Fast Method for Multiple Human Face Segmentation in Color Image

Rudy Adipranata[1], Eddy[2], Cherry G. Ballangan[3], Ronald P. Ongkodjojo[4]
Informatics Dept., Petra Christian University, Surabaya, Indonesia
E-mail : rudya@petra.ac.id[1], yong_jenz@yahoo.com[2], cherry@petra.ac.id[3], ronaldpaska@gmail.com[4]

Abstract

Face segmentation has been used in many different applications, such as face tracking, video surveillance in security control system and human computer interface. In this paper we propose a two-stage face segmentation system that consists of face position detection and facial feature extraction. From the experimental results, the system produced a facial detection success rate of approximately 82.7 percent.

Keywords: *face segmentation, face position detection, facial feature extraction*

1. Introduction

Human face detection and segmentation is an active research area until recently. This field of research plays an important role in many applications such as face identification system, face tracking, video surveillance and security control system, and human computer interface. Those applications often require segmented human face which is ready to be processed. There are many factors that influence the success of human face detection and segmentation. Those factors include complex color background, condition of illumination, change of position and expression, rotation of head, and distance between camera and subject. Many methods have been proposed and developed for human face segmentation, such as using neural networks [1, 4, 5], Eigen faces with background learning [2], statistical approach [5], fuzzy pattern matching [10], and color information and geometric knowledge [6, 7, 8, 9, 11, 12].

In this paper we propose a fast method for human face detection and segmentation that is modified from the methods proposed by Hsu [8] and Wang [11]. Our method consists of two stages: face position detection and facial feature extraction. Face position detection is to obtain face candidates region which later will be applied to the next stage. Facial feature extraction is to obtain the area of facial feature such as the two eyes and the mouth. Our contribution is the addition of rechecking convex hull in order to achieve better result.

2. Face Position Detection

The overall process of the segmentation system can be seen at Figure 1.

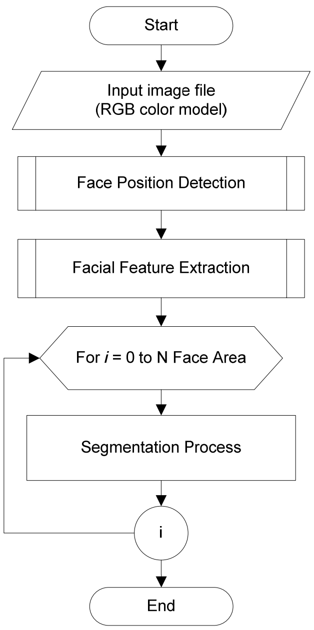


Figure 1. Flowchart of the segmentation system

The initial process to obtain the location of face candidates is skin color detection. In this process, analysis will be done in two colors model, the normalized RGB (Red, Green, Blue) and HSV (Hue, Saturation, Value). In general, the color from a pixel is obtained from the combination between values R (Red), G value (Green), and value B (Blue), where each of the valuable component is in the range of 0 to 255. Normalization value from every obtainable color component is then obtained by using the following formula:

$$r = \frac{R}{I} \quad g = \frac{G}{I} \quad b = \frac{B}{I} \qquad (1)$$

$$I = R + G + B$$

R, G, B = original color of red, green and blue respectively.

r, g, b = normalized color of red, green and blue respectively where r + g + b = 1

Hereinafter, analysis process is carried out in HSV (Hue, Saturation and Value) format. HSV color model selected to be used in this analysis process because it is assumed to be the closest approach of human perception for a color, in this case the skin color. For the conversion from RGB color model into HSV color model, we used equation (2) as follow.

$$H1 = \cos^{-1}\left\{ \frac{0.5\left[(R-G)+(R-B)\right]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\},$$

$$H = \begin{cases} H1, & if \;\; B \le G \\ 360° - H1, & if \;\; B > G \end{cases} \qquad (2)$$

$$S = \frac{Max(R,G,B) - Min(R,G,B)}{Max(R,G,B)},$$

$$V = \frac{Max(R,G,B)}{255}$$

From the result of analysis by using combination between two color models, we can determine whether a part of a picture is including skin-color or not. Acceptable boundaries for human skin color used in this method is as follow [11]:

$$\begin{aligned} &0.36 \le r \le 0.465, \quad 0.28 \le g \le 0.363, \\ &0 \le H \le 50, \ 0.20 \le S \le 0.68, \ 0.35 \le V \le 1.0 \end{aligned} \qquad (3)$$

After determining the skin color, the next process is grouping or merging the parts to form face candidate region by using 8-connectivity neighbourhood. A region is considered as a face candidate region if the height to width ratio falls within the range of golden ratio $= (\sqrt{5}+1)/2 \pm tolerance$ while the tolerance value is 0.65 [12]. For every face candidate region found, we keep the region coordinate for later processing. We show the face position detection flowchart in Figure 2.

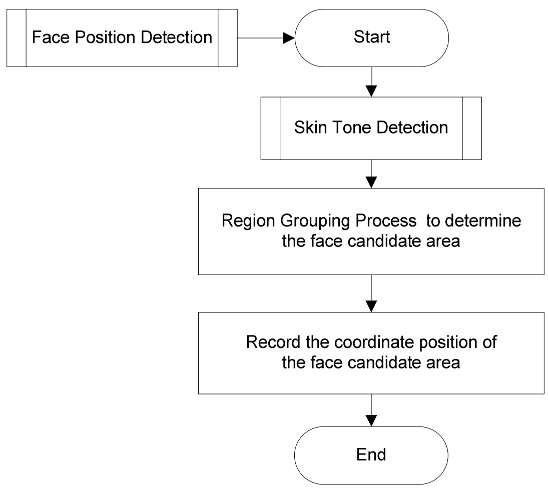


Figure 2. Flowchart of face position detection

3. Facial Feature Extraction

Facial feature extraction is a stage to obtain region candidates from facial feature, in this case the two eyes and the mouth. The eyes and the mouth are selected as facial feature because they have different characteristics compared to other part of face. Eyes have circular form and their colors are darker compared to human face skin color, and also mouth has a shape like an ellipse and its color is different from the skin color. After obtaining some regions for eye and mouth candidates, there will be a selection process to determine true eye and mouth region. The facial feature extraction will be performed for each face candidate region obtained from the face position detection phase.

Before extracting the facial feature, we mask the face candidate region using convex hull algorithm. The face mask is used to achieve a faster facial feature extraction process. The example of this face mask can be seen in Figure 3.

(a) (b) (c)

Figure 3. (a) Face candidate region; (b) Mask; (c) Result of face masking

In addition to convex hull algorithm, which is used to obtain a convex area of the face candidates, there is an optional process to solve some problem that cannot be solved by ordinary convex hull algorithms. This process is called rechecking convex hull process.

Rechecking convex hull is a process to recheck every pixel value on the mask region area which is resulted from the ordinary convex hull, whether the pixel is included in the human skin tone criteria or not by using equation 3. As long as the rechecking convex hull process does not find any pixels that are included in human skin tone criteria, the process will replace those pixels with a certain value of color which has been determined before. If the pixel has exactly the same value with the background pixel, the process will replace it with nothing. This rechecking convex hull process will stop if it has found a pixel which is included in human skin tone criteria. Rechecking convex hull process has two types of direction, vertical direction and horizontal direction.

The flowchart of the facial feature extraction can be seen in Figure 4.

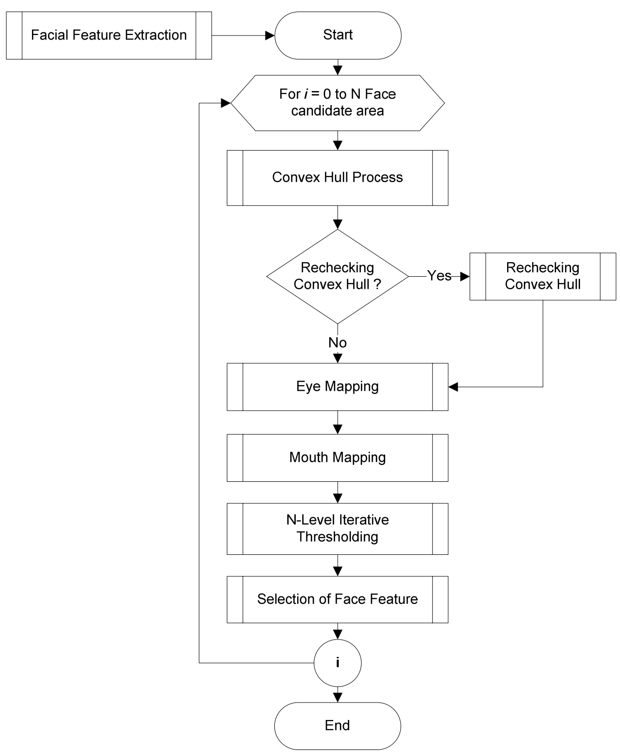


Figure 4. Flowchart of facial feature extraction

To perform the facial feature extraction, the map of eyes and mouth are used [9]. The eye maps are constructed from luminance and chrominance of the image using YCbCr color model. For conversion from RGB to YCbCr color model, we use equation (4).

$$\begin{aligned} Y &= 0.299 * R + 0.587 * G + 0.114 * B \\ Cr &= (R - Y) * 0.713 + delta \\ Cb &= (B - Y) * 0.564 + delta \end{aligned} \qquad (4)$$

The formula to construct eye maps are as follows [8]:

$$EyeMapC = \frac{1}{3}\left\{ (Cb^2) + (\tilde{C}r^2) + (Cb / Cr) \right\} \qquad (5)$$

$Cb^2, \tilde{C}r^2, Cb/Cr$ are normalized between 0 and 255, while $\tilde{C}r^2$ is the negative value of Cr ($\tilde{C}r^2 = 255 - Cr$).

$$\begin{aligned} EyeMapL &= \frac{Y(x,y) \oplus g_\sigma(x,y)}{Y(x,y) \Theta g_\sigma(x,y) + 1} \\ g_\sigma &= \begin{cases} |\sigma|.(|1 - (R(x,y)/\sigma)^2|^{1/2}) - 1 & R \le |\sigma| \\ -\infty & R > |\sigma| \end{cases} \\ R(x,y) &= \sqrt{x^2 + y^2} \\ \sigma &== \left\lfloor \sqrt{W.H} \middle/ (2.Fe) \right\rfloor \end{aligned} \qquad (6)$$

g_σ is a hemispheric structuring element which will be applied as the structuring element in gray-scale dilation and gray-scale erosion. R(x,y) is the euchlidean distance from the central point towards every point of the structuring element. W and H are width and height of face candidate area respectively, while Fe is maximum ratio between the size of the eye and the face. The eye map from lumninance and chrominance are then combine by using AND operation.

Mouth map is constructed by using equation (7).

$$\begin{aligned} MouthMap &= Cr^2.(Cr^2 - \eta.Cr / Cb)^2 \\ \eta &= 0.95.\frac{\frac{1}{n}\sum_{(x,y)\in G} Cr(x,y)^2}{\frac{1}{n}\sum_{(x,y)\in G} (Cr(x,y)/Cb(x,y))^2} \end{aligned} \qquad (7)$$

$Cb^2, Cb/Cr$ are normalized between 0 and 255, while η is ratio of the Cr^2 average value to Cr/Cb average value. n is the number of pixel in the face candidate area.

The next process is to select the true eye and mouth from the candidates. In general, the selection can be divided into two parts, iterative thresholding and calculation of the eye-mouth triangle weight [8].

For iterative thresholding, there are important factors that need to be considered: the level of iteration and the threshold value. The level of iteration can be calculated using the formula:

$$\begin{aligned} L &= Max\left\{ \lceil \log_2(2\sigma) \rceil, \lfloor \log_2(Min(W,H)/Fc) \rfloor \right\} \\ \sigma &= \left\lfloor \sqrt{W.H} \middle/ (2.Fe) \right\rfloor \end{aligned} \qquad (8)$$

W and H are wide and height of the face candidate area respectively, while Fe is the maximum ratio between the size of the eye and the face. Fc is the minimum expected size of face candidate.

Threshold value used in the iterative thresholding can be calculated using the equation (9).

$$T = \frac{\alpha}{n} \sum_{(x,y)\in FG} Map(x,y) + (1-\alpha) . \max_{(x,y)\in FG} Map(x,y) \quad (9)$$

Map(x,y) is the value from facial feature map, in this case is EyeMap and MouthMap which has been obtained before. While α is a parameter value for calculating the threshold, where the value would be 0.5 for initial threshold and 0.8 for the stopping threshold.

After iterative thresholding, the second part for determining the true eyes and mouth is carried out by calculating the triangle weight of the two eyes and the mouth as follows [8]:

$$tw(i,j,k) = emw(i,j,k) . ow(i,j,k)$$
$$emw(i,j,k) = \frac{ew(i,j) + mw(k)}{2} \quad (10)$$
$$ew(i,j) = \frac{EyeMap(X_i,Y_i) + EyeMap(X_j,Y_j)}{2 . EyeMap(X_m,Y_m)} \quad i > j;\ i,j \in [1, N_{eye}]$$
$$mw(k) = \frac{MouthMap(X_k,Y_k)}{MouthMap(X_m,Y_m)} \quad k \in [1, N_{mouth}]$$
$$ow(i,j,k) = \prod_{r=1}^{2} e^{-3(1-\cos^2(\theta_r(i,j,k)))}$$

i and j are the i-th and j-th eye candidates, while k is the k-th mouth candidate

$tw(i,j,k)$ is the triangle weight

emw(i,j,k) is the eye-mouth weight

ew(i,j) is the eye weight

mw(k) is the mouth weight.

ow(i,j,k) is the face-orientation weight

$EyeMap(X_m,Y_m)$ is the eye map value from eye candidate having highest response or most significant eye candidate

$MouthMap(X_m,Y_m)$ is the mouth map value from mouth candidate having highest response or most significant mouth candidate

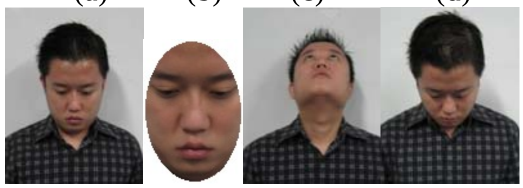
The combination between two eyes candidates and one mouth candidate that have the biggest triangle weight will be selected. And the face candidate region that has those selected combination of two eyes and mouth candidate will be selected as the true human face area.

4. Experimental Result

The experiments were carried out using 25 pictures with various number and positions of face, background and illumination. These pictures also involve faces that wear eyeglasses.

The first experiment was performed to one object with different face position either horizontal or vertical. Some of the results are shown in Figure 5.

(a) (b) (c) (d)



(e) (f) (g) (h)

Figure 5. (a,b,e,f) Succesful segmentation; (c,d,g,h) Failed to be segmented

The result of those experiments indicates that face will be detected if the maximum rotation is 45 horizontal or vertical degrees as seen in Figure 5.

Other experiment was done to objects that wear eyeglasses as shown in Figure 6. This experiment results show that faces can be segmented successfully.

Figure 6. Experiment object wearing eyeglasses

Some experiments on pictures with strong illumination were also carried out as can be seen in Figure 7.

Figure 7a. Experiment with strong illumination

Figure 7b. Experiment with normal illumination

The result of experiment using strong illumination (Figure 7a) still can do segmentation successfully, but face detected is imperfect. On the other side, the result of experiment using normal illumination (Figure 7b) has the better result of face segmentation. The illumination factor affects the result of skin tone detection process and the mapping of facial features, which are the Eye Mapping and Mouth mapping.

The next experiment was done at objects having background color similar to skin color. Result of this experiment indicates that face can be detected successfully as shown in Figure 8.

Figure 8. Experiment using background similar to skin color

Other experiment was done using multiple objects and different background as shown in Figure 9. This figure shows that the method failed to detect one of the eight face objects..

Figure 9. Experiment using multiple objects

On the next experiment, we will show the benefit of using rechecking convex hull method compared to the ordinary one.

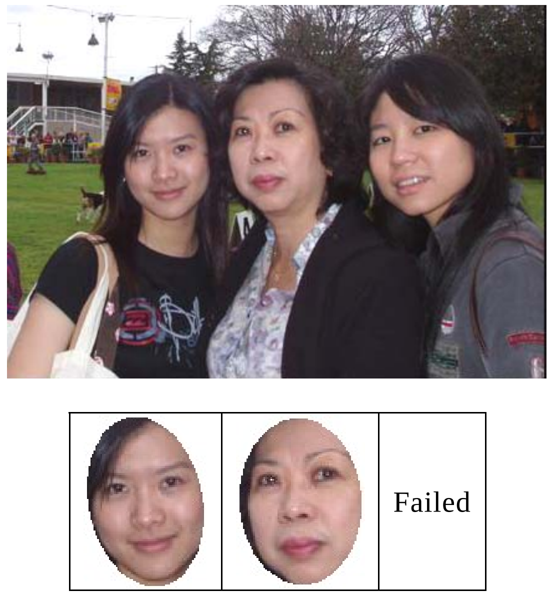


Figure 10a. Experiment using ordinary convex hull

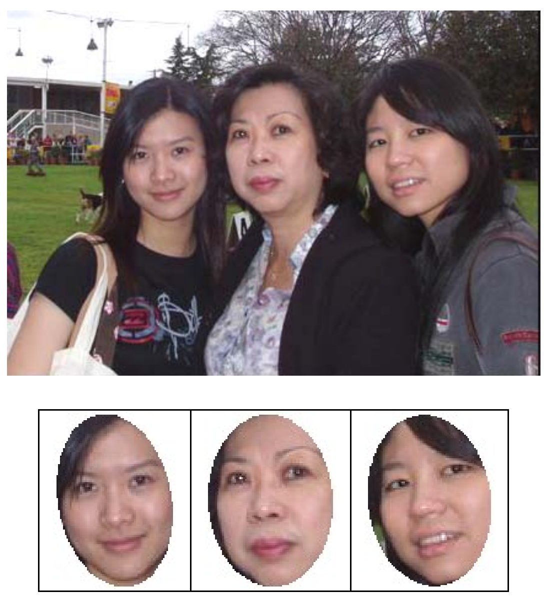


Figure 10b. Experiment using convex hull with additional rechecking process

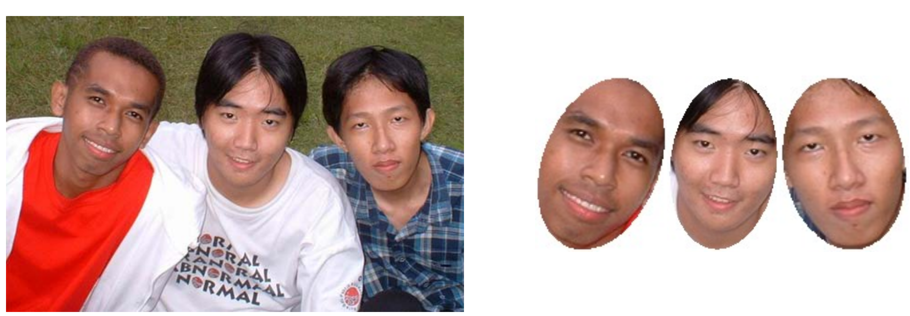


Figure 11. Another examples of face segmentation process

From all experiments that involved 75 faces from 25 images, the face detection with ordinary convex hull method was able to detect 39 faces, 10 faces detected were wrong and 26 faces were failed to be detected. The successful rate of eye detection is 52.00 percent and mouth detection is 60.00 percent. If the experiments were tested using the detection with rechecking convex hull method, there were 62 faces that is successfully detected, 3 face detected were wrong and 10 faces were failed to be detected. The successful rate of eye detection is 81.33 percent and mouth detection is 85.33 percent. Those experiments were tested on Pentium D 3.40 GHz, 1 GHz of RAM using Microsoft Windows XP Home Edition, also showed that the average time of the detection process using convex hull methods is 8.69 seconds (with total time 339.197 seconds) to find one valid face object. If the detection process were done using the rechecking convex hull method, the average time will be 5.43 seconds (with total time 337.142 seconds). We show the complete experiment result of face segmentation using convex hull and rechecking convex hull in Table 1 and Table 2.

Table 1. Experiments result of face segmentation using convex hull

No.	Source Image			FOUND				
	∑ Face	∑ Eye	∑ Mouth	∑ FACE Valid	∑ FACE Invalid	∑ Eye	∑ Mouth	Time Elapsed
1	2	4	2	2	0	3	1	10.61
2	3	6	3	3	0	5	3	8.984
3	3	6	3	3	0	6	3	10.078
4	2	4	2	1	1	3	2	15.044
5	3	6	3	2	0	4	2	4.903
6	3	6	3	1	1	3	2	24.766

No.	Source Image			FOUND				
	∑ Face	∑ Eye	∑ Mouth	∑ FACE		∑ Eye	∑ Mouth	Time Elapsed
				Valid	Invalid			
7	8	16	8	0	1	0	1	39.406
8	2	4	2	2	0	4	2	5.891
9	4	8	4	0	0	0	0	10.234
10	3	6	3	2	1	4	2	34.675
11	2	4	2	2	2	4	2	20.562
12	3	6	3	3	0	5	3	8.406
13	3	6	3	2	0	2	2	7.012
14	1	2	1	0	1	1	1	1.843
15	1	2	1	1	0	2	1	7.297
16	1	2	1	0	1	1	1	2.656
17	4	8	4	4	0	8	4	5.479
18	2	4	2	0	0	0	0	13.875
19	2	4	2	0	0	0	0	19.322
20	4	8	4	0	0	0	0	14.382
21	5	10	5	3	0	6	3	3.868
22	6	12	6	2	0	4	2	14.781
23	2	4	2	2	0	4	2	18.358
24	4	8	4	2	2	5	4	36.218
25	2	4	2	2	0	4	2	0.547
Total	**75**	**150**	**75**	**39**	**10**	**78**	**45**	**339.197**

Table 2. Experiments result of face segmentation using rechecking convex hull

No.	Source Image			FOUND				
	∑ Face	∑ Eye	∑ Mouth	∑ FACE		∑ Eye	∑ Mouth	Time Elapsed
				Valid	Invalid			
1	2	4	2	1	0	2	1	10.203
2	3	6	3	3	0	5	3	8.219
3	3	6	3	3	0	6	3	10.797
4	2	4	2	2	0	4	2	15.156
5	3	6	3	3	0	6	3	5.203
6	3	6	3	3	0	6	3	25.5
7	8	16	8	6	0	12	6	40.406
8	2	4	2	0	2	2	1	5.375
9	4	8	4	3	0	6	3	10.575
10	3	6	3	3	0	6	3	34.813
11	2	4	2	2	0	3	2	19.563
12	3	6	3	3	0	4	3	7.656
13	3	6	3	3	0	6	3	7.75
14	1	2	1	1	0	2	1	1.281
15	1	2	1	1	0	2	1	7.818
16	1	2	1	0	0	0	0	2.984

No.	Source Image			FOUND				
	∑ Face	∑ Eye	∑ Mouth	∑ FACE Valid	∑ FACE Invalid	∑ Eye	∑ Mouth	Time Elapsed
17	4	8	4	0	0	0	0	2.875
18	2	4	2	2	0	4	2	13.984
19	2	4	2	1	1	3	2	19.922
20	4	8	4	4	0	8	4	14.765
21	5	10	5	5	0	10	5	3.719
22	6	12	6	6	0	12	6	13.703
23	2	4	2	2	0	4	2	18.235
24	4	8	4	4	0	7	4	35.703
25	2	4	2	1	0	2	1	0.937
Total	**75**	**150**	**75**	**62**	**3**	**122**	**64**	**337.142**

7. Conclusion

From the experimental results, the successful rate of the system to detect face with the convex hull method is approximately 52.00 percent and the false negative of the system is approximately 13 percent. The successful rate of the system to detect face with rechecking convex hull method is approximately 82.7 percent and the false negative of the system is approximately 4 percent. The system could not detect a face that has rotation larger than 45 horizontal or vertical degrees. Illumination is one of the factors that affect the result of human skin detection and facial feature extraction. The image source with strong or weak illumination needs to be preprocessed to obtain normal illumination before the face detection process is performed.

8. References

[1] R. Feraud, O.J. Bernier, E. Viallet and M.Collobert, "A Fast and Accurate Face Detection Based on Neural Network", *IEEE Trans. Pattern Analysis and Machine Intelligence* vol. 23, no. 1, 2001, pp.42-53.

[2] R. K. Singh and A. N. Rajagopalan, "Background learning for robust face recognition", *Int'l Conference on Pattern Recognition*, 2002.

[3] J. Yang and A.Waibel, "A Real-time Face Tracker", *Proceedings of WACV'96*, 1996, pp. 142–147.

[4] H.A. Rowley, S. Baluja and T.Kanade, "Neural Network-Based Face Detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, Jan 1998, pp. 23-38.

[5] H.A. Rowley, S. Baluja and T.Kanade, "Rotation Invariant Neural Network-Based Face Detection", *Proc IEEE Conf. Computer Vision and Pattern Recognition*, 1998, pp 38-44.

[6] C.H. Lee, J.S. Kim, K.H. Park, "Automatic Human Face Location in a Complex Background Using Motion and Color Information", *Pattern Recognition*, vol. 29, no. 11, 1996, pp. 129-140.

[7] K. Sobottka, I. Pitas, "A Novel Method for Automatic Face Segmentation, Facial Feature Extraction and Tracking", *Signal Process. Image Communication*, vol. 12, no. 3, 1998, 263-281.

[8] Hsu, Rein-Lien, Abdel-Mottaleb, Mohamed, Jain, and Anil K. "Face Detection in Color Images". *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, 2002, pp. 696-706.

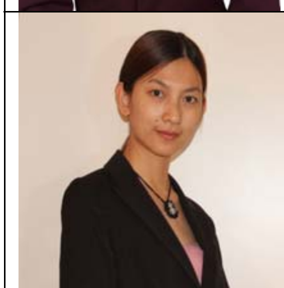
[9] N. Herodotou, K.N. Plataniotis, A.N. Venetsanopoulos, "Automatic Location and Tracking of The Facial Region in Color Video Sequences", *Signal Process. Image Communication*, vol. 14, no. 5, 1999, pp. 359-388.

[10] H. Wu, Q. Chen, M. Yachida, "Face Detection from Color Images using a Fuzzypattern Matching Method", *IEEE Trans. Pattern Analysis and Machine Intelligenc*, vol. 21, no. 6, 1999, 557-563.

[11] Wang, Yanjiang and Yuan, Baozong. "A Novel Approach for Human Face Detection from Color Images under Complex Background". *IEEE Computer Society*. Elsevier Science Ltd., 2001.

[12] Sandeep, K and Rajagopalan, A.N, “Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information”, *ICVGIP*, 2002.

Authors

	**Rudy Adipranata** received his B.Sc in Electrical Engineering from Petra Christian University, Surabaya, Indonesia in 1997, and master degree from Graduate School of Software, Dongseo University, Busan, South Korea in 2004. Currently, he is working as a Lecturer in Informatics Department, Petra Christian University, Surabaya, Indonesia. His research interests are image processing and computer vision.
	**Eddy Soewandojo** was born on 29th January, 1985 in Surabaya, Indonesia. He received his bachelor degree in Informatics Department from Petra Christian University, Surabaya, Indonesia in 2007. Since 2007, he joined Petra Christian University as Lecturer in Informatics Department and takes the responsibilities on some subject, such as Algorithms and Programming, Data Structure and Object Oriented Programming. His current research interests are in the areas of image processing, computer vision and biometrics.
	**Cherry Galatia Ballangan** received her B.Sc in Mathematics from Bogor Agricultural University (IPB), Indonesia, in 1999 and Master degree in Applied Information Technologies from the University of Sydney, Australia in 2005. She was a Lecturer in Informatics Department, Petra Christian University, Surabaya, Indonesia and currently working towards a Ph.D degree at the University of Sydney, Australia. Her research interests include image processing, medical image registration and medical image segmentation.
	**Ronald P. Ongkodjojo** was born on 20th April,1984 in Surabaya, Indonesia. He graduated from Informatics Department, Petra Christian University, Surabaya, Indonesia in 2007. Currently he is working as Software Engineer in iWow Communication Pte. Ltd, Singapore. His responsibility are establishing the communications between Locosto and Renesas chip using EMIF to MFI protocol, doing the driver portion for Locosto software such as charging, power consumption and serial port communication, and providing the driver to support Java application.