

Motion Detection and Face Recognition For CCTV Surveillance System

Ade Nurhopipah*¹, Agus Harjoko²

¹ Department of Informatics Technology, STMIK Amikom Purwokerto, Indonesia

² Department of Electronics and Computer Science, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *adenoer@gmail.com, aharjoko@gmail.com.

Abstrak

Closed Circuit Television (CCTV) telah dipakai dalam kehidupan sehari-hari untuk berbagai keperluan. Pada perkembangannya, pemakaian CCTV berubah dari sistem pengawasan pasif sederhana menjadi sistem pengawasan cerdas terintegrasi. Pada penelitian ini dilakukan deteksi gerak dan pengenalan wajah pada video CCTV yang diharapkan menjadi dasar pengambilan keputusan untuk menghasilkan sistem pengawasan yang terintegrasi, efektif dan efisien.

Pengolahan video CCTV ini memberikan tiga keluaran berupa informasi deteksi gerak, informasi deteksi wajah dan informasi identifikasi wajah. Metode Accumulative Differences Images (ADI) digunakan untuk deteksi gerak, dan Haar Cascade Classifiers digunakan untuk deteksi wajah. Ekstraksi fitur dilakukan dengan menerapkan Speeded-Up Robust Features (SURF) dan Principal Component Analysis (PCA). Fitur tersebut kemudian dilatih dengan Counter-Propagation Network (CPN).

Pengujian dilakukan terhadap 45 video CCTV secara offline. Hasil pengujian menunjukkan tingkat keberhasilan deteksi gerak sebesar 92,655%, keberhasilan deteksi wajah sebesar 76%, dan keberhasilan identifikasi wajah sebesar 60%. Hasil tersebut menunjukkan bahwa deteksi dan identifikasi wajah melalui video CCTV ini belum memperoleh hasil optimal. Penerapan proses deteksi gerak sudah ideal diterapkan secara real-time, namun kombinasinya dengan pengenalan wajah menyebabkan delay waktu yang cukup signifikan.

Kata kunci— ADI, Haar Cascade Classifiers, SURF, PCA, CPN

Abstract

Closed Circuit Television (CCTV) is currently used in daily basis for a wide variety of purposes. The development of CCTV has transformed from a simple passive surveillance into an integrated intelligent control system. In this research, motion detection and facial recognition in CCTV video is used as the basis of decision making to produce automatic, effective and efficient integrated system.

This CCTV video process provides three outputs, motion detection information, face detection information and face identification information. Accumulative Differences Images (ADI) method is used for motion detection, and Haar Classifiers Cascade method is used for face detection. Feature extraction is done with Speeded-Up Robust Features (SURF) and Principal Component Analysis (PCA). Then, these features are trained by Counter-Propagation Network (CPN).

Offline tests are performed on 45 CCTV video. The result shows 92.655% success rate on motion detection, 76% success rate on face detection, and 60% success rate on face detection. It shows that this faces detection and identification through CCTV video have not been able to obtain optimal results. The motion detection process is ideal to be applied in real-time conditions. Yet if it's combined with face recognition process, it causes a significant time delay.

Keywords— ADI, Haar Cascade Classifiers, SURF, PCA, CPN

1. INTRODUCTION

CCTV has been used daily for many purposes such as crime investigation, traffic control, chemical process record, production control, and security surveillance. Public places such as offices, hospitals, shops, tourism resorts and also houses install CCTV for Security surveillance. In its development, the use of CCTV has transformed from passive surveillance into integrated intelligent surveillance system [1].

One of CCTV's drawbacks is the operator has to control record the whole time and identify faces manually. It also produces a large amount of data since the surveillance is done continuously [2]. Its drawback in less effective manual surveillance and its big need of memory can be reduced by setting an intelligent system which is integrated with CCTV in order to automatically produce surveillance system and data control.

Office is one of the places where CCTV is used at most for surveillance system as the main purpose. The intelligent surveillance system using CCTV video data in offices can be installed for some purposes, for example motion surveillance, data storage control, access control, warning alarm control, and employee's attendance. In order to make CCTV data the basis of the decision, it is necessary to process CCTV data to provide various information. This research is done to process CCTV videos to gain information on motion, face detection, and face identification in places monitored by CCTV.

Face detection process is an inseparable part of face recognition process. Face detection using Viola-Jones method is widely used because it produces high accuracy on face detection. One of the research using this method is research[3]. Research [4] also uses AdaBoost algorithm approach in face detection process to count the number of the faces in a classroom.

Face recognition system [5] is built by implementing Speeded-Up Robust Features (SURF) for face pattern extraction and Support Vector Machine (SVM) for classification method. The result of this research shows that the system succeeds in dealing with variation in illumination, perspective, expression and scale. Research [6] and [7] uses face pattern recognition for attendance system application. Research [6] develops attendance system by capturing students' images using the camera set in the front part of the classroom. In research [7], the camera is set on the classroom door to record their faces naturally. These research applies Viola-Jones method for detection and Eigenface for identification.

Research [8] develops web-basis application for face detection at *real time background* used as employee's attendance. In face identification process, PCA is used for face pattern recognition and Haar Cascade method is used for face detection. This system has been tested and makes 68 % accuracy of face recognition. It is also used for security surveillance system. Research [9] builds access control system automatically. It uses Viola-Jones method for detection process, and PCA for face identification. Research [1] develops an intruder warning system for house safety. If an excessive movement is detected, system will detect faces with Eigenface algorithm. If it can't identify faces, it will send a warning and pictures to the house owner's mailbox. This face recognition test system makes 62% accuracy.

This research uses motion detection and face recognition for surveillance system through CCTV. This process involves *Accumulative Differences Images* (ADI) method for movement detection, *Cascade Classifiers* (*Haar Cascade Classifiers*) for face detection, *Speeded-Up Robust Features* (SURF) algorithm and *Principal Component Analysis* (PCA) for feature extraction and reduction, and *Counter-Propagation Network* (CPN) algorithm for data training and testing in face identification proces.

ADI method used by comparing image differences on some sequential frame can minimize error better than a method that can only count motions from two frame on each process. This method is preferred because of its simplicity that can save computation time. *Haar Cascade Classifiers* as a face detection method is believed to be able to process images fast and

produce a good detection success [10]. The use of this method is expected to produce a good face accuracy as in research [3].

SURF algorithm used in feature extraction has some benefits. It can detect feature fast, describe a feature in a unique way and has invariance toward transformation and noise. SURF implementation method is expected to be able to overcome variation in illumination, perspective, expression and scaling, as in research [5]. It is hoped that the research using *Principal Components Analysis* (PCA) which is started with extraction by SURF algorithm can give better accuracy on face detection than that in research [1] and [8]. It is also anticipated that PCA can extract prominent and influential specific features and eliminate data redundancy. In training and testing process of face identification test, CPN algorithm is used for it has high accuracy in pattern recognition process and is built with simple algorithm to save computation[11].

2. METHODS

2.1 Research Design

This research is designed with four main processes : motion detection process, face detection process, data training process, and face identification process. In general, the architectural design of the research can be seen in diagram of Figure 1.

The first step is by acquiring video images from CCTV. Those images will be used for motion detection process. If a motion is detected, the information of time stamp and images with detected motion will be stored. Then, the motion value will be compared to face detection threshold. If the motion exceeds it, the face recognition process will be committed. Images with detected motion will be the input of face detection process. This process will decide if the face object exists. If it exists, detection information will be recorded and continued to face identification process.

In face identification process, the face features is tested by using the value gained from data training and matched with data basis. The data training process is done before testing the face identification by using the same image processing. Face identification process produces decision and record the identity of the identified face.

The collected data in this research is in the form of video file and was taken in PT. Mitranet Software Online Purwokerto for five days. The videos were taken indoor by CCTV at 07:45 – 08:45 IWST, 12:00 – 13:00 IWST, and 15:15 – 16:15 IWST. The CCTV video resolution is 960 x 720 and 640 x 480 with 10 fps frame per second. There are 90 video files with 10 minute duration each. Of those 90 files, 45 videos are trained and the rest is tested.

2.2 Image Enhancement

Image Enhancement is a process of image manipulation using certain operations to gain quality based on a specific application purpose. The enhancement that will be applied are histogram equalization and gamma correction. Histogram equalization is good to enhance the contrast on image, but sometimes the intensity produced are too dark or too light. To overcome this problem, enhancement continued with gamma correction.

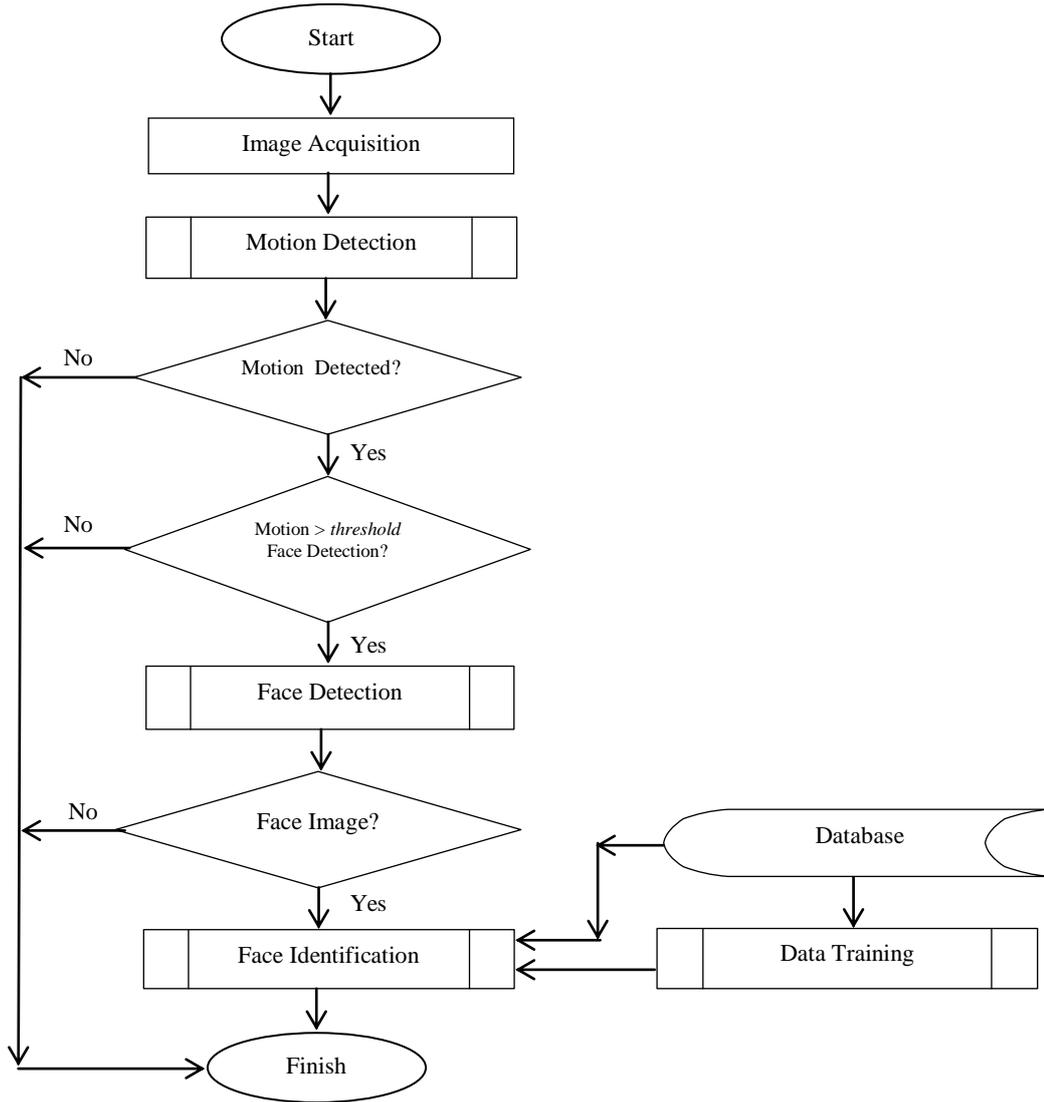


Figure 1 The Design of Motion Detection and Face Identification

2.3 Accumulative Differences Images (ADI)

In this research, motion segmentation is carried on using ADI. In this method, not only one but some images are compared to references. Comparison results between the referent images with tested images will be accumulated and compared to certain threshold [12]. For example, $f(x, y, t_1) = R(x, y)$ is the referent image and $f(x, y, t_k) = f(x, y, k)$ is the k -image where $k > 2$, therefore, ADI value can be defined in equation (1).

$$A_k(x, y) = \begin{cases} A_{k-1}(x, y) + 1, & \text{if } |R(x, y) - f(x, y, k)| > T_A \\ A_{k-1}(x, y), & \text{otherwise} \end{cases} \quad (1)$$

Next, determine the threshold which sets the accumulation limit. This limit determines the presence of motion. If it exceeds the threshold, the image has motion.

2.4 Haar Cascade Classifier

Haar Cascade Classifier algorithm is a learning machine approach to detect an object which is able to process images fast and results high detection. This method uses three key following techniques:

1. Integral image
The size and location of each kernel is used to calculate Haar feature value. This procedure will produce a large number of feature value. We need to use integral image method to count it fast. With this method, this operation only takes 4 main pixels whatever the amount of the pixel in one image is.
2. AdaBoost-based algorithm learning
All image features calculated with integral image are mostly irrelevant. A feature may be a good feature for a region with suitable property. For other areas, the same feature might be irrelevant. The best way to choose a feature can be done by using AdaBoost algorithm.
3. Cascade method for merging classifiers
Cascade structure increases the speed of the detector by focusing on areas where there are objects in the image. The best features will form a strong classifier which classifies faces into positive and negative images. Here, classifier is applied in order to reject the non-face sub-window. The whole process of this detection process is aimed to grow a decision tree called cascade as shown in Figure 2.

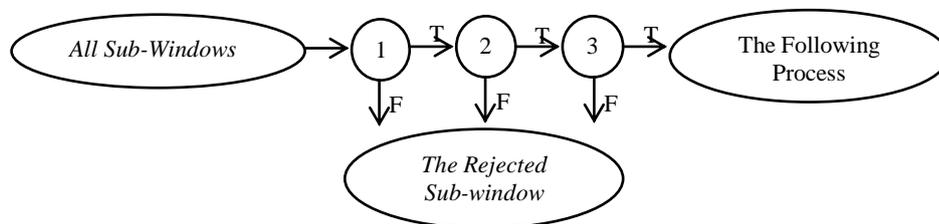


Figure 2 Detection scheme with cascade classification

2.5 Speeded-Up Robust Features (SURF)

There are three main steps when using SURF, they are integral image, feature detection and feature description. Feature detection uses integral image by applying blob-like feature and is done as follows.

1. Do convolution using the second partial derivative box filter approximation of Gaussian. Use bigger size when doing convolution to shape a pyramid image. The result of convolution will form Hessian matrix. It is used because it has a good performance on both computation and accuracy speed.
2. In order to bear up the detected features from scaling, find the Hessian matrix extrema
3. Do localization for the feature candidate on each space scale using non-maximum suppression method. The extrema of Hessian matrix determinant is interpolated with quadratic 3D which shows Hessian function.

SURF descriptor needs definition on feature orientation in order to bear up rotation. The value of SURF descriptor is *Haar Wavelet* toward x and y directions which refers to dx and dy . It is a region of 20s. Each region is split into 4x4 sub-region. Each sub-region is explained by *HaarWavelet* response based on sample 5x5 with a four-component vector. Finally, there are 64 Surf descriptors in each feature.

2.6 Principal Component Analysis (PCA)

PCA is a statistic procedure using orthogonal transformation which converses correlated variables into uncorrelated fewer variables called Principal Component. PCA method can extract dominant specific feature and eliminate data redundancy.

The steps of PCA method are as follow:

1. Prepare a set of testing data in the form of vector
2. Find the median of the data
3. Find the difference between the tested data and the median
4. Count the covarian matrix value
5. Calculate the eigen vector and eigen value of covarian matrix
6. Find the main component based on the biggest *eigenvalue* and use it to transform data to different subspace.

2.7 Counter-Propagation Network

This research will use *Counter-Propagation Network* (CPN) model which is a combination of *unsupervised learning* method with Kohonen layer and *supervised learning* with Grossberg layer. There are two types of CPN architecture, they are *Full CPN* and *Forward-only CPN*. This research uses Full CPN. There are 3 layers in CPN architecture. The first layer is the input layer, the second is cluster layer, and the third is the output layer representing the value of input value approach and output target. Figure 3 shows *full CPN* architecture.

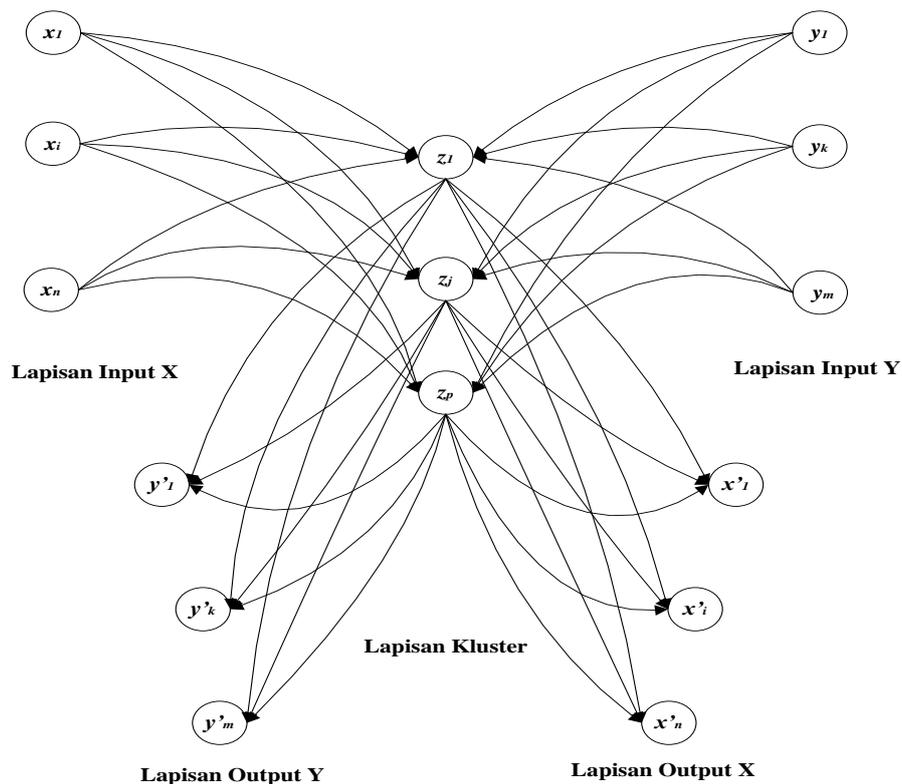


Figure 3 Architecture of Fully Counter-Propagation Network

There are two phases of CPN training. Phase one is *unsupervised learning* where the units of cluster layer are competing. Phase two is *supervised learning* where the winning unit is processed to determine the output of *output* layer. The CPN training algorithm is as follow [13].

S0 : Initialize weights, learning rate, and other variables.

Phase 1

S1 : While stopping condition for phase 1 training is false, do step 2 until step 7.

S2 : For each training input pair of x and y , do step 3 until step 5.

S3 : Apply activation of input layer X to vector x and input layer Y to vector y .

S4 : Find the winner cluster unit, mark it with J index. The winner node has the smallest z_in net input, shown in equation (2)

$$z_in_j = \sum_i (x_i - v_{ij})^2 + \sum_k (y_k - w_{kj})^2 \quad (2)$$

S5 : Update weights into Z_j winner unit by using equation (3) and (4)

$$v_{ij}(new) = (1 - \alpha)v_{ij}(old) + \alpha x_i \quad (3)$$

$$w_{kj}(new) = (1 - \beta)w_{kj}(old) + \beta y_k \quad (4)$$

$$i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m.$$

S6 : Reduce learning rates α and β

S7 : Test stopping condition for phase 1 training.

Phase 2 :

S8 : While stopping condition for phase 2 training is false, do step 9 until step 15.

(Note : α and β value are small constant value during phase 2)

S9 : For each training input pair of x and y , do step 10 to 14.

S10 : Apply activation of input layer X to vector x and input layer Y to vector y .

S11 : Find the winner cluster unit by using equation (2), mark index winner unit with J . The winner unit has the smallest z_in .

S12 : Update weights into Z_j , by using equation (3) and (4).

S13 : Update weights from Z_j unit to output layers by using equation (5) and (6).

$$u_{jk}(new) = (1 - a)u_{jk}(old) + a y_k \quad (5)$$

$$t_{ji}(new) = (1 - b)t_{ji}(old) + b x_i \quad (6)$$

$$i = 1, 2, \dots, n \text{ and } k = 1, 2, \dots, m.$$

S14 : Reduce α and β learning rate.

S15 : Test stopping condition for phase 2 training.

3. RESULTS AND DISCUSSION

The result of this research has three output information: motion detection, face detection, and face identification. The summary of the test result is as shown in Table 1. Next, some factors related to test success value are analyzed.

Table 1 The Result of Motion Detection and Face Recognition

Result	Success (%)	Time (second)	
Motion detection	92.655	Time of Detection	1.115
		Process per <i>frame</i>	0.02
Face detection	76	0.166	
Face identification	60	0.036	

The research implementation of motion detection and face identification for surveillance system using CCTV is shown in Figure 4.

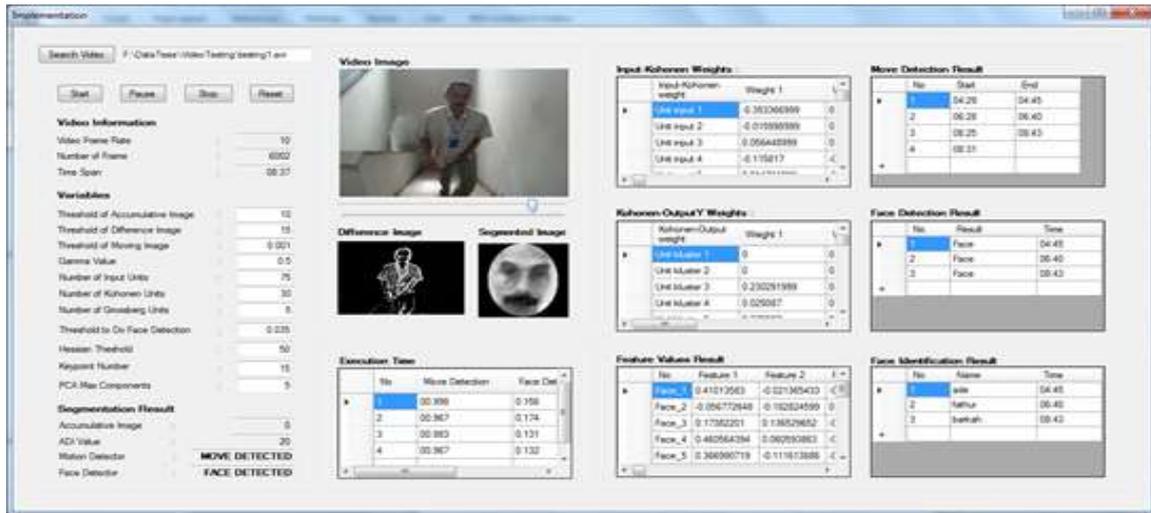
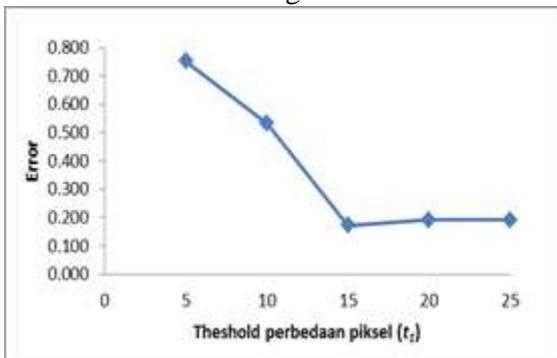


Figure 4 : The implementation of motion detection and face identification for surveillance system using CCTV video

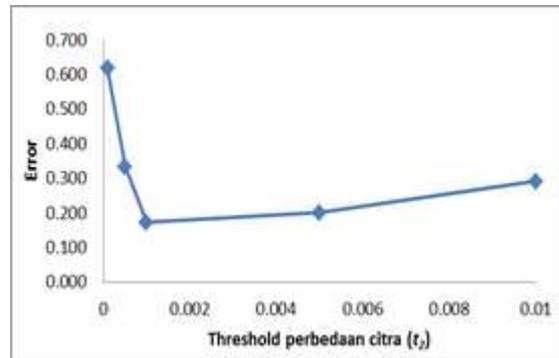
3.1 The Result of Motion Detection

The face detection result explains an output information on time detection about when the motion starts and ends. This method involves three types of threshold: pixel difference threshold (t_1), image difference threshold (t_2), and motion threshold (t_3).

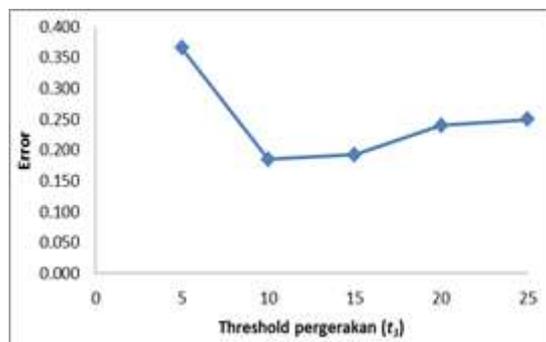
This research uses 5-video simulations to determine the best threshold value which will be used for motion detection. Face detection test is done by applying value $t_1=15$, $t_2=0.001$ and $t_3=10$ which is the best threshold of the three. The simulation chart of the three motion detection threshold is shown in Figure 5.



(a) Simulation result of pixel difference threshold (t_1)



(b) Simulation result of image difference threshold (t_2)



(c) Simulation result of motion threshold (t_3)

Figure 5 Simulation chart to motion detection threshold

3.2 *The Result of Face Detection*

In this process, there are limitations in order to decrease computation load :

1. The motion threshold limit is 3.5 % of the total pixel,
2. Take only 3 frames to be analyzed per second
3. The detected face, which is the frontal size, is 70 x 70 up to 140 x 140 pixel in size.
4. The detected face is one per frame. If there is more than one face image in a frame, take the biggest face image.

Next, initiate detection to each frame that meets the limit. There are things that can be produced and analyzed from face detection:

1. The total number of the detected face is 125. There is 95 images contain correct detection information and the rest is incorrect. The success rate of face detection is 76%.
2. The false face detection can be negative or positive. *False positive* detection occurs when face is detected but in fact, face is absent. On the contrary, *false negative* detection occurs when face is not detected but in fact, there's face.
3. *False positive* occurs when Haar feature value of the detected face meets Haar feature threshold on Haar Cascade Classifier decision tree. It can also happen when the face taken from one frame is in one the biggest region. Meanwhile, it is very possible that the face area meant here is smaller than false positive.
4. Small image resolution, non-ideal face position (non-frontal face), blocked face, lighting, or blurred image by fast motion are the causes of false negative.

After face is detected, crop the face area, resize to synchronize the face image, and mask in ellips to reduce the noise around it.

3.3 *The Result of Face identification*

Image enhancement is applied to face image gained before extraction process. Then, extract the data using SURF method and it will produce keypoint on face image. Each keypoint has 64 descriptors. The number of keypoint on each detected image is different. This research takes 15 keypoints. PCA application is applied to find 5 main components of the 64 descriptors. These 5 main components are considered representing all keypoint.

In previous training data, PCA algorithm is applied to 2940 training data and produces reduced data. Besides, the results of PCA process on data training are the average rate, eigen value and eigen vector which are kept as the basis to determine the projection of data reduction for testing data.

3.3.1 *The Result of Training data*

Training data is processed with CPN algorithm in which 2940 data used. There are things that can be produced and analyzed from face detection:

1. The training consists of two phases. The result of Phase 1 training results 0.045528 MSE on the 164th iteration. Meanwhile, phase 2 results 0.045525 on the 158th iteration. The tested training data with the gained load achieved 94.28% success.
2. The total cluster unit which is too big causes longer computation time. Meanwhile, the small unit causes the network unable to train data better.
3. Big learning rate value makes it possible for the network to modify loads to produce small error even though more iteration are relatively needed.
4. The time needed for this training data process is 61.793 seconds.

3.3.2 *The Result of testing data*

Face identification in a motion period with the same face, maybe produce different identification result. For example, the face image shows A's image in frame 1, while in frame 2, the face image is B's. The conclusion is taken based on the accumulation of the most identification result per motion period. There are things that can be produced and analyzed from face detection:

1. The number of the whole detected face is 65. There is 39 correct face identification, while 26 is incorrect face identification. So, the success rate of face identification is 60%.
2. False identification can happen because of some factors:
 - a. Small face image resolution makes the face pattern can't be well extracted.
 - b. The enhancement process is less optimal to overcome lighting problem, noise, and so on.
 - c. Face with variation in position and illumination can make data extraction unstable.
 - d. The network load resulted in training process can't separate the testing data well.

The result of face identification is 60% and is not better than the previous research which uses almost similar system as in [1] and [8]. Research [1] gained 62% of face identification and reseach [8] reached 68%. Eventhough producing smaller accuracy, this research has smaller limitations related to video taking for testing data. In research [1], web camera is set in better lighting and the employees have to face the camera. Their faces have similar figure to the data basis. In reseach [8] the test is done in simulations, in which 13 people are told to walk in front of the camera. From this result, it can be stated that face identification process using CCTV video in this research hasn't reached maximum result. To perform better, it is necessary to do further and deeper research on related variables that may influence significantly.

3.4 *Execution Time*

In this research, the average of execution time is calculated to initiate motion detection, face detection and identification on 10 video samples. The summary of program execution time can be seen in Table 2.

Tabel 2 Time of Program Execution Table

Time	Motion Detection (second)	Face Detection (second)	Face identification (second)
Minimum	0.967	0.105	0.015
Maximum	3.806	0.211	0.080
Average	1.115	0.166	0.036

Formal standard limit to determine ideal time on a real time surveillance system hasn't been known as well as the ideal process time limit in this research. However a 1.115 second of average detection time is a relatively adequate time for the user to detect a motion as soon as possible.

Real time system performance criteria is restricted by the environmental needs of each system. In this research, ideal time can be determined by the processing delay time. Tested video has a 10 frame per second rate. It means that the ideal time to process the video is under $1/10=0.1$ second per frame. The time needed to detect and recognize a face in this research is

$0.166+0.036 = 0.202$ second. It shows a significant time delay. Therefore if this research is conducted in real time, the time delay should be handled.

4. CONCLUSIONS

This research is conducted by designing and testing motion detection and face recognition on a CCTV video. Motion detection using ADI method shows 92.655% success rate. Average time needed to take a decision on motion detection process is 1.115 second. Facial detection using Haar Cascade Classifier produces a 76% success rate. Training data using CPN algorithm results 0.0455 MSE with a 94.286% success rate. Face identification by applying training data value and pattern extraction using SURF and PCA results 60% success rate.

The ideal time for processing is under 0.1 second per frame, while in this research it takes 0.202 second. It can be concluded that the total time of face recognition process of this research results a significant time delay. The testing result, both accuracy and time which resulting delay, shows that this research needs more improvement.

REFERENCES

- [1] Padaruth, S., Indiwarsingh, F. & Bhugun, N., 2013, A Unified Intrusion Alert System using Motion Detection and Faces Recognition, *2nd International Conference on Mechine Learning and Computer Science (IMLCS)*, Kuala Lumpur.
- [2] Namrata, Pradeep & Sagar, R., 2013. Cognitive Security System Based on Image Comparison and Motion Detection with Able Memory Usage. *International Journal of Advances in Engineering & Technology*, VI(2), pp.850-61.
- [3] Putro, M.D., Adji, T.B. & Winduratna, B., 2012, Sistem Deteksi Wajah dengan Menggunakan Metode Viola-Jones, *Proseding Seminar Nasional "Science, Engineering and Technology"*. Malang.
- [4] Santoso, H. & Harjoko, A., 2013. Haar Cascade Classifier dan Algoritma Adaboost untuk Deteksi Banyak Wajah dalam Ruang Kelas. *Jurnal Teknologi*, VI(2), pp.108-15.
- [5] Anand, B. & Shah, P.K., 2016. Face Recognition using SURF Features and SVM Classifier. *International Journal of Electronics Engineering Research*, VIII(1), pp.1-8.
- [6] Balcoh, N.K., Yousaf, M.A., Ahmad, W. & Baig, M.I., 2012. Alghoritm for Efficient Attendance Management: Face Recognition Based Approach. *International Journal of Computer of Science Issues (IJCSI)*, IX(4), pp.146-50.
- [7] Febrianto, A.J., 2012, Pengenalan Wajah Dengan Metode Principle Component Analysis (PCA) Pada sistem Absensi Real Time, *Tesis*, Magister Teknik Elektro, Universitas Gadjah Mada, Yogyakarta.
- [8] Tharanga, J.G.R., Samarakoon, S.M.C. & Karunarathne, T.A.P., 2013, Smart Attendance using Real Time Face Recognition (Smart-FR), *SAITM Research Symposium on Engineering Advancement*, Sri Lanka.
- [9] Lwin, H.H., Khaing, A.S. & Tun, H.M., 2015. Aotomatic Door Access System Using Face Recognition. *International Journal of Scientific & Technology Research* , 4(06), pp.294-99.
- [10] Lienhart, R. & Maydt, J., 2002, An Extended Set of Haar-like Features for Rapid Object Detection, *International Conference on Image Processing.*, Institute of Electrical and Electronics Engineers (IEEE).

- [11] Hecht-Nielsen, R., 1987. Counterpropagation networks. *Optical Society of America*, XXVI(23), pp.4979-84.
- [12] Gonzalez, R.C. & Woods, R.E., 2008, *Digital Image Processing*, 3rd ed Prentice Hall, New Jersey.
- [13] Fausett, L., 1994, *Fundamentals of Neural Networks Architectures, Algorithms and Applications*, Prentice Hall, New Jersey.