, Pp. 324-332. © 2017-The authors. Published by Atlantis Press.

[4] Dipankar Narendra Arya, Sivanji K.L.V., Raghunadha Reddy, Sivanantham S and Sivasankaran K, "A Face Detection System Implemented on FPGA based on RCT Colour Segmentation", 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Department of Micro and Nanoelectronics School of Electronics Engineering VIT University Vellore – 632014, Tamilnadu, India

[5] M. P. Ooi, "Hardware implementation for face detection on Xilinx Virtex-II FPGA using the reversible component transformation colour space", Electronic Design, Test and Applications. Third IEEE International Workshop, Jan 2006.

[6] A. M. G. B. Reinhard, E. and P. Shirley, "Color transfer between images," IEEE Computer Graphics and Applications, vol. 21, pp. 34 – 41, 2001.

[7] Chi-Jeng Chang, Pei-Yung Hsiao, Zen-Yi Huang (2006). 'Integrated Operation of Image Capturing and Processing in FPGA', IJCSNS International Journal of Computer Science and Network Security,

[8] S.Vidya Dharan, Mohamed Khalil-Hani, Nasir Shaikh-Husin, 2015 IEEE Student Conference on Research and Development (SCOReD), VeCAD Research Laboratory, Faculty of Electrical Engineering,Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Malaysia

[9] Sharifara, Ali, et al. "A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection." Biometrics and Security Technologies (ISBAST), 2014 International

[10] Monali Chaudhari, Gauresh Vanjare, Dhairya Thakkar, Malay Shah and Amit Kadam,‖ Intelligent Surveillance and Security System‖, Vol. 3, Issue 3, March 2015, pp. 2291- 2299.

[11] Paul Viola, Micheal Jones, "Rapid object detection using a Boosted Cascade of Simple features" CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2001

[12] Cordiner, A.; Ogunbona, P.; Wanqing Li, "Face detection using generalised integral image features," Image Processing (ICIP), 2009 16th IEEE International Conference on , vol., no., pp.1229,1232, 7-10 Nov. 2009.

[13] Lienhart and J. Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. IEEE ICIP 2002

[14] 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, pages118–183, 2003.

[15] W.-C. Hu, C.-Y. Yang and D.-Y. Huang, "Feature-based Face Detection against Skin-color Like Backgrounds with Varying Illumination", Journal of Information Hiding and MultimediaSignal Processing, 2011

[16] Z. Li, L. Xue and F. Tan, "Face detection in complex background based on skin color features and improved AdaBoost algorithms", Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on, vol. 2, 2010, December 10-12, pp.723, 727.

[17] E. Hjelmas and B. K. Low, "Face Detection: A Survey," Computer Vision and Image Understanding, vol. 83, 2001, April, pp. 236-274.