Human activity recognition using support vector machine for automatic security system

Taufiq Bagaskara Supriyatna\textsuperscript{1}, Surya Michrandi Nasution\textsuperscript{2} and Ratna Astuti Nugraheni\textsuperscript{3}

Departement of Computer Engineering, Telkom University, Bandung West Java, Indonesia.

E-mail:\textsuperscript{1}bagastopx@student.telkomuniversity.ac.id, \textsuperscript{2}michrandi@telkomuniversity.ac.id, \textsuperscript{3}ratnaan@telkomuniversity@ac.id

Abstract. Home is a place for human beings to live and also to socialize with the society. In this rapid technology development, a home security system is a must to prevent crimes. One of the innovation is home automation which enables controlling the doors automatically. RGB-D camera can detect human body movements which later can be used to recognize the activity through the Support Vector Machine (SVM) algorithm. The result of activity prediction can be used as home automation input for the home security system. In this paper, a security system analysis has been carried out which is beneficial for the users in maintaining the security system through combining the RGB-D Camera and Skeleton Tracking which then classified by using SVM algorithm. The result shows that the optimal data resulted through 1 meter distance with 100% accuracy, while testing variable C of SVM resulting in C = 2 as the best score for C variable with 92% accuracy.

1. Introduction
Human negligence causes the thievery of goods occurred. This paper aims to implement a human activity recognition using RGB-D Camera located in a room or bedroom, and the result of skeleton tracking is then classified using Support Vector Machine (SVM) algorithm. The training data for SVM classification are stored in a computer. From the detected activity, the doors can be locked and unlocked by the system automatically. The process of the program is very sensitive and depends on the accuracy of the Skeleton Tracking detection which is done in real time. Training data is made through capturing the user’s gestures. SVM algorithm is used for classification, and the results are forwarded to the microcontroller to control the automatic door lock. Users using SVM method as in related research on balance recognition system obtained accuracy value of 80% \cite{2}. By using this system, the user create an automated home security system towards crimes.

2. Related work
First, we review some theory related to human activity recognition based on RGB-D camera. Then, we provide a brief description of skeleton tracking.
2.1 RGB-D camera

RGB-D Camera has a function to record videos which are beneficial for recording the gesture displayed through capturing frame with screenshot in live stream [1] and other human activities detection feature by detecting three colour elements such as Red, Green, and Blue. One colored image is expressed as 3 grayscale matrices in the form of matrix for Red (R-layer), matrix for Green (G-layer) and matrix for Blue (B-layer) [7].

![Figure 1. RGB-D camera sensors [8].](image)

The RGB-D camera is similar with webcam, which can capture video at 640 X 480 resolutions with 32 bit colour at 30 frames per second [6]. Figure 1 shows the RGB-D camera used is kinect camera XBOX 360. The kinect camera technology used consists of infrared transmitters, an RGB-D camera and specially for detecting an object and human motion in three dimensional scale [4].

2.2 Kinect camera

Kinect depth sensor works according to the principle of structured light, which projects the familiar skeleton pattern to area in front of it, then kinect camera deduces the depth value from the distance calculation between the detected pattern [10]. The depth sensor sees the object as a collection of small dots. The infrared projector constantly projects these points along the range of view. These points are arranged in a pseudo-random pattern embedded in the sensor’s kinect camera. Infrared projector functions find out what pattern is formed, and how the point is drawn. From these points the infrared projector compares the distance of the image from the infrared camera to the pattern it produces, then uses the difference between them to calculate the distance of each point from the sensor [5]. If the object detected by kinect has an equation with the joint data possessed by kinect, then the object will later be recognized by kinect as a human [14].

2.3 Skeleton tracking

Skeleton tracking is the depth image processing to establish the position of the human skeleton joints [15]. With a skeleton tracked by Kinect, the first feature that can be extracted is the position of the joint. Because, each joint has 3 values of 3 coordinates and a frame consisting of 20 joints. Thus, the feature vector has 60 dimensions [16]. Skeleton tracking is human body joint detection that uses depth frame in processing the data from 3 positions namely, X axis, Y axis and Z axis. This property has a set of joints [5]. Skeleton tracking is a feature provided by Kinect SDK, and this feature can trace the main joint of human body. Skeleton tracking is not separated from the use of depth sensor. Depth sensor maps the objects based on the distance to be compared with previous training data [3].
Skeleton tracking is able to assume the coordinate points of the frame and the distance of the object from the camera. Skeleton tracking condition of the right hand \((1,0,0)\) and left hand \((-1,0,0)\) is divided from X axis, Y axis and Z axis. The axis is the coordinate point obtained from the kinect SDK feature. This paper uses six coordinate points which can be seen on Figure 2.

**2.4 Depth sensors**

Depth sensors consist of a combination of infrared laser projector and monochrome CMOS sensor takes video data in 3D regardless of the kinect camera's light conditions. Infrared in Kinect cannot be seen in plain view and is not harmful to the human body. Infrared is able to send thousands of rays that bounce to find out the object in front of it [3]. The infrared light generated by the CMOS sensor monochrome camera and measures the time of the existing light after reflected by the object in front of it [9].

**2.5 Support vector machine**

Support Vector Machine (SVM) is a set of guided learning methods that analyze data and recognize patterns, used for classification and regression analysis. The original SVM algorithm was created by Vladimir Vapnik and the current standard derivative is Soft Margin. Standard SVM takes the training data set and predicts for each input given, the possibility of input is a member of one of the two classes that exist, which makes an SVM a binary linear nonprobabilistic classifier. Because SVM is a classifier, then given a training set, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether new data falls into one category or another [12]. SVM learning supports various class problems by calculating hyperplants between each class and the rest uses SVM with a linear kernel to classify and analyze human body regression [13]. Determining the kernel function used will greatly affect the prediction results. Suppose \({x_1, \ldots, x_n}\) is a dataset and \(y \in \{+1, -1\}\) is the class label of data \(x_i\). Data that is right on the dividing plane is called support vector [11].

**3. Research method**

This paper describe the scenario of the activities carried out from the beginning of the program to the end of the program in the forms of flowchart. It starts from introducing the human activities detection by using RGB-D camera, capturing the data and stored it in the computer, and also the SVM classification. The following is the flow of the implementation system and designing the system till becoming the mechanism in accordance.
The flowchart as stated in the Figure 3 appears to start from the initiation of RGB-D camera whether it has been detected by the computer. After the camera initiation is done, the user body will be captured by RGB-D camera. From the gesture, the specified skeleton coordinates are stored in the computer data will be compared using the SVM method. The user can see the dataset in the show the dataset menu to see how many data for both sleeping and not sleeping activity.

### 3.1 Training data

The process in capturing data is shown in Figure 4, in which the process is started with obtaining training data needed. After that the user need to determine data to record by choosing an activity sleeping or not sleeping on data capturing menu beforehand.
The following is the explication of training data process shown in Figure 4:

- In the process of User Gesture, the system will track the user’s skeleton and determine its coordinate point.
- In the process of Capturing Data, the system will record input; sleeping and not sleeping.
- In the process of calculating using SVM, the training data will be saved separately and limited by class on Support Vector.
- The process of Show Dataset will display user’s coordinate points result that have been recorded.

**Figure 4.** Flowchart process system training data.
Figure 5. Not sleeping activities detection (Up); Sleeping activities detection (down).

In Figure 5 activities detected is the result of predicting activities obtained by using SVM (Support Vector Machine) method; sleeping and not sleeping activities. If the user’s activities detection is successful, the program will send command through serial port toward hardware to control the automatic lock.
3.2 Classification of support vector machine

The classification of SVC (Support Vector Classification) method is a process to find the model or function that explains or differentiates concept or class of data for predicting class that is not recognized from certain object. The measurement of SVM method classification is done by comparing the training data with testing data [2]. The following are the concept of classification:

- Inputting the training data based on the training data saved in the computer that are normalized based on class (0 and 1).
- Determining dot product of every data by using linear kernel function. Data transposed by using multiplication matrix $A \times A^T$. Data $X_i = \{1, 2, 3, \ldots, 70\}$ are the data used, and Class $Y_i= \{0, 1\}$ are class of sleeping and not sleeping.
- In the kernel model, every data will be compared with a series of data. Data will be compared with the data itself and other data. The explication can be seen in the example of table below:

| $X_1$ | $X_2$ | $X_n$ |
|-------|-------|-------|
| $K[X_1, X_2]$ | $K[X_1, X_2]$ | $K[X_n, X_n]$ |
| $K[X_2, X_1]$ | $K[X_2, X_1]$ | $K[X_n, X_n]$ |

- The calculation of testing and transpose data used kernel formula to get value dot product below:
  \[ K(X_i, X_j) = X_i \cdot Y \]  

- Calculating matrix formula by using the following formula:
  \[ D_{ij} = y_i y_j (K(x_i, x_j)) + y^2 \]  

- After obtaining all data resulted from matrix calculation, the next step is finding error value by using formula:
  \[ E_i = \sum_{j=1}^{i} a_j D_{ij} \]  

- Counting value delta alpha by using the following formula:
  \[ \delta \alpha_i = \min \{ \max [y_i (1 - E_i), -\alpha_i] , C - \alpha_i \} \]  

- From the previous delta alpha value, it can found new alpha value by using the following formula:
  \[ \text{new } \alpha_i = \alpha_i + \delta \alpha_i \]  

- Finding bias value:
  \[ b = -\frac{1}{2} (<w, x_1> + <w, x_2>) \]  

- After finding $\alpha$, $w$, and $b$, testing data will be tested to count dot product value with training data by using kernel function. $K(x, y) = x \cdot y$.
- Where $x$ is testing data and $y$ is training data. And soon until all testing data are multiplied by all training data, thus the result of multiplication of testing data with dot product data found.
- Next, doing the calculation of decision function as follow:
  \[ f(x) = w \cdot x + b \]  
  \[ f(x) = \sum_{i=1}^{m} \alpha_i y_i K(x_i, y_i) + b \]

4. Experiment result and analysis

The research in this paper has an input in the form of Video streaming of skeleton object of human body obtained from RGB-D camera and skeleton that is processed by using SVM algorithm resulting in Home Automation to activate automatic door. After testing, the results obtained are as follow.

5. Testing to variation of distance

Systems testing toward variation of distance are carried out by recording 70 training data (35 data of sleeping, and 35 data of non-sleeping) with different distances. Experiment was done 30 times for each class.
Table 2. Test result of distance’s variation.

| Distance | TP  | FP  | FN  | TN  | Accuracy |
|----------|-----|-----|-----|-----|----------|
| 1 meter  | 30  | 0   | 0   | 30  | 100%     |
| 2 meter  | 25  | 5   | 0   | 30  | 91%      |
| 3 meter  | 19  | 11  | 0   | 30  | 81%      |

Table 2. Test result shows the accuracy level at a distance ranging from 1 to 3 meters between the user and kinect camera, using 35 data of sleeping and 35 data of not sleeping. The correctly classified sleeping data is labeled as TP (True Positive), while FP (False Positive) is for incorrectly classified sleeping data. The correctly classified non-sleeping data is labeled as TN (True Negative), while FN (False Negative) is for incorrectly classified non-sleeping data.

The level of accuracy vary for the parameters of 1 meter, 2 meters and 3 meters. The highest accuracy obtained is 100% at a distance of 1 meter, and it is because skeleton training is not fully detected between data testing and data training, causing the activity classification to always be successful. The distance of 2 meters obtained 91% accuracy, while 3 meters distance obtained 81% for the accuracy.

6. Testing of variation data and distance

This experiment is to see the effect of the various distance and also various ratio of sleeping and not sleeping data. With the total of 70 training data used for both sleeping and not sleeping activity, the experiment was done 30 times for each class.

Table 3. Test result of data ratio and distance variation.

| Distance | Data Ratio (Sleeping : Non-Sleeping) | TP  | FP  | FN  | TN  | Accuracy |
|----------|-------------------------------------|-----|-----|-----|-----|----------|
| 1 meter  | 20 : 50                             | 9   | 21  | 0   | 30  | 65%      |
|          | 50 : 20                             | 30  | 0   | 3   | 27  | 95%      |
| 2 meter  | 20 : 50                             | 10  | 20  | 2   | 28  | 63%      |
|          | 50 : 20                             | 30  | 0   | 2   | 28  | 96%      |
| 3 meter  | 20 : 50                             | 10  | 20  | 1   | 29  | 65%      |
|          | 50 : 20                             | 30  | 0   | 2   | 28  | 96%      |

Table 3 shows the different accuracies given by each scenario. The optimum result for the distance of 1 meter obtained when using 50 data of sleep and 20 data of not sleep, the accuracy reach 95%. As for the distance of 2 and 3 meters, the optimum scenario was also obtained when using 50 sleep data and 20 not sleep data, which resulted in 96% accuracy. If being compared with data variations using 20 data of sleep and 50 not sleep, both distance of 1 and 3 meters got the accuracy of 65% and the distance of 2 meters obtained an accuracy of 63%. This indicates the more training data of sleep activity, the better the accuracy value, if the ratio of data to sleep more than the data is not sleeping, then the level of sensitivity of research on sleep activity is better and more accurate.

System setting experiment variable C value on SVM (Support Vector Machine) method using 70 data at a distance of 2 meters with 35 sleep data and 35 data not sleeping. Tests were conducted 20 times using a linear kernel. The following table of test results:
Table 4. Conclusion of value c.

| Value C | Successful | Failed | Accuracy |
|---------|------------|--------|----------|
| 2       | 37         | 3      | 92%      |
| 5       | 32         | 8      | 80%      |
| 10      | 35         | 5      | 87%      |
| 50      | 20         | 20     | 50%      |
| 75      | 14         | 26     | 35%      |

Based on the experiment above can be concluded that testing the value of C variable affects the value of the accuracy system. The table shows that C=2 got the best accuracy of 92% compared to other value of C. Value of C is computation value of the system in processing data, the greater value of C that is changed causes longer computation and decreases accuracy.

7. Conclusion

This paper explains the human activities detection experiments using SVM algorithm which is in line with the specified activities for automation of security doors. The optimal data training variation on the system is 50:20 for sleep and not sleep. The experimental system on C score variable of SVM resulting in C = 2 as the best score for C variable. SVM algorithm obtained optimal data at 1 meter distance with 100% accuracy, 91% accuracy for the distance of 2 meters, and 81% accuracy for 3 meters, for further research users can develop using methods other than SVM.

References

[1] Gunawan, Alexander AS and Ashadi Salim 2013 Pembelajaran Bahasa Isyarat Dengan Kinect Dan Metode Dynamic Time Warping Journal Mat Stat 13 2 77-84
[2] Jayanti, Ni Ketut Dewi Ari, Kadek Dwi Pradnyani Novianti and I. Wayan Sumalya Implementasi Metode Support Vector Machine Pada Sistem Pengenalan Jejaitan. Semmasteknomedina online 5.1 2017: 1-2
[3] Hartono, Hartono, Liliana Liliana and Rolly Intan 2015 Pendeteksian Gerak Menggunakan Sensor Kinect for Windows Journal Infra 3.2 pp-375
[4] Budiman, Rendi, Imam Kuswardayana and Dwi Sunaryono 2012 Integrasi kinect dan unreal development kit menggunakan kerangka kerja openNI pada studi kasus game berbasis interaksi gerakan.Technical Journal ITS 1.1 A208-A210
[5] Setyani Fatima, R Rizal Isnanto and Kurniawan Teguh Martono 2014 Pengenalan Bahasa Inggris untuk Anak Melalui Permainan Edukasi" The Zoo" Berbasis Kinect Journal of Technology and Computer Systems 2.1 136-142
[6] Hartono H, Liliana L and Intan R Pendeteksian Gerak Menggunakan Sensor Kinect for Windows Journal Universitas Kristen Petra pp 2-3
[7] Mulyawan Hendra, Hadi Samsono and M Zen Setiawardhana 2017 Identifikasi dan Tracking Objek Berbasis Image Processing secara Real Time Journal Campus PENS-ITS
[8] Wang, Youwen, et al 2012 Kinect based dynamic hand gesture recognition algorithm research Intelligent Human-Machine Systems and Cybernetics (IHMSC), 4th International Conference vol 1
[9] Utami, Dina Budhi and Muhammad Ichwan 2017 Pengenalan Pose Tangan Menggunakan HuMoment Journal Infotel 9.1 100-107
[10] Hadi, Sabitha Wildani, Gembong Edhi Setyawan and Rizal Maulana 2017 Sistem Kendali Navigasi Ar. Drone Quadcopter Dengan Prinsip Natural User Interface Menggunakan Microsoft Kinect Journal Information Technology Development and Computer Science e-ISSN 2548 2017: 964X
[11] A Muis Imelda and Affandes Muhammad 2015 Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet. Journal of Science, Technology and Industry vol 12 pp 189 – 197
[12] Kartika Lidya, Syahfitri Opim Salim and Sitompul Efendi, Syahril 2015 Sentiment Analysis
pada Teks Bahasa Indonesia menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN) Journal National Seminar on Information and Communication Technology

[13] Patsadu O, Nukoolkit C and Watanapa B 2012 Human Gesture Recognition using Kinect Camera Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference pp 28-32

[14] Indarto, Elbert Teguh, Edi Leksono, and Eko Mursito Budi 2015 Design of Human Presence Detection System using Sensor Kinect Journal Control Automation and Instrumentation 7.1 27

[15] Wibowo, Achmad Teguh, Erni Yudaningtyas and Sunaryo Sunaryo 2013 Natural User Interface Technology Using Kinect as a Trigger of Hardware Work Based on Fuzzy Inference System Journal EECCIS 7 1 1-6

[16] Le T L and Nguyen M Q 2013 Human posture recognition using human skeleton provided by Kinect 2013 International Conference Computing, Management and Telecommunications (ComManTel) pp 340-345