Kinect - Based Application System Design for Pencak Silat Movement using Support Vector Machines (SVMs)

Ernest Caesar Omar Syarif, Zener Sukra Lie*, Winda Astuti
Automotive and Robotics Engineering Program, Computer Engineering Department
BINUS ASO School of Engineering, Bina Nusantara University
Jakarta 11480, Indonesia
ernest.syarif@binus.ac.id, lie.sukra@binus.edu, wastuti@binus.edu

Abstract
Pencak Silat is an Indonesian martial art that has been recognized by UNESCO as a World Cultural Heritage object in 2019. In the practice, it is very difficult to know the accurate movement because no one has control over the movement except with instruction from their coach. Therefore, using Kinect – based study, we developed an application that could estimate body positions compared to stored model positions. By using a skeleton stream which has vector data, we can estimate the angle between two vectors and save it as a reference for further assessment with the user model. In the classification, Support Vector Machines (SVMs) was used which the implementation was done in LabView program. This application has an accuracy rate more than 90% for performing assigned tasks.

Keywords
Pencak Silat, Support Vector Machine, Kinect, LabVIEW

1. Introduction
Pencak Silat is one of the styles of martial arts from Indonesia that have gained traction in the international community, as we can find traces of this style in films and international competitions such as Asian Games. In general, Pencak Silat techniques consists of two important components, that is the basic techniques and the attack techniques. The basic technique consists of four main parts, the stance, tide attitude, stride and slam; and attack technique also has four main parts namely punch, kick, rebuttal, and kickback (D.F. 2000). So, in the process of learning, this specific martial art is necessary to reiterate the correct technique. In the development process, it was realized that traditional guidance (face-to-face) has its own limitation from the trainer. Therefore, we require innovation in the technology aspect for this martial art specifically for training. So that, the student can continue without a coach being present but still have the right guidance for the techniques.

Technologies in the world of sports are widely used for game development techniques, strategy analysis, and security. For example, the EPTS (Electronic Performance & Tracking System) technology used in soccer to analyze the physical performance of players, IoT application for monitoring the physiological parameters in the rehabilitation athlete training (Jiang 2020), athlete's heart rate monitor (Xiao 2020), and training systems using augmented reality (AR) (Soltani and Morice 2020). However, the use of technology in pencak silat has been only implemented for the competition, namely, using VAR (Video Assistant Referee) technology to help referees and judges make the decisions. The other words, the technology to help the process of pencak silat movement techniques training has yet to exist. In this research, a system to help users train and understand the correct stance positions is being developed using Kinect V2 and LabVIEW (Laboratory Virtual Instrument Engineering Workbench) as the platform.

Kinect is a motion-sensing peripheral created by Microsoft that can detect human gestures for various purposes, for example for a controller using body movements, education, games, and also rehabilitation. This sensor is able to detect human movements based on skeleton tracking which is able to detect the shape of the human body using a depth sensor and an RGB camera (J.S. 2013). In the implementation, Kinect sensors has been used in various systems such as motion capture system (Iliukhin et al. 2017), badminton movement recognition and analysis system (Ting et al. 2015), golf swing reorganization and segmentation system (Zhang et al. 2012) and so on.

LabVIEW is a graphical programming platform from National Instruments that is capable of data acquisition, data processing, data analysis, to instrument and equipment control (Travis and Kring 2007). This platform is commonly used for data acquisition because of its robustness, easy to use and a vast selection of application-specific libraries (Larsen 2011).

From these technologies using Kinect and LabVIEW, able to detect the position or movement point are expected. Then, the system would be informed the error in the user positions or movement. By implementing the kinectone Makerhub library, this system can be developed on LabVIEW platform to read the user's skeletal data (specified angle of the joint points) from the Kinect V2 sensor. After that, the user's position data will be logged and then classified using Support Vector Machine (SVM). This system is expected beginners and athletes can train and study the correct stances without a trainer or coach being present.

2. Methods and Literature Reviews

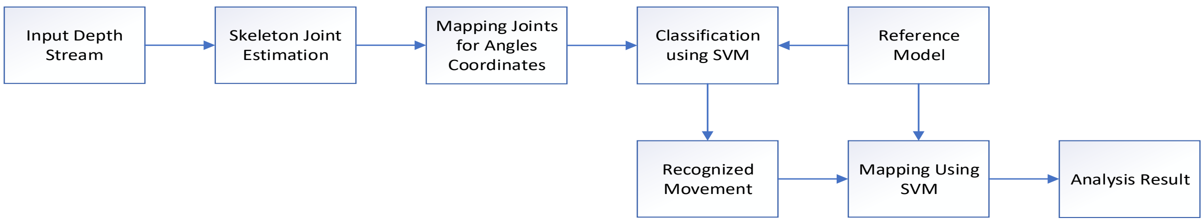
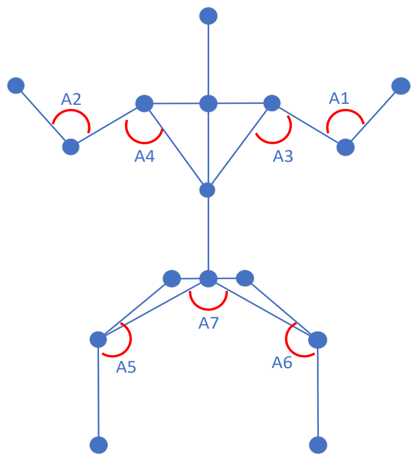


Fig.1 An Overview of the Proposed System

Figure 1 illustrates a schematic of our proposed system, which consists of movement recognition and analysis frameworks. For the pencak silat stance and sample testing are acquired by Microsoft Kinect V2 sensor and skeletal joints are estimated subsequently. Then, the tracked joints are calculated to get the angles coordinates. Finally, recognition and analysis results are obtained by mapping two time-series data using Support Vector Machine (SVM) algorithm to get the information of the stance form and the error of the angles. The use of SVM methods is due to the mapping idea is non-linearly training data into higher-dimensional feature space through the kernel function. Besides based on the results show that the SVM method give the scalable and accurate image annotation results (Xiong et al. 2011).

For greater specification, Kinect V2 was used in this system due to higher depth image precision, and accuracy than its predecessor (O. and D. 2017), where the depth image is essential in the process of skeletal tracking. For raw skeletal data, the Microsoft Kinect sensor has embodied the data in BodyFrameSource. BodyFrameSource gives data frames roughly in sync with other Data Sources, each one detailing potential and directional data about the body joint (Rahman 2017). In Kinect V2 for windows, the skeletal model, which also knows as stick model, encompasses 25 body joints. In this research, several body joints used to give information for seven angles. Seven angles are Right Hand (Shoulder Right - Elbow Right – Wrist Right), Left Hand (Shoulder Left- Elbow Left – Wrist Left), Right Arm (Spine Mid – Shoulder Right – Elbow Right), Left Arm (Spine Mid – Shoulder Left – Elbow Left), Right Leg (Hip Right – Knee Right – Foot Right), Left Leg (Hip Left – Knee Left – Foot Left), Legs (Knee Left – Spine Base – Knee Right) which illustrated in **Figure 2**.



A1- Right Hand (Shoulder Right - Elbow Right – Wrist Right),
A2- Left Hand (Shoulder Left- Elbow Left – Wrist Left)
A3- Right Arm (Spine Mid – Shoulder Right – Elbow Right)
A4- Left Arm (Spine Mid – Shoulder Left – Elbow Left)
A5- Right Leg (Hip Right – Knee Right – Foot Right)
A6- Left Leg (Hip Left – Knee Left – Foot Left)
A7- Legs (Knee Left – Spine Base – Knee Right)

Figure 2. Several joints from skeletal model in Kinect Sensor V2, which used to present the angles. Right Hand (A1), Left Hand (A2), Right Arm (A3), Left Arm (A4), Right Leg (A5), Left Leg (A6), Leg (A7)

After the angles have been determined, the angle calculation will be performed using 2018 32-bit LabVIEW program with kinectone makerhub library. With kinectone makerhub library, the reference of the camera sensor Kinect V2 is opened and initiated. Calculation of each angle will be done using 3 joint points which have been determined. It should be noted that the points taken have 3-dimensional elements so calculations must be made on the x, y, and z axes. Here is a mathematical model if you want to calculate the angle at B from 3 joint ABC points:

$$\cos x = \left(\frac{\sum(\overline{BA} \times \overline{CB})}{|\overline{BA}| \times |\overline{CB}|}\right) \quad (1)$$

$$\frac{x}{\pi} \times 180° = \theta \quad (2)$$

Where θ is angle in radian, x is angle in degree, $\overline{BA}$ is the coordinate vector between A and B, $\overline{CB}$ is the coordinate vector between B and C, $|\overline{BA}|$ is magnitude for coordinate vector between A and B, and $|\overline{CB}|$ is magnitude for coordinate vector between B and C.

In the process, four movements of the stances, namely front, back, middle, and side stance as shown in **Figure 3** were collected for the basic data. A front stance is a posture where one leg is in the front and the other leg is on the back, while the whole-body weight is supported by the leg in front. Two feets must form a 30 degrees angle. A back stance is a posture where one leg is in the front and the other leg is on the back, while the whole-body weight is supported by the leg on the back. the front foot must be positioned straight and back foot forming a 60 degrees angle. A middle stance is a stance with two feet aligned straight with shoulders and body weight is equally supported by one bent leg. two aligned feet must form a 30 degrees angle. And a side stance is a stance where both legs are open out parallel to the body with the bodyweight is supported by a bent leg, two aligned feet must form a 30 degrees angle.

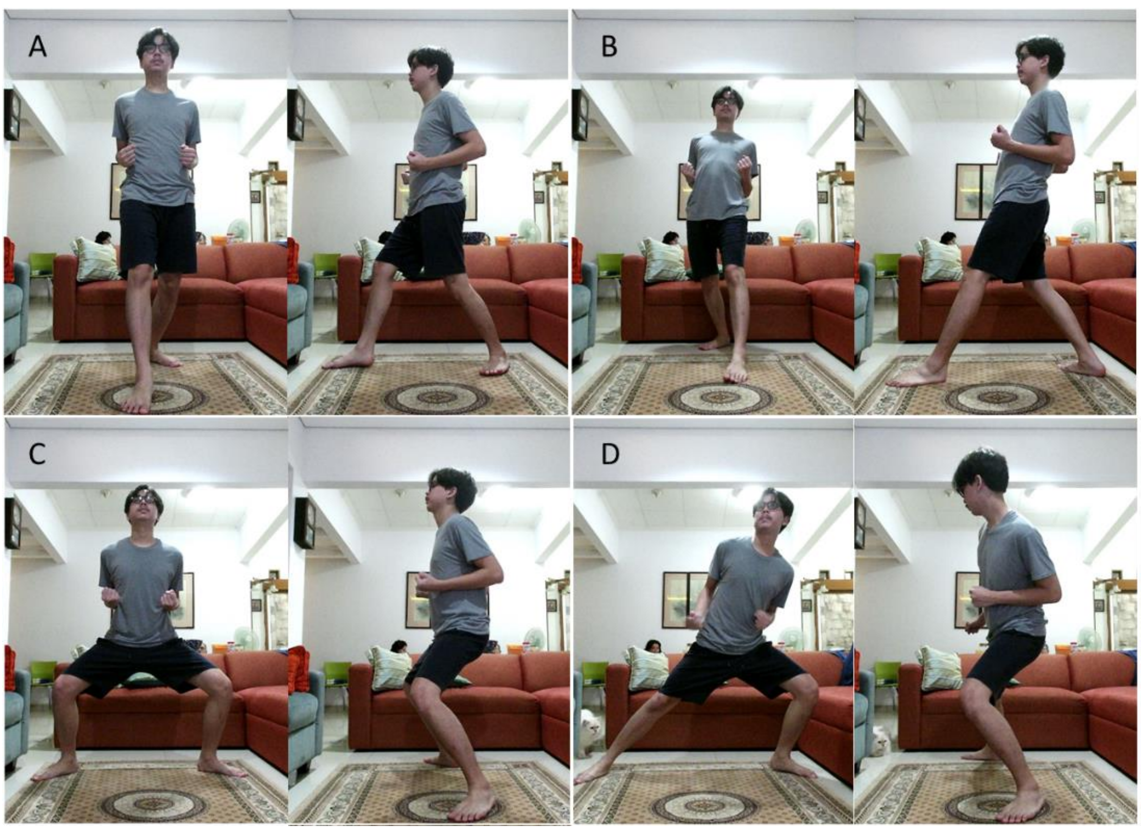


Figure 3. Front and side view of the four-basic stance, front stance (A), back stance (B), middle stance (C), side stance (D)

The amounts data of basic data used for training and testing are approximately 60 data for each form. Data retrieval positions that performed, are the distance between the subject and the camera by 200-240 cm, the height of the camera 53.5 cm from the floor, and the subjects have height around 175-185 cm. Because the inability of the Kinect V2 platform of using multiple Kinect sensors (C. et al. 2019), The subjects have to face 45° from the camera during the process to reduce the probability of the Kinect sensor from unable to trace the joints is not detected by the sensor

To identify, the SVMs, which the input is coming from the sensor data, were used. SVM is a learning machine technique, based on statistical learning theory. There are two important parts that have to be considered. The first is discriminative classification and the second is optimization via Lagrange multipliers. In discriminative classification, binary classification task with y=±1 labels must be considered. SVMs seek a globally optimized solution to avoid over-fitting, so that, it will have the ability to deal with a large number of features. In the linear separable case, there exists a separating hyperplane which function is;

$$\mathrm{w}\cdot \mathrm{x} + b = 0 \quad \mathrm{w} \in R^N, b \in R, \tag{3}$$

This implies;

$$yi(\mathrm{w}\cdot \mathrm{x} + b = 0) \geq 1, i = 1,\ldots,N \tag{4}$$

Optimized linear division constructs a hyperplane and separates into two classes (J.Salomon 2001). By minimizing ||w|| subject to the constrain, the SVMs approach tries to find a unique separating hyperplane. Here ||w|| is the Euclidean norm of w, and the distance between the hyperplane and the nearest data.

The points for each class are 1/||w||. With introducing the Lagrange multiplier *i*, the SVMs training procedure amounts used to solve a convex quadratic problem (QP). The solution is a unique globally optimization result, which has the following properties

$$\mathrm{w} = \sum_{i}^{N} \alpha_i y_i x_i \tag{5}$$

If *i* is not equal to zero, x_i are called support vectors. When SVMs is trained, the decision can be obtained by comparing every new example *x* with only the support vector {x_i}, *i* Є SV;

$$\mathrm{y} = \mathrm{sign} \sum_{i \in SV} \alpha_i y_i (x_i^T \cdot x) + b \tag{6}$$

In the real problem, a Non-linear separable case is often used incorporation into the linear separable case. Figure 3 shown a Non-linear separable data by introducing miss-classification data. In a non-linear separable case, SVMs perform a non-linear mapping of the input vector x from the input space into a higher dimensional feature space, where the mapping data is determined by the kernel function. The three typical kernel functions often used in data classification are listed in Table 1, in which the choice of the kernel is based on the characteristic of the data. The difference kernel function can be selected to obtain the optimal classification results. The choice of the degree in the polynomial kernel and weight in the Gaussian RBF kernel also depends on the characteristic of data.

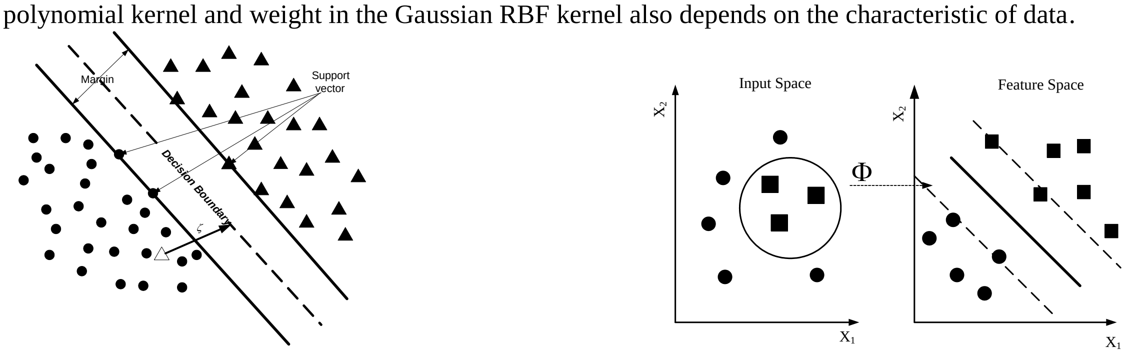


Figure 3. SVMs with linearly separable data.

Table 1. Formulation for kernel function (S.Y.Kung 2014)

Kernel	$K(x,x_i)$
Linear	$x^{T} \cdot x_j$
Polynomial	$(x^{T} \cdot x_j + 1)^{d}$
Gaussian RBF	$\exp(-\|x-x_j\|2/2\sigma^2)$

The extension of the SVM to solve the problem is applied by multi-class classification which used in this paper. This method is based on the binary classification problem namely, one-on-one multi-class classification. The classifier was constructed from the data combination of class classifications. The decision is based on the largest class which they most identify**.**

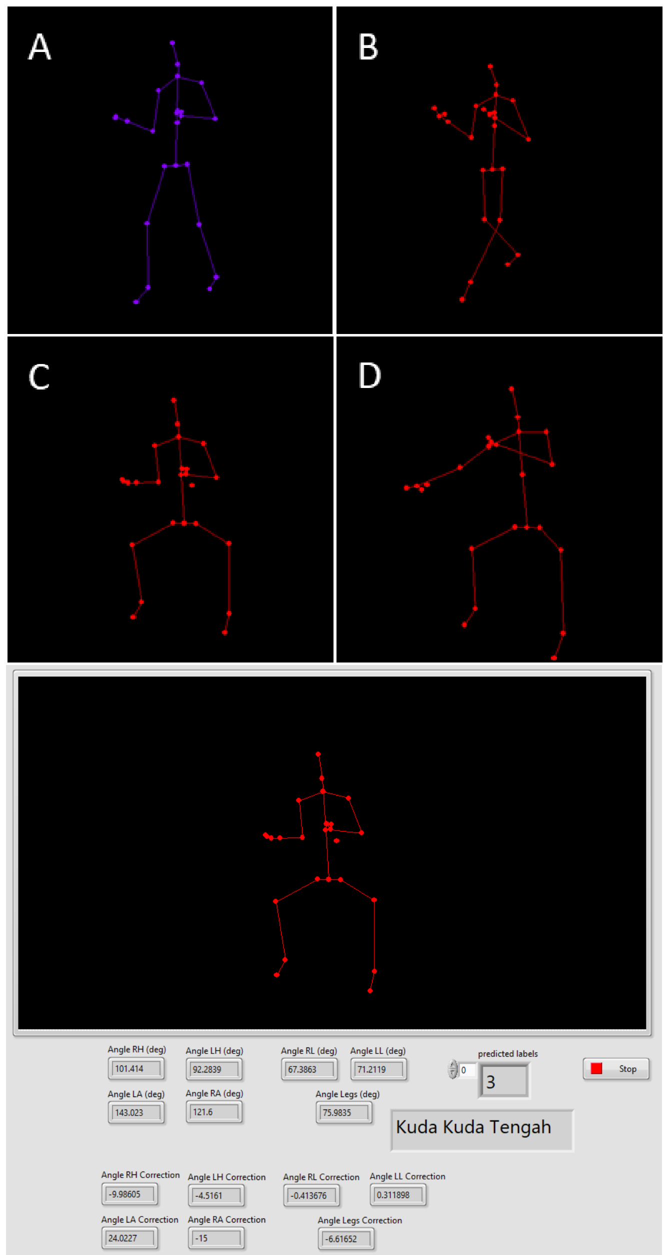


Figure 4. Skeletal view for 4 movements of the stances: a) front, b) back, c) middle, d) side forms and e) GUI (Graphical User Interface) of the system running.

3. Results and Discussion

Recognition of the object which converted into skeletal form is the first step in this study. It is shown in **Figure 4**(a, b, c, d) four skeletal images shown in real-time in the LabVIEW program are the images for the position of the stance from front, back, middle and side. From the results of this skeletal images, the calculation process is performed. It is useful to inform the angle of the user's position during the training process.

The GUI (Graphical User Interface) of the system shown in **Figure 4e**. A GUI displays objects that convey information and represent actions that can be taken by the user.in this system. There are have 5 important parts. Namely:

1. Skeletal data stream same as figure 4 (a,b,c,d)
2. Seven angles which are calculated and displayed in real-time
3. Seven angles with the correction shown how big the angle of error is from the reference angle
4. The label and the name of prediction stances. In the figure shown "kuda-kuda tengah" which mean the "middle stances"
5. A stop button to stop the program.

Table 2. Data optimization from datalogging which used for the benchmark of angle for each stance

Stance	Degree Right Hand	Degree Left Hand	Degree Right Arm	Degree Left Arm	Degree Right Leg	Degree Left Leg	Angle Legs
Front	112.6172	89.9732	127.448	115.6992	9.2512	13.7036	127.5372
Back	122.2044	110.4992	132.652	119.7392	24.3204	47.6928	145.1824
Middle	111.4227	96.84967	136.6757	119.0137	67.82067	70.92933	82.60333
Side	116.9068	25.40393	88.24929	112.3532	45.51786	18.93643	97.15464

In the recognition process, the Support Vector Machine (SVM) algorithm with the polynomial kernel function is used for the classification. Namely, the polynomial kernel function has the highest training accuracy model, that is 98%. Whereas the linear kernel function has the training accuracy model of 94,2% and Gaussian RBF kernel function has the accuracy model of 96.1%. Besides the recognition process, the system also performs data logging, and the data will be optimized. It is used as a benchmark for every movement. The benchmark data shown in **Table 2** is an averaged data logging for each angle of the stances, conducted for each movement of the stances.

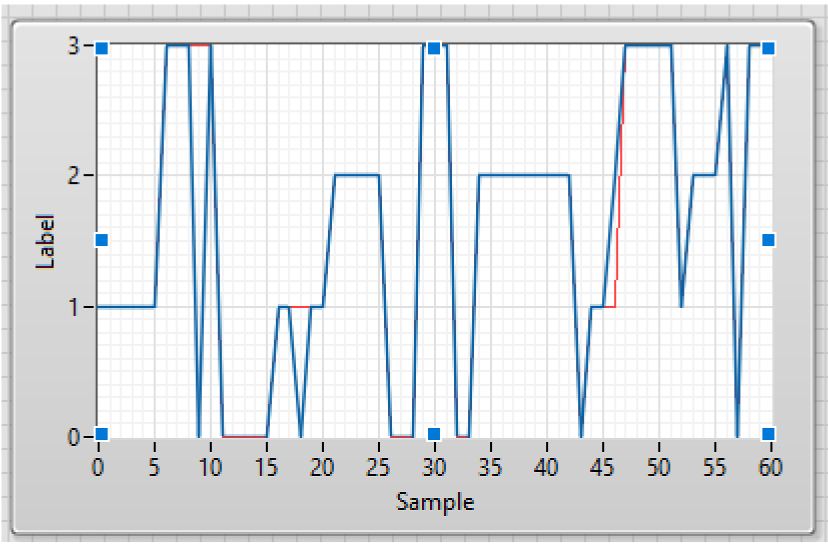


Figure 5. The examples of testing data which predict the motion of stances. Label 0, 1, 2, 3 for the prediction of front, back, middle and sides stances.

After classification, the process continues with testing. Results that we obtained from the process are shown in **Figure 5** where the x-axis is a number for the sample performed and the y-axis shows the movements made where 0 is for the front stances, 1 is for the back stances, 2 for the middle horses and 3 for rear stance. The graph showed that there are have 2 lines namely the blue line which shows the prediction data obtained and the red line shows the real data. From the experiments, the accuracy for every movement is 100% for the front stances, 86% for the rear stances, 100% for the middle stances, 94% for the side Stances. So, the average accuracy for the four models is 95%. Some of the imperfect case results are because of the frequent appearance of disturbances during the training process, imperfect movement position, and wrong movement.

4. Limitation and Future Development

The limitation of this experiment is that the system has only scratch the surface on the techniques of pencak silat which can only calculate the error of four different stance position. It is known that pencak silat have a lot of different styles of complex position and movement. However, given the time span of this experiment takes place can only make a classification model for four stances only, the future possibility for expanding to classify a wide horizon of positions and movements is attainable. This system is designed to calculate the error of one person, considering the fact that the Kinect platform is able to acquire six skeletal data in the same time. Another limitation, the number of samples are less diverse. The sample used has almost the same height of 175-185 cm which incidentally is the average height of an adult male.

It is already stated in the Methods and Literature Reviews section that the Kinect V2 has the probability of unable to trace the subject's skeletal data whenever the sensor cannot track the user's joints. However, the Kinect platform have already evolved with its newest iteration of camera, the Azure Kinect. The latest Kinect has a higher resolution of depth sensor than the V2 model, in which a higher resolution of depth sensor will impact the accuracy of skeletal tracking. The ability to connect multiple Kinects in the latest platform will increase the accuracy of the data and be able to record more complex positions and movements.

Aside from newer technology, for future development, we will conduct the testing with a more diverse number of samples. Furthermore, adding several features outside the stances such as punch, kick, rebuttal, and kickback.

5. Conclusion

In this paper, a Kinect based application was approach to detect the form of pencak silat movement. Using Support Vector Machine algorithm, the classify and mapping movement of the pencak silat stances was done, and experimental results have clearly shown the promising performance with an average recognition accuracy of 95%. In this paper, the angle calculation of stances form was being performed using 2018 32-bit LabVIEW which the information is very useful to conduct the pencak silat players to periodically benchmark their performances. In the future, this system design can be improving by adding several features for example the attack techniques and other basic techniques.

Reference:

D. F. Draeger, *The weapons & fighting arts of Indonesia*. Charles E. Tuttle Company; Boston: Mass.,2000.

Jiang, Y. *Combination of wearable sensors and internet of things and its application in sports rehabilitation. Computer Communications*, Computer Communication, 150, p167-176. doi: 10.1016/j.comcom.2019.11.021, 2020.

Xiao Ningning, *Combination of wearable sensors and internet of things and its application in sports rehabilitation*, Measurement, 108102. https://doi.org/10.1016/j.measurement.2020.108102, 2020.

Soltani, P., & Morice, A. H. P. *Augmented reality tools for sports education and training. Computers & Education*, Computers & Education, 155, 103923, https://doi:10.1016/j.compedu.2020.103923, 2020.

J. S. Jean, *Kinect hacks*. Beijing: OReilly, 2013.

Iliukhin, V., Mitkovskii, K., Bizyanova, D., & Akopyan, A. *The Development of Motion Capture System Based on Kinect Sensor and Bluetooth-Gloves*. Procedia Engineering, 176, 506-513. doi: 10.1016/j.proeng.2017.02.350, 2017.

Ting, Huong Yong & Sim, kok swee & Abas, Fazly. *Kinect-based badminton movement recognition and analysis system.* 14. 25-41, 2015.

Zhang, Lichao & Hsieh, Jui-Chien & Li, Shaozi. *A Kinect Based Golf Swing Reorganization and Segmentation System.* 843-847. 10.1007/978-3-642-37015-1_73, 2012.

Travis, J., & Kring, J. *LabVIEW for everyone: Graphical programming made easy and fun*, 2007.

Larsen, R. W. *LabVIEW for engineers*. Upper Saddle River, NJ: Prentice Hall/Pearson, 2011.
O. Wasenmüller and D. Stricker, *Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision*, Computer Vision – ACCV 2016 Workshops Lecture Notes in Computer Science, pp. 34–45, 2017.
Xiong Hanchen, Szedmark Sandor and Piater Justus, *Scalable, accurate image annotation with joint SVMs and output kernels*, Neurocomputing, 269 205-214. https://doi.org/10.1016/j.neucom.2014.11.096, 2011.
Rahman Mansib, *Beginning Microsoft Kinect for Windows SDK 2.0: Motion and Depth Sensing for Natural User Interface*, Apress, Canada, 2017.
C. Sielużycki, J. Maśliński, P. Kaczmarczyk, R. Kubacki, W. B. Cieśliński, and K. Witkowski, *Can Kinect aid motor learning in sportsmen? A study for three standing techniques in judo*, Plos One, vol. 14, no. 2, 2019.
J. Salomon, *Support Vector Machines for Phoneme Classification*, 2001.
S. Y. Kung, *Kernel Methods and Machine learning*. Cambrige University Press, United Kingdom 2014
N.Cristianini, T.J.Shawe, *An introduction to Support Vector Machine and other kernel-based learning methods*. Cambridge University Press, 2000. (N. and T.J. 2000)

Biographies:

Ernest Caesar Omar Syarif was born in Jakarta, Indonesia on January 23, 1997. Ernest Caesar Omar Syarif have volunteered in D'BASE Shell Eco-Marathon Team from 2017 to 2019 and was an intern at PT. Halia Teknologi Nusantara which the company specializes in the LabVIEW and integration system consultant. Currently, he is a student in BINUS ASO School of Engineering, majoring in Automotive and Robotics Engineering.

Zener Sukra Lie was born in Jakarta, Indonesia on March,14, 1984. Zener Sukra Lie completed his undergraduate degree in Electrical Engineering from University of Indonesia, Jakarta, Indonesia 2006. In the same year, He joined the graduate program in Electrical Engineering with majoring microelectronics at Bandung Institute of Technology, Bandung, Indonesia. In 2012, He got his doctorate degree on Nuclear Power and Energy Safety Engineering Department from University of Fukui, Fukui, Japan. Currently He is a lecturer in the department of Automotive and Robotics, BINUS ASO School of Engineering (BASE), Bina Nusantara University and the same time he works as a voluntary researcher for M3RC. His research and lecture revolve around Laser-Induced Breakdown Spectroscopy, and Intelligent System.

Winda Astuti received the B.S. Degree in Electrical, Control and Computer engineering from National Institute of Technology, Bandung, Indonesia, in 2000 and the M.S. degree in Electrical and Computer Engineering from International Islamic University Malaysia, Malaysia, in 2008. She is finishing the Ph.D. degree in Mechatronics engineering at International Islamic University Malaysia, Malaysia. Previously, she was a lecturer at Electrical Department at University of Jenderal Soedirman (2007-2009). Currently, she is a full lecturer in the Department of Automotive and Robotics, Binus Aso School of Engineering (BASE), Bina Nusantara University. Her main research interests include control systems theory and applications, intelligent systems, signal and image processing.