Study of Viola Jones Face Detection on Color Image based on Skin Pigmentation Level

Nur Afny C. Andryani

[1] Tanri Abeng University, email: n_afny@tauniversity.ac.id

Abstrak

Metode deteksi wajah secara otomatis menjadi sebuah tantangan tersendiri dalam dunia penelitian. Hal ini mengingat representasi wajah yang tidak sederhana karena bukan merupakan sesuatu hal yang bersifat tegas. Banyak penelitian yang mengarah pada pengajuan algoritma baru untuk melakukan deteksi wajah otomatis. Algoritma Viola Jones seringkali dijadikan tolok ukur dikarenakan algoritma ini dianggap titik awal algoritma deteksi wajah otomatis yang efisien. Algoritma Viola Jones dikembangkan oleh Paul Jones dari Microsoft R&D dan Michael J.Jones dari Mitsubishi R&D sejak tahun 2004. Bebrapa publikasi ilmiah sebelumnya telah banyak menyajikan studi mengenai performa algoritma Viola Jones dengan beberapa kriteria gambar wajah. Akan tetapi, belum pernah ada yang mebahas bagaimana performa algoritma Viola jones berdasarkan pigmentasi kulit wajah yang dideteksi. Tulisan ini menyajikan studi atas algoritma Viola Jones dengan kriteria pembatas jenis pigmentasi kulit wajah. Untuk mengindikasikan pigmentasi kulit wajah, digunakan elemen L* pada ruang warna CIELAB. Pigmentasi kulit wajah diklasifikasikan menjadi gelap, coklat dan terang. Hasil simulasi dan uji statistic menunjukkan bahwa performa algoritma Viola Jones cenderung menurun ketika mendeteksi wajah dengan nilai pigmentasi tinggi (kulit gelap). Beberpa uji hipotesa dilakukan untuk memberikan validasi pada hasil simulasi.

Kata kunci: *Computer Vision, Face detection, Viola Jones Algorithm, L* on CIELAB*

Abstract

Automatic face detection has been very complex and challenging research topic due to the complexity of faces' characteristics that is not rigid object. There have been many works on proposing robust algorithm on image detection. Many researcher use Viola Jones algorithm as their initial point and benchmark. The Viola-Jones face detection itself is the most popular and recent applicable algorithm that has been developed since 2004 by Paul Jones from Microsoft R&D and its co-inventor, Michael J. Jones from Mitsubishi R&D. Many previous works present the study on the Viola Jones algorithm subject to frontal face with no consideration on the skin pigmentation level. This paper presents study on The Viola Jones performance on color image that consider skin pigmentation level. To indicate the skin pigmentation level, the L* element on CIELAB color space is used. The skin pigmentation level is clustered into dark skin, brown skin and fair skin. The simulation result show that the Viola Jones performance tends to decrease when the skin pigmentation getting high (dark skin). Some hypotheses test had been done to support the claim.

Keywords: *Computer Vision, Face detection, Viola Jones Algorithm, L* on CIELAB*

1. Introduction

Face Detection is normally a first part on a series of face recognition. Face detection is an activity to find the location and the size of face object on a given arbitrary image containing an unknown number of faces [6]. Besides its own unique characteristics, the challenging issue on face detection is more commonly stimulated by the uncontrollable environment that could be coming from the image condition (illumination, camera characteristics), occlusion, facial expression (smiling, laughing, crying, and others), or its structural component (presence or absence of beards, moustaches, glasses or other element) [7].

The current demand on face detection algorithm is high accuracy and efficient computational cost. A lot of methods of face detection have been proposed by many researchers. In 2000, Yongmin Li et al proposed SVM (Support Vector Machine) on solving face detection problem combining with Eigen face modelling [8]. Similar work proposed by P.Wang and Q.Ji on more complex scene [9]. As overall, this method performs more accurate detection compared to Eigen space modelling that had been proposed previously by Turk & Petland [10]. However, the drawback of this method is on running time, it is slower [8]. Another approach using Neural Network had been proposed by other researchers [11] [12] [13]. Generally Neural Network approaches achieve high accuracy due to its powerful discrimination ability between face and non-face object. However, the high accuracy can be achieved by having high number of training data [13]. Besides, the method need to process all possible position of different resolution of the original image. These issues lead to the high computational cost. In another hand, method proposed by Turk & Petland performs very fast detection but low in accuracy [13].

Among all proposed method, The Viola Jones face detection algorithm is the most well-known and applicable [6], [2]. It performs high accuracy with quite fast detection. There are three main contribution of this algorithm. The integral image to represent the original image, Adaboost classifier for feature selection and its cascade structure of classifier to reject the non-face object [2]. These three contributions that leads to many other face detection methods [2].

2. Background Theory

2.1   Face Detection

Face Detection is normally a first part on a series of face recognition. Face detection is an activity to find the location and the size of face object on a given arbitrary image containing an unknown number of faces [6]. Besides its own unique characteristics, the challenging issue on face detection is more commonly stimulated by the uncontrollable environment that could be coming from the image condition (illumination, camera characteristics), occlusion, facial expression (smiling, laughing, crying, and others), or its structural component (presence or absence of beards, moustaches, glasses or other element) [7].

The current demand on face detection algorithm is high accuracy and efficient computational cost. A lot of methods of face detection have been proposed by many researchers. In 2000, Yongmin Li et al proposed SVM (Support Vector Machine) on solving face detection problem combining with Eigen face modelling [8]. Similar work proposed by P.Wang and Q.Ji on more complex scene [9]. As overall, this method performs more accurate detection compared to Eigen space modelling that had been proposed previously by Turk & Petland [10]. However, the drawback of this method is on running time, it is slower [8]. Another approach using Neural Network had been proposed by other researchers [11] [12] [13]. Generally Neural Network approaches achieve high accuracy due to its powerful discrimination ability between face and non-face object. However, the high accuracy can be achieved by having high number of training data [13]. Besides, the method need to process all possible position of different resolution of the

original image. These issues lead to the high computational cost. In another hand, method proposed by Turk & Petland performs very fast detection but low in accuracy [13].

Among all proposed method, The Viola Jones face detection algorithm is the most well-known and applicable [6], [2]. It performs high accuracy with quite fast detection. There are three main contribution of this algorithm. The integral image to represent the original image, Adaboost classifier for feature selection and its cascade structure of classifier to reject the non-face object [2]. These three contributions that leads to many other face detection methods [2].

2.2   Viola Jones Algorithm

The Viola Jones algorithm is a method which had been proposed by Paul Viola (Microsoft Research Center) and Michael J.Jones (Mitsubishi Electric Research Center) in 2001 and scientifically published on International Journal of Computer Vision in 2004 [2]. Among all proposed automatic face detection method, Viola-Jones method is the most well-known and most applicable [6]. It dramatically influences the recent research on this field [6].

There are three main contribution of this method, Integral image for feature evaluation, Adaboost classifier for feature selection and cascade structure for rejecting on non-face object [6] [2]. The viola jones methods utilize Haar Like feature to represent the image. It prefers to use feature rather than the pixel directly for more efficient detection reason [2].

The rectangle feature is computed very rapidly using integral image. To calculate the value of the above feature is by subtracting the pixel on the white area with the pixel on the black area. To help calculating the value of the whole feature on the representative image, The viola Jones proposed Integral Image. The integral image at location x, y contain sum of the pixels above and to the left of *x, y*:

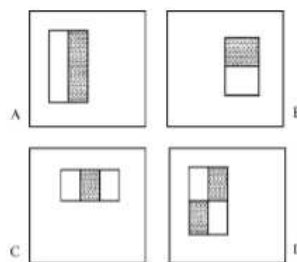$$ii(x, y) = \sum_{x' \le x, y' \le y} i(x, y) \qquad (1)$$

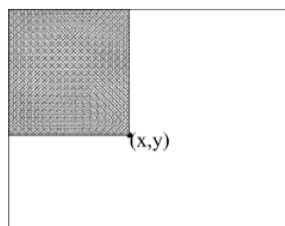**Figure 1. Example of Haar Like feature [2]**

Figure 2. Value of Integral Image at point (x,y) is the sum of all the pixels above and to the left [8]
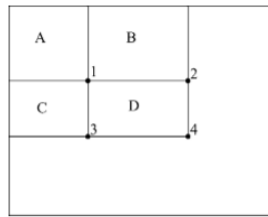
Figure 3. Area for calculating pixel at point (x,y) [2]

Table 1. Sample Segmentation

L*	Group	Number of Sample
18-56	Dark	24
57-69	Brown	26
70-90	Fair	44

Table 2. Performance of Viola Jones Algorithm on Color Image

L*	Group	Number of Sample	Detection Rate	False Positive
18-56	Dark	24	50%	1
57-69	Brown	26	77%	0
70-90	Fair	44	89%	3

$ii(x,y)$ is the integral image and $i(x,y)$ is the original image [8]. Integral image is an image which its pixel value is the sum of the pixel from the left side up to the bottom right side.

To determine the average of pixel value at point (x,y), divide the grey area on figure 2 into four parts [2].

The sum of pixel at location 2 is A+B, the sum of pixel 3 is A+C, the sum of pixel at location 4 is A+B+C+D. The sum of pixel within D is (4+1)-(2+3).

Second contribution of this method is implementing Adaboost Classifier to determine the feature and threshold. The Adaboost classifier accumulate several weak classifier to build strong classifier. Adaboost classifier is a set of filter that can be utilized to segment the image [14]. The characteristics of Viola Jones face detection method is on its cascade structure of the classifier. The cascade structures consist of 3 stages. The objective of each stage is to reject the non-face object. On the first stage, every sub image will be classified by certain Haar Like feature. The first stage result is True if the image satisfy the Haar like feature and False otherwise [14]. On this stage it may revealed around 50 % out of sub images. On the second stage, the classification criteria is the integral image process. The output of the second stage of classifier is true when the sub image satisfy the integral image process and false otherwise [14]. On the last stage, the output of the sub image is true if it satisfies the Adaboost classifier and false otherwise [14].

2.3   Images in RGB and CIELAB

Images can be represented in any color space. It is needed to specify every color of the image in a standard way. RGB and CIELAB are two of the color space.

In RGB color space, any color is represented by intensity values of three primary colors; Red, Green and Blue. Thus, every primary color on an image represented by RGB consist of those three component images [15]. Digital Camera captures and stores the image in RGB format.
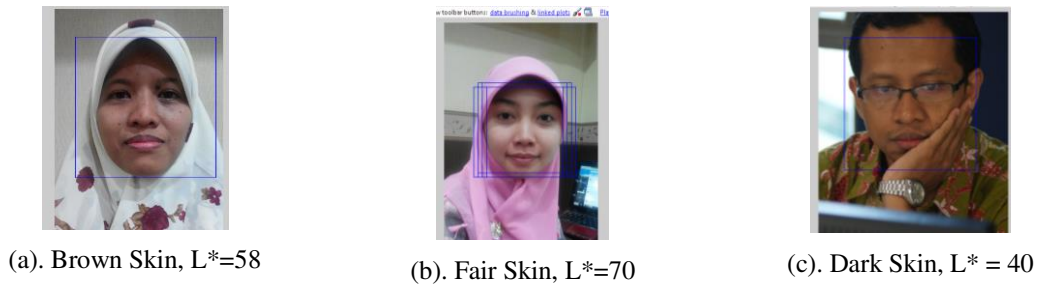
(a). Brown Skin, L*=58   (b). Fair Skin, L*=70   (c). Dark Skin, L* = 40

Figure 4. Successful face detection

(a). Dark Skin, L*=38

(b). Fair Skin, L* = 71

(c). Fair Skin, L* = 65

Most of the eyebrow covered by hair

Figure 5. Failure Detection

Table 3. Information of Sample

Group	Number of Sample	Success (x)	Detection Rate (p)	n(1-p)	np
Dark	24	12	50%	12	12
Brown	26	20	77%	6	20
Fair	44	39	89%	5	39

CIELAB is a color space that was proposed by CIE in 1976. An image in CIELAB color space is represented by three component images; L*, a* and b*. L* represents degree of lightness ranging from 0 to 100, a* represents degree of the redness-greenness (negative value for greenness and positive value for redness), and b* represents degree of yellowness-blueness (negative value for blueness and positive value for yellowness) [16]. Pigmentation level is inversely linear with the L* value (high pigmented skin or dark skin has low L* value and vice verca). Thus it can be utilized to indicate the skin pigmentation level.

3. Simulation set-up

For examining the Viola Jones, 94 color images was taken, both by using Lenovo Camera P780 and randomly from the internet with maximum pixel size 512 x 512. Since the picture is taken randomly from the internet, the characteristics of the image acquisition is vary. The purpose of this study is to examine the performance of The Viola Jones Face Detection algorithm on dark, brown and fair skin. Thus, the RGB picture convert into CIELAB spectrum and taken the L* value to indicate the pigmentation level. The sample segmentation are described as stated on Table 1.

All the taken picture are frontal face with at least 85% are facing forward, single face and no additional accessories such hat, or spectacles.

Simulation are running on MATLAB 2009a with utilizing Haar Cascade feature in OpenCV.

4. Result And Discussion

There are two parameters used for describing the Viola Jones Face Detection algorithm performance; detection rate and false positive. The algorithm is implemented into all the taken sample image and had been analyzed based on the skin pigmentation level. The summary of the simulation result is presented on Table 2.

The simulation result shows that the Viola Jones face detection are having moderate strong ability to do face detection by having very low number of false positive and average of detection rate around 72% for three different skin pigmentation level. However, it can be seen that the detection rate tends to decrease when the L* getting smaller. In another words, the detection ability on the dark skin face is inferior compared to the fair and the brown skin color.

Figure 4 shows some sample that successfully identified by Viola Jones Algorithm for various skin pigmentation level. In another hand, Figure 5 presents some faces that cannot be identified. Early analysis on the unsuccessful identification factors could be the incomplete face component besides the skin pigmentation itself. The incomplete face component can be represented by the covered eyebrow by hair (Figure 5(c)). However, further studies are needed to prove that the incomplete face component influence the detection performance besides the skin level pigmentation.

5. Hypotheis test

To support the claim that the performance of Viola Jones Face Detection tends to decrease subject to high pigmentation level (low L*), some sets of hypotheses are done. The hypothesis test will give population insight of comparing the Viola Jones Face detection performance on between fair and brown skin, fair and dark skin, brown and dark skin. The statistics parameter to be measured is comparing proportion of two population. The set-up and statistics calculation of hypotheses are described below in Table 3 and Table 4.

It is presented on Table 3 that all np and n(1-p) for sampling from three different pigmentation level is above or equal to 5. It means that the statistics P_1-P_2 is approximating normally distributed. Thus z-statistics test is implemented.

Given the Z- statistics and Pooled Proportion as followed [7]:

$$P = \frac{x_1 + x_2}{n_1 + n_2} \tag{2}$$

$$Z = \frac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{3}$$

Derived the pooled proportion and the Z-statistics as presented on Table 4.

The hypotheses set up consists of three proportion comparison to see that the detection rate of lower pigmentation level is better compared to the higher pigmentation level. By using 90% confidence level, or 10% significance level, we can see that the three hypotheses fall in the rejection area. By having $Z_{0.10}$ = 1.28, all the z-statistics value are above $Z_{0.10}$. It means that the null hypothesis is rejected for the three hypotheses set-up.

Table 4. Statistical Test

Set	Hypothesis	Pooled Proportion (P)	Z-statistics
Set 1	$H_0 : P_{fair} = P_{brown}$ $H_1 : P_{fair} > P_{brown}$	0,0842	1,379
Set 2	$H_0 : P_{fair} = P_{dark}$ $H_1 : P_{fair} > P_{dark}$	0,75	3,548
Set3	$H_0 : P_{brown} = P_{dark}$ $H_1 : P_{brown} > P_{dark}$	0,64	1,985

By using 90% confidence, it can be concluded that there are sufficient evidence to support that the Viola Jones face detection performance on fair skin is better compared to brown skin and dark skin. In addition, there is also sufficient evidence to support that the performance of Viola Jones face detection of brown skin is better compared to dark skin.

6. Conclusion

A study on Viola Jones Face Detection algorithm performance on color image based on skin pigmentation level has been done. The skin pigmentation level is indicated by the L* component of CIELAB color space. Based on the simulation result shows that Viola Jones Face Detection algorithm has 89% detection rate when it is applied on fair skin faces (low pigment, high L*) and tends to decrease when the skin pigmentation getting high (low L*, dark skin) . Since the skin pigmentation level in this study is indicated by the L* value on CIELAB color space which stands for the lighting, it can be concluded that the Viola Jones Face Detection algorithm is sensitive to the skin pigmentation level or the degree of lightness. This performance sensitivity subject to pigmentation level is also supported by the hypothesis test. It shows that as overall the Viola Jones Face Detection performance tends to decrease when the pigmentation getting higher or low L*. Thus, more improvement to handle face detection on high pigmented skin (dark skin) is needed.

Some of the simulations indicate that the performance is not only influenced by the pigmentation level but it is also influenced by the completeness of the face components. Some sample shows that even though the skin pigmentation is considered as fair but the algorithm fail to detect the face area. In fact the failure detected image has some face's component missing, such as the covered eyebrow by the hair. The idea of the impact of the face component completeness into the detection rate is still very immature and need to have further investigation.

Daftar Pustaka

[1]   M. Julian, P. Vlad and P. T. Jean, "Face Detection with Boosted Gaussian Features," *PATTERN RECOGNITION, The Journal of The Pattern Recognition Society,* vol. 40, pp. 2283-2291, 2007.

[2]   P. Viola and M. J. Jones, "Robust Real TIme Face Detection," *International Journal of Computer Vision*, pp. 137-154, 2004.

[3]   Faizan, N. Asima and A. Zeeshan, "Image Based Face Detection and Recognition "State of Art"," *International Journal of Computer Science*, vol. 9, no. 6, pp. 169-180, 2012

[4]   Y. Yen, L. Tyng and S. F. Chiou, "Face Detection with Occlusion," in *The 8th European Conference on Computer Vision*, Prague, Czech Republik, 2004.

[5]   M, B. P and R. S. G, "Face Recognition using Neural Network," *International Journal of Scientific and Research Publication,* vol. 3, no. 3, 2013.

[6]   M. Zervos, "Multi Camera Face Detection and Recognition applied to People Tracking," Master Thesis, School of Computer and Communication Science, Ecole Polytechnique Federale de Lausanne, Switzerland, 2013.

[7]   T. Barbu, "An Automatic Face Detection System for RGB Images," *International Journal of Computer, Communication and Control,* vol. 6, no. 2011, pp. 21-32, 2011

[8]   Y. Li, S. Gong, J. Sherrah and H. Liddell, "Multiview Face Detection Using Support Vector Machine and Eigenspace Modeling," in *Forth International Conference on Knowledge based Intelligent Engineering System & Allied Technology*, 2000.

[9]   W. P and J. Q, "Multi view Face Detection Under Complex Scene Based on Combined SVM," in *International Conference on Pattern Recognition*, Cambridge, 2004

[10]   T. M and P. A, "Eigenfaces for Recognition," *Journal of COgnitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.

[11]   F. R, B. O, V. J and C. M, "A Fast and Accurate Face Detector based on Neural Network," *IEEE Transaction on Pattern Analysis and Machine Intelligent,* vol. 1, pp. 42-53, 2001

[12]   Y. Weng and J. Jiang, "Face Detection based on Skin Color in Image by Neural Network," in *International Conference on Intelligent and Advance System*, Kuala Lumpur, 2007.

[13]   L. Huang, A. Shimizu and H. Kobatake, "Face Detection Using a Modified Radial Basis Function Neural Network," in *International Conference on Pattern Recognition*, 2002.

[14]   F. Comaschi, S. Stuijk, T. Basten and H. Corporaal, "RASW a Run Time Adaptive Sliding Window to improve Viola Jones Object Detection," in *7th International Conference on Distributed Smart Camera*, Palm Spring, 2013

[15]   Dani Ihtatho, " Objective Assesment of Area and Erythema of Psoriasis Lesion using Digital Imaging and Colourimetry", *MSc Thesis*, August, 2008

[16]   K.N.Plataniotis and A.N, Venetsanopoulos, "Color Image Processing and Application", *Berlin Springer-Verlag,* 2000.

[17]   Bowerman, O'Connel, Murphree and Orris, "Essential of Business Statistics", *Forth Edition, McGraw-Hill International Edition,* 2012