Human activity recognition using skeleton data and support vector machine

Komang Mandira G A¹, Surya Michrandi N² and Ratna Astuti N³

¹²³School of Electrical Engineering, Telkom University, Bandung, Indonesia

E-mail: ¹mandiragiri@student.telkomuniversity.ac.id,
²michrandi@telkomuniversity.ac.id, 
³ratnaan@telkomuniversity.ac.id

Abstract. In this paper, we propose a method for recognizing human activities using skeleton data by RGB-D camera, namely Kinect device. The human activity recognition is a learning in the field of computer vision. In its application, the recognition of human activity can be used for a sign language learning, human-computer interaction, surveillance of the elderly, image processing and etc. Our approach is based on skeleton data with coordinate value of each joints in human body, that will be classified using support vector machine algorithm when performing a movement to predict the activities name. Experiments were performed with a new training data that we’ve create manual from capturing movement while human target are doing activities. Experiments result show that the system best average accuracy is 93.75% of all activities prediction with the optimal distance of object to the devices is 2 meters.

1. Introduction

Human activity recognition is one of the most popular research on computer vision area. This is due to some of the promising applications in areas such as visual surveillance, human performance analysis and computer-human interfaces [1]. Human activities contain movement and body position when doing some activities. Tracking the human body position can be done with RGB-D camera. RGB-D Camera contain Red, Green, Blue and Depth sensor. The example of RGB-D camera is Kinect Devices. The advantages of using Kinect devices are Kinect devices can track human joints position, can be used in low light condition, match for using in the indoor place, reliability, and competitive cost.

Kinect device can recognize human activities by capturing skeleton data. Skeleton data can obtain coordinate value of each joint detected by the device. The data will be used for classifying human activity based on movement done. We choose 8 different activities; waiting, drinking, playing handphone, reading, writing, watching tv, dozing, and sleeping. Position of the activities will be decided by three, there is stand up, sit down, and lie down. The activities will be used for training and testing data.

The activities will be classified using support vector machine method. This method based on author [1] [3] [5] that obtain high accuracy and best performance. The support vector machine is selected because it performs best in large feature space, can handle large classification of data better, and match for classifying non-linear data. Human activity recognition need coordinate data of each human joints position. This coordinate is a large data with variative value of each joints. Support vector machine is the best method for classifying a large data that’s why we use this method to recognize human activity. The basic process of recognizing human activity can be done with system block diagram on Figure 1.
Skeleton data are used for training and testing data, then support vector machine will classify and predict the activity from testing data. The output is activity name done by human target based on testing data.

![System block diagram](image)

**Figure 1.** System block diagram.

2. **Related work**

Trinh Hoai An Et al. [2] proposed the fallen motions recognition by humans. Researchers used the Support Vector Machine (SVM) method to detect fallen human movements. Detection of fallen human movements using Kinect sensors that can initiate posture and human skeletons with depth sensors. Detection of fallen human movements is divided into four, falling back, falling forward, falling to the left side and falling to the right side. In the test, the researchers compared the movement by using the SVM method to know the normal movement and abnormal movement (fall). In addition, there are three tests conducted that is by comparing the results of the visuals that have never demonstrated the falling motion and who have never demonstrated. The result is quite high that is 93.5% the level of accuracy with the viewer who ever made the movement and 85.3% in the view that has never made a falling motion.

Megha D. Bengalur [1] proposed the human activities recognition using body poses with the Support Vector Machine method to get interesting results. The recognition of the body pose is done by using the Kinect camera as a means to capture data in the form of images captured by the camera. Images containing human body poses with depth sensor feature allow the camera to detect skeletal. Skeletal is used as data to recognize actions / activities performed by humans. The method of classifying it using the Support Vector Machine by comparing human activity with the dataset / training data of human activities stored on the system. In this research it is explained that the test is done at different place that is indoor (indoor) and outdoor (outside the room). Activities tested in the form of drinking, sleeping, reading, stretching the hand, writing for indoor activities. While the activity outside the tested room is clapping, jumping, stretching, running, walking, shaking hands, hugging, and drinking while walking. The use of SVM as a method to classify human activity with datasets provides accurate results. Multi-class SVM in this study aims to compare real-time activities with many activities contained in dataset / training data. The results obtained are 89% of activities performed by humans can be accurately detected by the system.

Salvatore Gaglio Et al. [3] uses two methods: Support Vector Machine (SVM) and Hidden Markov Model (HMM) in the recognition of human activity. Data collection is done with Kinect camera where Kinect camera is a camera that has many features such as, RGB Camera and depth sensor. Kinect camera can capture images with low light conditions and Kinect camera can know the distance of objects caught by reflecting infrared light to the object and then analyze the distance of the object. In this study, Kinect cameras are used as data retrieval in the form of human body postures caught on camera and then processed by the system so as to bring out a kind of human joints that are selected. Furthermore, the human posture is classified with SVM so as to obtain the results of movement captured system. HMM here served as a determinant of activities conducted by humans. Determination of human activity is done by comparing activity with training data that has been stored before. The training data used are two KARD (Kinect Activity Recognition Dataset) and Cornell Activity Dataset (CAD-60). The results of the accuracy of the SVM and HMM method with KARD and CAD-60 training data show excellence in KARD training data.

Orasa Patsadu Et al. [4] proposed the recognition of gestures / gestures using Kinect cameras by comparing four methods to obtain different degrees of accuracy in the recognition of human gestures. Kinect camera is used as a data retrieval tool by taking data from the posture of the human body caught
on camera. The methods used in this research are Backpropagation Neural Network (BNN), Support Vector Machine (SVM), Decision Tree (DT) and Naïve Bayes (NB). The test is to recognize three gestures of human standing, sitting and lying down. Then the recognition is compared with the training data that has been stored that is by using data training KNIME. KNIME is a data analyst that is open source. In this study, this training data is used to classify human gestures. The results obtained with the four methods produce different accuracy. In this research, the highest level of accuracy was obtained by BNN method that is 100% and followed by SVM method with 99.75% accuracy.

Haitao Wu Et al. [5] proposed the recognition of human activity by combining two methods of Support Vector Machine (SVM) and Hidden Markov Model (HMM). Human activities are captured by Kinect cameras. Kinect cameras provide information about the depth of images that can be used to represent human gestures and provide skeletal joints on recorded data. The data is then processed to be classified by the SVM method and the activity obtained is determined by HMM. The result of the combination of these two methods gained a high level of accuracy in the recognition of human activity. In this research, the comparison of three methods, the first method of SVM with 94.9% accuracy, HMM method with 93.58% accuracy rate and the combination of both SVM and HMM methods obtained 98.22% precision.

3. Methodology
We created a system for human activity recognition that use support vector machine algorithm for classifying and predicting activity name based on human movement. Figure 2 shows our flowchart for training and testing steps. The input is skeleton data from Kinect device. For the training step we must input the human body position and activity name manually. For testing data we must test it in real time. The data information on training steps contains name of activity, body position (stand up, sit down or lie down), time created the training data, and coordinate joints each parts of human body. We used supervised learning that must contain training data for classifying testing data to get the prediction result.

![Figure 2. Flowchart system.](image-url)
3.1. **Skeleton joints position**

Human skeleton can be done with Kinect for Windows Software Development Kit. This feature name is skeleton basic. Skeleton basic feature can track humans with various joint points. Skeleton basic can track up to six users and up to 20 joints for each skeleton, but only two users can be tracked in detail [6]. Which means that the sensor can track twenty points of the skeleton joints.

![Skeleton joints position](image)

*Figure 3. Skeleton joints position [11].*

That 20 joints is head, center of shoulder, spine, center of hip, left hip, right hip, right shoulder, right elbow, right wrist, right hand, left shoulder, left elbow, left wrist, left hand, right knee, right ankle, right foot, left knee, left ankle, left foot. In this paper we used all joints position for best performance and precision. In Figure 3 is the illustration of skeleton joint positions.

3.2. **Support vector machine**

The Support Vector Machine (SVM) was first introduced in 1995 by Cortes and Vapnik [7] to solve classification and regression problems. The idea of SVM is searching for best line that divided two class and decided the support vector [8]. This line called hyperplane. Support vector machine used for linear classification but SVM is developed in order to use for non-linear classification by inserting kernel trick on feature space. The nonlinear concept of SVM is that data $x^*$ is mapped by the function $\Phi(x^*)$ to a vector space with a higher dimension [9].

![Mapping data to new feature space](image)

*Figure 4. Mapping data to new feature space [9].*
Figure 4 show that the input space change to feature space to reconstruct the data. It shows that yellow and red data on input space can’t separate by linear solution. function $\Phi$ maps each data in the input space to a new vector space with a higher dimension (3 dimensions), where both classes can be separated linearly by a hyperplane. The math notation for this mapping is,

$$\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^q \quad d < q$$ (1)

This mapping is done by maintaining the data topology, in the sense that the two data that are close to the input space will also be close to the feature space, whereas the two data that are far away in the input space will also be far away in the feature space [9].

Furthermore, the learning process in SVM in finding support vector points only depends on the dot product of the data that has been transformed in a new space with a higher dimension, namely $\Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$. Because generally this transformation is unknown, and is very difficult to understand easily, then the dot product calculation according to Mercer’s theory can be replaced by K function $K(\vec{x}_i, \vec{x}_j)$ which defines implicitly the transformation $\Phi$ [9]. This is referred to as the Kernel Trick, which was formulated with,

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$$ (2)

In this paper we using Gaussian Radial Basis Function kernel (RBF). Considering all the circumstances, regardless of the low-dimensional, high-dimensional, large sample or small sample, the Gaussian kernel function is considered as an ideal choice, it also has a wider domain of convergence [10]. The expression of RBF kernel is,

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\frac{||\vec{x}_i - \vec{x}_j||^2}{2\sigma^2}\right)$$ (3)

3.3. Testing scenario
The testing scenario will be testing in real time activity recognition performed by single person with several distance 1, 2, and 3 meters. Each activity will be tested 30 times to obtain the correct and incorrect result prediction of activity name. The accuracy performed based on the number of correctly predicted from the classification set model and the number of incorrectly predicted from the same classification set model. This relation can be calculated by this function,

$$\text{Accuracy} = \frac{\text{Correct Test Data}}{\text{Total Test Data}} \times 100\%$$ (4)

4. Result
4.1. Training data
We created 8 activities for training data by capture human skeleton while doing each activity in several position. The list of activity is waiting, drinking, playing handphone, reading, writing, watching tv, dozing, and sleeping. The data is coordinate value each joints position for each activity. Table 1 show that all activity with skeleton images.
Table 1. Training data list.

| Position | Activity Name   | Images          |
|----------|-----------------|-----------------|
| Stand up | Waiting         | ![Standing Waiting Images](image1) |
| Stand up | Drinking        | ![Standing Drinking Images](image2) |
| Stand up | Playing handphone | ![Standing Playing Handphone Images](image3) |
| Sit down | Reading         | ![Sitting Reading Images](image4) |
| Sit down | Writing         | ![Sitting Writing Images](image5) |
| Sit down | Watching tv     | ![Sitting Watching TV Images](image6) |
| Sit down | Dozing          | ![Sitting Dozing Images](image7) |
| Lie down | Sleeping        | ![Lying Sleeping Images](image8) |
4.2. Testing data

The experiments performed 30 times for each activity to decide the correct and incorrect prediction. Training data used in this paper is 10 for each activity. The result of each activity prediction accuracy is from Table 2, 3, and 4. Testing data on stand up position and waiting activity obtain high error on 1 meter distance because Kinect device cannot track the joints of left hip, right hip, right knee, right ankle, right foot, left knee, left ankle, and left foot. The distance between Kinect device and the human body is too close. The best distance for this testing data is 2 meter.

Table 2. Testing data with 1 meter of distance.

| Position | Activity name | Prediction | Accuracy |
|----------|---------------|------------|----------|
|          |               | Correct    | Incorrect |          |
| Stand up | Waiting       | 19         | 11       | 63.33%   |
| Stand up | Drinking      | 30         | 0        | 100%     |
| Stand up | Playing handphone | 25   | 5        | 83.33%   |
| Sit down | Reading       | 30         | 0        | 100%     |
| Sit down | Writing       | 30         | 0        | 100%     |
| Sit down | Watching tv   | 30         | 0        | 100%     |
| Sit down | Dozing        | 28         | 2        | 93.33%   |
| Lie down | Sleeping      | 16         | 14       | 53.33%   |
| **Total** |              | 208        | 32       | 86.67%    |

Table 3. Testing data with 2 meter of distance.

| Position | Activity name | Prediction | Accuracy |
|----------|---------------|------------|----------|
|          |               | Correct    | Incorrect |          |
| Stand up | Waiting       | 30         | 0        | 100%     |
| Stand up | Drinking      | 30         | 0        | 100%     |
| Stand up | Playing handphone | 30  | 0        | 100%     |
| Sit down | Reading       | 29         | 1        | 96.67%   |
| Sit down | Writing       | 27         | 3        | 90%      |
| Sit down | Watching tv   | 29         | 1        | 96.67%   |
| Sit down | Dozing        | 29         | 1        | 96.67%   |
| Lie down | Sleeping      | 21         | 9        | 70%      |
| **Total** |              | 225        | 15       | 93.75%    |

Table 4. Testing data with 3 meter of distance.

| Position | Activity name | Prediction | Accuracy |
|----------|---------------|------------|----------|
|          |               | Correct    | Incorrect |          |
| Stand up | Waiting       | 29         | 1        | 96.67%   |
| Stand up | Drinking      | 28         | 2        | 93.33%   |
| Stand up | Playing handphone | 27  | 3        | 90%      |
| Sit down | Reading       | 28         | 2        | 93.33%   |
| Sit down | Writing       | 27         | 3        | 90%      |
| Sit down | Watching tv   | 30         | 0        | 100%     |
| Sit down | Dozing        | 24         | 6        | 80%      |
| Lie down | Sleeping      | 22         | 8        | 73.33%   |
| **Total** |              | 215        | 25       | 89.58%    |
5. Conclusion
In this paper, human activity recognition has been introduced by support vector machine with radial basis function kernel. The average result is 86.67% accuracy for each activity in 1 meter distance, 93.75% accuracy for 2 meter distance and 89.58% accuracy in 3 meter distance. The best distance between Kinect devices and the object is 2 meter with highest accuracy result. For future work, the designed system can be tested in more complex activity which has a high similarity in the human shape and movement.

References
[1] Megha D Bengalur 2013 Human Activity Recognition Using Body Pose Features And Support Vector Machine International Conference on Advances in Computing Communications and Informatics (ICACCI) pp 1970-1975
[2] Trinh Hoai An, Truong Quang Phuc, Nguyen Thanh Hai and Tran Thanh Mai 2015 Support Vector Machine Algorithm For Human Fall Recognition Kinect-Based Skeletal Data 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science
[3] Gaglio S, Re GL, Morana M Human 2015 Activity Recognition Process Using 3-D Posture Data IEEE Trans. Human-Machine Systems pp 586-97
[4] Orasa Patsadu, Chakarida Nukoolkit and Bunthit Watanapa 2012 Human Gesture Recognition Using Kinect Camera Ninth International Joint Conference on Computer Sience and Software Engineering (JCSSE) pp 28-32
[5] Haitao Wu, Wei Pan, Xingyu Xiong and Suxia Xu 2014 Human Activity Recognition Based on the Combined SVM&HMM International Conference on Information and Automation.pp 219-224
[6] Jana A 2012 Kinect for windows SDK programming guide Packt Publishing Ltd
[7] Vapnik V and Cortes C 1995 Support Vector Networks Machine Learning 20 273-297
[8] Pusphita Anna Octoviani, Yuciana Wilandari and Dwi Ispriyanti 2014 Penerapan Metode Klasifikasi Support Vector Machine (Svm) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang Jurnal Gaussian vol 3 4 pp 811 – 820.
[9] Anto Satriyo Nugroho, Arief Budi Witarto, and Dwi Handoko 2003 Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika
[10] Ye Zhihua and Li Honglian 2012 Based on Radial Basis Kernel Function of Support Vector Machines for Speaker Recognition 2012 5th International Congress on Image and Signal Processing (CISP) pp 1584-1587
[11] David Catuhe 2012 Programming with the Kinect for Windows software development kit Pearson Education