Emotion Recognition using Fisher Face-based Viola-Jones Algorithm

Kartika Candra Kirana [1], Slamet Wibawanto [2], Heru Wahyu Herwanto [3]
Department of Electrical Engineering
State University of Malang
Jl. Semarang No 5, Malang, Indonesia
kartika.candra.ft@um.ac.id [1], slamet.wibawanto.ft@um.ac.id [2], heru_wh@um.ac.id [3]

***Abstract*—In the form of the image integral, this primitive feature accelerates the performance of the Viola-Jones algorithm. However, the robust feature is necessary to optimize the results of emotion recognition. Previous research [11] has shown that fisher face optimized projection matrix in the low dimensional features. This feature reduction approach is expected to balance time-consuming and accuracy. Thus we proposed emotion recognition using fisher face-based Viola-Jones Algorithm. In this study, PCA and LDA are extracted to get the fisher face value. Then fisher face is filtered using Cascading AdaBoost algorithm to obtain face area. In the facial area, the Cascading AdaBoost algorithm re-employed to recognize emotions. We compared the performance of the original viola jones and fisher face-based viola jones using 50 images on the State University of Malang dataset by measuring the accuracy and time-consuming in the fps. The accuracy and time-consuming of the Viola-Jones algorithm reach 0.78 and 15 fps, whereas our proposed methods reach 0.82 and 1 fps. It can conclude that the fisher face-based viola-jones algorithm recognizes facial emotion as more accurate than the viola-jones algorithm.**

Keywords—facial, emotion recognition, fisher face, FLDA, Viola-Jones algorithm.

I. INTRODUCTION

The Viola-Jones algorithm is often thought of as a rapid processing for face detection. In this algorithms, AdaBoost algorithms classify the rectangular features in "cascade" stages. The rectangle feature is computed rapidly using an integral theorem every shifting of sub-window. In the form of the image integral, this primitive feature accelerates the performance of the Viola-Jones algorithm. [1] The rate of face detection proceeded in 15 frames per second (fps), thus it supports real-time processing.

Despite image integral can be calculated quickly, the representation of integral is less responsive to changes in face angle[2] [3] [4]. These problems decrease the accuracy. In addition, image integral requires a very large derivative feature to find face boundaries, even for thumbnails [5]. In other research contexts, image integrals cannot identify a problem optimally. For example, [6] explains that river image characteristics are almost identical to other images, such as the ocean. Therefore [6] identifies the river using the modified image integrals based on hydromorphology features. This research segment river from trees, roads, roofs, shore and the sky using hydromorphology feature until 70%. However, this system distinguishes between sea and river in an accuracy of 68%.

In emotion recognition research, image integrals cannot identify a problem optimally. [7] proposes the Viola-Jones algorithm to detect human emotion in video sequences. Their results show that proposes the Viola-Jones algorithm only reaches the accuracy in 70%. However, their proposed methods can be computed rapidly. [8] also mentioned that the feature integral result does not work well on emotional recognition, thus they combines the Viola-Jones and Edge-Histogram of Oriented Gradient as feature descriptor identify the emotions of the patients. However, Edge-Histogram of Oriented Gradient cannot describe the frequency distribution with the open class.

Image integrals cannot identify the facial emotion optimally because the differences of human expression in showing emotion. For example, uninterested emotions can be marked by yawning (eyes still open). Some others can be marked by falling asleep (eyes closed). Both of these examples have different image integral values, so we need another robust feature in recognizing emotions.

[9] compared textured features in the emotion recognition using Support Vector Machine (SVM). They show that fisher face presents the emotion more represent than gradient and wavelet. However, SVM computes the multi-class classification slowly. [10] proposed the 2D fisher face which is hoped to find the best projection direction matrix. The result shows that 2D fisher face more accurate than Principal Component Analysis (PCA) and 2D-PCA. However, 2D fisher face computes slower than PCA. [11] implements artificial neural networks and fisher face features to recognize emotions in the learning context. They reach an accuracy of 81%. However, they confirmed that the applying of a complex neural network architecture produce an accurate system that computes slowly. Although fisher face optimized projection matrix in the low dimensional features. This feature needs a longer time than the integral features.

The above research [9][10][11] have been shown that the fisher face feature is a robust feature in the facial emotion recognition. However, they consume a longer time than the primitive features. On the other sides, the Cascaded Adaboost algorithm in Viola-Jones algorithms handles the computing

978-1-5386-8402-3/18/$31.00 ©2018 IEEE

time with reject unnecessary area quickly using hard limit [2][3]. The combination of fisher face and the Cascaded Adaboost algorithm is expected to balance time-consuming and accuracy in facial emotion recognition, thus we propose emotion recognition using fisher face-based viola-jones algorithm. In this study, our contribution is modifying the features of fisher face as a substitute for feature rectangle to detect faces and emotions simultaneously.

II. Facial Emotion

The emotion can be performed using, gestures, speech, control, and facial expression. However, the most of popular approach in emotion recognition is the acquisition of facial-based features [12]. The emotion is divided into the positive emotion and the negative emotion. The emotion is also divided into neutral, happiness, sadness, anger, fear, surprise and disgust [12]. In this research, we applied our proposed method in the learning environment, thus the basic emotion cannot present the learning specifically. There are two popular learning emotion approaches consisting of Russell's Circumplex approach and the affective and constructive learning-based approach. Russell's Circumplex Approach classify emotion into interest, engagement, confusion, frustration, satisfaction, hopefulness, boredom, and disappointment [13]. In addition, the affective and constructive learning-based approach determine emotions more simply. It separated the emotion into four types. First, the student feels interested in a matter of learning. Second, the student began to have confusion and difficulties. Thirst, student emotions start negative when feeling there is a frustration in learning. Fourth, shows students gain new insights as the basis for the search for new ideas. [13] [14] [15].

III. FISHER FACE

Fisher faces or named as fisher linear discriminant is the combination of Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). PCA is an unsupervised algorithm, whereas LDA is an unsupervised algorithm. PCA keeps the distribution information but cannot project the optimal matrix. LDA project the optimal matrix under Fisher criterion, but the dimension of the input space is greater than the number of training images, thus it cannot be applied directly. [9] [10] [11]

Projection PCA of the matrix is computed by Equation 1 and projection LDA of the matrix is computed by Equation 2.

$$PCA(\varphi) = \varphi^T S_T \varphi, \tag{1}$$

$$LDA(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T S_w \varphi}, \varphi \neq 0, \tag{2}$$

where φ^T is the transpose of the matrix φ. Total scatter matrix S_T is the summation of the S_b between-class scatter matrix and S_w within-class scatter matrix.

S_b between-class scatter matrix is shown as follows:

$$S_b = \frac{1}{N}\sum_{i=1}^{C} N_c(\mu_i - \mu)(\mu_i - \mu)^T, \tag{3}$$

and S_w within-class scatter matrix is shown as follows:

$$S_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{Nc}(x_{ij} - \mu_i)(x_{ij} - \mu_i)^T \tag{4}$$

where if N present the total of data and C present the total of class, then N_C present the total of data in class C. *Meanwhile,* x is the vector and μ is the mean of the vector.

IV. Viola-Jones Algorithm

In Viola-Jones algorithm, In this algorithms, AdaBoost algorithms classify the rectangular features in "cascade" stages. The rectangular features are calculated using an integral representation which is calculated using Equation 5.

$$W_e = \sum_{y_i \neq k_m(x_i)} w_i^m, \tag{5}$$

where W_e is the integral of sub window's and w_i is integral to pixel-i. A number of M filters are looped at each cascaded stage. The weak classifier (k_m) is selected from the pool of the classifier. The weight of the classifier w_t is updated on each iteration.

V. Methodology

A. Dataset

We captured ten UM's student (six men and four woman with age 18-24 years) in the class using CANON EOS 700D. The camera is placed facing the student at the top of the presentation screen, while the camera and the student sitting position have the different angle of inclination. Video resolution is set 720X1280 pixels in 25 frames per seconds (fps). These videos are converted into the image in 25 frames per minutes, thus there is 100 image of student facial expression. 50 images are set as training data and 50 is set as testing data. In this research, we only capture the two emotion consist of interest and bored in the real class. These both emotions are most commonly found in real class. 'Interest' represents the positive emotion which is described as a state of willingness to process and understand information. Nevertheless, 'bored' represent negative emotion which is described as a state of unwillingness to process and understand information.

B. Facial Emotion Recognition using Fisher Face-based Viola-Jones Algorithm

The pseudocode Fisher face-based Viola-Jones algorithm is shown in Figure 1. Our contribution is modifying the features of fisher face as a substitute for feature rectangle to detect faces and emotions simultaneously.

```
Input: Face image
Output: Emotion{Interest,Bored}
for i ← 1 to num of sub window's shift do
 for j ← 1 to num of cascaded stages do
   for k ← to num of filter do
    Calculate We fisher face Using Equation (6)
    update wface using Equation(8)
    update wemotion = Equation (8)
    if we < wface then
     break for k loop
    elseif we < wemotion then
     output = bored
    else
     output = interest
    end_if
   end_for
  end_for
 end_for
```

Fig. 1. Pseudocode of Fisher Face-based Viola-Jones Algorithms for Facial Emotion Recognition

Like the original Jones-Viola algorithm, the features on each shifted window are calculated. In this study, we do not present rectangular features, but we modified the Viola-Jones algorithm using the fisher-face feature. Fisher face is a composite of PCA and LDA calculated using Equations 1 and 2 that can be simplified into Equation 6.

$$w_e = J(\varphi) = \varphi_{Fisher}\,\Lambda \tag{6}$$

At Equation 6, the feature of sub window's w_e are projected by the matrix φ of fisher's face J. It calculated using a diagonal eigenvalue matrix Λ that equal to Equation 1 and φ_{Fisher}. φ_{Fisher} eliminates zero eigens and sort in descending order.

A number of M filters are looped at each cascaded stage. The weak classifier (k_m) is selected and the weight of classifier α_m is calculated by:

$$\alpha_m = \frac{1}{2} ln\left(\frac{1 - e_m}{e_m}\right). \tag{7}$$

where e_m is the weight-normalized and it is set as follows :

$$e_m = \frac{W_e}{W}. \tag{8}$$

where W is a maximum weight. The updated threshold w_t is updated on each iteration $m+1$ by:

$$w_t^{m+1} = \begin{cases} w_t^m e^{\alpha_m} & k_m(x_i) \text{ is a miss} \\ w_t^m e^{-\alpha_m} & otherwise \end{cases}, \tag{9}$$

In this study, there is two threshold w_t consist of w_{face} and $w_{emotion}$. w_{face} is the threshold of the face. w_{face} is compared with the weight of the sub window's w_e. If w_e are less than w_{face}, the sub-windows is rejected as the face. Whereas if w_e are more than w_{face}, the sub window is processed in emotion recognition. The procedure of emotion recognition is similar to face detection, However, w_e are compared with $w_{emotion}$. $w_{emotion}$ is the threshold of emotion. If w_e are less than $w_{emotion}$, the sub-windows is set as bored. Whereas if w_e are more than $w_{emotion}$, the sub window is set as interest.

C. Evaluation

Series of experiments are conducted to compare the performance of Fisher Face-based Viola-Jones Algorithm and original Viola-Jones Algorithm for the facial emotion recognition. We used original Viola-Jones Algorithm version [7] as a comparison. The emotions ground truth is analyzed manually by an expert. The estimated results are compared with ground truth to obtain the accuracy, precision, and recall which are calculated as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$precision = \frac{TP}{TP + FP} \tag{11}$$

$$recall = \frac{TP}{TP + FN} \tag{12}$$

where TP is the number of the interest emotion that classified correctly, TN is the number of the bored emotion that classified correctly, FN is the number of interest emotion that classified as bored emotion, and FP is the number of bored emotion that classified as interest emotion.

VI. RESULT AND DISCUSSION

In this study,we compare the performance of Fisher Face-based Viola-Jones Algorithm and original Viola-Jones Algorithm for the facial emotion recognition. In this study, we initialize the variables based on [1] and [7] research shown in Table I. We use a base resulution of 24x24 pixels that this bounding box was enlarged up to 50%. We also set 38 layer cascade of classifiers without comparing the best layer values. In addition we implemented 3 stages which are more one stage than [1]. Stage I is used to reject unface area. Stage II is used to update the weight of facial emotion, whereas stage III is used to estimate the emotion using the updated weight.

We evaluated the accuracy, precision, recall, and time-consuming of the system using 50 data of the University of Malang's learning dataset. There are 25 images of students with 'interest' emotions and 25 others have 'bored' emotions in testing data. Moreover, we initialize the 'interest' emotions as a positive class and 'bored' emotions as a negative class. The comparison of evaluation is shown in Table II.

In Table II, we compare our proposed algorithm and the original Viola-Jones algorithms [1][7]. The results show that original Viola-Jones algorithm achieves accuracy, precision, recall, and time up to 0.78, 0.77, 0.80 and 15 fps respectively, whereas our proposed method achieves 0.82, 0.78, 0.84, and 1 fps. These results show that the fisher face-based Viola-Jones algorithm recognizes facial emotion more accurate and precision than the Viola-Jones algorithm. However, our algorithm is 15x slower than the original Viola-Jones. At the same time, the original Viola-Jones algorithm is able to execute 15 frames per second, our algorithm is only able to extract 1 frame every second. The original Viola-Jones system works faster because it uses simple summation based features, while the features we use work twice in PCA and LDA calculations. however, our system is more accurate because the fisher face is able to produce the best vector dimension to represent the face and emotions as described by [9] [10] and [11].

TABLE I. THE INITIALIZATION

Variables	Value
Base resolution Layer cascade of classifiers Stages	24x24 pixels[a] 38[a] 3

[a] The value set is the same as [1]

TABLE II. THE COMPARISON OF EVALUATION

The Algorithm	Results							
	TP	*TN*	*FP*	*FN*	*Accuracy*	*precision*	*Recall*	*Time*[c]
Original VJ[b]	20	19	6	5	0.78	0.77	0.80	15 fps
Fisher face-based VJ[b]	21	20	6	4	0.82	0.78	0.84	1 fps

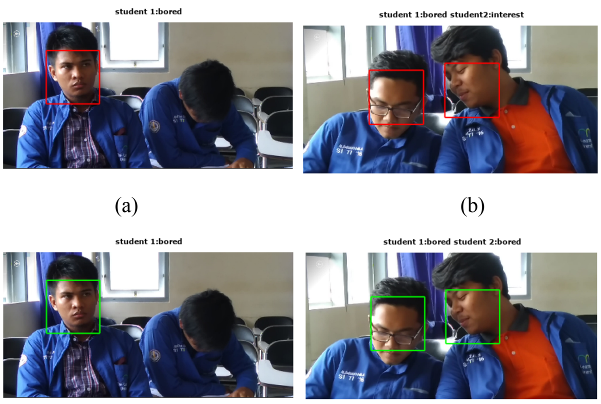
[b] VJ = Viola-Jones algorithm

[c] time in frame per second (fps)

The example of comparison results in our tests is shown in Figure 2. Figure 2 (a) and Figure 2 (b) show the results of the original Viola-Jones algorithm, while Figure 2 (c) and Figure 2 (d) show the results of the fisher face-based Viola-Jones algorithm In Figure 2 (a) and 2 (c), both methods can not detect faces that are not facing the camera. However, if the angle of the face is not extreme, both methods can detect the face.

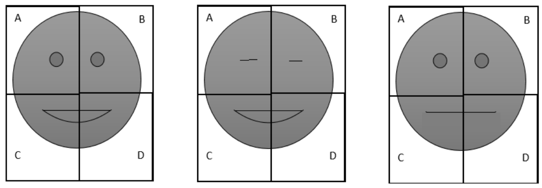
In Figures 2 (b) and 2 (d), both methods can detect two faces that are not facing the camera because the face angle is not extreme. In Figure 2 (b), the viola jones algorithm detects one interested student and one student bored, while in Figure 2 (d), our algorithm detects both students bored. Based on the ground truth by the experts, the two students have an indication of boredom in learning. This is shown by the eyes that did not focus on presentation slides and talk with other students. There is a possibility that the original jones viola detect the emotion of the 2nd student emotion as the interest because there is a lot of data of interest trainer which is marked with a smile. Especially for emotional recognition, we find the dominance of the integral area. The analogy of integral dominance is shown in Figure 3.

In the original Viola-Jones algorithm, lip expression is more dominant than the eye expression because the lip changes increase the integral value greater than the eye. This is shown from the comparison between Figure 2 (a) to Figure 2 (b) as changes in eye expression and Figure 2 (a) to Figure 2 (c) as changes in lip expression. While in fisherface, we take the statistical feature on the pixels set as the face, so there is no dominance of certain areas on the face.



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

Fig. 2. The Example of Result Comparison (a),(b) Result of [7] Algorithm (c),(d) Result of Our Proposed Method



(a) (b) (c)

Fig. 3. The analogy of integral dominance

VII. CONCLUSION

Series of experiments are conducted to evaluate the performance of Fisher Face-based Viola-Jones algorithms. The results show that original viola-jones algorithm achieves accuracy, precision, recall, and time up to 0.78, 0.77, 0.80 and 15 fps respectively, whereas our proposed method achieves 0.82, 0.78, 0.84, and 1 fps. These results shows that the fisher face-based viola-jones algorithm recognizes facial emotion more accurate and precision than the viola-jones algorithm. However, our algorithm is 15x slower than the original viola jones.

In further research, speed improvements are needed. This can be achieved by combining the integral and fisher face concepts which the integrals can be calculated quickly and the fisher face detects facial emotion accurately. In addition, the process of shifting windows is still conventional which is start from left to right. The process of finding the shortest path to the face area can be optimized by a heuristic algorithm.

ACKNOWLEDGMENT

This research is fully supported by the State University of Malang using PNBP funding. We also thank the Faculty of Engineering for providing the object of study.

REFERENCES

[1] P. Viola and M.J. Jones, “Robust Real Time Face Detection” in International Journal of Computer Vision. vol. 57 (2004), pp.137-154.

[2] D. Peleshko and K. Soroka, “Research of usage of Haar-like features and AdaBoost algorithm in Viola-Jones method of object detection” 2013 12th International Conference on the Experience of Designing and

Application of CAD Systems in Microelectronics (CADSM), pp. 284-286. Feb. 2013.

[3] T. Ephraim, T. Himmelman , and K. Siddiqi, “Real-Time Viola-Jones Face Detection in a Web Browser”, 2009 Canadian Conference on Computer and Robot Vision, pp. 321-328. 25-27 May 2009.

[4] M.V. Alyushin, V.M.Alyushin, and L.V. Kolobashkina, “Optimization of the Data Representation Integrated Form in the Viola-Jones Algorithm for a Person's Face Search,” 8th Annual International Conference on Biologically Inspired Cognitive Architectures, Procedia Computer Science, vol. 123 (2018), pp. 18–23.

[5] A. D. Egorov , D. U. Divitskii, A. A. Dolgih, and G. A. Mazurenko “Some cases of optimization face detection methodes on image (Using the Viola-Jones method as an example)”, 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), pp. 1075-1078, Jan. 29 2018-Feb. 1 2018.

[6] J. Cuevas, A. Chua, E. Sybingco, and E. A. Bakar, “Identification of river hydromorphological features using Viola-Jones Algorithm”, 2016 IEEE Region 10 Conference (TENCON), pp. 2300-2306. November 2016.

[7] P. Ithaya Rani and K. Muneeswaran, “Robust real time face detection automatically from video sequence based on Haar features”, 2014 International Conference on Communication and Network Technologies, pp. 276-280. Dec. 2014.

[8] H. Candra, M. Yuwono, R. Chai, H. T. Nguyen, and S. Su “Classification of facial-emotion expression in the application of psychotherapy using Viola-Jones and Edge-Histogram of Oriented Gradient”, 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 423-426, Aug. 2016.

[9] G. K. Verma and B. K. Singh, "Emotion recognition based on texture analysis of facial expression," 2011 International Conference on Image Information Processing, Himachal Pradesh, 2011, pp. 1-6.

[10] Y. Li, W. Yuan, H. Sang and X. Li, "Combination recognition of face and ear based on two-dimensional fisher linear discriminant," 2013 IEEE 4th International Conference on Software Engineering and Service Science, Beijing, 2013, pp. 922-925.

[11] S. Wibawanto and K.C.Kirana, “Recognition of Student Emotion based on Matrix-1 Median Fisher’s Face and Backpropagation Algorithm,” 2017 International Conference on Electrical Engineering and Informatics, pp. 117-122. October 2017.

[12] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion Recognition using Facial Expressions," International Conference on Computational Science, pp. 1175–1184, June 2017.

[13] L. Darling, S. Orcutt, K. Strobel, E. Kirsch, I. Lit, and D. Martin., The Learning Classroom: Emotion and Learning, California: Stanford University School of Education, 2010, pp. 90-104.

[14] A.R Faria, A. Almeida, C. Martins, R. Gonçalves, J. Martins, F. Branco, “A global perspective on an emotional learning model proposal,” Telematics and Informatics, 2016, in press.

[15] P. Vela, P.A. Vela, R.J. Jensen, Emotions, Technology, Design, and Learning: Robots, Emotions, and Learning, Elesevier Inc, 2016, pp.183-197.