A Frontal Pose Face Detection and Classification System Based on Haar Wavelet Coefficients and Support Vector Machine

Iwan Setyawan and Ivanna K. Timotius, *Member, IACSIT*

***Abstract*— This paper presents an integrated face detection and classfication system for faces with frontal pose. The face detection sub-system is based on Haar wavelet coefficients and the face classification sub-system is based on support vector machines. The proposed system is trained using the VISiO multi-view face database and is tested using the commonly used test sets. Our experiments show that the proposed face detection sub-system has a 94.8% detection rate while the face classification sub-system has a 68.1% classification rate.**

***Index Terms*— Face detection, face classification, haar wavelet, support vector machine.**

I. Introduction

Nowadays, there are a lot of applications that rely on the classification of human faces. For example, such a system can be employed as a part of a security system that allows access to a certain area only to persons that are members of a certain group. Another example is a surveillance system that can give an alert to law enforcement agencies of the presence of people that are known to belong to terrorist groups. Each of these applications relies on the integration of a face detection system and a face classification system. The face detection system is used to locate the facial area within the input image while the face classification system is used to determine to which group the detected person belongs.

Ideally, a face detection system should be able to locate all the faces in a digital image, regardless of position, scale, orientation, age, expression, illumination conditions and image content [1]. There are many discriminating features that can be used to detect faces proposed in the literature. For example, one can use human skin colour detection to locate faces [1][2]. Another approach is to use template matching methods [3]. The authors in [2] have also proposed another possible approach, by detecting the presence of objects, like nose and nostrils, which are normally present in human faces. Finally, another approach that has been proposed is the use of statistical models of the facial area and non-facial areas [4]. In our previous works [5][6], we have shown that the use of 1-dimensional Haar wavelet coefficients as a discriminating feature to detect faces can give a satisfactory result which are robust againts variations in background, illumination and subject expression.

Manuscript received November 14, 2011; revised November 22, 2011.

Authors are with the Department of Electronic Engineering, Satya Wacana Christian University, Salatiga, Indonesia (e-mail: iwan.setyawan@ieee.org); (e-mail: ivanna_timotius@yahoo.com).

Face classification system is a machine that is used to classify people based on their face. Support Vector Machines (SVM) is one of the possible methods used to classify faces. SVM is well-known because this method utilized optimalization approach to construct a separating hyperplane as a decision surface [7]. This hyperplane is constructed in a high-dimensional feature space by using a nonlinear mapping. By the means of SVM, the face classification system is expectantly able to construct a decision surface which maximizes the margin of separation between the face images among the two groups of people.

This paper presents a system that could detect the areas of an image which contain faces and subsequently classify the detected faces into one of two groups (i.e., "members" and "non-members"). This system is designed to detect faces with frontal pose only. Furthermore, the system is not designed to establish the identity of each individual subject. In other words, the system is not designed as a face recognition application.

The remainder of this paper is organized as follows. Section 2 will review the Haar wavelet transformation. Section 3 will summarize the SVM classifier. Section 4 will discuss the proposed face detection and classification algorithms. Section 5 will discuss the experimental setup and results that we obtained. Finally, Section 6 will present conclusions and some pointers to our future work.

II. Haar Wavelet

Haar wavelet transform is the oldest wavelet transform. It is also the wavelet transform with the simplest basis transform [8], consisting simply of 2 step functions. This basis function can be scaled so that it spans either a large portion of the image (i.e., more spatial extent or lower frequency resolution) or so that it covers only a small portion of the image (i.e., less spatial extent or higher frequency resolution).

At its highest frequency resolution, the basis function covers 2 adjacent pixels. The horizontal and vertical wavelet representations can thus be calculated by simply using the following equations [9]:

$$I_h(i,j) = I(i+1,j) - I(i,j) \tag{1}$$

$$I_v(i,j) = I(i,j+1) - I(i,j) \tag{2}$$

In Equations (1) and (2), I_h and I_v refers to the horizontal and vertical 1-dimensional Haar wavelet representations of an input image I, respectively. The indices i and j refers to the spatial position of the image pixels.

Similar to other wavelet transforms, Haar wavelet transform is widely used in the field of digital image compression [8],[10]. This transformation can also be utilized in pattern recognition applications, particularly in face detection and recognition applications. This is due to the fact that Haar wavelet coefficients can capture the features of the input image. The authors in [1] describe the use of Haar-like features in face detection problem. The author in [9] also uses Haar wavelet coefficients as one discriminating feature to detect faces. Equations (1) and (2) can be used to capture the horizontal and vertical features of facial regions. Facial features, such as eyes, nose and mouth can be captured in this fashion because they form the high-frequency components of the facial regions (i.e., these features introduce rapid changes in luminance values). An example of such facial features is shown in Figure 1. As this Figure shows, important details to identify facial regions (mouth, nose, eyes, etc.) can be nicely captured.

Fig. 1. Horizontal (center) and vertical (right) facial features captured using Haar wavelet transform

III. Support Vector Machine

Basically, SVM is a linear machine which is used as a supervised classification method between two classes. This SVM use an optimization method to construct a hyperplane as a decision surface which maximize the margin of separation between the positive and negative classes. By using a nonlinear mapping to the input vectors, this SVM is generalized into a nonlinear machine which aims to construct an optimal separating hyperplane in a high-dimensional feature space.

The constrained optimization problem of SVM is defined as follow. Given the training samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where $\mathbf{x}_i$ is a training vector and y_i is its class label being either +1 (for positive class) or -1 (for negative class), SVM wants to find the optimum weight vector $\mathbf{w}$ and the optimum bias b of the separating hyperplane such that [7][11]:

$$y_i(\mathbf{w}^T\varphi(\mathbf{x}_i)+b) \geq 1-\xi_i, \quad \forall i \qquad (3)$$
$$\xi_i \geq 0, \quad \forall i$$

with $\mathbf{w}$ and the slack variables ξ_i minimizing the cost function:

$$\Phi(\mathbf{w}, \xi_i) = \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i \qquad (4)$$

The slack variables ξ_i represent the error measures of data, the parameter C is the penalty assigned to the errors, and the fucntion $\varphi(\cdot)$ is a nonlinear mapping which maps the data into a higher dimensional feature space.

The constrained optimization problem stated above is called the primal problem [7]. This problem might be solved by using the method of Lagrange multipliers. By using this method, the dual problem could be formulated as follow. Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ be the training samples, find the Lagrange multipliers $\{\alpha_i\}_{i=1}^{N}$ that maximize the function [7]:

$$Q(\boldsymbol{\alpha}) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \qquad (5)$$

subject to

$$\sum_{i=1}^{N}\alpha_i y_i = 0 \qquad (6)$$

and

$$0 \leq \alpha_i \leq C \qquad \text{for } i = 1, 2, \ldots, N \qquad (7)$$

The parameter C is the penalty assigned to the errors which is specified by users. $k(\mathbf{x}_i, \mathbf{x}_j)$ is the nonlinear inner product kernel which defined as follow:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)\cdot\varphi(\mathbf{x}_j) \qquad (8)$$

For an unseen data $\mathbf{z}$, its predicted class can be obtained by the decision function:

$$D(\mathbf{z}) = \text{sgn}(\mathbf{w}^T\varphi(\mathbf{z}) + b) \qquad (9)$$

By having the Lagrange multiplier, the decision function is equal to:

$$D(\mathbf{z}) = \text{sgn}\left(\sum_{i=1}^{N}\alpha_i y_i k(\mathbf{x}_i, \mathbf{z}) + b\right) \qquad (10)$$

IV. Proposed System

The proposed system consists of two main sub-systems: the face detection sub-system and the face classification sub-system. The block diagram of the system is given in Figure 2. The input of the system is an image that may contain one or more facial areas. The input image is processed by the face detection sub-system to determine the locations of the face(s) in the image. The output of this sub-system is a cropped image(s) containing only the facial area(s). This output is then fed into the face classification sub-system. The face classification sub-system evaluates this image and decides whether the subject in the input picture belongs to the "member" class or the "non-member" class. In this section, we will discuss each sub-system in greater detail.

A. Face detection sub-system

The face detection sub-system starts with creating reference Haar wavelet coefficients. This reference is constructed from training samples of facial regions. First, we compute the average face sample from the available training samples. These training samples have to be of the same pose. Then we use Equations (1) and (2) to compute the horizontal and vertical Haar wavelet representation vectors, respectively, of the average face sample. We shall call these vectors $\mathbf{R}_h$ and $\mathbf{R}_v$ for the horizontal and vertical reference representations respectively. Then we normalize the vectors by using the following equation.

$$\mathbf{R}' = \frac{\mathbf{R} - \mu}{\sigma} \qquad (11)$$

$\boldsymbol{R}'$ is the normalized version of $\boldsymbol{R}$. The terms μ and σ represent the mean and standard deviation of $\boldsymbol{R}$, respectively. We shall call the normalized vectors $\boldsymbol{R}_h'$ and $\boldsymbol{R}_v'$,

respectively.

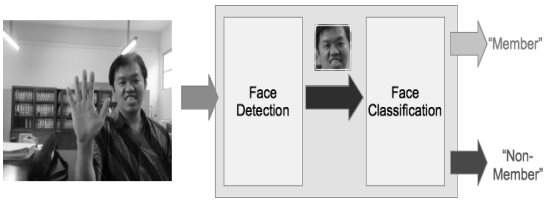


Fig. 2. Block diagram of the proposed system

An input image, that may contain facial regions, is processed using a moving window, S_{MN}, of size $M \times N$. In this paper, we chose $M = N = 48$. In each iteration, we compute the horizontal and vertical 1-dimensional Haar wavelet representation vectors of the window called $\mathbf{S}_h$ and $\boldsymbol{S}_v$. These vectors are normalized using Equation (11), resulting in normalized vectors $\boldsymbol{S}_h'$ and $\boldsymbol{S}_v'$. We then compare these Haar wavelet representations to the previously constructed reference representations, by computing the normalized cross-correlation values between $\boldsymbol{R}_h'$ and $\mathbf{S}_h'$ (and respectively between $\boldsymbol{R}_v'$ and $\boldsymbol{S}_v'$), as follows:

$$C_h = \frac{\sum_x \mathbf{R}_h'(x)\mathbf{S}_h'(x)}{\sqrt{\sum_x \mathbf{R}_h'^2(x)\sum_x \mathbf{S}_h'^2(x)}} \tag{12}$$

$$C_v = \frac{\sum_x \mathbf{R}_v'(x)\mathbf{S}_v'(x)}{\sqrt{\sum_x \mathbf{R}_v'^2(x)\sum_x \mathbf{S}_v'^2(x)}} \tag{13}$$

In Equations (12) & (13), C_h and C_v are the cross-correlation values of the horizontal and vertical representation vectors, respectively. We declare that S_{mn} contains a facial region if both $C_h > T_h$ and $C_v > T_v$. Otherwise, we consider that S_{mn} does not contain a facial region. T_h and T_v are the thresholds for the horizontal and vertical representation correlation values respectively, and are determined empirically (in our experiments, we use $T_h = T_v$). The chosen values of these thresholds are a compromise between the probability of missed detections and false alarms. A high threshold value will reduce the probability of false alarm, but increase the probability of missed detections and vice versa. In this paper the thresholds are chosen such that the system is biased towards false detections. The aforementioned process is repeated until all areas of the input image have been processed. Since facial region can have various sizes, we have to perform the detection process at various spatial scales. We do this by scaling the input image by a scaling factor. This scaling factor, ζ, is also determined empirically in our experiments. The proposed face detection algorithm can be summarized in the flowchart shown in Figure 3.

B. Face classification sub-system

The face classification sub-system aims to classify face images into two classes: member and non-member as shown in Figure 2. The input images for the face classification sub-system are 64 × 64 pixel 8-bit grayscale images taken from the output of the face detection sub-system. The Gaussian kernel [4] was used as the nonlinear mapping $\varphi(\cdot)$:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|}{\sigma}\right) \tag{14}$$

The penalty to the error C used in this paper is 50 and the parameter of the kernel function σ used in this paper is 90. These parameter values were determined empirically by means of 2-fold cross validation using 30 subjects from the Video, Image, and Signal Processing (VISiO) laboratory Satya Wacana Christian University (SWCU) multiview face database that will be further explained in Section V.

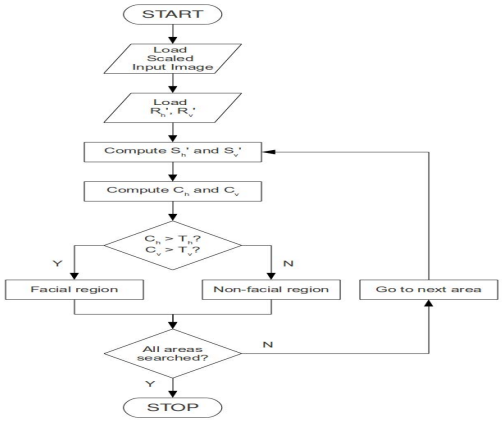


Fig. 3. Flowchart of the proposed face detection system

C. The problem of multiple detections

Since an area can be declared as a facial area if the correlation values C_h and C_v are larger than a certain value, it is possible that during the detection process multiple detections on the same face occur. An example of this multiple detection is shown in Figure 4.

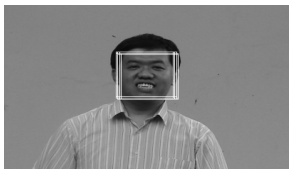


Fig. 4. Example of multiple detections on a single face

This problem is solved by choosing the detected facial area with the strongest detector response and then eliminating the other detected areas that lie within a certain distance from the area with the strongest response. The rationale behind this choice is that we can safely assume that faces in the input image will not overlap. In our experiments, the distance is chosen to be 48 pixels from the center of the area with the strongest response. We call this 96 × 96 pixel area the "proximity area". This is shown in Figure 5. In this figure, the area with the strongest response is represented by the thick square. The dashed square represents its proximity area. The thin squares represent multiple detections of the same face, and is therefore ignored (eliminated) since they lie within the proximity area.

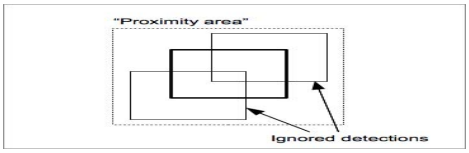


Fig. 5. Elimination of multiple detections

Ideally, the process of eliminating multiple detection areas should be performed completely within the face detection sub-system. However, in the proposed system we chose to perform this elimination process after the classification process. We did this to prevent the possibilty of eliminating areas that, when processed by the face classification sub-system, will give a true positive decision. In cases where overlapping areas produce mixed decision (i.e., some areas are classified as "member" and other areas are classified as "non-member") we chose to keep the areas classified as "member" and eliminate the "non-member" areas. In other words, our system is biased towards false positive response.

V. Experiment Setup and Results

The system described in Section IV is trained using the images taken from the Video, Image, and Signal Processing (VISiO) laboratory Satya Wacana Christian University (SWCU) multiview face database [7]. This database currently contains face images of 100 subjects. The subjects are evenly distributed in gender (i.e., 50 subjects are female and 50 subjects are male). The age varies between 19 and 69 years. The images are taken under controlled condition in our laboratory. Each subject is photographed against a uniform white background using an identical setting. For each subject, 105 pictures are taken with a variation of pose, expression and facial accessories. Thus, in total the database contains 10500 images. The resulting images are cropped around the facial area. The images in this database are resampled as 64 × 64 pixel 8-bit grayscale images. Examples of the images contained in the database are shown in Figure 6.

Fig. 6. Examples of the images in the VISiO face database

The images used to build the reference feature vectors for the face detection sub-system is taken from the VISiO database. In this case we use a subset of the database, containing 100 frontal face images with neutral expression and no facial accessories. Furthermore, we perform a pre-processing on the images. The pre-processing is as follows. The first step of the pre-processing is to align and adjust the scale of the images. We did this by aligning the position of the eyes of the subjects. This step is necessary since the misalignment and scale variations in the original images will yield a virtually "featureless" average face image. The second step is to crop the images into 48 × 48 pixel 8-bit grayscale images. This is done to minimize the areas outside the desired facial area. An example of the pre-processed face image and the average face sample is shown in Figure 7.

Fig. 7. Pre-processed face sample (left) and the average face sample (right)

For the face classification sub-system, 30 subjects were chosen from the database: 6 subjects were designated as members of VISiO laboratory and 24 subjects were designated as people that are not members of the VISiO laboratory. The pictures of the subjects belonging to the "members" class are shown in Figure 8. All images of the subjects (i.e., including all pose, expression and accessories) are used in the training of the face classification sub-system.

Fig. 8. Subjects designated as "members"

The performance of the proposed system is evaluated by using a set of test images. In our experiments, we used 114 different images. In total, these images contain 135 faces of which 60 faces belong to subjects classified as "members". Some images contain both "members" and "non-members". The test images are obtained using various means. Some images are specifically taken for this purpose. The other images are taken from the Combined MIT/CMU Test Images [4][9] or downloaded from the internet. The test images are all grayscale images (or are colour images converted into grayscale) but have various spatial resolutions and quality. In particular, images containing the faces of "members" are taken under various conditions and are not part of the training images for the face detection and classification sub-systems. Therefore, it can be said that the proposed system has to deal mostly with images that are "new" to it. Examples of the test images used in this paper are shown in Figure 9.

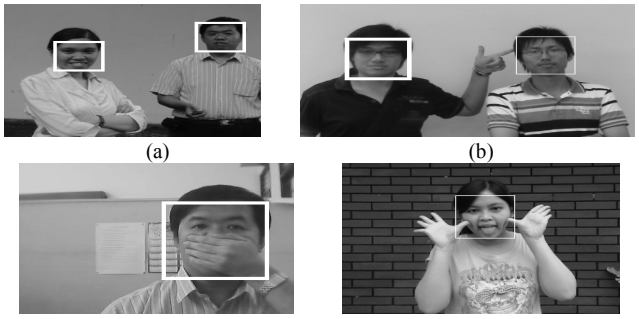
Fig. 9. Examples of the test images

Our experiments show that the face detection sub-system can detect 128 faces with 7 missed detections and 8 false detections, giving a detection rate of 94.8%. The face classification sub-system have 33 true positives, 59 true negatives, 19 false positive and 24 false negatives, yielding a classification rate of 68.1%.

Examples of the output of the proposed system are shown in Figures 10 and 11. In these Figures, "members" are highlighted using thick squares while "non-members" are highlighted using thin squares. It should be noted that in these examples, the faces are detected using different scaling parameters. Figure 10 shows the example of correct detection and classification results. It can be seen from these examples that the system can detect and classify human faces in various conditions. In cases where both a "member" and a "non-member" are present (Fig. 10(b)) the system can still correctly classify each subject. The robustness of the proposed system is shown in Fig. 10(c), where the system is still able to correctly detect and classify a subject although the subject covers his mouth.

Figure 11 shows examples of cases in which the proposed

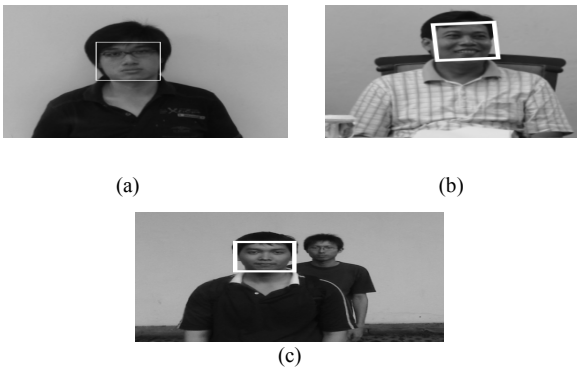
system gives incorrect detection and/or classification results. Figures 11(a) and 11(b) shows an example of correct face detection, but an incorrect classification (the classification results are respectively a false negative and a false positive). Figure 11(c) shows an example of a missed detection (the subject on the right is not detected). However, in this case the subject on the left is both correctly detected and classified.



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

Fig. 10. Examples of correct detection and classification results.

The results of the experiments show that the face detection sub-system performs well, giving a high detection rate with low false alarms and missed detection rates. It is robust to variations of lighting conditions, subject expressions and slight pose. However, we also find out that this sub-system is sensitive to image scaling. The choice of the scaling parameter *(ζ)* is critical on the successful detection of faces. An example of such sensitivity is shown in Figure 11(c). In this example, the subjects were standing at different distances from the camera. The resulting difference in apparent face sizes requires different choices of ζ. Therefore, the scaling parameter appropriate for the subject on the left (closer to the camera, hence larger apparent size) results in the missed detection of the subject on the right (smaller apparent size).



(a) (b) (c)

Fig. 11. Examples of incorrect detection and/or classification results

The face classification sub-system is shown to be moderately successful. We observed that this sub-system is sensitive to the background of the images. In cases where the background is dark, the system tends to produce false negative results. This behaviour occurs because the training samples of the face classification are taken with white (i.e., light) background. As we can see in Figures 4 and 6, this background is visible in the training samples and hence influences the behaviour of the system.

VI. Conclusion

In this paper, we have presented an integrated face detection and classification system. The face detection sub-system gives a satisfactory result while some improvements to the face classification sub-system should be implemented. There are at least two different improvement strategies that can be implemented. The first strategy is by building a more comprehensive training set with more background variations. The other strategy that can be implemented is by removing the influence of the background of the current training set. This can be achieved by, for example, defining a Region of Interest (RoI) that contains only the face of the subject, excluding the areas that represent hair and background. In our future work, we will investigate and implement the aforementioned strategies to improve the performance of the system.

References

[1] S.Z. Li and A.K. Jain, "*Handbook of Face Recognition"*, Springer Science+Business Media, Inc. 2005.

[2] R.S. Feris, T.E. de Campos, and R.M. Cesar, Jr, "*Detection and Tracking of Facial Features in Video Sequences, Lecture Notes in Artificial Intelligence 1793"*, Springer-Verlag Press, 2000

[3] C.W. Park and M. Park, "Fast Template-based Face Detection Algorithm using Quad-tree Template", "*Journal of Applied Sciences"*, Vol. 6, No. 4, pp. 795 – 799, 2006

[4] H. Schneiderman and T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars", in *Proc. "IEEE Conf. Computer Vision and Pattern Recognition"*, 2000, pp. 746 – 751

[5] I. Setyawan, I.K. Timotius, and A.A. Febrianto, "Face Detection System using 1-dimensional Haar Wavelet Transform Coefficients" in "*Proc. 6th Int. Conf. Information & Communication Technology and Systems"*, Surabaya, Indonesia, 2010, pp. VI-35 – VI-40

[6] I. Setyawan, I.K. Timotius, and A.A. Febrianto, "A face detection and classification system using haar wavelet coefficients and support vector machine", in "*Proc. 4th Int. Conf. Computer and Electrical Engineering"*, Singapore, 2011, pp. 577 – 581

[7] S. Haykin, "*Neural Network: A Comprehensive Foundation"*, New Jersey: Prentice-Hall, 1999

[8] R.C. Gonzales and R.E. Woods," *Digital Image Processing"*, 3rd Ed., Pearson Education, Inc., 2010

[9] C. Liu, "A Bayesian Discriminating Features Method for Face Detection", "*IEEE. Trans. Pattern Analysis and Machine Intelligence"*, Vol. 25, No. 6, pp. 725 – 740, 2003

[10] J.C. Russ, "*The Image Processing Handbook"*, 5th ed., CRC Press, 2007

[11] V. Vapnik, "*Statistical Learning Theory"*, New York: Springer Berlin Heidelberg, 1998

[12] I.K. Timotius, I. Setyawan, and A.A. Febrianto, "Face Recognition between Two Person using Kernel Principal Component Analysis and Support Vector Machines", "*International Journal on Electrical Engineering and Informatics"*, Vol. 2, No. 1, PP. 53 – 61, 2010

Iwan Setyawan received his Bachelor's and Master's degrees from the Department of Electrical Engineering, Bandung Institute of Technology, Indonesia in 1996 and 1999, respectively, and his Ph.D degree from the Department of Electrical Engineering, Delft University of Technology, The Netherlands in 2004.

He is currently an Assistant Professor at the Department of Electronic Engineering of the Satya Wacana Christian University, Indonesia. His research interests are in image and video processing. Dr. Setyawan is a Senior Member of the IACSIT since 2011 and a Member of the IEEE since 2009.

Ivanna K. Timotius received her Bachelor degree from Department of Electronic Engineering, Satya Wacana Christian University, Salatiga, Indonesia in 2003, and her Master degree from Department of Electronic Engineering, Chung Yuan Christian University, Chungli, Taiwan in 2009. She is currently a Lecturer at Department of Electronic Engineering of the Satya Wacana Christian University, Salatiga, Indonesia. Her research interests are in pattern recognition and image processing.

Ms. Timotius is a Member of the IACSIT since 2011.