Lip Image Feature Extraction Utilizing Snake's Control Points for Lip Reading Applications

Faridah, Balza Achmad, Binar Listyana S
Departement of Engineering Physics, Gadjah Mada University, Yogyakarta, Indonesia

Article Info

Article history:

Received Jan 17, 2015
Revised May 4, 2015
Accepted May 26, 2015

Keyword:

Feature Extraction
Lip Feature
Lip Reading
Lip Segmentation
Snake's Countour

ABSTRACT

Snake is an active contour model that catches and locks image edges, then localizes them accurately. The simplest Snake consists of a set of control points that are connected by straight lines to form a closed loop. This paper discusses the application of Snake to find the visual feature of lip shapes. In most previous papers, visual feature of lip shapes is represented by Snake's contour. In this paper, the feature of lip shapes is represented by six control points on lip Snake's contours. By simply utilizing six control points representing one lip Snake's contour, it is expected to reduce the burden on pattern recognition stage. To demonstrate the performance of this method, some analysis has been conducted on the effect of lip conditions and illumination. The results shows that the overall lip feature extraction using the proposed method is better for lips that have more contrast to the surrounding skin, optimum room illumination that gives the best result is in the range of 330-340 lux.

Corresponding Author:

Faridah,
Departement of Engineering Physics,
Gadjah Mada University,
Jalan Grafika 2 Yogyakarta, Indonesia
Email: faridah@ugm.ac.id

1. INTRODUCTION

Communication is very important in our life. Without communication, human beings will not develop as advanced as at present, and a lot of information will not conveyed properly as well. As one form of communication, visual communication becomes important when audio communication is not possible, for example in an environment with a large noise, or when the audio can not be easily heard, such as for deaf people. For such cases, visual communication can be performed by reading the speaker's lip movements. Each syllable pronounced by a person will form a pattern of different lip shape [1].

Current development of digital image processing and pattern recognition technologies allows us to recognize certain objects utilizing visual data to be translated into corresponding information to understand the character of the object, such as in automatic lip-reading system [2]-[6]. Extraction of information or visual traits that exist in lip images is an important part which still becomes a major focus of research and development in automatic lip reading system. The extracted visual information should represent lip movement patterns corresponding to the words spoken by the speaker. The challenge in developing feature extraction method on this system is the difficulty that arises from external disturbances, such as lighting, lip condition, and the way the speaker pronounce words.

Two approaches were used in extracting visual feature of lip movement patterns, namely image-based and model-based approaches [7]. Image-based approach typically uses image information such as size, color and coordinates of pixels, and used as a feature vector. The advantages of this method are this method is relatively quick and it is easy to obtain feature vectors of an image. However, the dimension of the feature vector is usually large, and consequently, it becomes a huge burden in the recognition process. As for the second approach, model-based method is a method that utilizes a model of lip patterns in pronouncing words. In this method, a model is constructed of several model parameters, usually in the form of parameter space. The model needs to be able to provide complete picture of the actual lip shapes as well as lip shape changes while talking only using as few parameters as possible. Some example of model-based method is Snake and Active Contour Models [8], [9], Active Shape Models [10], [11], and Deformable Models [12], [13].

Snake is a simple model-based method to obtain the contours of an object. This method was first developed by Kaas et al [8] which had been applied to extraction of visual characteristics of lips [8], [14]. Snake is an active contour model that catches and locks image edges, then localizes them accurately. The simplest Snake consists of a set of control points that are connected by straight lines to form a closed loop. This paper will discuss the application of Snake to find visual feature of lip shapes. The feature of lip shapes is not the same as lip contour models mentioned in previous papers, but will be represented by six control points on lip contours. This paper will give an overview to the reader, a simple method that can be used to obtain visual features of lip shapes. By simply utilizing six control points representing one lip contour, it is expected to reduce the burden on pattern recognition stage. To demonstrate the performance of this method, some analysis will be conducted on the effect of illumination and lip conditions.

2. RESEARCH METHOD

The proposed method is illustrated in Figure 1. The method consists of three steps, namely lip segmentation, contour extraction, and feature extraction.

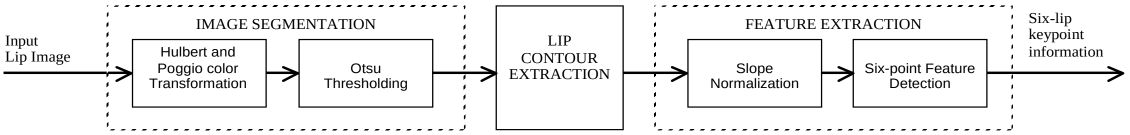


Figure 1. Schematic diagram of the proposed method

2.1. Image Segmentation

Image segmentation method needs to be performed before the extraction process. Image segmentation aims to separate the object (lip, in this case) from the background (skin). The method used in this paper is a combination of Hulbert and Poggio color transformation and Otsu thresholding.

Hulbert and Poggio color transformation is based on differences in color composition between lip as the object and skin as the background. Skin colors are marked more on color composition compare to brightness, even on different people. Color compositions of skins are remarkably constant even when exposed by a lot of illumination. An example of histograms depicting RGB color compositions of lips and skin can be seen in Figure 2 [15]. It can be seen that the difference between red and green for lips is greater than that for skins. Hulbert and Poggio [16] defined the value of the pseudo hue to illustrate this difference, as follows.

$$h(x,y) = \frac{R(x,y)}{G(x,y)+R(x,y)} \tag{1}$$

Otsu thresholding [17], commonly referred to as adaptive threshold, is an automatic thresholding technique. Otsu thresholding is needed to perform image binarization to the image resulted from Hulbert and Poggio color transformations. Otsu method calculates the value of threshold (T) for segmentation based on the input image. The processed image is usually in grayscale format consists of two important parts, namely object and background. Otsu thresholding technique seeks the optimal threshold value to separate object from background by maximizing the variance between classes (object and background) while minimizing the variance within classes. The maximum value of the variance between classes is defined in equation (2).

$$\begin{aligned} \sigma_w^2(T) &= \max\nolimits_t \sigma_w^2(t) \\ \sigma_w^2(t) &= w_1(t)w_2(t)(\mu_2(t)-\mu_1(t)) \end{aligned} \tag{2}$$

Where w_1 and w_2 are the probability of pixels in each class, while μ_1 and μ_2 are the mean grayscale of each class. The probabilities and means for each class are updated iteratively.

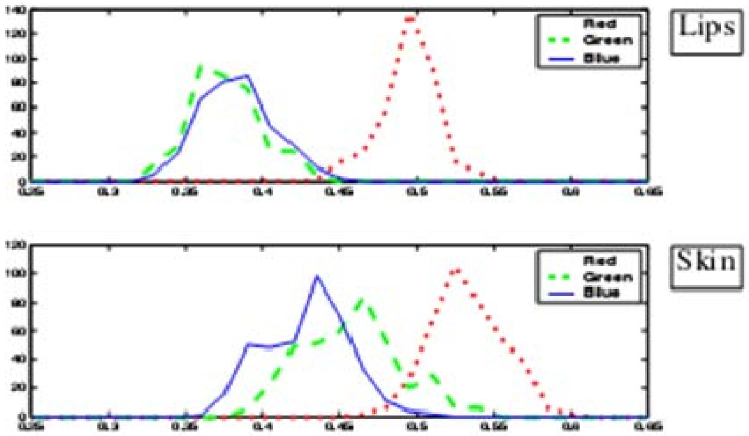


Figure 2. Comparison of skin and lip RGB histograms [15]

2.2. Lip Contour Extraction

Snakes, first developed by Kass et al [8], is a method that uses active contour models to detect certain features in an image. The features are flexible surface curves that can adapt dynamically to the boundary edge of an object. This system consists of a set of points that are interconnected and controlled by splines, as shown in Figure 3. The determination of the object in an image through active contour is an interactive process. The user must estimate the initial contour which is usually set nearly similar to the object features. Furthermore, the contour will be pulled towards the features in the image due to the influence of the internal energy that construct the image.

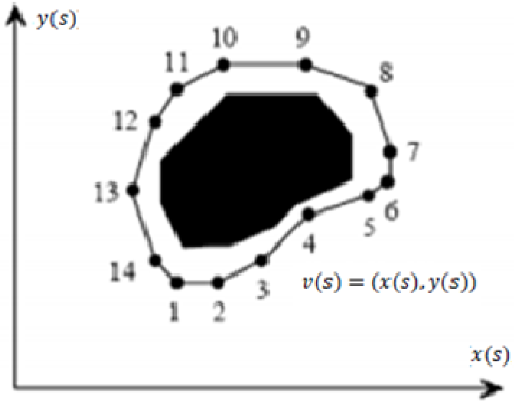


Figure 3. Basic form of an active contour [18]

Active contour is a set of control points in a contour, which parameter are defined as [9],

$$\vec{v}(s) = (\vec{x}(s), \vec{y}(s)) \tag{3}$$

where $x(s)$ and $y(s)$ are coordinates of the control points in the contour, and s is the index of the control points (Figure 3). The Snakes parameter can be expressed as a function of energy, which consists of three kind of energies, namely internal energy ($E_{internal}$), image energy (E_{image}), and constraint energy (E_{const}) [9],

$$E^{*}_{Snake} = \int_0^1 E_{Snake} = \int_0^1 \{E_{internal}(v(s)) + E_{image}(v(s)) + E_{const}(v(s))\} ds \tag{4}$$

Internal energy is due to the elasticity and the bending of splines constructing the Snake.

$$E_{internal} = E_{Elastic} + E_{bending}$$
$$E_{internal} = \alpha(s)\left|\frac{dv(s)}{ds}\right|^2 + \beta(s)\left|\frac{d^2v(s)}{ds^2}\right|^2 \qquad (5)$$

where α is an elasticity constant and β is a bending constant of the contour. The value of α makes the splines act as membranes, whereas β determines the stiffness of the splines. Setting zero to α makes the snake does not care of the length of each spline, while setting zero to β permits the splines to form straight corners.

$$E_{image} = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \qquad (6)$$

where w is the weight of each feature of the object image. The snake will stuck on these features which are tipically the actual contour of the object. The effect of external energy is controlled by a parameter, γ.

2.3. Feature Extraction Lip Image

In this paper, we take six points from outer lip-border as the feature, as illustrated in Figure 4. These points are basically the leftmost, rightmost, upmost, and bottommost points of the lip. The feature points are taken from the Snake points obtained from the contour extraction step. We use 40 control points for the Snake, which gives proper the six-point feature.

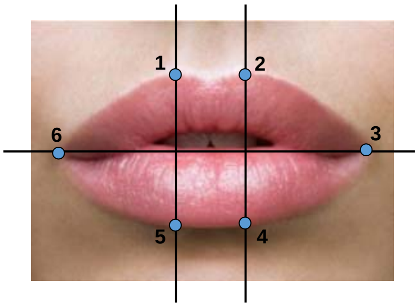


Figure 4. Six-point feature representing lip pattern

In the case of lip that is not upright, it is necessary to untilt the image. This tilt normalization is done by calculating the slope, θ, between the leftmost (point 6) and the rightmost (point 3) points,

$$\theta = \arctan\frac{x2-xt}{y2-yt} \qquad (7)$$

and then rotating the image according to the slope as shown in Figure 5.

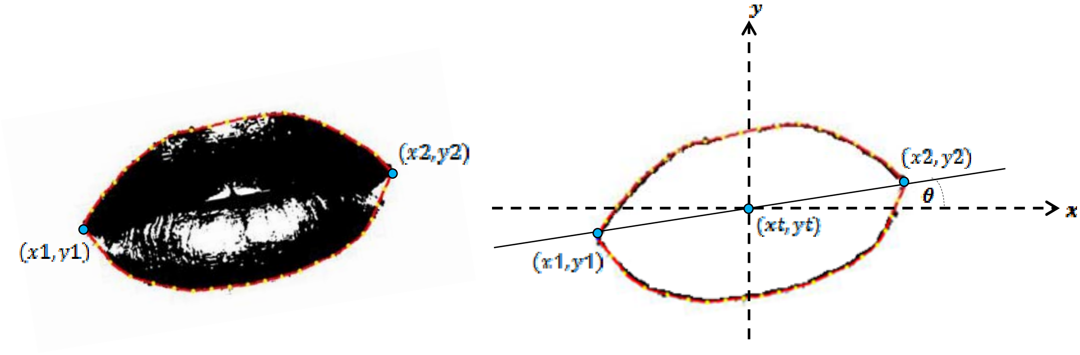


Figure 5. Slope Normalization

2.4. Testing and Analysis

The performance of the proposed feature extraction is expressed by the deviation of the six-point feature to the actual respective coordinates, which is called as extraction error. The extraction error is calculated from the vertical deviation of point 1, 2, 4 and 5, as well as horizontal deviation of point 5 and 6. The measurement method is shown in Figure 6. The extraction error is represented by the average deviation of the six points.

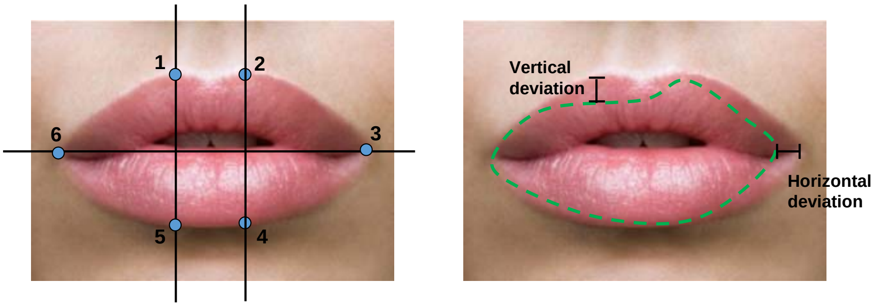
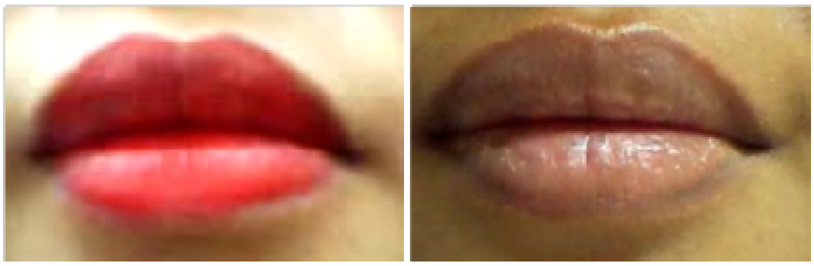


Figure 6. The method of calculating feature extraction error

The analysis will focus on the extraction error under varied light illumination and lip conditions. The lip conditions includes red lips (lips that contrast with the color of the skin) and pale lips (lips with color closes to the color of surrounding skin). An example of each lip condition is given in Figure 7. The room lighting is varied as follow, (a) 100-110 lux, (b) 180-190 lux, (c) 230-240 lux, (d) 330-340 lux, and (e) 380-390 lux.



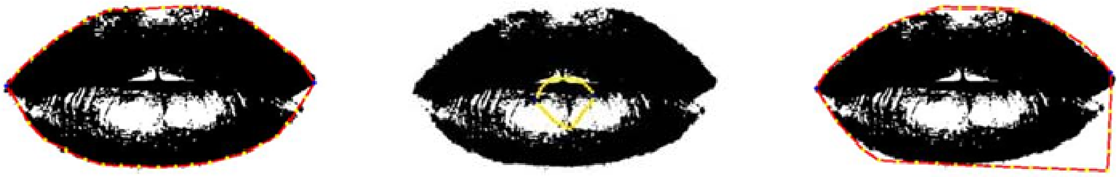
(a) (b)

Figure 7. Two conditions of lips, red (a) and pale (b)

3. RESULTS AND ANALYSIS

3.1. Results of Lip Contour Extraction

Lip contour extraction results are shown in Figure 8. Extraction of the contour is done using Snakes with 40 control points. The success of finding the exact contours in this method depends on the selection of initial contour and control parameters, namely α, β and γ. In this paper, we use ellipses as contour initialization, which is close to the shape of lips. Meanwhile, the control parameters α, β and γ are varied between 0.3 and 1, to find the optimal parameter values. The results indicate that contour extraction can be done properly by the Snake with the same values for all control parameters.

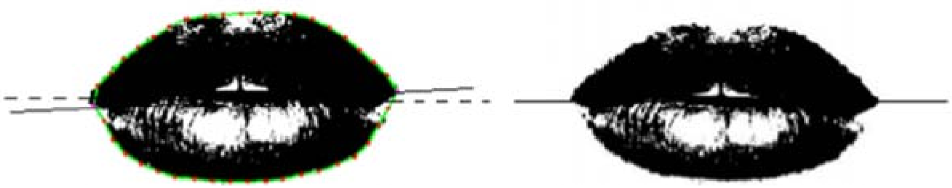


(a) (b) (c)

Figure 8. Contour extraction result by varying the values of α, β, and γ, (a) 0.3; 0.3; 0.3 properly extracted, (b) 0.3; 0.7; 0.3 not properly extracted, (c) 0.3; 0.3; 0.7 not properly extracted

3.2. Results of Lip Feature Extraction

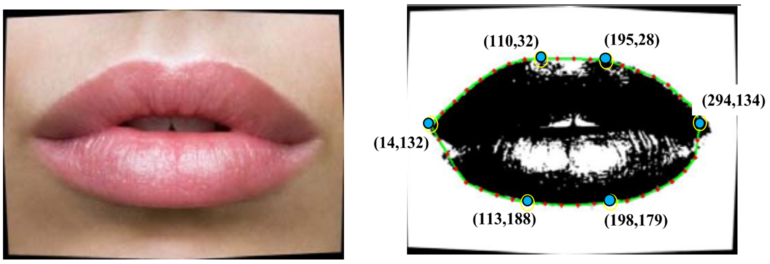
Visual feature extraction is done with the information obtained from the lip contour extraction by utilizing Snake control points. Visual feature extraction starts with slope normalization, with results as shown in Figure 9.



(a) (b)

Figure 9. (a) Tilted lip, (b) uptilted or slope normalized lip

The number of snake points used in this study was as many as 40 points. Out of the 40 Snake control points, six control points are selected as lip feature, as shown by Figure 10.



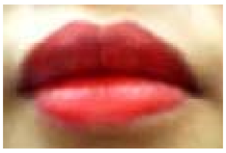
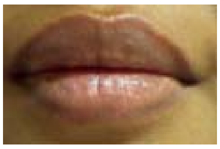
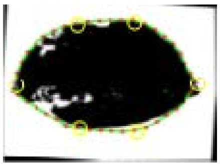
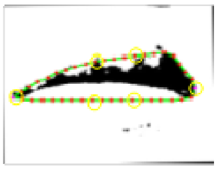
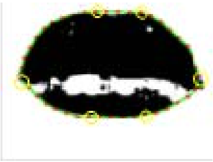
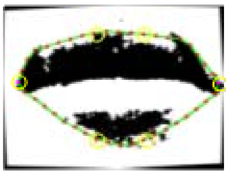
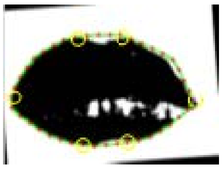
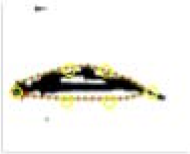
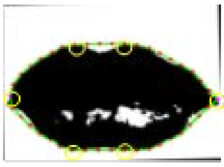
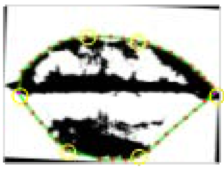
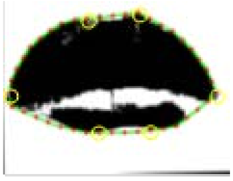
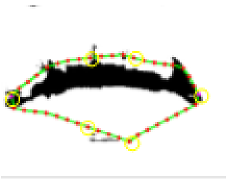
(a) untilted original image (b) extraction result

Figure 10. Example of six points lip feature taken from Snake control points

3.3. Test Results and Discussion

Lip feature extraction has been applied to red and pale lips under different room lighting. The examples of such conditions can be seen in Table 1. Table 1 shows that the proposed feature extraction works better for red lip despite the illumination of the room. Segmentation process is able to properly differentiate the lip from the skin, hence enable the Snakes to detect the edge of the lip and provide control points that are good representation of the lip feature. For pale lips, the segmentation process still can not provide good result for some room illumination. However, the Snake is still able to find fairly good result in 180-190 and 330-340 lux room illumination, although the obtained six-point features are not optimal ones.

Tabel 1. Example of Lip Feature Extraction under Different Lip and Room Lighting Conditions

	Lip condition	
	Red	**Pale**
Room illumination (lux)		
100-110		
180-190		
230-240		
330-340		
380-390		

Quantitatively, the performance of lip feature extraction represented by extraction error can be seen in Table 2. As expected, the extraction errors of red lips are lower than those of pale lips, amounting 5.4 pixels compare to 25 pixels for pale lips. Room lighting also has less effect on the case of red lips.

Tabel 2. Extraction Error Under Different Room Lighting Conditions (In Pixel Unit)

Room illumination (lux)	**Lip condition**		**Extraction error due to room illumination**
	Red	Pale	
100-110	7.0	34	36.3 ± 30.6
180-190	5.9	21	24.3 ± 20.3
230-240	5.7	40	28.2 ± 19.5
330-340	3.2	8.1	20.4 ± 25.7
380-390	5.3	22	21.8 ± 16.4

4. CONCLUSION

Overall lip feature extraction using the proposed method is better for lips that have more contrast to the surrounding skin, with extraction error of 5.4 pixels, compare to 25 pixels for pale lips. Optimum room illumination that gives the best result is in the range of 330-340 lux with extraction error of 20.4 pixels. Manual quantification method for ground truth has uncertaintues of 0.6 in horizontal direction and 2.4 in vertical direction.

ACKNOWLEDGEMENTS

This paper a is part of research grant on Lips Motion Pattern Similarity Analysis on Portable Device for Deaf Kids Speech Therapy in Indonesian Language funded by Universitas Gadjah Mada (2013).

REFERENCES

[1] A. Balza, *et.al.*, "Lip Motion Recognition for Indonesian Syllabie Pronounciation Utilizing Hidden Markov Model Method", *Telkomnika*, 13(1), pp. 173-180, 2015.
[2] E.D. Petajan, "Automatic Lipreading to Enhance Speech Recognition", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 40–47, 1985.
[3] U. Meier, *et.al.*, "Adaptive Bimodal Sensor Fusion for Automatic Speechreading", *Proc. of the International Conference on Acoustics, Speech, and Signal Processing ICASSP'96*, 1996.
[4] K. Mase, *et.al.*, "Automatic Lipreading by Optical Flow Analysis", *Systems and Computer*, 22(6), pp. 67-76, 1991.
[5] C. Bregler, *et.al.*, "Improving Connected Letter Recognition by Lipreading", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 557–560, 1993.
[6] R. Mustafa, *et.al.*, "An Efficient Lip-reading Method Using K-Nearest Neighbor Algorithm", *Telkomnika*, 13(1), pp. 180-186, 2015.
[7] A.W.C. Liew, *et.al.*, "Lip Contour Extraction from Color Images using a Deformable Model", *Pattern Recognition*, 35, pp. 2949-2962, 2002.
[8] M. Kass, *et.al.*, "Snakes: Active Contour Model", *Int. Journal Computer Vision*, 4, pp. 321–331, 1988.
[9] R. Kaucic, *et.al.*, "Accurate, Real-time, Unadorned Lip Tracking", *Proceedings of the 6th International Conference on Computer Vision*, pp. 370 –375, 1998.
[10] J. Luettin, *et.al.*, "Visual Speech Recognition Using Active Shape Models and Hidden Markov Models", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 817–820, 1996.
[11] T.F. Cootes, *et.al.*, "Use of Active Shape Models for Locating Structures in Medical Images", *Image Vision Comput.,* 12(6), pp. 355–365, 1994.
[12] A.W.C. *Liew*, *et.al.*, "Region-based Approach to Robust Lip Contour Extraction", *IEE Electron. Lett.*, 36(15), pp. 1272–1274, 2000.
[13] A.L. Yuille, *et.al.*, "Feature Extraction from Faces Using Deformable Templates", *Int. Journal Computer Vision*, 8(2), pp. 99–111, 1992.
[14] P. Delmas, *et.al.*, "Automatic Snakes for Robust Lip Boundaries Extraction", *IEEE International Conference on Acoustic, Speech, and Signal Processing (ICCASP'99)*, 1999.
[15] N. Eveno, *et.al.*, "A New Color Transformation For Lips Segmentation", *IEEE Workshop on Multimedia Signal Processing (MMSP'01)*, 2001.
[16] A. Hulbert, *et.al.*, "Synthesizing a Colour Algorith from Examples", *Science*, 239, pp. 482-485, 1998.
[17] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram", *IEEE Transaction On Systems, Man, and Cybernetics. Vol. SMC-9, Pontificia Universidance Catolica Do Rio De Janeiro*, 1979.
[18] D. Cremers, et.al., "Diffusion Snakes: Introducing Statistical Shape Knowledge into the Mumford-Shah Functional", *Int Journal of Computer Vision*, 50(3), pp. 295-313, 2002.

BIOGRAPHIES OF AUTHORS

Faridah, is senior lecturer at the Department of Engineering Physics, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia. She received her Bachelor degree in Engineering Physics from Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, and Master degree in Microelectronics from Nanyang Technological University, Singapore. Her research and teaching interests are in Instrumentations and Visual Sensors. She has published research papers in various journals and Conferences.

Balza Achmad, is senior lecturer at the Department of Engineering Physics, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia. His research interests are in Instrumentations, Robotics and Visual Sensors. He is a member at the Center for Robotics and Automation (CentRA) as well as the Integrated and Smart Green Building (INSGREEB), Universitas Gadjah Mada. He has published research papers in various journals and Conferences.

Binar Listyana S, is an engineer at Technology Centre, PT Dirgantara Indonesia. She received her Bachelor degree in Engineering Physics from Universitas Gadjah Mada, Yogyakarta, Indonesia. His research interests are in Instrumentations and Visual Sensors.