Real-Time Facial Feature Point Detection and Tracking using Rule-based Approach and Integral Projection

Joel C. De Goma, **Joel Roy V. Jamias**, **Anthony H. Kwong**, **Jessica Mae S. Salvador**

Abstract: The face is a basis of how one can describe emotions of another as it is a medium of communication. Through looking at the face of the person, one can detect what kind of feeling he is portraying. It may be a case for humans; however, computers cannot detect the emotion of humans simply by looking at their faces, computers rely on the positioning of the feature points of the eyes, and mouth in order to determine the emotion. In this connection, this study aims to create a system wherein it can detect the features and track points of the face in a real-time video. The authors attempt to combine different algorithms in order to detect and track the feature points in the face. Based on the results of the conducted tests, the setup has recorded an overall success rate of 44.56 %. This is due to the large skin range that was used in the program.

I. INTRODUCTION

Face detection is the process of determining the certain location and sizes of the face within an image or scenery. It only detects facial features and will ignore everything else such as buildings, trees bodies, and others.

During the past, Subudhi used a simple yet effective algorithm for detecting faces which considers the skin tone of the subject in the image [1]. In Khandait's study [2] facial feature detection is just simply locating a human face in an image. Although locating a face in an image is simple, there is a challenge in it when it comes to occlusions and lighting conditions. In the previous studies, simpler methods were used to locate human faces in an image or video. They used a camera and a grabber to extract a human face on a video an in an image. Although it was proven effective for still images or images that are normalized, this method proved to be weak against faces that are not facing directly into the camera based on Chen's study [3].

Color processing was also a solution for effective facial detection. Although it is faster compared to method which involves heavy mathematical equations, it is vulnerable against images and videos that have irregular illuminations based on Kotropoulos [4].

Revised Manuscript Received on March 08, 2019.

Joel C. De Goma, School of Information Technology, Mapua University, 333 Sen Gil J. Puyat Ave., Makati City, jcdegoma@mapua.edu.ph.

Joel Roy V. Jamias, School of Information Technology, Mapua University, 333 Sen Gil J. Puyat Ave., Makati City, Philippines.

Anthony H. Kwong, School of Information Technology, Mapua University, 333 Sen Gil J. Puyat Ave., Makati City, Philippines.

Jessica Mae S. Salvador, School of Information Technology, Mapua University, 333 Sen Gil J. Puyat Ave., Makati City, Philippines.

Most face detection systems attempt to extract a fraction of the whole face; thereby, eliminating most of the background and other areas of an individual's head such as hair that is not necessary for the face recognition task. This process involves static images, which unfortunately has large search space for possible locations of the face. This would create a problem on how to assume that the detected face is accurate when it comes to size based on Balasuriya's study [5].

In Padilla's study, they used the Haar classifier in identifying the feature points of the face, mainly the eyes. They made the eyes the primary basis for finding the other features of the face; and if one eye is not visible, then the nose will be made as the primary basis for detecting the other region of interest. They have also used a YCbCr for the detection of the skin segmentation [6].

The main significance of this study is to be able to address the problems encountered on the previous study to integrate the extraction of the facial feature point and track it. One problem is the low detection rate of the face features due to lighting condition and to differentiate the human skin tone from the background. Another problem is to detect other facial feature point even if occlusions and change in orientation occurs.

The study aims to track the facial feature in real-time videos by (1) changing the pixel value to reduce illumination; (2) increase the color range and segmenting the facial features or region of the subject's face correctly; and (3) determining the feature points in the event of occlusions and change in pose orientations.

The study will form a combination of different algorithms which will include face detection, region segmentation, feature points classifier and a facial feature point tracker and establishing different rules. The output should be able to accurately detect the face and track the feature points in it in a real-time video.

For this research, the authors will combine four different methods to be able to further the study in the face detection algorithm and as to solve the problems specifically in terms of the accuracy and speed. The authors intend to use the

Face detection module, the ROI segmentation which will be a projectile based for finding the gradients, double threshold and the edge tracking by hysteresis; the authors also intend to use the facial feature point classification in which the authors will be using the knowledge-based or rule-based system wherein the authors will set up ruled in determining each of the key points of the

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

face. After determining where the facial feature points are, they will now then track them using the projection based approach algorithm.

For this research, the authors have decided to use the face detection module in which YCbCr will also be used [7] for skin segmentation and the ellipse matching for getting the face region of the subject. After that, the authors have chosen the projectile face segmentation for the ROI process [8]. After defining the edge of the features, next is where the authors will determine if the objects detected by the previous steps are indeed a facial feature point; for this step, the Authors will be using a rule-based process wherein the authors will set up the rules in which the system will check the objects and if they will fit the criteria, then they will be considered a facial point and will be marked as one [9]. Finally, the feature points detected will be tracked in the video and the authors have decided to choose the projection based algorithm for this.

II. METHODOLOGY

The authors manipulated factors on the test data such as the skin color, background, luminance, occlusion and the angles for the pitch and yaw movement. These are manipulated in order for the authors to test if the methods could still perform its functions correctly.

To support the study, the authors are going to use database from “FERET face Database” by P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi or “AR Face Database” by Aleix Martinez and Robert Benavente. The database, a collection of different faces, will be used for testing. The test image will be processed by our proposed algorithm for facial feature point detection.

The proposed facial feature point detection and tracking algorithm is composed of four (4) main modules. These modules are the face detection, Region of Interest (RoI) segmentation or facial feature segmentation, facial feature point classification rules and facial feature point tracking. The input image size is a 320x240 resolution. There should only be one face exist on the input image. The face angle must be yaw or pitch as long as the face features are visible, and a small rotation angle from -5^{o} to $+5^{o}$ for roll.

Data Set

In order for the authors to conduct experimentation, they collected videos based on skin tone, background color, luminance, occlusion and angles of movements. To identify on which skin tone the test subject belong, the Authors used the Von Luschan’s Chromatic Scale. The Von Luschan scale has five types of skin classification and composed of 36 colors [10]

For the real-time processing, the group started the webcam and processed it simultaneously. The test subjects were asked to face the webcam while performing the occlusions (Eyeglass, Sunglass, Right Eye, Left Eye, Nose and Mouth), in addition to these occlusions. The test subjects were also asked to face the webcam without having any occlusions.

Although this is a test for the Real-time detection and tracking, the images from the real-time video was saved once the face and the feature points are properly aligned. The saved images were used for checking if the prototype was able to identify the points correctly per frame.

Face detection Algorithm

The first module of the system is face detection. Authors proposed an effective approach on detecting the face region on the image based on the study in [11]. This method was proven to be effective for face detection method in true-color images and can also execute with very low computational demands thus having faster results. The module is consisting of three (3) sub-modules: skin color segmentation, grayscale conversion and shape matching.

Region of Interest (R.O.I)

In our proposed method, the Authors will use projection [12]. However, they aim to detect six (6) regions of the face such as the left and right eyebrows, left and right eyes, nose and mouth regions. The input in this module is the ellipse face Egray. It will be then processed to vertical projection to output the four (4) rectangular regions which are Veyebrows, Veyes, Vnose and Vmouth. These four (4) regions will undergo to horizontal projection which will produces six (6) rectangular regions which are HLeyebrow, HReyebrow, HLeye, HReye, Hnose and Hmouth. The peak of the graph indicates that there is a facial feature on that area. They noted the locations of each of the areas which represent each of the regions.

Feature Point Detection

In getting the feature point of the in the input image, the authors used the knowledge-based method. The authors used spatial rules similar to Serhan’s study [12], which provided a robust tracker for facial expression tracking in challenging scenarios involving head movements and occlusions.

Facial Feature Point Tracking

The approach used in this study is the projection [13]. Because of the results and conclusion about their method, wherein the algorithm manages to lessen the tracking problem into three (3) main steps, namely the vertical, horizontal and orientation parameters which are independently estimated, the authors used it for the face tracking module.

Experiment

In the study, the authors used the AR Face Database. The AR Face Database contains a total of 126 people's faces from seventy (70) men and fifty-six (56) women. The resolution of each image is 320x240. The images have frontal view faces with different facial expression, illumination condition and occlusions (sunglasses). For our test setup, real-time snapshots from a webcam and fifty (50) percent of the whole database will be used.

For the occlusions, the authors considered the full occlusion of the right and left eye, the occlusion in the nose and mouth, the wearing of sunglasses and eyeglasses. For the background color, they considered the colors red, brown,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

black, white and complex backgrounds. The illuminations that they considered are fluorescent light and dim lighting. For the pose orientation of the face, they used the pitch and yaw having the angles of -45, 0 and 45 degrees. Figure 1 shows the setup illustration for this work.

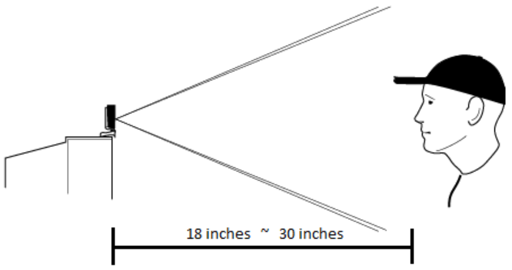

Fig. 1 Setup for Testing

Feature Points Accuracy

The authors reviewed the recorded video images (Real-time) and manually locate and record the location of the eighteen (18) expected feature points. Figure 2 shows the location of the 18 points.

The detected feature points by the prototype were compared to the expected feature points and each image and each point were manually analyzed. The distance of the detected point and the expected point should be less than 3.175mm (12 pixels) for it to consider as valid feature point. The authors used Microsoft Paint to measure the distance between points.

They followed the basic percentage formula for getting the accuracy of each eighteen (18)feature points. The formula is shown as Equation 1:

$$Point\ Accuracy = \frac{Total\ Valid\ Points}{Total\ Video\ Frame} \times 100 \quad (1)$$

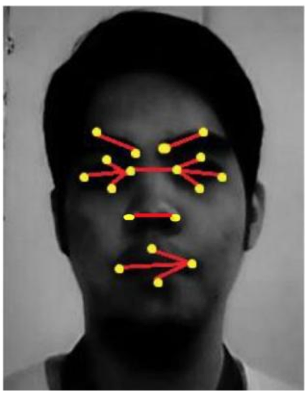

Fig. 2 18 point testing

III. RESULT AND DISCUSSION

The Authors Also Made A Separate Testing Process For The Accuracy Of The Feature Points Detected Within A Real-Time Video. The Test Consisted Of The Same Set Of Occlusions That Were Considered From The Previous Tests (Glasses, Shades, Occlusion On The Right Eye, Left Eye, Nose And Mouth) And Also The One Wherein There Are No Occlusions. The Backgrounds For The Videos, Like In The Real-Time Tests, Are Not Controlled Variables. As Per Previous Tests, They Conducted The Following With Test Subjects That Have Different Type Of Skin Tones (Brown, White And Yellow) From The Other. The Authors Were Able To Acquire A Success Rate Of 28.01% For The Entire Test Data. The Test Data Is Composed Of Test Subjects Having The Skin Color Of White, Brown And Yellowish. They Used Different Backgrounds Such As Red, Black, White, Brown And Complex Backgrounds. They Achieved A Success Rate Of 27.48 % For The Skin Color White, 15.79 % For The Skin Color Of Brown And 17.26% For The Skin Color Yellow.

The following test results are mainly focus on the feature points of the face, namely the: Left point of the left eyebrow, Right point of the left eyebrow, Left point of the right eyebrow, Right point of the right eyebrow, Top point of the left eye, Bottom point of the left eye, Left point of the left eye, Right point of the left eye, Top point of the right eye, Bottom point of the right eye, Left point of the right eye, Right point of the right eye, Left Nostril, Right Nostril,Top point of the mouth, Bottom point of the mouth, Left point of the mouth, Right point of the mouth. Figure 3 shows the illustration for some of the region investigated. Table 1 shows the summary of detected points for no occlusions. Table 2 shows the summary of detected points for with occlusions and Table 3 shows the overall summary detection points.

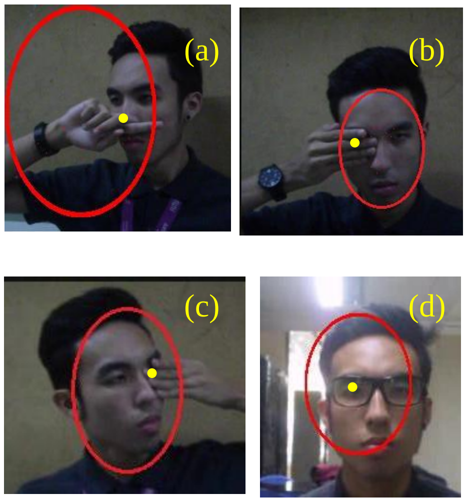

Fig. 3 Detection region investigated(a) nose b) full occlusion (right eye) (c) full occlusion (left eye) (d) eyeglasses

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

Table. 1 Summary of Detected Points (No Occlusions)

Summary (Without Occlusion) Detected Points	Rate %
Left point of the left eyebrow	73.00%
Right point of the left eyebrow	86.62%
Left point of the right eyebrow	94.96%
Right point of the right eyebrow	93.24%
Top point of the left eye	71.80%
Bottom point of the left eye	72.53%
Left point of the left eye	68.46%
Right point of the left eye	68.80%
Top point of the right eye	83.39%
Bottom point of the right eye	82.63%
Left point of the right eye	83.14%
Right point of the right eye	82.88%
Left Nose	65.80%
Right Nose	75.70%
Top point of the mouth	75.84%
Bottom point of the mouth	72.90%
Left point of the mouth	67.39%
Right point of the mouth	70.06%

Table. 2 Summary of Detected Points (With Occlusions)

Summary (with Occlusion) Detected Points	Rate %
Left point of the left eyebrow	50.93%
Right point of the left eyebrow	53.38%
Left point of the right eyebrow	51.54%
Right point of the right eyebrow	50.63%
Top point of the left eye	59.01%
Bottom point of the left eye	59.02%
Left point of the left eye	53.88%
Right point of the left eye	50.36%
Top point of the right eye	58.52%
Bottom point of the right eye	57.81%
Left point of the right eye	55.01%
Right point of the right eye	52.02%
Left Nose	26.90%
Right Nose	27.69%
Top point of the mouth	27.13%
Bottom point of the mouth	26.44%
Left point of the mouth	25.24%
Right point of the mouth	25.51%

Table. 3 Overall Summary of Detected Points

Summary (Overall) Detected Points	Rate %
Left point of the left eyebrow	61.97%
Right point of the left eyebrow	70.00%
Left point of the right eyebrow	73.25%
Right point of the right eyebrow	71.93%
Top point of the left eye	65.40%
Bottom point of the left eye	65.77%
Left point of the left eye	61.17%
Right point of the left eye	59.58%
Top point of the right eye	70.95%
Bottom point of the right eye	70.22%
Left point of the right eye	69.07%
Right point of the right eye	67.45%
Left Nose	46.35%
Right Nose	51.70%
Top point of the mouth	51.49%
Bottom point of the mouth	49.67%
Left point of the mouth	46.31%
Right point of the mouth	47.78%

IV. CONCLUSION

Overall, the large skin range is the main source of why the detection and tracking fails and succeeds. The large skin range is able to identify the skin from other colors but when it comes to backgrounds that have similar value to skin tones, that is the time when the algorithm is having a hard time on identifying which is a part of the skin and which is a background. The authors recommend that the skin segmentation would need more attention and improvement to be able to attain a higher accuracy rate of detection and tracking. For future work, they intend to model other properties like facial hair that still cause some problems as well. Aside from the occlusion with the eyes, there are also possible other factors that might have been the reason for the low accuracy result of the detection or tracking process. In which should be considered for further improvement of the study.

REFERENCES

1. A.K. Jain and S.Z. Li, *Handbook of Face Recognition* (Springer, 2011).
2. S.P. Khandait, P.D. Khandait, and R.C. Thool, Int. J. Recent Trends Eng. **2**, 179 (2009).
3. J.H. Lai, P.C. Yuen, W.S. Chen, S. Lao, and M. Kawade, in Recognition, Anal. Track. Faces Gestures Real-Time Syst. 2001. Proceedings. IEEE ICCV Work. (IEEE, 2001), pp. 168–174.
4. C. Kotropoulos and I. Pitas, in *Image Process. 1997. Proceedings., Int. Conf.* (IEEE, 1997), pp. 105–108.
5. L.S. Balasuriya, (2000).
6. R. Padilla, C.F.F. Costa Filho, and M.G.F. Costa, World Acad. Sci. Eng. Technol. **64**, (2012).
7. H.K. Saini and O. Chand, Skin (Los. Angeles). **3**, 1781 (2013).
8. D. Vukadinovic and M. Pantic, in *Syst. Man Cybern. 2005 IEEE Int. Conf.* (IEEE, 2005), pp. 1692–1698.
9. L. Zhao and J. Tao, in Int. Conf. Image Vis. Comput. Hamilton, New Zeal. (2007), pp. 7–12.
10. A.R. Dixon and E.E. Telles, Annu. Rev. Sociol. (2017).
11. Q. Wu and W. Yang, in *Comput. Vis. Concepts, Methodol. Tools, Appl.* (IGI Global, 2018), pp. 397–420.
12. S. Coşar and M. Çetin, Image Vis. Comput. **29**, 335 (2011).
13. Z. Huang, R. Wang, S. Shan, and X. Chen, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2015), pp. 140–149.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication