Volume 9 Issue 1 February 2016 ISSN 2088-7051

Jurnal Ilmu Komputer dan Informasi

Journal of Computer Science and Information

Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/1 (2016), 1-8
DOI: http://dx.doi.org/10.21609/jiki.v9i1.287

PERFORMANCE COMPARISON BETWEEN SUPPORT VECTOR REGRESSION AND ARTIFICIAL NEURAL NETWORK FOR PREDICTION OF OIL PALM PRODUCTION

Mustakim[1], Agus Buono[2], and Irman Hermadi[2]

[1]Information System Department, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau
Jl. H.R. Soebrantas No. 155 Km. 18, Pekanbaru, 28293, Indonesia
[2]Computer Science Department, Faculty of Mathematic and Natural Science, Institut Pertanian Bogor
Jl. Meranti Wing 20 Level 5-6, Bogor, 16680, Indonesia

E-mail: mustakim@uin-suska.ac.id

Abstract

The largest region that produces oil palm in Indonesia has an important role in improving the welfare an economy of the society. Oil palm production has increased significantly in Riau Province in every period. To determine the production development for the next few years, we proposed a prediction of the production results. The dataset were taken to be the time series data of the last 8 years (2005-2013) with the function and benefits of oil palm as the parameters. The study was undertaken by comparing the performance of Support Vector Regression (SVR) method and Artificial Neural Network (ANN). From the experiment, SVR resulted the better model compared to the ANN. This is shown by the correlation coefficient of 95% and 6% for MSE in the kernel Radial Basis Function (RBF), whereas ANN resulted only 74% for R^2 and 9% for MSE on the 8^{th} experiment with hidden neuron 20 and learning rate 0,1. SVR model generated predictions for next 3 years which rose 3%-6% from the actual data and RBF model predictions.

Keywords: *Artificial Neural Network (ANN), Palm Oil, Prediction, Radial Basis Function (RBF), Support Vector Regression (SVR)*

Abstrak

Daerah penghasil kelapa sawit terbesar di Indonesia mempunyai peranan penting dalam peningkatan kesejahteraan dan ekonomi masyarakat. Produksi kelapa sawit mengalami peningkatan yang signifikan di Provinsi Riau dalam setiap kurun waktu, untuk menentukan perkembangan produksi beberapa tahun ke depan, kami mengusulkan suatu prediksi dari hasil produksi. Dataset yang diambil adalah data time series dari data yang diperoleh selama 8 tahun terakhir (2005-2013) dengan fungsi dan manfaat kelapa sawit sebagai parameter. Dalam implementasinya peramalan dilakukan dengan membadingkan kinerja metode *Support Vector Regression* (SVR) dan *Artificial Neural Network* (ANN). Dari percobaan, SVR menghasilkan model terbaik dibandingkan dengan ANN yaitu ditunjukkan dengan koefisien korelasi sebesar 95% dan MSE 6% pada kernel *Radial Basis Function* (RBF), sedangkan ANN hanya menghasilkan R^2 sebesar 74% dan MSE 9% pada percobaan ke-8 dengan *hidden neuron* 20 dan *learning rate* 0,1. SVR model menghasilkan prediksi untuk 3 tahun kedepan yang memiliki kenaikan antara 3%-6% dari data aktual dan prediksi model RBF.

Kata Kunci:*Artificial Neural Network (ANN), Kelapa Sawit, Prediksi, Radial Basis Function (RBF), Support Vector Regression(SVR)*

1. Introduction

Riau is a province in the central of Sumatra, Indonesia that has 8.91 million hectares of area. Riau consists of 12 districts and 142 sub-districts. In 2013, Riau was recorded as a province that has the largest area of oil palm in Indonesia, with 2.26 million hectares. The average production of oil palm in Riau is 6.93 million tons per year spread in 10 sub-districts [1]. The production of oil palm in Riau is increasing every year for both its production and its plantation area. Information that was released by Riau Central Bureau of Statistics showed that there was a decreasing value of certain area. It was due to the change and replanting oil palm that has reached the limit of its age production.

The amount of oil palm production in Riau illustrates its benefits toward the prosperity level of a region [2]. In addition, oil palm also contributed to the sustainability of three main different industries. First of all, the production of Crude Oil palm (CPO) [3]. Secondly, it affects downstream industries derived from waste oil [3]. Lastly and the most important for Riau, it is used as a raw materials for

the development of renewable energy with the composition of waste that has been prescribed for each part such as shells, fibers, and oil palm's empty bunch, to over-come the electricity crisis [4-7].

A broad view and production of oil palm was also used as a decision making reference for Steam Power Plant development in Riau with simulation of extraction calculation 50% oil palm waste [8-9]. On the other hand, oil palm that has spread throughout Riau at this time have become a phenomenon among investors in terms of both production and waste. The local government is also seeking a way to develop energy using the raw material of oil palm as an alternative of fossil energy. This is in accordance with the mandate of the law No. 30/2007 concerning about chapter 20 verse 4. It is stated that is the provision and utilization of new and renewable energy should be enhanced by the central and local governments appropriate with their authority. One of the renew-able energy which is mentioned in the law is biomass which is made from oil palm [10]. The problem is the condition of oil palms in Riau in a long term, whether the result of production can always provide the raw material supply of alternative energy or vice versa. It must be dependent from local government policy.

Several studies had discussed topic related to forecasting of oil palm production both in term of production statistics or based on past data. In 2009, Hermantoro was predict oil palm [11]. By comparing determiner parameters, he concluded that oil palm production will increase. The study was conducted by using a machine learning technique called Artificial Neural Network (ANN). However, he did not mention the accuracy of the prediction result. Mustakim [12] also studied another predictiton of oil palm using a different method called Support Vector Regression (SVR). This is done by using time series data Riau from 2005 to 2013. The Research concluded the best model accuracy of SVR is 95% and 6% for error in the kernel of Radial Basis Function (RBF).

Therefore, this study will discuss the performance comparison between the best model of SVR and best model of ANN to predict the oil palm production in Riau by utilizing last 8 years data (2005-2013). SVR is used to overcome several data over-fitting from the data set. The expectation of this study is to provide conclusions related to the best model in predicting the production of oil palm for the coming years.

2. Methods

This research was conducted with multiple steps including data collection, data selection, SVR modelling, ANN modelling, and analysis of performance comparison between SVR and ANN. Several literatures that compare SVR and ANN often conclude that SVR is better than ANN. This research will also prove some statement best algorithm SVR modelling than ANN modelling. For more details, methodology can be seen in Figure 1.

Data Collection

The data that were used in this research are the production and productivity of oil palm. The data ware originated from Central Bureau of Statistics and Department of Estate Crops in Riau 2013. The data consists of 32 data points, and were recorded from 2005 to 2013. The data was filtered into 74 subdistricts based on Production Minimum Standard (PMS). SVR was able to overcome some of the data over-fitting in a data set. According to Christodoulos's research the minimum data required for prediction is 16 up to 20 data points [13].

Data Selection

Data selection was done by performing pre-processing all of the data, several companies, and department determined that the PMS which was used as a target should be 1.000 ton/period or an average minimum production of 1.000 ton/ year. There are only 74 from 142 sub-districts that fulfil this Production Minimum Standard (PMS). After establishing and obtaining the data points that will be used to make a prediction the next step is to divide the data into two parts: training dataset and testing dataset. The division was based on k-fold cross validation by randomly dividing the data into k subsets and all the data were used for both testing data and training data [14]. All of the data will also be normalized. To obtain the same weight from all data attributes and to obtain less variation. In other words, there are no attributes which more dominant or considered as more important than others from the result of its weighting [15].

Support Vector Regression (SVR)

SVR is the application of Support Vector Machine (SVM) for the case of regression. In the case of regression, output is in real or continuous numbers. SVR is a method that can solve over-fitting. Therefore, it will produce a good performance [16] and provide conclusions about the superiority and accuracy results [17].

It could also be applied to various cases with continuous data [18]. In 2003, Smola and Scholkopf explained about SVR by giving example of a condition which there is λ training dataset (x_j, y_j) with $j = 1,2, \dots, \lambda$ with input. $x = \{x_1, x_2, x_3\} \subseteq RN$ and output concerned $y = \{y_i, \dots, y_\lambda\} \subseteq RN$. By using SVR, a function of $f(x)$ will be found.

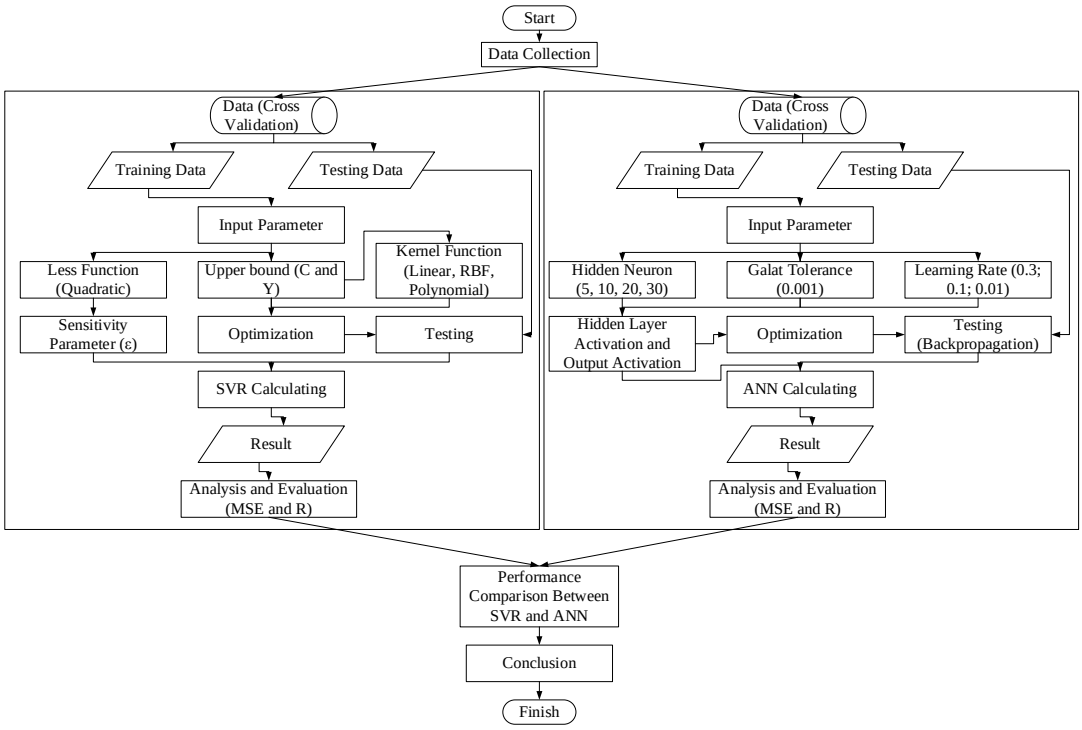


Figure 1. Research methodology

The function has the biggest deviation ε from the actual target for all training data. Then by using SVR, when the value of ε is equal to 0, perfect regression will be obtained. Based on the data, the SVR wanted to find a regression function of $f(x)$ that can approximate output to an actual target, with error tolerance of ε, and minimal complexity. Regression function of $f(x)$ can be stated by the following formula [19]:

$$f(x) = w^T \varphi(x) + b \quad (1)$$

Where $\varphi(x)$ indicates a point within a higher dimension feature space and the result of mapping the input of vector x in a lower dimension feature space. Coefficients w and b are estimated by minimizing the risk function that is defined in the equation(2) and (3):

$$min \frac{1}{2}\|w\|^2 + C\frac{1}{\lambda}\sum_{i=1}^{\lambda} L_e(y_i, f(x_i)) \quad (2)$$

$$\begin{matrix} y_i - w\varphi(x_i) - b \leq \varepsilon \\ w\varphi(x_i) - y_i + b \leq \varepsilon, i = 1,2,\dots,\lambda \end{matrix} \quad (3)$$

where,

$$L_\varepsilon\big(y_i, f(x_i)\big) = \begin{cases} |y_i - f(x_i)| - \varepsilon |y_i - f(x_i)| \geq 0 \\ 0, other \end{cases} \quad (4)$$

There are three kernel functions on SVR models. They are Linear, Polynomial and Radial Basis Function (RBF). These 3 kernel functions are in LIBSVM [20]:

Linear Kernel

$$k(x, y) = x^T y + C \quad (5)$$

Polynomial Kernel

$$k(x, y) = (\alpha x^T y + C)^d \quad (6)$$

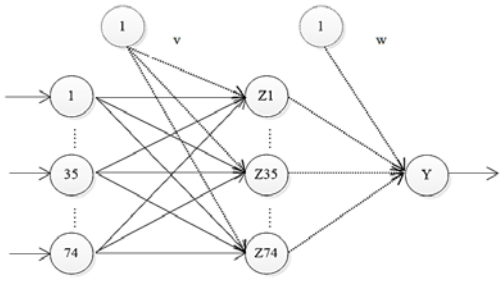
Radial Basis Function (RBF) Kernel

$$k(x, y) = exp(-\gamma \|x - y\|^2) \quad (7)$$

Artificial Neural Network (ANN)

ANN is a network of small processing unit group that is modelled based on human neural tissue. The ANN has an adaptive system that can change its structure to solve problems based on external or internal information that flows through the network [21]. In its development, ANN architecture is divided into two parts; Single Layer Network and Multiple Layer Network [22]. Models of Multiple Layer Network's category such as backpropagation [23].

Backpropagation trains a network to get a balance between the network's ability to recognize patterns that are used during training as well as network's ability to give the correct response toward input pattern which are similar (but not equal) with



Input Neuron [Production Data 2005-2013] | Hidden Neuron [5, 10 and 15] | Output Neuron [Prediction of Production]

Figure 2. ANN architecture

the pattern that are used during training [24]. Backpropagation network has 3 phases: advance phase, reverse phase, and weight modification phase to decrease error that might occur [25].

Backpropagation architecture consists of input neuron/layer, hidden neuron/layer and output neuron/layer. Each layer consists of one or more artificial neuron. The network architecture that is used in this research can be seen in Figure 2.

Comparative Analysis SVR and KNN

This analysis was done by comparing the best results between models of SVM and ANN that were calculated based on the error size and terminated coefficient. If y_i is the claimed predicttion value for the i-data and $\hat{y}_i$ is the actual output value of the i-data and m is amount of data, then the error size that is often used is Mean Squared Error (MSE).

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2 \qquad (8)$$

3. Results and Analysis

SVR Experiment

SVR requires appropriate kernel parameters to conduct the training. To obtain the optimum kernel, optimization was done by using grid search while training. There are two parameters that are optimized using grid search. They are parameter C and parameter γ. Polynomial γ parameter is part of α. Parameter C is the penalty value toward error model of SVR, whereas parameters γ was used as an input to kernel functions that will be used. RBF kernel and polynomial require parameter C and γ, whereas linear kernel only required parameter C [26]. To search for the optimum value from parameter C and γ, a combination of training and testing process experiment for RBF was conducted

TABLE 1
MSE AND R^2 ON THE KERNEL RESPECTIVELY

Kernel	MSE	R^2
Linear	0,10	0,92
Radial Basis Function	0,06	0,95
Polynomial	0.18	0,62

220 times. 55 combination experiment were for the linear kernel and 220 experiments combine polynomial with various value of parameter C and γ, so that an optimal model was produced. Other than parameter C and γ, testing was done by applying parameter New-SVR with a value of 4. The performance kernel function model can be known through the correlation coefficient (R) value and the value of MSE. The best model is a model with the largest value of R (approaching 1) and the smallest value of MSE (close to 0). R and MSE is a simple method that is often used and have been verified in measuring errors.

Simulations which had been performed to find the best accuracy on RBF kernel. The polynomial with a C combination is between 2^{-6} and 2^5 and a γ combination is between 2^{-1} up to 2^4. Likewise, for the linear kernel the C combination is between 2^{-6} and 2^5. This kind of combination was also conducted by Hendra Gunawan [27] to find the best accuracy in the case of rice production predictiton in 2012 that resulted the accuracy above 95%.

Some phases and steps that were done at linear kernel were optimized in parameter C. In accordance with previous studies. Linear kernel is the simplest one compared to other kernels. Experiment combination that ranges from 2^{-6} up to 2^5 produced minimum MSE of 0,053308 or 5% with a maximum R^2 value of 0,921253 or 92%. RBF kernel will optimize the value of γ that ranges from 2^{-1} up to 2^4. Parameter C at the same range to linear kernel can obtain the smallest error value of 1.4%, on fold 2. The largest determination coefficient is obtained on fold 1 with 95%. Similar to RBF, polynomial kernel optimize value of γ and C at the same range on RBF.

The best experiment in polynomial with error value of 18% and determination coefficient of 62% is in fold 1. The parameter γ and C range between 2^{-1} and 2^0. The value of error and the determination coefficient from those three kernels can be seen on Table 1. Based on experiment from the three kernels, the relationship between observation and prediction can be seen and are shown in Figure 3, 4 and 5.

On polynomials, experiments that were conducted illustrate the inverse curve between actual and predicted it taken based on experiments with the smallest error without considering other aspects. In the linear kernel, making the conclusion of an

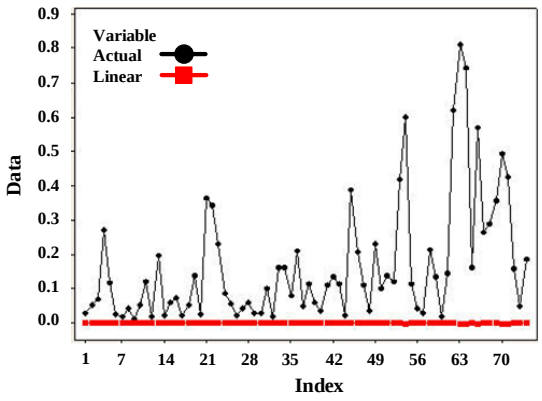


Figure 3. Comparison of linear kernel prediction result with observation on the production of oil palm in normal form

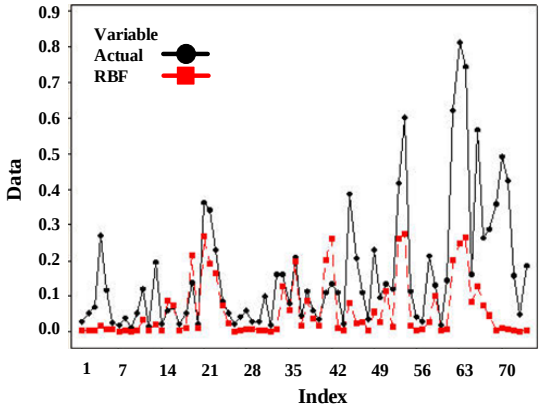


Figure 4. Comparison of RBF kernel prediction result with observation on oil palm production in normal form

experiment series was also based on the value of the smallest error. From the comparison of linear and polynomial kernels, the linear is more optimum. The results caused some experienced overfitting data.

From the experiment, prediction models that showed the highest correlation level and the lowest value of error was the one that was done by using RBF kernel. This is appropriate with SVM guide that states RBF kernel is more superior in many cases of machine learning [28].

Experiment of ANN

ANN experiment was done by using the same data based on 32 data points and Hidden Neuron comparison with a learning rate of 4 cross validation. Moreover, the experiments were comprised of 12 ANN models. Table II and III showed the characteristics and specifications that were used for ANN architecture and the best experiments result, respectively.

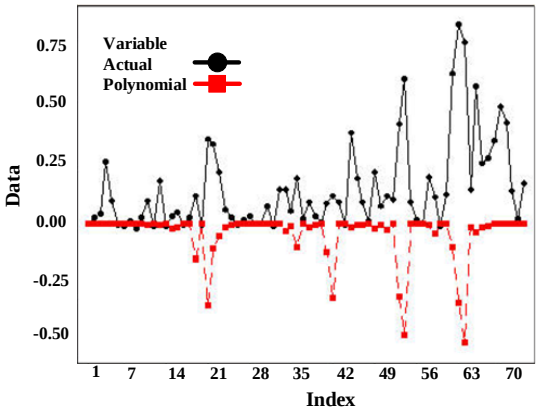


Figure 5. Comparison of polynomial kernel prediction result with observation on oil palm production in normal form

TABLE 2
CHARACTERISTIC AND SPECIFICATION USED

Characteristic	Specification
Architecture	1 hidden layer
Hidden Neuron	2, 10, 20 and 30
Neuron Output	1 (Prediction Production of Palm Oil)
Activation Hidden Layer	Sigmoid Binary
Activation Output	Linear
Galat Consideration	0,001
Learning Rate	0,3; 0,1 and 0,01
Maximum Epoch	1000

TABLE 3
THE BEST EXPERIMENT RESULT OF ANN MODEL

Try	Hidden Neuron	Learning Rate	R^2	MSE
1	2	0,3	0,57	0,11
2	2	0,1	0,43	0,08
3	2	0,01	0,66	0,10
4	10	0,3	0,49	0,14
5	10	0,1	0,62	0,15
6	10	0,01	0,51	0,31
7	20	0,3	0,53	0,12
8	20	0,1	0,74	0,09
9	20	0,01	0,44	0,13
10	30	0,3	0,63	0,17
11	30	0,1	0,50	0,22
12	30	0,01	0,69	0,19

From Table 3, it can be seen that the experiment that used ANN had the best model. It was known from the 8th experiment that the ANN has a determination coefficient value of 74% and error value of 9%. Likewise, for the second experiment, it had the lowest error value between among other experiments with 8%. However the second experiment only has a determination coefficient of 43%. Therefore, from the result of best R^2 and best MSE, it can be concluded that the 8th experiment with hidden neuron 20 and learning rate 0.1 was the best model of a series model which was produced to predict the relationship between observed data and

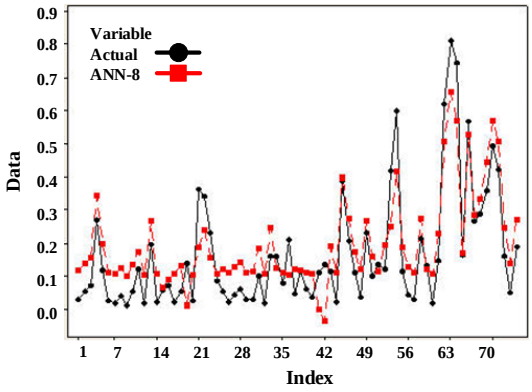


Figure 6. Time graph series comparison between actual data and result of best ANN model prediction

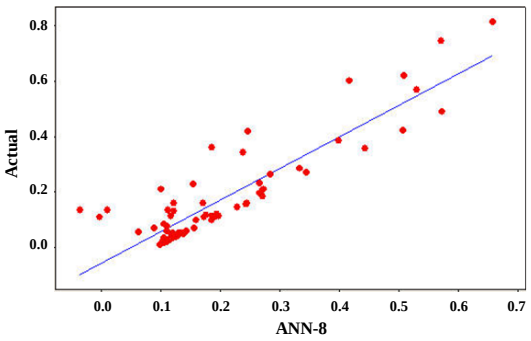


Figure 7. Regression graph that is produced by the ANN model for actual data and prediction

best model of ANN. This can be seen at time series Figure 6 plot and scatter plot Figure 7.

Performance Comparison between SVR and ANN

From the experiment, the method that produced the best model for oil palm production is the SVR model. The model has a determination coefficient of 95% and error value of 6%. From the percent-age, it can be seen that the two methods produced very much differences on value R^2.

Prediction of Best Model

From the best SVR model the prediction result that was obtained for three years ahead can be described based on estimated actual data prediction and oil palm production prediction in the next year.

Figure 9 shows that the average increase for each recording period is 3%-6% in normalization form. It will have the same pattern for year 2017, if the pattern data used are still the same as the actual data and the prediction results. Nature is not a factor that will be used as references in this study. The

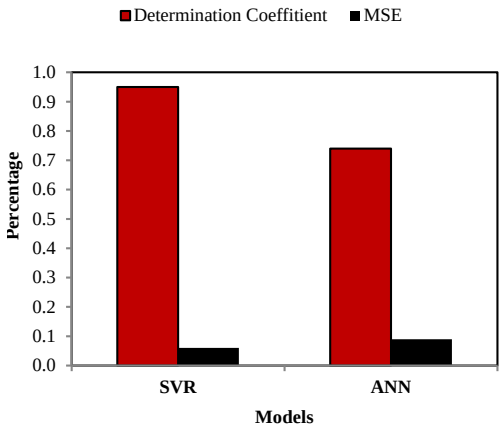


Figure 8. The best model comparison between SVR and ANN that is shown in size of correlation coefficient and MSE

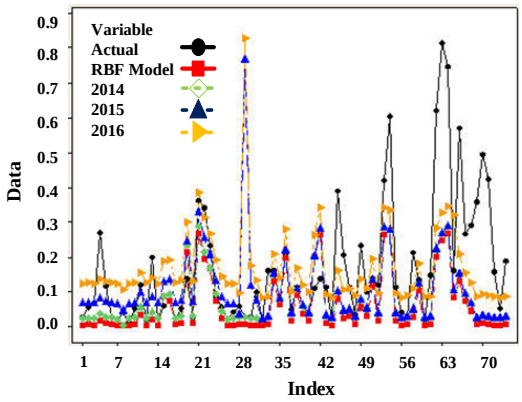


Figure 9. Comparison graph between actual data and RBF model and oil palm production prediction for 3 years ahead

factors that will be used as references in this study is only based on the final data.

4. Conclusion

From the conducted research, it can be concluded that SVR model is better than ANN model for oil palm production prediction's case in Riau. ANN Model got the best value of determination coefficient ($\boldsymbol{R^2}$) 74% with galat error 9% on the 8th experiment, while SVR on the RBF kernel produced a smaller error i.e. 6% and also a bigger $\boldsymbol{R^2}$ i.e. 95%. A very huge difference of determination coefficient value proved that by using time series data, SVR model is more superior compared to ANN model. Prediction results for next three years gradually in normal form as many as 3%-6%. Prediction results do not reckon the nature or other factors in the field that could affect production in each period.

References

[1] Riau Central Bureau of Statistics (BPS). "Riau in Figures of 2006-2013 *(Riau dalam Angka 2006-2013)* and District in Figures of 2006-2013 *(Kabupaten Dalam Angka 2006-2013)"*, 2013.

[2] Elinur,"Analysis of Energy Consumption and Supply in Indonesian Economy (*Analisis Konsumsi dan Penyediaan Energi dalam Perekonomian Indonesia*)", Master Thesis, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia, 2011.

[3] A.R. Pudyantoro,"Effects of Fiscal Policy and The Economy Gas Upstream Sector in Riau Province (*Dampak Kebijakan Fiskal dan Sektor Hulu Migas Terhadap Perekonomian Provinsi Riau*)", Master Thesis, Faculty of Agricultural Technology, Institut Pertanian Bogor, Indonesia, 2012.

[4] E. Mahajoeno, "Development of Renewable Energy from Waste Liquid Palm Oil Mill (*Pengembangan Energi Terbarukan Dari Limbah Cair Pabrik Minyak Kelapa Sawit*)". Master Thesis, Faculty of Agri-cultural Technology, Institut Pertanian Bogor, Indonesia, 2008.

[5] Saepudin, "Renewable Energy (Biogas) from Palm Oil Waste. Research Center for Electric Power and Mechatronics (*Energi Terbarukan (Biogas) dari Limbah Kelapa Sawit. Pusat Penelitian Tenaga Listrik dan Mekatronik*)". *Indonesia Institution of Science*, 2010.

[6] D. Partogi, M.N. Amin, and S.T. Kasim, "Analysis Electricity Production Cost Per KWh Using Liquid Waste Biogas Fuels Oil Palm (PKS PLTBGS Tandun Application) *(Analisis Biaya Produksi Listrik Per KWh Menggunakan Bahan Bakar Biogas Limbah Cair Kelapa Sawit (Aplikasi pada PLTBGS PKS Tandun)*". *Singuda Ensikom*. 3 (1): 17-22, 2013.

[7] M.S. Nur, "Characteristics of Palm Oil As Raw Materials Bioenergy (*Karakteristik Kelapa Sawit Sebagai Bahan Baku Bioener-gi*)", San Design, Bogor, 2014.

[8] Mustakim, "Support Vector Regression and Multi-Attribute Decision Making for Productivity Prediction and Regional Renewable Energy Development Ranking *(Support Vector Regression dan Multi-Attribute Decision Making untuk Prediksi Produktifitas dan Perankingan Wilayah Pengembangan Energi Terbarukan)"*. Master Thesis, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia, 2015.

[9] B. Sunarwan, and R. Juhana, "Utilization of Waste Oil for Fuel and Renewable Energy (*Pemanfaatan Limbah Sawit untuk Bahan Bakar Energi Baru dan Terbarukan*)", *Jurnal Tekno Insentif Kopwil* 4. 7(2): 1-14, 2013.

[10] D. Kusdiana, "Real conditions of Energy Requirements and Sources of Renewable Alternative Energy in Indonesia (*Kondisi Rill Kebutuhan Energi di Indonesia dan Sumber-sumber Energi Alternatif Terbarukan*)". General of Electricity and Energy Utilization Department of Energy and Mineral Resources, 2008.

[11] R.W.P. Hermantoro, "Palm Oil Production Based on Prediction of Soil Quality Using Artificial Neural Model Network (ANN) (*Prediksi Produksi Kelapa Sawit Berdasarkan Kualitas Lahan Menggunakan Model Artificial Neural Network (ANN)*)". *Jurnal Agroteknose*. 4(2), 2009.

[12] C. Christodoulos, "Forecasting with limited data: Combining ARIMA and diffusion models", *Technological Forecasting & Social Change*. 77 (2010) 558–565, 2010.

[13] M. A. Agmalaro, "Statistical Downscaling GCM Data Modeling Using Support Vector Regression to Predict Monthly Rainfall in Indramayu (*Pemodelan Statistical Down-scaling Data GCM Menggunakan Support Vector Regression untuk Memprediksi Curah Hujan Bulanan Indramayu*)", Master Thesis, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia, 2011.

[14] V.R. Patel, and R.G. Mehta, "Impact of Outlier Removal and Normalization Approach in Modified K-Means Clustering Algorithm". IJCSI International Journal of Computer Science Issues. 8(5), 2011.

[15] E. Piantari, "Hyperspectral data Feature Selection for Rice Productivity Prediction with Genetic Algorithm Support Vector Regression (*Feature Selection Data Hiperspektral Untuk Prediksi Produktivitas Padi dengan Algoritme Genetika Support Vector Regression*)", Essay, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia 2011.

[16] N. Ibrahim, and A. Wibowo, "Support Vector Regression with Missing Data Treatment Based Variables Selection for Water Level Prediction of Galas River in Kelantan Malaysia". *Wseas Transactions on Mathe-matics*. 13(1), 2014 E-ISSN: 2224-2880, 2014.

[17] J. F. De Paz, B. Pérez, A. González, E. Corchado, and J. M. Corchado, "A support vector regression approach to predict carbon dioxide exchange," in Distributed Computing and Artificial Intelligence. Springer Science + Business Media, 2010, pp. 157–164.

[18] Smola, B. Schölkopf, "A Tutorial on Support Vector Regression: NeuroCOLT, Technical

Report NC-TR-98-030", *Royal Holloway College, University of London*, UK, 2003.

[19] Santosa, "Data Mining Techniques Use of Data for Business Purposes (*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis)"*, Graha Ilmu, Yogyakarta, 2007.

[20] V. Sharma, "A comprehensive Study of Artificial Neural Network", International Journal of Advanced Research in Computer Science and Software Enginering India, 2012.

[21] J.K. Siang, "Artificial Neural Network and Its Programming using MATLAB (*Jaringan Saraf Tiruan dan Pemrogramannya Menggunakan MATLAB*)", ANDI, Yogya-karta, 2004.

[22] S. Kusumadewi, and S. Hartati, "A Neuro-Fuzzy Integration of Fuzzy system and Neural Network (*NEURO-FUZZY Integrasi Sistem Fuzzy dan Jaringan Saraf*)", Graha Il-mu, Yogyakarta, 2006.

[23] D.O. Maru'ao, 2010, "Neural Network Implementation in Foreign Exchange Kurs Prediction", International Seminar Information Technology, 2010.

[24] A.A. Adebiyi, A. Charles, A. Marion, S. Otokiti Sunday, "Stock Price Prediction using Neural Network with Hybridized Market Indicators". *Journal of Emerging Trends in Computing and Information Sciences*. 3(1): 1-9, 2012.

[25] G. Adhani, A. Buono, A. Faqih, "Support Vector Regression modelling for rainfall prediction in dry season based on Southern Oscillation Index and NINO3.4", *International Conference on Advanced Computer Science and Information Systems*. 2013: 315-320, 2013.

[26] H. Gunawan, "Hyperspectral Band Selecti-on Using the Recursive Feature Elimination for Rice Production Prediction using Support Vector Regression (*Seleksi Hyperspectral Band Menggunakan Recursive Feature Elimination untuk Prediksi Produksi Padi dengan Support Vector Regression*)", Essay, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia, 2012.

[27] Hsu in R. Hidayat, "Toddler Nutrition Status System Prediction Using Support Vector Regression (*Sistem Prediksi Status Gizi Balita dengan Menggunakan Support Vector Regression*)", Essay, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia, 2013.

Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/1 (2016), 9-16
DOI: http://dx.doi.org/10.21609/jiki.v9i1.367

WALL-FOLLOWING BEHAVIOR-BASED MOBILE ROBOT USING PARTICLE SWARM FUZZY CONTROLLER

Andi Adriansyah[1], and Shamsudin H. Mohd. Amin[2]

[1]Departement of Electrical Engineering, Faculty of Engineering, Universitas Mercu Buana
Jl. Meruya Selatan, Jakarta Barat, 11650, Indonesia
[2]Center of Artificial Intelligent and Robotics (CAIRO), Universiti Teknologi Malaysia,
Skudai, Johor Bahru, 81310, Malaysia

E-mail: andi@mercubuana.ac.id[1], sham@fke.utm.my[2]

Abstract

Behavior-based control architecture has been broadly recognized due to their compentence in mobile robot development. Fuzzy logic system characteristics are appropriate to address the behavior design problems. Nevertheless, there are problems encountered when setting fuzzy variables manually. Consequently, most of the efforts in the field, produce certain works for the study of fuzzy systems with added learning abilities. This paper presents the improvement of fuzzy behavior-based control architecture using Particle Swarm Optimization (PSO). A wall-following behaviors used on Particle Swarm Fuzzy Controller (PSFC) are developed using the modified PSO with two stages of the PSFC process. Several simulations have been accomplished to analyze the algorithm. The promising performance have proved that the proposed control architecture for mobile robot has better capability to accomplish useful task in real office-like environment.

Keywords: *behavior-based robot; wall-following behavior; fuzzy logic; PSO; PSFC;*

Abstrak

Arsitektur pengendali robot berbasis perilaku telah secara efektif menunjukkan kompetensinya dalam pengembangan teknologi robot bergerak. Karakteristik sistem logika fuzzy adalah salah satu solusi yang dapat diandalkan untuk menyelesaikan beberapa problem pada perancangan perilaku robot. Akan tetapi, terdapat kesulitan untuk dapat menala parameter fuzzy secara manual. Oleh karena itu beberapa studi dilakukan untuk mengintroduksi kemampauan pembelajaran pada sistem logika fuzzy. Tulisan ini membahas pengembangan arsitektur pengendali robot berbasis perilaku dengan memanfaatkan Particle Swarm Optimization (PSO). Perilaku robot mengikuti dinding berbasiskan Particle Swarm Fuzzy Controller (PSFC) dibangun menggunakan PSO yang telah dimodifikasi dengan dua tahap proses PSFC. Beberapa pengujian telah dilakukan untuk menganalisa performansi algoritma tersebut. Hasil pengujian menunjukkan bahwa perancangan tersebut memiliki performansi yang menjanjikan bahwa robot dapat menyelesaikan tugasnya dengan baik pada suatu lingkungan tertentu.

Kata Kunci: *robot berbasis perilaku; perilaku pengikut dinding; logika fuzzy; PSO; PSFC;*

1. Introduction

Emerging a mobile robot is a remarkable task. Usually, the mobile robot should face unpredictable environment, perceive inaccurate sensor and act with unsatisfactory actuator in high speed response [1,2]. Behavior-based control architecture is an alternative method suitable to address these problems [3-7]. The architecture is able to act with fast real-time response, provides for higher level deliberation and has confirmed its reliable results in standard robotic activities. However, a kind of soft computing is needed to complete two key problems in behavior-based systems, such as generating optimal individual behavior and coordinating multiple behaviors.

Currently, several methods that hybrid fuzzy system with evolutionary algorithms has been offered in behavior-based mobile robot, such as Genetic Algorithm (GA) [8,9], Genetic Programming [10] to overcome the behavior-based issues. However, the current evolutionary algorithms used have several drawbacks [11], such as not easy to be implemented and computationally expensive [12], require process should be completed and parameters should be adjusted, have slow convergence ability to find near optimum solution, and dependent heuristically to genetic operators [13].

Fortunately, Kennedy and Eberhart presented the Particle Swarm Optimization in 1995 [14, 15]. PSO is one of evolutionary computation technique to find the best solution by acting like social behavior of groups such as fish schooling or bird flocking. There are several benefits of the PSO as compared to other evolutionary computation methods. The PSO is not difficult to be implemented and is computationally reasonable since its memory and CPU speed requirements are low. Additionally, the PSO requires only a few process should be completed and parameters to be adjusted. On the other side, the PSO has quick convergence ab-ility to find optimum or near optimum solution. Generally, PSO has demonstrated to be an effective method for numerous wide ranging optimization problems. Moreover, in some cases it does not suffer from the problems encountered by other evolutionary computation [11-13].

This paper addressed the problems of developing control architecture of mobile robot with behavior-based system, especially in wall-following. The problem solving is related to the specification of mobile robot tasks, the development of mobile robot behaviors, the interpretation of the environment, and the validation of the final system. This paper uses and develops soft computing, making extensive use of Fuzzy Logic and Particle Swarm Optimization (PSO) named as Particle Swarm Fuzzy Controller (PSFC). The use of PSO is to tune fuzzy membership function and to learn fuzzy rule base for wall-seeking behavior. This fuzzy tuning and learning is performed to accomplish the best behavior-based system.

2. Methods

Wall-following behavior steer the robot to follow wall in order to help goal accomplishment. Based on some distances measured between the mobile robot and the walls, the mobile robot would maintain some fixed distance between both robot and the wall even at edges [16-18]. In this work, MagellanPro mobile robot is used for verification and performance analysis of the proposed algorithm. The MagellanPro is a rounded mobile robot from iRobot, Real World Interface (RWI), the recognized industry leader in the exciting field of pioneering mobile robotic. The dimension of the robot is as follow: D= 40.64 cm, H= 25.4 cm, r=5.7 cm, W= 36 cm and M= 18.2 kg, where D is diameter, H is height, r is the radius of wheels, W is distance between two wheels, and M is mass.

Figure 1 illustrated a model of MagellanPro mobile robot for simulation exercises for the proposed algorithm. The mobile robot is positioned on a two dimensional Cartesian workplace, in wh-ich a global coordinate $\{X, O, Y\}$ is defined. The robot

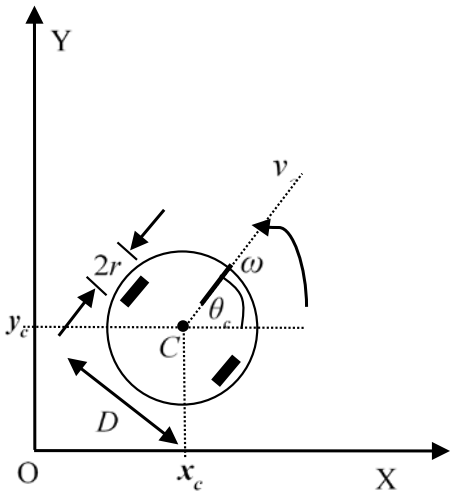


Figure 1. Model of MagellanPro mobile robot

has three degrees of position that are represented by a posture $p_c = (x_c, y_c, \theta_c)$, where (x_c, y_c) indicate the spatial position of the robot guide point in the coordinate system and θ_c is the heading angle of the robot counter-clockwise from the x-axis.

The mathematical model for the robot movement can be obtained with differentially steered drive system or known as differential drive system [19]. Based on this system, the robot can move to different positions and orientations as a function of time. The derivatives of *x*, *y* and θ can be obtained as equation(1) to equation(3).

$$\frac{dx}{dt} = v_c \cos\theta_c \tag{1}$$

$$\frac{dy}{dt} = v_c \sin\theta_c \tag{2}$$

$$\frac{d\theta}{dt} = \omega_c \tag{3}$$

where ω_c is the angular velocity of the robot and where v_c is the linear velocity of the robot.

By applying the current position of the robot, $p_c = (x_c, y_c, \theta_c)$, the next position of the mobile robot is shown in equation(4) to equation(6).

$$x_{c+1} = x_c + v_c \cos\theta_c * \Delta t \tag{4}$$

$$y_{c+1} = y_c + v_c \sin\theta_c * \Delta t \tag{5}$$

$$\theta_{c+1} = \theta_c + \omega_c * \Delta t \tag{6}$$

Then, as assuming the value of Δt is a unit time step, the next position of the robot which also can be written as $p_{c+1} = (x_{c+1}, y_{c+1}, \theta_{c+1})$, in a simple form is given by the equation(7) to equation(9).

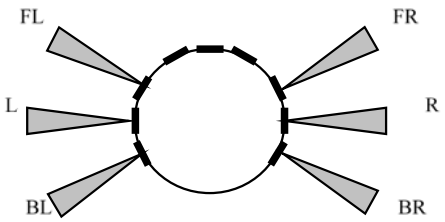


Figure 2. Sonar Configuration

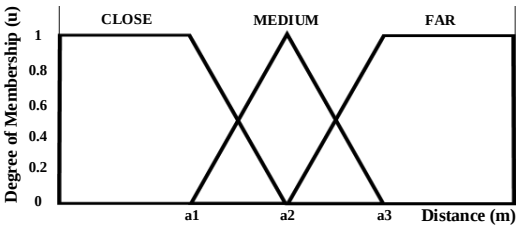


Figure 3. The membership functions of distances and angle

$$x_{c+1} = x_c + v_c \cos\theta_c \tag{7}$$

$$y_{c+1} = y_c + v_c \sin\theta_c \tag{8}$$

$$\theta_{c+1} = \theta_c + \omega_c \tag{9}$$

In order to make the robot able to follow the wall then six sonar have been mounted on them. These sensors would measure the distance between positions of sonar in mobile robot and the object accordingly. The positions of sonars are illustrated in Figure 2.

There are two zones of sensor, namely: Left Zone Sensors and Right Zone Sensors. Left Zone Sensors comprise Back Left (BL), Left (L), and Front Left (FL). Meanwhile, Right Zone Sensors consist of Front Right (FR), Right (R) and Back Right (BR). The angle between sonars is 22.5 degree. Each zone of sonar's is according to a behavior. The behavior constitutes: Left Zone Sensors with left wall following behavior and Right Zone Sensors with right wall following behavior. Distances obtained by sonar's process in each zone are used as the input on behalf of a behavior.

FLC structure based on Mamdani technique is used in this system. Each left wall following behavior and right following behavior have three inputs. The inputs are front left distance (FL), left distance (L) and back left distance (BL) used for left wall following behavior, and front right distance (FR), right distance (R) and back right distance (BR) used for right wall following behavior. All the distances are obtained by sonar sensor.

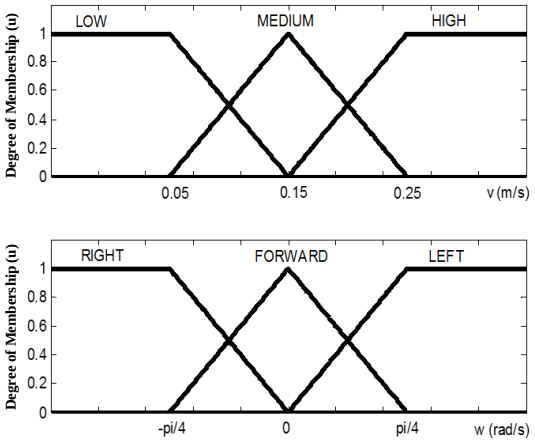


Figure 4. The membership function of linear velocity and angular velocity

The fuzzy input has three linguistic terms, which are CLOSE, MEDIUM and FAR for distances and RIGHT, FORWARD and LEFT for angle, as depicted generally in Figure 3. Three linguistic terms is chosen on behalf of the minimal number for fuzzy system.

In this work, linear velocity v and angular velocity ω are applied as outputs of all fuzzy behavior modules. The linguistic terms used are LOW, MEDIUM and HIGH for linear velocity, and, RIGHT, FORWARD, and LEFT for angular velocity. The fixed membership functions of v and ω is shown in Figure 4.

Basically, Particle Swarm Fuzzy Controller is an FLC improved by a tuning or learning process based on PSO. In PSFC, PSO is applied to explore for an optimized Knowledge Base (KB) of a fuzzy system for a specific problem and to ensure those parameter values are suitable with respect to the design principles. The KB parameters establish the optimization space, which is then transformed into suitable position on which the explore process operates. Figure 5 shows the concept of a PSFC scheme where PSO design and fuzzy processing are the two fundamental parts.

At the beginning of the process, the initial populations comprise a set of particles that are scattered all over the search space. The initial population may be randomly generated or may be partly supplied by the user. However, in this works, the populations are randomized initially.

Afterward, one particle is taken and decoded to the actual value of the wall-following fuzzy parameter. These sets of fuzzy controller parameters are then used to control the fuzzy behavior where it undergoes a series of tracking response of multi-step reference set point.

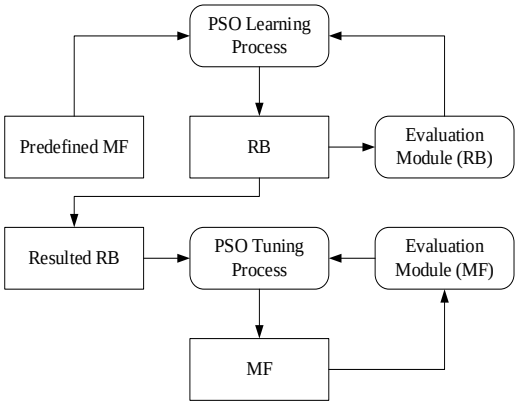


Figure 5. The concept of a PSFC

Based on the various state of the control system, the performance of the system is calculated by using a predefined fitness function.

In the first stage, PSO starts to learn fuzzy rule base with predefined fuzzy membership function. In the next stage, PSO continues to tune fuzzy membership functions based on the results from the fuzzy rule base. By means of these two stages, ideal fuzzy parameter could be reached without human intervention. PSO is then used to tune the fuzzy controller parameters to minimize the fitness function. The assignment of the fitness function serves as a guidance to lead the search toward the optimal solution.

In the beginning, to learn fuzzy rule base, each rule is prearranged into integer codes that are created on number in linguistic terms of output membership function. Consequently, there are '1', '2', and '3' for LOW, MEDIUM, and HIGH for linear velocity, and RIGHT, FORWARD, and LEFT for angular velocity, respectively. The coded parameters for each behavior are arranged to form particles of the population.

Furthermore, fuzzy membership functions are tuning with equation(10) and equation(11).

$$C_{x+1} = C_x + k_i \tag{10}$$

$$D_{x+1} = D_x + j_i \tag{11}$$

where k_i and j_i are adjustment coefficients, C_x, and D_x are set of center and width of each fuzzy membership function, respectively. The adjustment coefficients take any real positive or negative value. Therefore, k_i makes each center of membership function shift to the right or left and the membership functions shrinks or expands through j_i, as shown in Figure 6. The shifting coding strategy will simplify searching computation, because there is no necessity to sort the value of membership functions in ascending manner.

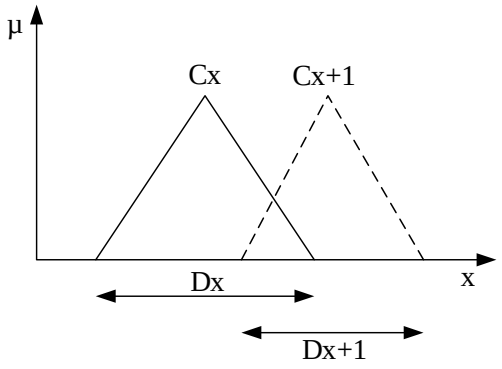


Figure 6. Principle in tuning of membership function

The PSO procedure starts with randomly prearranged initial populations. Then, all populations of particles are assessed based on fitness function to decide the *pbest* and *gbest*. Based on several initial investigations and trials and errors, the fitness functions for wall-following can be obtained as equation(12).

$$f_{wall} = \sum_{i=0}^{I} \sum_{k=0}^{K} (100 e_d^2(k) + 0.1/v(k)) \tag{12}$$

where I is the entire number of start position, K is the number of step simulation for each start position, e_d is the distance error and $v(k)$ are the *linear velocity* at k, respectively.

In this work, a Sigmoid Decreasing Inertia Weight (SDIW) is used to provide faster speed of convergence and better accuracy of optimized value [20]. Consequently, PSFC would generate optimal and reliable wall-following behavior of the mobile robot.

3. Results and Analysis

Some experiments have been performed. Some steps of experiments have been designed. Firstly, a PSFC optimization processes is conducted to find the optimized value of fuzzy parameters. Then, simulations of the mobile robot based on the PSFC are analyzed to investigate the wall-following control behavior. Results of fuzzy behavior that are obtained manually, obtained by GA, called as Genetic Fuzzy Controller (GFC) from previous works are used as comparison [21].

PSO and GA processes for wall-following behavior are shown in Figure 7, where, evolutions of the best fitness value against generation are illustrated. At the beginning, the process tended to have more global search ability because of large inertia weight. It was shown that the fitness value over all generations is converging quickly. After that, the process tended to have more local search ability

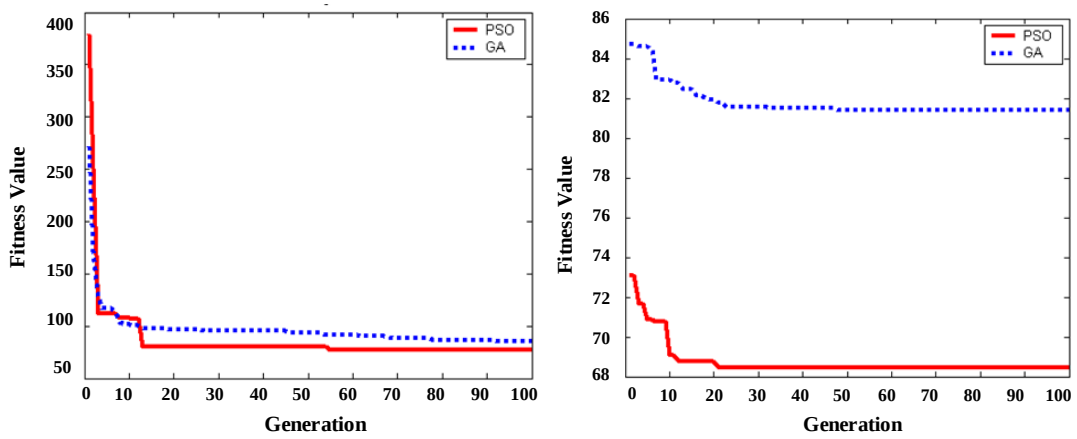


Figure 7. Comparison of PSO vs. GA process (a) Rule base learning, (b) Membership function tuning

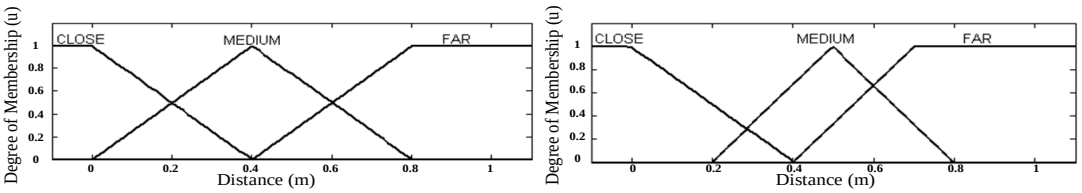


Figure 8. Membership function of left wall following behavior before and after optimization

TABLE 1
COMPARISON OF FITNESS VALUES

Process	Fitness Value
FLC (manually)	254.1924
GA (Stage 1)	84.9731
GA (Stage 2)	81.4125
PSO (Stage 1)	76.705
PSO (Stage 2)	68.4668

caused by the small value of inertial weight. This phenomenon was illustrated by the value of fitness function updated towards the lowest one. Moreover, the figures showed that the kind of searching in RB learning process is wider than in MF tuning process. It showed that learning the rule base of FLC is very complicated than tuning the membership. It was also evidenced that the best optimized fitness value could be obtained using two stages of PSO process.

Furthermore, a comparison between Particle Swarm Optimization and Genetic Algorithm was investigated. Table I listed the fitness value based on the process to show a comparison between learning or tuning fuzzy parameters manually, GA process and PSO process. It is noted that PSO and GA provided better results than FLC, but PSO had higher convergence speed and obtained better optimized value than GA. Figure 8 depicted membership function change for wall-following behavior generated by PSO process.

The target of left wall following or right wall following behavior is that the mobile robot could maintain the distance between the robot and the wall while detecting the wall in its left side or right side, respectively. Left wall following is use-d here for experiments showed. As described bef-ore, the wall was represented as a sequence of poi-nts and assumed as a thin wall. In this work, the maintained distance is 0.4 m. Hence, a left edge wall was plan-ed and placed with 3 m long vertically and 2 m long horizontally in a room. The mobile robot movements from different starting position were demonstrated and the time responses, which are left distance, linear velocity and angular velocity, were presented.

Figure 9 demonstrated the mobile robot movements from position (6.4, 2.5, π/2), the initial distance from the wall is 0.2 m and with t = 20 s. As previous experiments, the fuzzy behavior was obtained using FLC, GFC and PSFC. Furthermore Figure 11 (a) presented time response for each algorithm, accordingly. Figure 10 and Figure 11 (b) demonstrated that the three algorithms provided different control performance. Fuzzy behavior obtained by FLC had the worst performance. The response time to maintain the distance was very slow. Furthermore, the algorithm was also very sensitive to disturbance. It was shown that angular velocities fluctuated and the linear velocity reduced when the mobile robot runs in the left edge situation. However, the fuzzy behavior obtained by PSFC had better performance as compared to GFC. For both al-

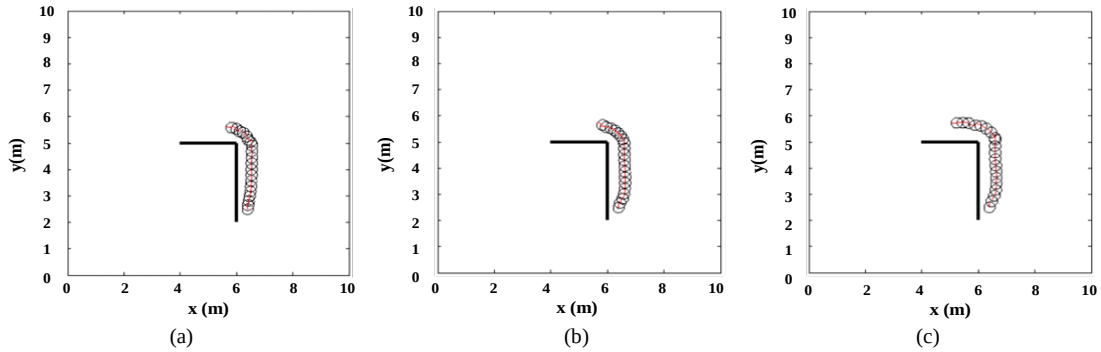


Figure 9. Mobile Robot movements for left wall following behavior starting from (6.4, 2.5, π/2) (a) FLC, (b) GFC, and (c) PSFC

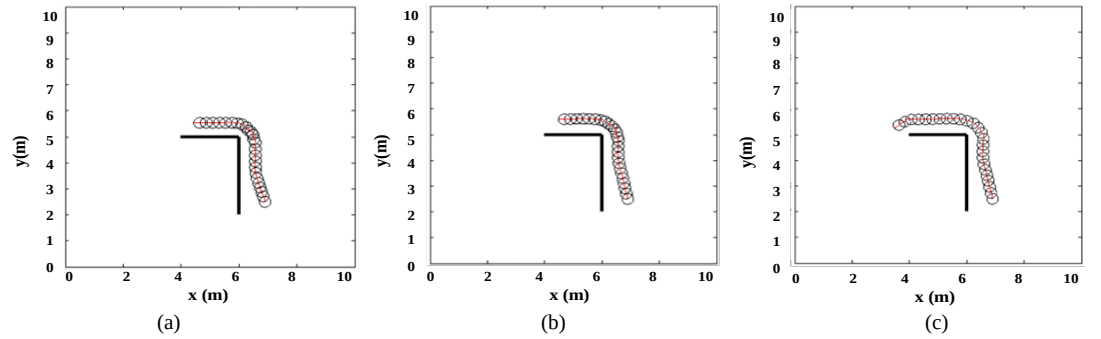


Figure 10. Mobile Robot movements for left wall following behavior starting from (6.9, 2.5, π/2) (a) FLC, (b) GFC, and (c) PSFC

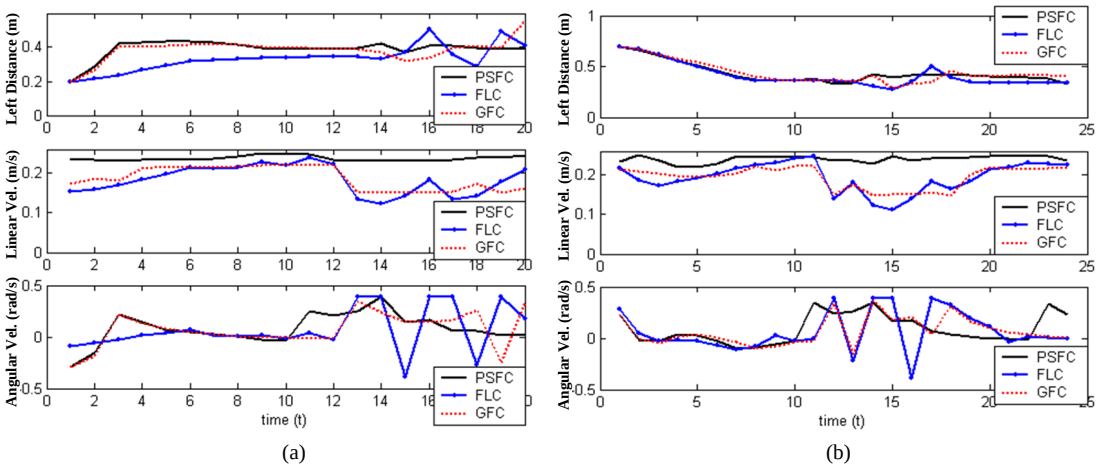


Figure 11. Time response of left wall following using FLC, GFC, and PSFC, starting from (a) (6.4, 2.5, π/2) and (b) (6.9, 2.5, π/2)

gorithms in spite of providing the same response of distance, the PSFC was less sensitive to disturbance than the GFC when the mobile robot runs in the left edge situation. It is noted that the PSFC can perform the left edge situation with higher linear velocity and with relatively less change in angular velocity.

Different starting positions and different wall structure were presented to further investigate the performance of the left wall following fuzzy behaviors. Figure 10 showed the mobile robot movements from position (6.9, 2.5, π/2), the initial distance from the wall is 0.7 m and t = 25 s with the same shape of wall. The time response for each algorithm, accordingly, was shown in Figure 11 (b). Moreover, the mobile robot movements form position (6.4, 2.5, π/2), the initial distance from the wall is 0.2 m and t = 25 s with unstructured wall was depicted in Figure 12. Figure 13 showed the time response for each algorithm, accordingly. From Figures 12 and 13, it is noted that PSFC had better performance than other fuzzy behaviors, which are FLC and GFC. The fuzzy behavior obtained by PSFC had the fastest response time, able to main-

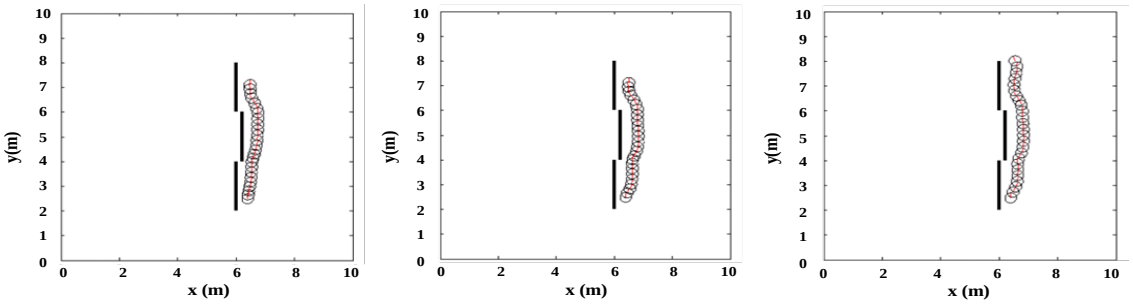


Figure 12. Mobile Robot movements for left wall following behavior with unstructured wall starting from (6.4, 2.5, π/2) (a) FLC, (b) GFC, and (c) PSFC

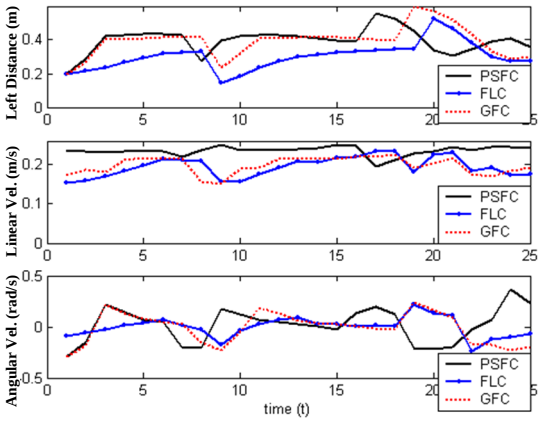


Figure 13. Time response of left wall following with unstructured wall starting from (6.4, 2.5, π/2) using FLC, GFC, and PSFC

tain the distance successfully with higher speed and less changed in the direction of the mobile robot, especially when the robot faced unstructured wall.

4. Conclusion

Wall-following behavior-based control architecture has successfully demonstrated their competence in mobile robot development. Fuzzy Logic Systems appear to be very useful to develop the high reliable and effective behavior-based system. However, there are difficulties to tune Fuzzy Sys-tem manually. This paper presents the development of fuzzy wall-following behavior-based control architecture using PSO for MagellanPro mobile robot. The work has been done in some tasks: behavioral designing of the mobile robot.

Based on the experiment results, the mobile robot is able to deal with wall-following behaviors. Generally, it is noted that the proposed control architecture has the good ability to be applied in MagellanPro mobile.

Acknowledgement

The author wishes to express his sincere recognitions to the Directorate General of Higher Education, Ministry of Research, Technology and Higher Education which awarded the grant funding for Fundamental Research (Project No. 2013 0263/E5/ 2014).

References

[1] Widodo, N. S, Rahman, "A. Vision Based Self Localization for Humanoid Robot Soccer", *TELKOMNIKA*, Vol. 10(4), pp. 637-644, 2010.

[2] Wicaksono, H, Khoswanto, H., and Kuswadi, S. "Teleautonomous Control on Rescue Robot Prototype", *TELKOMNIKA*, Vol.10 (4), pp. 621-628, 2012.

[3] Dongshu, W., Yusheng, Z., and Wenjie, S. "Behavior-Based Hierarchical Fuzzy Control for Mobile Robot Navigation in Dynamic Environment", *Chinese Control and Decission Control* (CCDC 2011), China, pp. 2419-2424, 2011.

[4] Parasuraman, S., Ganapathy, V., and Shirinzadeh, B. "Behaviour Based Mobile Robot Navigation Technique AI System: Experimental Investigation on Active Media Pio-neer Robot", *IIUM Engineering Journal*, Vol. 6 (2), pp. 13-25, 2005.

[5] Bao, Q.Y., Li, S.M., Shang, W.Y., and An, M.J. "A Fuzzy Behavior-based Architecture for Mobile Robot Navigation in Unknown Environments", *International Conference on Artificial Intelligence and Computational Intelligence* (AICI 2009), Shanghai, pp. 257-261, 2009.

[6] Mo, H., Tang, Q., and Meng, L. "Behavior-based Fuzzy Control for Mobile Robot Navigation", *Mathematical Problems in Engineering*, pp. 1-10, 2010.

[7] Khatoon, S, Ibraheem. "Autonomous Mobiler Robot Navigation by Combining Local and Global Techniques", *International Journal of Computer Applications*. Vol. 37(3), pp. 1-10, 2012.

[8] Sim, K. B., Byun, K.S., and Lee, D. W. "Design of Fuzzy Controller using Schema Coevolutionary Algorithm". *IEEE Transaction of Fuzzy System*, Vol. 12(4), pp. 565-570, 2004.

[9] Merchan-Cruz, E. A., Moris, A. S. "Fuzzy-GA-based trajectory planner for robot manipulators sharing a common workspace", *IEEE Transaction On Robotics*, Vol. 22(4), pp. 613-624, 2006.

[10] Tunstel, E.W, de Oliveira, M.A.A., and Berman, S. "Fuzzy Behavior Hierarchies for Multi Robot Control". *International Journal of Intelligent Systems*. Special Issue: Hierarchical Fuzzy Systems. Vol. 17(5), pp. 449-470, 2002.

[11] Saxena, A. and Saxena, A. "Review of Soft Computing Techniques used in Robotics Application", *International Journal of Information and Computation Technology*. Vol. 3(3), pp. 101-106, 2013.

[12] Hassan, R., Cohanim, B, and de Weck, O. "A Comparison of Particle Swarm Optimization and the Genetic Algorithm", *American Institute of Aeronautics and Astronautics*, pp. 1-13, 2004.

[13] Jones, K.O. "Comparison of Genetic Algorithm and Particle Swarm Optimization", *Proceeding of International Conference on Computer Systems and Technologies*. Bulgaria. Vol. IIIA, pp. 1-6, 2005.

[14] Eberhart, R.C. and Kennedy, J. "A new optimizer using particle swarm theory", *Proceeding of the Sixth International Symposium on Micro Machine and Human Science*. Nagoya, pp. 39-43, 1995.

[15] Kennedy, J. and Eberhart, R.C. "Particle Swarm Optimization", *Proceeding of IEEE International Conference on Neural Networks*. IV. Perth, pp. 1942-1948, 1995.

[16] Hsiao, C. W., Chien, Y. H., Wang,W. Y., "Wall Following and Continously Stair Climbing Systems for A Tracke Robot", *International Conference on Networking, Sensing and Control*, pp. 371-375, 2015.

[17] Dash, T., Nyak, T. and Swin, R., "Controlling Wall Following Robot Navigation Based on Gravitational Search and Feed Forward", *Proceeding of the 2nd International Conference on Perception and Machine Intelligence*, pp. 196-200, 2015.

[18] Juang, C. F. Chen, Y. H. and Jhan, Y. H., "Wall-following Control of Hexapod Robot using a Data-Driven Fuzzy Conroller Learned through Differentiial Evolution", *IEEE Transactions on Industrial Electronics*, Vol. 62 (1), pp. 611-619, 2015.

[19] Dudek, G. and Jenkin, M. Computational Principles of Mobile Robotics. 1st ed. Cambridge, MA: Cambridge University Press. 2000.

[20] Adriansyah, A., Amin, S HM. "Analytical and Empirical Study of Particle Swarm Optimization with A Sigmoid Decreasing Inertia Weight", *Regional Postgraduate Conference on Engineering and Science*. Johor Bahru, pp. 247-252, 2006.

[21] Adriansyah, A. and Amin, S.H.M. "Knowledge Base Tuning using Genetic Algorithm for Fuzzy Behavior-based Autonomous Mobile Robot", *Proceeding of 9th International Conference on Mechatronics Technology*. Kuala Lumpur, pp. 20-125, 2005.

Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/1 (2016), 17-25
DOI: http://dx.doi.org/10.21609/jiki.v9i1.362

WEB NEWS DOCUMENTS CLUSTERING IN INDONESIAN LANGUAGE USING SINGULAR VALUE DECOMPOSITION-PRINCIPAL COMPONENT ANALYSIS AND ANT ALGORITHMS

Arif Fadllullah[1,2], Dasrit Debora Kamudi[1,3], Muhamad Nasir[1,4], Agus Zainal Arifin[1], and Diana Purwitasari[1]

[1]Department of Informatics Engineering, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember, Kampus ITS Sukolilo, Surabaya, 60111, Indonesia
[2]Universitas Borneo Tarakan, Jl. Amal Lama 1, Tarakan, 77115, Indonesia
[3]Politeknik Negeri Nusa Utara, Jl. Kesehatan 1 Tahuna, Sangihe, 95812, Indonesia
[4]Politeknik Negeri Bengkalis, Jl. Bathin Alam-Sungai Alam, Bengkalis, 28711, Indonesia

E-mail: arif14@mhs.if.its.ac.id

Abstract

Ant-based document clustering is a cluster method of measuring text documents similarity based on the shortest path between nodes (trial phase) and determines the optimal clusters of sequence document similarity (dividing phase). The processing time of trial phase Ant algorithms to make document vectors is very long because of high dimensional Document-Term Matrix (DTM). In this paper, we proposed a document clustering method for optimizing dimension reduction using Singular Value Decomposition-Principal Component Analysis (SVDPCA) and Ant algorithms. SVDPCA reduces size of the DTM dimensions by converting freq-term of conventional DTM to score-pc of Document-PC Matrix (DPCM). Ant algorithms creates documents clustering using the vector space model based on the dimension reduction result of DPCM. The experimental results on 506 news documents in Indonesian language demonstrated that the proposed method worked well to optimize dimension reduction up to 99.7%. We could speed up execution time efficiently of the trial phase and maintain the best F-measure achieved from experiments was 0.88 (88%).

Keywords: *web news documents clustering, principal component analysis, singular value decomposition, dimension reduction, ant algorithms*

Abstrak

Klasterisasi dokumen berbasis algoritma semut merupakan metode klaster yang mengukur kemiripan dokumen teks berdasarkan pencarian rute terpendek antar node (*trial phase*) dan menentukan sejumlah klaster yang optimal dari urutan kemiripan dokumen (*dividing phase*). Waktu proses *trial phase* algoritma semut dalam mengolah vektor dokumen tergolong lama sebagai akibat tingginya dimensi, karena adanya masalah *sparseness* pada matriks *Document-Term Matrix* (DTM). Oleh karena itu, penelitian ini mengusulkan sebuah metode klasterisasi dokumen yang mengoptimalkan reduksi dimensi menggunakan Singular Value Decomposition-Principal Component Analysis (SVDPCA) dan Algoritma Semut. SVDPCA mereduksi ukuran dimensi DTM dengan mengkonversi bentuk *freq-term* DTM konvensional ke dalam bentuk *score-pc Document-PC Matrix (*DPCM). Kemudian, Algoritma Semut melakukan klasterisasi dokumen menggunakan *vector space model* yang dibangun berdasarkan DPCM hasil reduksi dimensi. Hasil uji coba dari 506 dokumen berita berbahasa Indonesia membuk-tikan bahwa metode yang diusulkan bekerja dengan baik untuk mengoptimalkan reduksi dimensi hingga 99,7%, sehingga secara efisien mampu mempercepat waktu eksekusi *trial phase* algoritma se-mut namun tetap mempertahankan akurasi F-*measure* mencapai 0,88 (88%).

Kata Kunci: *klasterisasi dokumen web berita, principal component analysis, singular value decomposition, reduksi dimensi, algoritma semut*

1. Introduction

The emergence of internet has brought great changes in terms of the information presentation. Online sites become popular publications, for reliability in offering a variety of information quickly, hot, and plentiful with a wider range of readers. The fact that the number of articles is very large and disorganized would cause difficulty to the readers to find news relevant to the topic of the interest. For that, we need an automatic clustering for news web documents.

The problem of clustering in large datasets is a combinatorial optimization problem that is difficult to be solved with conventional techniques. Hierarchical and partition based clustering are the common clustering categories. 'K' means algorithm is the very popular partition based clustering algorithm. However, there are few limitations observed with this 'K' means algorithm from literature. The limitations are (a) It fails to scale with large datasets. (b) The quality of the algorithm results highly depends on the initial number of static clusters [1].

Ant algorithm is a universal solution which is first designed to overcome Traveling Salesman Problem (TSP). This was proposed to find the shortest path, which the ants will choose the town (node) to be traversed by the probability function between the distance of nodes that must be taken and how much the level of trace pheromones. Ant algorithms is applied on clustering of text documents, by analogying documents as graph nodes. Idea of finding shortest path of those document nodes has triggered a clustering method which is known as Ant-based document clustering method [2-3]. This method consist of two phases, i.e. finding documents most alike (trial phase) and clusters making (dividing phase). The advantages of this algorithm are the ability to classify documents in unsupervised learning and flexibility to determine the initial number of clusters [3].

Before entering into a stage of Ant algorithms, VSM is required. VSM would build the relationship between the number of documents and terms of all documents in DTM (Document-Term Matrix). DTM is a structured data matrix obtained by forming a number of terms as column and documents as row. If all the terms entered into DTM, then the size of the document vectors will become larger. It will take a long time and increase the complexity of the trial phase calculation in Ant algorithms. For that reason, we need a strategy to optimize execution time of the trial phase Ant algorithms. One common approach is dimension reduction of the stop word and stemming specific to a particular language.

Stop word is used to remove the terms that they appear that often occur in for the entire document, but cannot be a significant feature of each document. While stemming is used to change affixes-word into basic-word, so that a number of affixes-word which has the same basic-word can be grouped into a single term. Most of the stemmer frame-work which has been standardized for English language documents, while the attention of the Indonesian domain is still relatively low, and continues to undergo development. Unlike English, Indonesian language has more complex category of affixes, which includes prefixes, suffixes, infixes (insertions), and confixes (combinations of prefixes and suffixes).

Several Indonesian language stemmer had been developed by Nazief and Adriani (1996), Vega, et al.(2001), and Arifin and Setiono (2002) [4]. In addition, there is also confix stripping (CS) stemmer made by Jelita Asian (2007) [5], and enhanced confix stripping (ECS) stemmer was an improvement of CS stemmer by Arifin, Mahendra and Ciptaningtyas (2009) [6].

Stemmer ECS produced document vectors with reduced terms size up to 32.66% of the 253 news documents in Indonesian language. It could create document clusters in dividing phase with F-measure of 0.86 [6]. Only the complexity of trial phase in document vectors processing is still comparatively high, although the term has been reduced by ECS stemmer. This happens because not only many documents and terms are processed, but also the sparse terms problem in DTM. Sparseness occurs because not all terms are in all documents, so that most conditions of the DTM term value is zero.

In this paper, we proposed a document clustering method for optimizing dimension reduction using Singular Value Decomposition Principal Component Analysis (SVDPCA) and Ant algorithms. SVDPCA will reduce the size of the DTM dimensions by converting freq-term of conventional DTM to score-pc of Document-PC Matrix (DPCM) without sparseness based on cumulative proportion of variance. Then, Ant algorithms will create document clustering using the vector space model based on the dimension reduction result of DPCM. The proposed method is expected to overcome the dimension reduction problem because a lot of affixes word and sparseness in DTM. Therefore it would speed up the execution time of the trial phase Ant algorithms and improve F-measure news documents clustering in Indonesian language.

Enhanced Confix Stripping (ECS) Stemmer

ECS stemmer is a stemmer framework as the improvement of the previous confix stripping stemmer. It does stemming in the news documents in Indonesia language. In this method affixes-word can be recorded into basic words, so that a number of affixes-word which has the same basic-word can be grouped into a single term [6]. The steps in the process of ECS stemmer recording exactly the same as the previous stemmer [5]. However, this stemmer corrects some examples of words that failures stemming by stripping stemmer confix previous methods as follows:

1. No prefix removal rule for words with construction of "mem+p...", for example, "mem-

TABLE 1
DTM FROM THE COLLECTION OF NEWS DOCUMENTS IN INDONESIAN LANGUAGE

Document	Perang	tembak	…	emas	banjir	Bencana
1	2	3	…	0	0	0
2	0	0	…	4	0	0
…	…	…	…	…	…	…
505	0	3	…	0	0	1
506	0	0	…	0	1	2

promosikan", "memproteksi", and "memprediksi".

2. No prefix removal rule for words with construction of "men+s...", for example, "mensyaratkan", and "mensyukuri".
3. No prefix removal rule for words with construction of "menge+...", for example, "mengerem".
4. No prefix removal rule for words with construction of "penge+...", for example, "pengeboman".
5. No prefix removal rule for words with construction of "peng+k...", for example, "pengkajian".
6. Suffix removal failures – sometimes the last fragment of a word resembles certain suffix. For examples, the words like "pelanggan" and "pelaku" failed to be stemmed, because of the "-an" and "-ku" on the last part of the word should not be removed.

Weighting *Document-Term Matrix* (DTM)

DTM is a mathematical matrix that describes the frequency of term co-occurrence in documents. In DTM, a number of terms is defined as column and documents as row (see Table 1).

TF-IDF (combined with a term frequency-inverse document) method has shown better performance when compared with the binary method and frequency. In TF-IDF, w_{ij} represents weight of term i on document j which is expressed in equation(1) [7]:

$$w_{ij} = tf_{ij}.\log_2\left(\frac{N}{df_i}\right) \quad (1)$$

where tf_{ij} represents frequency of term i on document j. N represents number of processed documents and df_j represents number of documents that actually have term i at least one.

Singular Value Decomposition-Principal Component Analysis (SVDPCA)

Sparseness occurs because not all terms are in all documents, so that most conditions of DTM term value is zero (see Table 1). SVDPCA is a mechanism of DTM-dimensional transformation which sparse terms into a dimension of Document PC Matrix (DPCM) without sparse (not redundant and correlated) by utilizing the calculation of Principal Component Analysis (PCA), where eigenvalues and eigenvectors is not searched by covariance matrix, but utilizing the left eigen-vectors results of Singular Value Decomposition (SVD). Dimensional DPCM then reduced by selecting the number of score-pc when their cumulative proportion of variance is more than 50% (see Table 2). [8]

TABLE 2
DPCM WITH THREE PCS

Document	PC1	PC2	PC3
1	-7.05324	-1.24343	-3.675594
2	2.65687	4.434534	1.98538
…			
505	-6.2431	-1.23234	2.322449
506	-2.4321321	-3.129341	1.321393

PCA is an orthogonal linear transformation that maps given data onto the PC system, which consists of first PC, second PC, and so on, according to the order of their variance values [9]. PCA estimates the transformation matrix S as the equation(2).

$$X(N \times m) = S^T A(N \times p) \quad (2)$$

DTM A is transformed to DPCM X by S, which generates column dimensions m is smaller than p when using only conventional DTM. To implement classic PCA into A with dimensions $N \times p$, N has to be larger than p. If N is smaller than p, then we cannot achieve convergence calculations for PCA. Therefore, we add SVD (Singular Value Decomposition) to overcome this problem. In the SVD transformation, the original matrix can be decomposed (similar to PCA eigen-decomposition) into three matrix components with the same size. If they was multiplied again, so they will produce the same value as the original matrix [10]. The decomposition of A is shown in equation(3):

$$A = P.D.Q^T \quad (3)$$

P and Q are $N \times N$ and $p \times p$ orthogonal matrices, respectively. D is a $(N \times p)$ matrix with singular value d_{ij} shown in equation(4):

$$d_{ij} = \begin{cases} d_{ij} \geq 0, & i = j \; and \; i < r \\ 0, & i \neq j \; or \; i > r \end{cases} \quad (4)$$

The rank of A is r, the columns of P and Q are the left and right eigenvectors of XX^T and X^TX, respectively. The singular values on the dia-

gonal D is the square root of the eigenvalues of AA^T and A^TA. The classic PCA depends on the decomposition of covariance (or correlation) matrix S by PCA eigenvalues decomposition. In fact, the covariance matrix S calculation of DTM will obviously take a long time, due to the large dimensions of DTM. While SVDPCA, we do not need to carry out covariance matrix S execution, because its value can be extracted from the matrix P which is left eigenvectors of the SVD results as space data records used to find PCA scores. As a SVDPCA result, we are not restricted by the constraint that N being larger than p [8].

Ant algorithms

Ant algorithms typically used to solve the Traveling Salesman Problem (TSP) algorithm that is inspired by the behaviour of ant colonies in search for food sources (documents) by leaving a trail pheromone [11].

In the case of Euclidean TSP, d_{ij} is the euclidean distance between towns i to j, i.e. $\boldsymbol{d_{ij}} = \left[(x_i - x_j)^2 + (y_i - y_j)^2\right]^{1/2}$. Let *m* be the total number of ants. Each ant is a simple agent with the following characteristics: First, it chooses the town (node) that will be passed next, with a probability function between the distances that must be taken and how the level of existing pheromone trail. Second, it is given a memory of the towns that have been passed (by using tabu list) and the last after completing a full path, the path of the ants are given a number of pheromone [6].

Let $\tau_{ij}(t)$ be the amount of pheromone trail on edge (*i,j*) at time *t*. Pheromone trail on each edge is updated on each tour, cycle, or iteration according to the equation(5):

$$\tau_{ij}(t+n) = \rho.\tau_{ij}(t) + \Delta\tau_{ij} \quad (5)$$

where ρ is a coefficient such that (1- ρ) represents the evaporation of pheromone trail between time *t* and *t+n*. The range of coefficient ρ is (0.1). Δ is a segment that have been passed by ant *k* as part of the trajectory of the nest towards food sources. The decline in the amount of pheromone allows ants to explore different paths during the search process. It will also eliminate the possibility of choosing the bad path. In addition, it can also help limit the maximum value is achieved by a pheromone trajectory. The total amount of pheromone trail laid on edge (*i, j*) defined in equation(6).

$$\Delta\tau_{ij} = \sum_{k=1}^{m} \Delta\tau_{ij}^k \quad (6)$$

where $\Delta\tau_{ij}^k$ is the quantity per unit length of pheromone trail laid on edge (*i,j*) by the *k-th* ant between time *t* and *t+n*; it is given by the equation (7).

$$\Delta\tau_{ij}^k \begin{cases} Q/L_k, & \\ if\ k-th\ ant\ use\ edge(i,j) & \\ in\ its\ tour & \\ 0, \quad otherwise & \end{cases} \quad (7)$$

Q is a constant and L_k is the tour length of the *k-th* ant. Each ant equiped with a data structure called the tabu list $tabu_k$, that saves the towns already visited. After each complete cycle, the tabu list is used to compute the ant's current solution, to get the shortest path route and its length. The transition probability from town i to town j for the *k-th* ant is defined as the equation(8).

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum[\tau_{ik}(t)]^\alpha [\eta_{ik}]^\beta}, k \in allowed_k \quad (8)$$

Where η_{ij} is the visibility of $1/d_{ij}$, $allowed_k = \{N\text{-}tabu_k\}$, α and β are parameters that control the relative importance of pheromone trail and visibility where $\alpha \geq 0$ and $\beta \geq 0$.

Ant-Based Document Clustering

Ant-Based Document Clustering is divided into two phase, which are finding the shortest path between the documents (trial phase) and separate a group of documents alike (dividing phase) based on previous trial phase result [1-3] and [6].
The finding shortest path adopts Ant algorithms trial phase, where the ant *k* probability (p_{ij}^k) to select a node (document) *j* from a node *i* position is defined as the equation(9).

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha [S_{ij}]^\beta}{\sum[\tau_{ik}(t)]^\alpha [S_{ik}]^\beta}, k \in Z_k \quad (9)$$

Z_k represents set of nodes (documents) that is not visited yet by k-th ant, $\tau_{ij}(t)$ represents pheromone trail on edge (*i,j*), α represents parameter that control the importance of pheromone trail, and β represents visibility parameter. S_{ij} represents cosine distance between node (*i,j*) which is defined as the equation(10).

$$S_{ij} = \frac{\sum_k [w_{ki}] \cdot [w_{kj}]}{\|d_i\|^2 \cdot \|d_i\|^2} \quad (10)$$

where w_{ki} and w_{kj} are score-pc weighting of *k*-th pc on document *i* and *j* of DPCM. $\|d_i\|^2$ and $\|d_j\|^2$ are the length of document vector *i* and *j*. For exa-

mple, if $||d_i||^2 = (t_1^2 + t_2^2 + t_3^2 + ... + t_k^2)^{1/2}$, so t_k is *k*-th pc of d_i document vector.

After the search nodes by ants in one iteration is complete, the next step is to add the amount of pheromone on edge passed by each ant, and evaporated. In contrast to Lukasz Machnik [2] [3], Arifin, et al [6] research to modify the method of increasing the number of pheromone (see formula 12) on edge *i,j* ($\Delta\tau_{ij}$), which is defined by equation(11).

$$\Delta\tau_{ij} = \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \tag{11}$$

where $\Delta\tau_{ij}^{k}$ is the amount of phremone trail laid on *edge* (*i,j*) by the k-th ant, which is defined as the equation(12).

$$\Delta\tau_{ij}^{k} \begin{cases} N \times L_k, \\ if\ k-th\ ant\ use\ edge(i,j) \\ in\ its\ tour \\ 0, \quad otherwise \end{cases} \tag{12}$$

where *N* is the total documents and L_k is the total route length of k-th ant. The evaporation of pheromone trail is given by equation(13):

$$\tau_{ij}(t+n) = (1-\rho).\tau_{ij}(t) + \Delta\tau_{ij} \tag{13}$$

where ρ is the pheromone evaporation coefficient.

It should be noted that the best shortest path is route that has the largest total similarities. It is because the length of path is calculated using cosine distance similarity between documents.

In the next phase, dividing phase is the formation of clusters based on the order of the documents obtained from the previous trial phase. The first document of that order, is regarded as the centroid μ of the first cluster. The clustering process begins from μ and the next document in sequence (called a comparative document D) is defined as equation(14).

$$\delta < \cos(\mu, D) \tag{14}$$

where δ is the attachment coefficient that has value range at (0,1), and *cos(μ,D)* is cosine distance of document μ and *D*.

If the condition is true, then the document *D* into a group in centroid μ, and the next document in sequence into the next document *D*. If the condition is false, then the current document *D* becomes new centroid μ for the new cluster. The testing process is repeated to the next document until finished when the whole sequence of documents is done.

2. Methods

Documents clustering using SVDPCA and Ant Algorithms

Figure 1 shows order of the proposed method that consisting of some stages, where the first stage begins with the collection of corpus data as input is derived from web news documents in Indonesia language. Corpus data was consisted of 506 news

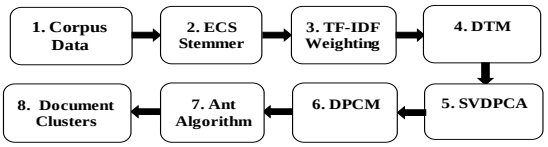


Figure 1. Design of the Proposed Method

TABLE 3
THE CORPUS CLASS WITH MANUALLY IDENTIFIED EVENTS

ID	Event	Medoid	Count file
1	Kasus sodomi Anwar Ibrahim	Tunangan Saiful Bicara Tentang Kasus Sodomi Anwar Ibrahim	52
2	Kasus suap Artalyta	Jadi "Sinterklas", Artalyta Bagi-bagi Makanan di Pengadilan	54
3	Bencana gempa bumi di China	Hampir 900 Terperangkap dan 107 Tewas di Gempa China	38
4	FPI dalam insiden Monas	Kasus Kekerasan oleh FPI	70
5	Konflik Israel-Palestina	Menteri Israel Tolak Gencatan Senjata di Gaza	50
6	Dua tahun tragedi lumpur Lapindo	Pengungsi Lapindo Tuntut Jatah Makan Dilanjutkan	42
7	Badai nargis di Myanmar	Yangon Rusak Parah, 351 Orang Tewas	48
8	Pengunduran jadwal Pemilu	Jadwal Pemilu Tidak Berubah Cuma Digeser	20
9	Perkembangan harga emas	Berkilaunya Investasi Emas	36
10	Kiprah Khofifah dalam Pilkada Jatim	Khofifah, Perempuan Lembut Setangguh Lelaki	30
11	Kematian mantan presiden Soeharto	Kematian Soeharto Tragedi Bagi Korbannya	56
12	Festival film Cannes	The Class Raih Penghargaan Tertinggi Cannes	10

from www.kompas.com, www.detik.com, www. wartakota.com and www.jawapos.com .Then, the corpus data processed by stop list and stemming using ECS stemmer. Stop word is used to remove the terms that they appear often for the entire document, but cannot be a significant feature of each document. While stemming is used to change affixes word into basic word, so that a number of affixes word which has the same basic word can be grouped into a single term. We also use stop list with 792 stop words and Indonesian language dictionary with 29337 basic-words of the proposed ECS stemmer.

After all terms of basic-word has been formed, then the term value processed by TF-IDF weighting on each document to make DTM. Stages of the fifth, sixth and seventh is proposed contribution of our research, which serves to overcome the sparseness problem when reducing the DTM dimensions, so that the executing time of cluster formation during process of finding the shortest path in the trial phase Ant algorithms can be optimized. SVD is used to find the set of orthogonal that divided into two spaces, right eigenvectors (Q) for a range of dimensions space and left eigenvectors (P) for a range of data records space. P values in the SVD is used as a data matrix S in the calculation of PCA. Then PCA transform S^T which maps original data X normalized (each term minus the average term) into new dimensions of data, by changing the TF-IDF weighting of term into a score-pc and reduce the dimensions S^TA based on significant score-pc. This selection is based on the largest proportion of variance of score-pc, so obtained a new DPCM without sparseness and dimensions have been reduced. In our research, the threshold limit of cumulative proportion of variance modifies the cumulative proportion limit of previous research [8], when it over 40%.

Furthermore, Ant algorithms divides the cluster process into two steps. The first step is to find the shortest path in the trial phase, where the results of the DPCM are converted to vector space model. Then search distances based on some rule of trial phase (see section Ant-Based Document Clustering) and the best shortest path that produces the greatest of total similarities. Due to the large number of parameters on how the real total similarities value or actual largest path, the parameters were tested regarding Marco Dorigo experiment in shortest path searching to the TSP using Ant algorithms [11]. The parameter values are: α= 2, β= 5, ρ= 0.5, ants=30, and the number of iterations or maximum rotation of 100. The second step of the dividing phase Ant algorithm that makes clusters based on the order of the documents obtained from the previous phase. The first document is regarded as the centroid μ of the first cluster. The clustering process begins from μ and the next document in sequence based on the condition δ (see formula 14) with parameter value is 0.008, until generated news documents clustering.

To verify whether documents clustering obtained by the proposed method is successful, so we require evaluation techniques. Evaluations of information retrieval are recall, precision and F-measure. Each cluster is generated by the proposed method is considered as the retrieval result. The manual-predefined class of documents as the ideal cluster that should be retrieved [4][6] with 12 manually identified news events from previous research [6] (see Table 3).

Specifically, for each predefined class *i* and cluster *j*, the recall and precision are defined as equation(15) and equation(16).

$$Rec(i,j) = R_{ij} = n_{ij}/n_i \tag{15}$$

$$Prec(i,j) = P_{ij} = n_{ij}/n_j \tag{16}$$

where n_{ij} is the number of documents of class *i* in cluster *j*, n_i is the number of documents of class *i*, and n_j is the number of documents of cluster *j*. The F-measure of class *i* on all cluster *j* is defined as the equation(17).

$$F_{ij} = (2 \times R_{ij} \times P_{ij})/(R_{ij} + P_{ij}) \tag{17}$$

Then, the overall F-measure is calculated by the following the equation(18).

$$F = \sum_i \frac{n_i}{n} \max\{F_{ij}\} \tag{18}$$

where *n* is the total number of documents, $\boldsymbol{n_j}$is the number of documents in class *i*, and $\boldsymbol{max\{F_{ij}\}}$ is the maximum value $\boldsymbol{F_{ij}}$ of class *i* on all cluster *j*. Because we will see increased performance of the proposed method, so the F-measure value and execution time of trial phase in our research will be compared with the previous method (Ant Algorithms without SVDPCA) [6].

3. Results and Analysis

Table 4 shows that ability ECS stemmer in reduce insignificant term. ECS stemmer could reduce the number of term up to 4758 terms, from a total of 7327 different terms in 506 corpus data. With the use of ECS stemmer, term had been reduced to 35%. Table 5 shows the standard deviation, proportion of variance (variance of each score-pc divided by the total variance), and the cumulative proportion of variance for each score pc results from SVDPCA. Score-pc sorted by descending,

TABLE 4
COMPARISON OF TOTAL TERM WITH AND WITHOUT ECS STEMMER

Stop list Removal	Stemmer	Total Term
No	Without *Stemmer*	7327
Yes	ECS *Stemmer*	4758

TABLE 5
VARIANCES OF THE TOP- FIFTEEN PCS

PCs	Standard Deviation	Proportion of Variance	Cumulative Proportion
PC1	19.703	0.050	0.050
PC2	19.394	0.048	0.098
PC3	18.944	0.046	0.144
PC4	15.635	0.031	0.175
PC5	15.134	0.029	0.205
PC6	14.545	0.027	0.232
PC7	13.826	0.025	0.257
PC8	13.379	0.023	0.279
PC9	13.088	0.022	0.301
PC10	12.465	0.020	0.321
PC11	11.556	0.017	0.339
PC12	11.295	0.016	0.355
PC13	10.758	0.015	0.370
PC14	10.503	0.014	0.384
PC15	10.258	0.014	**0.397**
PC16	9.671	0.012	0.409
PC…	…	…	…

so that PC1 has the highest standard deviation or variance from total of variance data.

PC2 and PC3 have the highest second and third standard deviation. And so on, thus obtained PCs limit which have a cumulative proportion of variance which reached 0.4 (40%), that was PC15. Selection score-pc could convert DTM which contains 506 documents x 4758 term become DPCM which contains 506 documents x 15 score-pc. This shows that the SVDPCA had been successfully optimize term into score-pc until 99.7%. DPCM form are converted into vector space model and then processed by Ant algorithms until obtain some cluster. In the proposed method, vector space model for each document did not use 4758 term but only 15 score-pc.

Table 6 shows the performance of the proposed method (SVDPCA and Ant algorithms) with the parameter values were set $\alpha = 2$, $\beta = 5$, $\rho = 0.5$, 30 ants, the number of iteration is 100, and $\delta = 0.008$, where the proposed method successfully created 14 clusters. It was more than two clusters from total of 12 classes defined manually. Because of the cluster label of the proposed method does not represent the same label with the class, so we need to identify clusters by comparing the members of each class to each cluster. From the comparison, a total of 6 classes (1, 3, 8, 9, 10, 12) defined manually has been identified in 6 different clusters (13, 3, 2, 11, 7, 14) of the proposed method. Even 4 from 6 clusters (13, 2, 7, 14) have exactly the same or similar all document members

TABLE 6
OVERALL F-MEASURE OF THE PROPOSED METHOD

Class	Cluster	Rec	Prec	F-measure	F-max	F-max Weighting
1	13	1	1	1	1	0.10
2	9	0.13	0.14	0.13	0.90	0.10
	10	0.81	1	0.90		
3	3	1	0.92	0.96	0.96	0.07
4	8	0.47	1	0.64	0.64	0.09
	9	0.52	0.61	0.56		
5	1	0.93	1	0.96	0.96	0.09
	9	0.04	0.04	0.04		
6	11	0.10	0.1	0.10	0.95	0.08
	12	0.91	1	0.95		
7	3	0.04	0.05	0.05	0.96	0.09
	4	0.92	1	0.96		
8	2	1	1	1	1	0.04
9	11	1	0.82	0.90	0.90	0.06
10	7	1	1	1	1	0.06
11	5	0.25	1	0.4		
	6	0.54	1	0.70	0.70	0.08
	9	0.11	0.02	0.03		
12	14	1	1	1	1	0.02
				Overall	F-	0.88

TABLE 7
COMPARISON PERFORMANCE OF METHODS

Description	SVD-PCA and Ant Algorithms	Ant algorithms (Arifin, etc, 2009)
SVDPCA	6 sec	-
Graph (Network)	5 sec	1 minutes 13 sec
The Search Shortest Path	8 minutes 56 sec	8 minutes 35 sec
Total Time of Trial Phase	9 minutes 7 sec	9 minutes 48 sec
Cluster Obtained	14 cluster	14 cluster
F-measure	0.88	0.78

with members of the class defined manually (1, 8, 10, 14), because each that clusters took the best value of recall and precision is 1. Furthermore, a total of 6 classes (2, 4, 5, 6, 7, 11) have the set document members spread to more than one cluster. If it happens, so we choose one cluster that has highest F-measure to became F-max for each class. The overall F-measure obtained by the sum of F-max weighting from all class where the result of this calculating performance of the proposed method was 0.88.

Table 7 shows comparison performance between proposed method (SVDPCA and Ant algorithms) and control method (only Ant algorithms) with the same parameter value. It shows the total time trial phase or finding the best path for 506 documents with the proposed method is faster

than the control method. This is due to the control method, terms that used to construct the vector graph is larger than the proposed method. It took 4758 terms from DTM for each document, so it needed 1 minutes 13 sec to calculate cosine distance between documents of that terms. While the proposed method only took 15 score-pc from DPCM for each document, so that the time required to calculate the cosine distance between documents just 5 sec, 1 minute 8 sec was faster than the control method.

However, the addition time of SVDPCA was not so affected because making vector graph (network) from matrix is very fast, so the total time trial phase of proposed method was 41 sec faster than the control method. This happens because the SVDPCA calculation can decrease the processing complexity for dimension reduction of DTM. If the classic PCA, we convert the DTM into DPCM in three stages. Firstly, we depend the decomposition of covariance (or correlation) matrix of DTM. Secondly, we have to calculate the eigenvalues using covariance matrix results, and the last we make eigenvectors based on eigenvalues that is used to get the DPCM matrix. While the SVDPCA, we just require one stage but gained three matrix at once, and choosing left eigenvector matrix to get the DPCM matrix, so that SVDPCA can support the efficiency of trial phase Ant algorithm.

In addition, with same parameters, both methods could make 14 clusters. However, the F measure performance of proposed method was 0.88 better than control method that only 0.78 (see table VII). The proposed method has proved that dimension reduction optimally by selecting the number of score-pc when their cumulative proportion of variance is more than 0.4. It has no effect to decrease in information quality of DPCM for making clustering documents. In fact, it could maintain F-measure is even better than control method with 30 ants and 100 iterations, although the limit cumulative proportion of variance suggested in previous study is 0.5 [8]. This is due to differences in the type and number of documents used previous study [8]. Although the control method did not reduce dimension of DTM, with parameters are 30 ants and 100 iterations, it could only reach the F-measure was 0.78. If the control method want to have the percentage of F-measure up to 80%, so it is still needed changes parameter values with more number of ants and iterations that would in turn be increase processing time when compared it with the proposed method.

The fact that the number of configuration parameters used in ant algorithm make difficult to determine the best configuration in order to make optimal cluster with the highest F-measure. Therefore, we plan to classifying news of documents with cluster algorithm that will develop classification method of documents based on Ant algorithms with supervised classification.

4. Conclusion

In this paper, we have proved that in order to speed up execution time of the trial phase Ant algorithms can be done by reducing sparse terms. SVDPCA could solve that sparse terms problem in DTM based acquisition term of ECS stemmer (taking 65% of all term) into a score-pc DPCM up to 99.7%, thus speeding up execution time of the trial phase Ant algorithms for finding the shortest path between documents. Then Ant algorithms could make document clusters with the best F-measure value achieved was 0.88 (88%) for the number of ants and iterations are not too large. The experimental results have shown that proposed method is more efficient and accurate.

References

[1] Vaijayanthi, P., Natarajan, A.M., & Murugadoss, R., *"Ants for Document Clustering,"* International Journal of Computer Science (IJCSI), vol. 9, no. 2, pp. 493-499. 2012.

[2] L. Machnik, *"Ants in Text Document Clustering"* In Proceeding of Advances in Systems, Computing Sciences and Software Engineering, pp. 209-212, 2006.

[3] L. Machnik, *"Documents Clustering Method based on Ant algorithms"* In Proceeding of the Internetional Multiconference on Computer Science and Information Technology, 2006.

[4] A. Z. Arifin and A. N. Setiono, "Klasifikasi Dokumen Berita Kejadian Berbahasa Indonesia dengan Algortima Single Pass Clustering" In Prosiding Seminar on Intelligent Technology and Its Application (SITIA), 2002.

[5] J. Asian, *"Effective Techniques for Indonesian Text Retrieval,"* PhD thesis School of Computer Science and Information Technology, RMIT, University Australia, 2007.

[6] A. Z. Arifin, I. P. A. K. Mahendra and H. T. Ciptaningtyas, *"Enhanced Confix Stripping Stemmer and Ant algorithms for Classifying News Document in Indonesia Language"* In Proceeding International Conference on Information and Communication Technology and Systems (ICTS), vol. 5, pp. 149-158, 2009.

[7] G. Salton, Automatic Text Processing, Addison Wesley, 1989.

[8] S. Jun, S.-S. Park and D.-S. Jang, *"Document clustering method using dimension reduction and support vector clustering to overco-*

me sparseness," Expert Systems with Applications, vol. 41, pp. 3204–3212. 2014.

[9] V. Cherkassky and F. Mulier, *"Learning from data: concepts, theory, and Methods"*, John Wiley & Sons, 2007.

[10] R. V. Ramirez-Velarde, M. Roderus, C. Barba-Jimenez and R. Perez-Cazares, *"A Parallel Implementation of Singular Value Decomposition for Video-on-Demand Services Design Using Principal Component Analysis"* In Proceeding of International Conference on Computational Science (ICCS), vol 29, pp. 1876-1887, 2014.

[11] M. Dorigo, V. Maniezzo and A. Colorni, *"The Ant System: Optimization by a Colony of Cooperating Agents,"* IEEE Transactions on Systems, Man, and Cybernetics-Part B, vol. 26, no. 1, pp. 29-41. 1996.

Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/1 (2016), 26-34
DOI: http://dx.doi.org/10.21609/jiki.v9i1.363

ROBUST INTEGER HAAR WAVELET BASED WATERMARKING USING SINGULAR VALUE DECOMPOSITION

Prajanto Wahyu Adi, and Farah Zakiyah Rahmanti

Faculty of Computer Science, Universitas Dian Nuswantoro (UDINUS), Jalan Imam Bonjol, No.207, Semarang, 50131, Indonesia

E-mail: prajanto@dsn.dinus.ac.id, farah_zakiyah@dsn.dinus.ac.id

Abstract

This paper proposed a hybrid watermarking method that used dither quantization of Singular Value Decomposition (SVD) on average coefficients of Integer Haar Wavelet Transform (IHWT). The watermark image embeds through dither quantization process on singular coefficients value. This scheme aims to obtain the higher robustness level than previous method which performs dither quantization of SVD directly on image pixels value. The experiment results show that the proposed method has proper watermarked images quality above 38dB. The proposed method has better performance than the previous method in term of robustness against several image processing attacks. In JPEG compression with Quality Factor of 50 and 70, JPEG2000 compression with Compression Ratio of 5 and 3, average filtering, and Gaussian filtering, the previous method has average Normalized Correlation (NC) values of 0.8756, 0.9759, 0.9509, 0.9905, 0.8321, and 0.9297 respectively. While, the proposed method has better average NC values of 0.9730, 0.9884, 0.9844, 0.9963, 0.9020, and 0.9590 respectively.

Keywords: *watermarking, SVD, integer wavelet transform, dither quantization*

Abstrak

Paper ini mengusulkan metode *watermarking* yang menggabungkan kuantisasi dither SVD pada nilai koefisien rata-rata IHWT. Citra watermark disisipkan melalui proses kuantisasi dither pada nilai koefisien singular. Skema ini dilakukan untuk memperoleh tingkat ketahanan yang lebih tinggi dari metode sebelumnya yang melakukan kuantisasi dither SVD secara langsung pada nilai piksel citra. Hasil percobaan menunjukkan bahwa metode yang diusulkan mempunyai kualitas citra ter-watermark yang layak diatas 38dB. Metode yang diusulkan mempunyai kinerja lebih baik dari metode sebelumnya dalam hal ketahanan terhadap beberapa serangan citra. Dalam kompresi JPEG dengan Quality Factor 50 dan 70, kompresi JPEG2000 dengan *Compression Ratio* 5 dan 3, *average filtering*, dan *gaussian filtering*, metode sebelumnya menghasilkan nilai *Normalized Correlation* (NC) rata-rata berturut-turut sebesar 0.8756, 0.9759, 0.9509, 0.9905, 0.8321, dan 0.9297. Sementara, metode yang diusulkan menghasilkan nilai NC rata-rata yang lebih baik sebesar 0.9730, 0.9884, 0.9844, 0.9963, 0.9020, dan 0.9590 berturut-turut.

Kata Kunci: *watermarking, SVD, integer wavelet transform, dither quantization*

1. Introduction

The rapid growth of the Internet and computer networks led to ease of distribution process of digital medium such as audio, video, and digital image [1-2]. It provides convenience in sharing digital content. However, the ease of copying medium without regard to the ownership, poses a threat on copyright protection [3]**.** Digital water-marking system provides a solution to protect the copyright of digital content [4]. It is able to hide proprietary information into digital medium in order to provide proof of ownership. In addition, the application of digital watermarking has been expanded to the several fields, such as: authentication, broadcast monitoring, and finger printing [3]. Generally, wa-termarking can be classified in-to several types. A widely used classification is based on domain that is used. Watermarking can be classified into spati-al and transform domain method. The spatial me-thods embeds watermark directly on image pixels which results in low computational complexity, but vulnerable against image processing attack. Instead, the transform domain methods have more resistance against many attacks [3-6].

Several studies about transform domain based watermarking have been carried out, such as Discrete Cosine Transform (DCT) [7-8], Discrete Wavelet Transform (DWT) [5,9], Lifting Wavelet

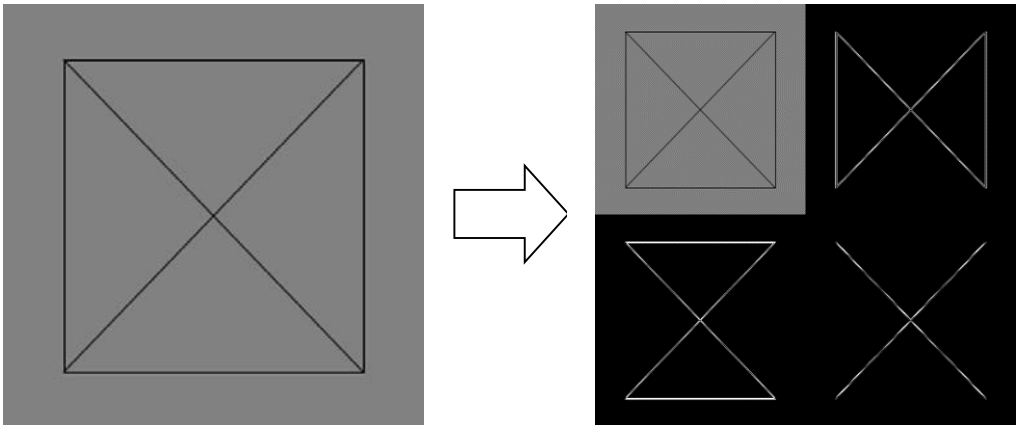


Figure 1. Integer Haar Wavelet Transformation HH

Transform (LWT) [4], and Integer Wavelet Transform (IWT) [3] and [10]. DCT is a popular method, but it can cause damage on image when inserting large size of watermark [11]. Alternatively, DWT is able to compute image features at multiple resolution level which can improve the imperceptibility and robustness of watermarked image [11]. However, the floating point coefficients of DWT will induce reversibility problem that occur after transformation [4]. The LWT and IWT are the second generation wavelet transform methods which developed from DWT through a lifting scheme. These methods can avoid the floating point problem of DWT [12] and able to make watermarking scheme more efficient [4]. This paper proposed the use of Integer Haar Wavelet Transform (IHWT) coefficients as the values of SVD, due to numerical advantages of IHWT [13].

The following sections of this paper are organized as follows: The process of embedding and extraction of the proposed method are explained in Section 2. In Section 3, the experiment results, performance comparison, analysis and discussion are presented. Finally, the conclusions are descrybed in Section 4.

IHWT is the second generation of wavelet transform which is developed from Discrete Haar Wavelet Transform (DHWT) via lifting scheme by Xu et al. [14]. Basically, the wavelet transformation on digital image is used to decompose image with size of *M* x *N* into four groups of wavelet coefficients with size of *M*/2 x *N*/2, namely LL, HL, LH, and HH. Those wavelets coefficients are often called as sub bands. The main idea of wavelet transformation on digital image is by applying low pass filter and high pass filter in horizontal and vertical orders. LL is produced from low pass filtering in both horizontal and vertical order. In HL, high pass filter is applied in horizontal order followed by low pass filter in vertical order and vice versa for LH. Afterwards applying high pass filter in both horizontal and vertical orders to generate HH. IHWT performs decomposition using 4 filters which generates from low pass and high pass filters as given by equation(1) until equation (4).

Figure 1 shows that the image is decomposed into four wavelet sub band namely: LL at the top left, HL at the top right, LH at the bottom left, and HH at the bottom right position of the wavelet sub bands. LL is the average coefficients values that contain the average value of image, this wavelet's sub band contain the significant values of image so it has high robustness in maintaining information contained therein [13,15]. Whereas, the HL, LH, and are the less significant coefficient value of that contain the edge are of the images. HL is used to removes the horizontal edge of image and let the vertical and diagonal edge of image. Otherwise, LH eliminates the vertical part and left the horizontal and diagonal parts. While, the HH is intended to removes both vertical and horizontal edge of the image.

Integer Haar Wavelet Transform (IHWT) decompose 2x2 non-overlapping block of image below into wavelet coefficients as given by equation(1) to equation(4).

$$\begin{bmatrix} \boldsymbol{I_{m,n}} & \boldsymbol{I_{m,n+1}} \\ \boldsymbol{I_{m+1,n}} & \boldsymbol{I_{m+1,n+1}} \end{bmatrix}$$

$$\boldsymbol{LL} = \left[\frac{\left[\frac{\boldsymbol{I_{m,n}} + \boldsymbol{I_{m,n+1}}}{\boldsymbol{2}}\right] + \left[\frac{\boldsymbol{I_{m+1,n}} + \boldsymbol{I_{m+1,n+1}}}{\boldsymbol{2}}\right]}{\boldsymbol{2}} \right] \quad (1)$$

$$\boldsymbol{HL} = \left[\frac{\boldsymbol{I_{m,n}} - \boldsymbol{I_{m,n+1}} + \boldsymbol{I_{m+1,n}} - \boldsymbol{I_{m+1,n+1}}}{\boldsymbol{2}} \right] \quad (2)$$

$$LH = \left[\frac{I_{m,n} + I_{m,n+1}}{2}\right] - \left[\frac{I_{m+1,n} + I_{m+1,n+1}}{2}\right] \quad (3)$$

$$HH = I_{m,n} - I_{m,n+1} - I_{m+1,n} + I_{m+1,n+1} \quad (4)$$

where, $I_{m,n}$ is a pixel value of every block of image at row *m* and column *n*.

The watermark is embedded on LL sub-band to gain high robustness; otherwise it can be embedded to HL, LH, or HH for high imperceptibility. After the embedding, the block of watermarked image (*I'*) below

$$\begin{bmatrix} I'_{m,n} & I'_{m,n+1} \\ I'_{m+1,n} & I'_{m+1,n+1} \end{bmatrix}$$

is reconstructed from wavelet coefficients as the following equation(5) to equation(8):

$$I'_{m,n} = LL + \left\lfloor \frac{LH+1}{2} \right\rfloor + \left\lfloor \frac{HL + \left\lfloor \frac{HH+1}{2} \right\rfloor + 1}{2} \right\rfloor \quad (5)$$

$$I'_{m,n+1} = I'_{m,n} - \left(HL + \left\lfloor \frac{HH+1}{2} \right\rfloor \right) \quad (6)$$

$$I'_{m+1,n} = LL + \left\lfloor \frac{LH+1}{2} \right\rfloor - LH + \left\lfloor \frac{HL + \left\lfloor \frac{HH+1}{2} \right\rfloor - HH + 1}{2} \right\rfloor \quad (7)$$

$$I'_{m,n+1} = I'_{m+1,n} - \left(HL + \left\lfloor \frac{HH+1}{2} \right\rfloor - HH \right) \quad (8)$$

Equation(5) until equation(8) shows that IHWT is able to maintain the lost information due truncateon process in equation(1) until (4) by the addition of integer value of 1 in the reconstruction process. This lifting scheme is able to restore the missing value in watermarked image.

2. Methods

In our method, the watermark bits are embedded on singular value of average coefficients using dither quantization as shown in Figure 2. This scheme aimed to obtain the higher robustness level than previous method by Mohan and Kumar [16], which performs dither quantization directly on singular value of image pixel.

Watermark Embedding

First of all, apply IHWT using equation(1) until equation(4) on original image to get LL, HL, LH, and HH. Then perform SVD transformation on each 8x8 block of average coefficients LL to get matrices U_b, S_b, and V_b of block *b* as given by equation(9).

$$SVD_b(LL_b) = U_b S_b V_b \quad (9)$$

where S_b is a singular matrices, U_b and V_b are the orthogonal matrices of block *b*.

Afterwards, calculate the lower bound and upper bound of each quantization step as formulated using equation(10) and equation(11).

$$l_i = \min S + q(i-2) \quad (10)$$

$$u_i = \min S + q(i-1) \quad (11)$$

where *q* is quantization step size, min *S* is minimum value of largest singular value S_{1b} on entire blocks, l_i and u_i are the upper bound and lower bound of quantization step *i*, for *i*=1, 2, …, *q*.

Once the bounds are calculated, generate a quantization table from l_i and u_i of each quantization step. Embed watermark by altering S_{1b} according to watermark bit value for $l_i \le S_{1b} < u_i$

$$S'_{1b} = \begin{cases} ((l_i + u_i)/2 + l_i)/2 & ,if\ w = 1 \\ ((l_i + u_i)/2 + u_i)/2 & ,if\ w = 0 \end{cases} \quad (12)$$

where $\boldsymbol{S'_{1b}}$ is the new largest singular value of block *b*, and *w* is the watermark bit value. Then perform inverse SVD on each block to get watermarked average coefficient *LL'*

$$LL'_b = U_b S'_b V_b^T \quad (13)$$

Where $\boldsymbol{V_b^T}$ is the conjugate transpose of V_b. Lastly, apply Inverse IHWT using equation (5) to equation(8) on LL', HL, LH, and HH to get watermarked image.

Watermark Extraction

Watermark extraction process is done by arranging the singular values of average coefficients into the corresponding quantization step in quantization table. It is then used to obtain the watermark bit value based on the difference of singular values toward upper bound and lower bound: 1) apply IHWT transformation using equation(1) until equation(4) on watermarked image to get LL, HL, LH, and HH; 2) perform SVD transformation on each 8x8 block of LL to get matrices U_b, S_b, and V_b; 3) use the quantization table as a key to extract watermark image; 4) extract watermark bit value on each block to get the watermark image using equation(14).

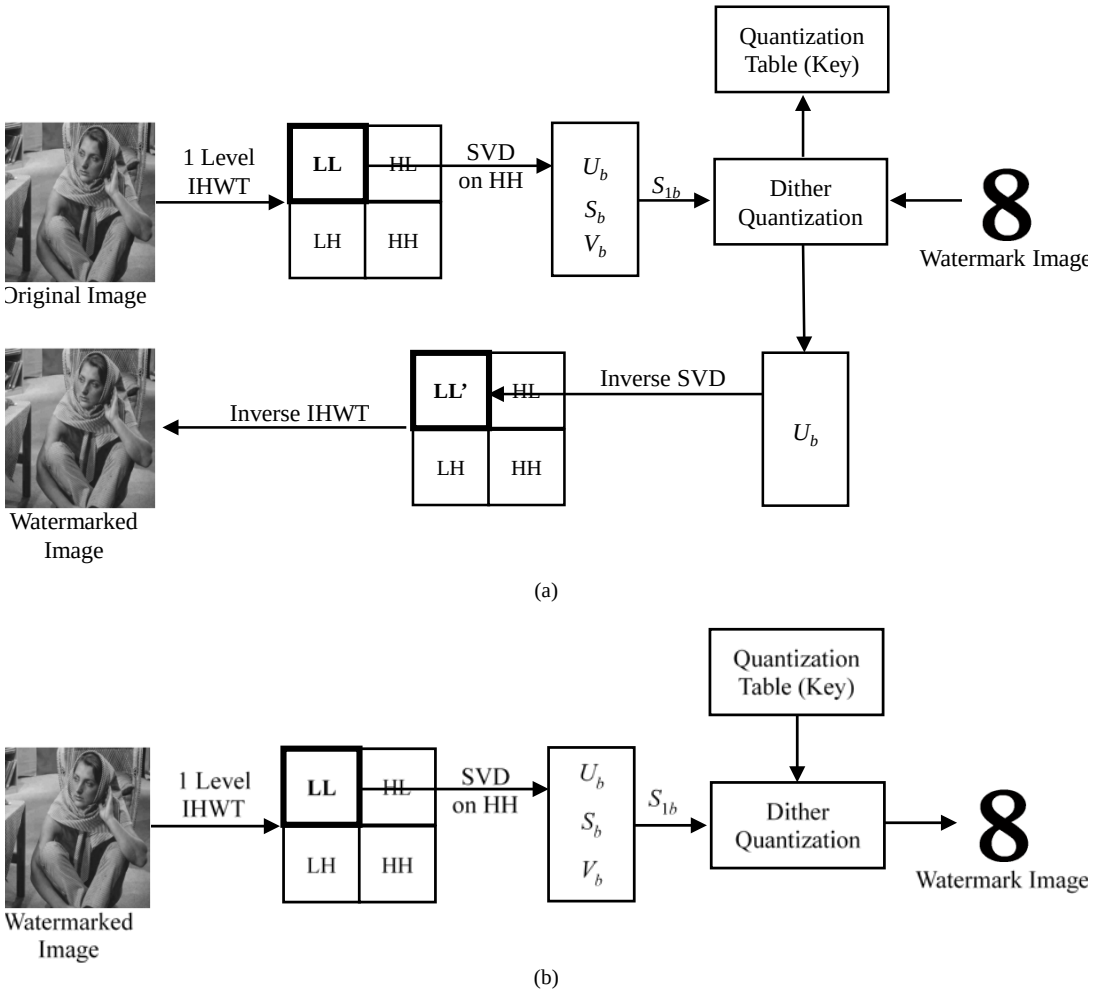


Figure 2. The proposed method: (a) embedding and (b) extraction

$$w_b = \begin{cases} 1 & ,if \quad l_i \le S_{1b} < (l_i + u_i)/2 \\ 0 & ,if \quad (l_i + u_i)/2 \le S_{1b} < u_i \end{cases} \tag{14}$$

Where w_b is the watermark value b, and S_{1b} is the largest singular value of block b for $l_i \le S_{1b} < u_i$.

3. Results and Analysis

This research uses 6 standard grayscale image within bitmap format (.bmp) and size of 512x512 pixels as original images and a binary image as watermark image as shown in Figure 3 and Figure 5.

Watermarked Image Quality

Based on previous research [16], watermark image is embedded using step size of 60 in order to get optimum results [16]. The quality of watermarked images are measured using standard Peak Signal to Noise Ratio (PSNR), and Structural Similarity (SSIM) [17].

Performance Comparison

The proposed method is compared with the previous method by Mohan and Kumar [16] in terms of the robustness of watermark images and the computational complexity of the methods in embedding and extraction.

Robustness

Several popular attacks on watermarked images, namely: JPEG compression, JPEG2000 compression, average filtering, and median filtering are used in this experiment. Performances of both methods are measured using Normalized Correlation (NC) which is widely used in field of watermarking. The first experiment is carried out to compare the robustness of the both methods against JPEG compression with the standard Quality Factor (QF) of 50 and 70.

Table 2 and Figure 6 show that the previous method has low robustness against JPEG compression attack with QF of 50. Meanwhile, the pro-

Figure 3. Original images

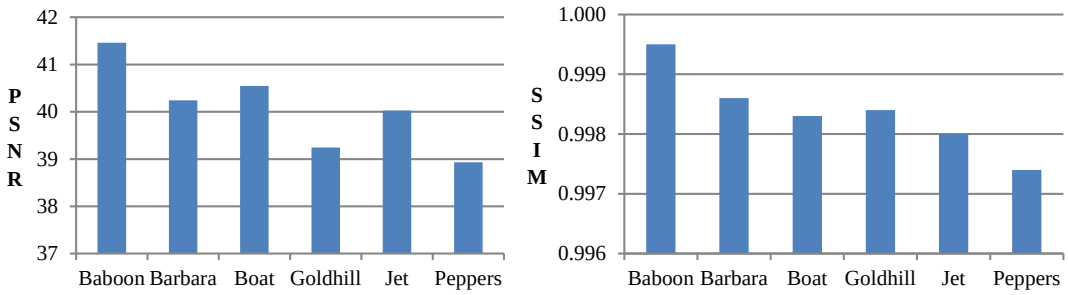


Figure 4. Watermarked Images Quality in (a) PSNR and (b) SSIM

Figure 5. Watermark image

posed method has higher robustness level than previous method. The proposed method even almost reached perfect robustness level at QF of 70.

The next comparison is deal with JPEG 2000 compression attack. The decent Compression Ratio (CR) of 5 and 3 are used to compress the water-marked images.

JPEG2000 attack does not give significant damage on watermark images as shown in Table 3 and Figure 7. Although the both methods have good robustness level, the proposed method still outperforms the previous method. The proposed method has perfect NC value of 1 in test of Gold-hill and Peppers images within compression ratio of 3. The last experiment is test the robustness against Average Filtering (AF) and Gaussian Filtering (GF) attacks. The average filter size used is 3x3, while the Gaussian filter have sigma value of 0.5 and filter size of 3x3 which is the default value of the filters. Table 4 and Figure 8 show that the both filtering attack gives significant dam-age on watermark images. It also shows that pro-posed method has better robustness level against average and Gaussian filtering attacks.

The proposed method has higher robustness level against such compression and filtering attacks. Table 5 shows that the extracted water-mark images of the proposed method has better visual quality than the previous method.

Computational Complexity

Test of computational time was conducted to measure performance of the proposed method against the previous method in embedding and extraction processes. The test was conducted in a laptop PC with second generation Intel Core i3- 2310M quad core processor 2.1 GHz of clock and 6GB of RAM. The execution times of the both methods are measured in milliseconds (ms).

TABLE 1
WATERMARKED IMAGES QUALITY

Images	PSNR	SSIM
Baboon	41.4615	0.9995
Barbara	40.2440	0.9986
Boat	40.5464	0.9983
Goldhill	39.2421	0.9984
Jet	40.0268	0.9980
Peppers	38.9337	0.9974

TABLE 2
ROBUSTNESS AGAINST JPEG COMPRESSION

Images	SVD[16]		Proposed	
	QF 50	QF 70	QF 50	QF 70
Baboon	0.8291	0.9643	0.9474	0.9846
Barbara	0.8579	0.9692	0.9743	0.9986
Boat	0.8622	0.9616	0.9631	0.9690
Goldhill	0.9047	0.9904	0.9891	0.9994
Jet	0.8832	0.9776	0.9697	0.9794
Peppers	0.9165	0.9922	0.9943	0.9994
Average	0.8756	0.9759	0.9730	0.9884

TABLE 3
ROBUSTNESS AGAINST JPEG2000 COMPRESSION

Images	SVD[16]		Proposed	
	CR 5	CR 3	CR 5	CR 3
Baboon	0.7881	0.9507	0.9192	0.9795
Barbara	0.9650	0.9960	0.9955	0.9994
Boat	0.9903	0.9989	0.9981	0.9994
Goldhill	0.9808	0.9992	0.9981	1.0000
Jet	0.9935	0.9992	0.9968	0.9994
Peppers	0.9877	0.9987	0.9987	1.0000
Average	0.9509	0.9905	0.9844	0.9963

TABLE 4
ROBUSTNESS AGAINST FILTERING

Images	SVD[16]		Proposed	
	AF	GF	AF	GF
Baboon	0.7328	0.8696	0.8494	0.9471
Barbara	0.8382	0.9270	0.8995	0.9618
Boat	0.8340	0.9283	0.9068	0.9451
Goldhill	0.8605	0.9608	0.9168	0.9703
Jet	0.8755	0.9368	0.9071	0.9524
Peppers	0.8873	0.9555	0.9325	0.9775
Average	0.8381	0.9297	0.9020	0.9590

The experiment results in Table 6 and Figure 10 shows that the proposed method has slower embedding process with average value of 1870ms, while the previous method has average embedding time of 740ms. However, the proposed method has faster extraction process with average time of 278ms.

IHWT have integer precision on the basic principle of wavelet transformation which able to perform down sampling as given by equation(1)

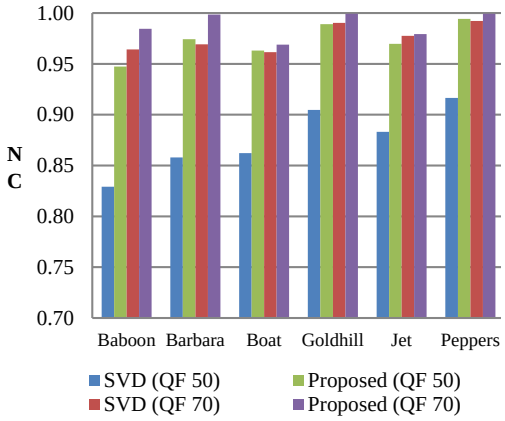


Figure 6. Robustness against JPEG Compression

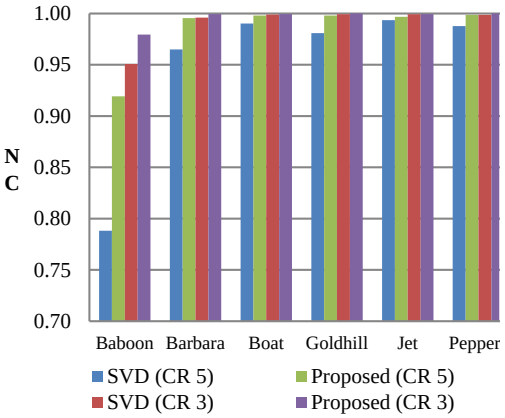


Figure 7. Robustness against JPEG2000 Compression

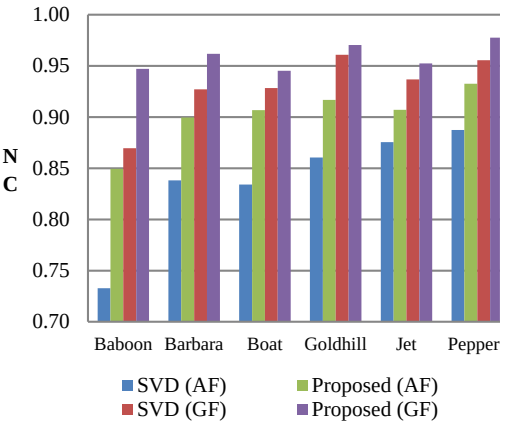


Figure 8. Robustness against Filtering

until (4). The image pixels value $I_{m,n}$, $I_{m+1,n}$, $I_{m,n+1}$, $I_{m+1,n+1}$ are down sampled into single coefficient value. The dither quantization of singular value of wavelet coefficients will reduce watermark image size from M x N into $M/2$ x $N/2$ due to down sampling scheme of IHWT. The smaller size of watermark will increase the density of coefficients that used as singular value in embedding process

as given by equation(12). It also reduces the computational complexity in extraction of watermark. Increasing number of coefficients used in SVD will improve the robustness of watermark image as well.

The watermark is embeds on LL coefficients which have lowest frequency among all wavelet coefficients. LL is derived from average value of image as given by equation(1), so it has low level of disparity or frequency. A low wavelet coefficient is able to maintain the value therein, thus increase the robustness of watermark image. The down sampling scheme and the use of low frequency coefficients are the two main factors that cause the superiority of the proposed method over the previous method by Mohan and Kumar.

TABLE 6
EMBEDDING AND EXTRACTION TIMES

Images	SVD[16]		Proposed	
	Embed	Extract	Embed	Extract
Baboon	748	336	1821	270
Barbara	719	339	1879	272
Boat	713	341	1857	274
Goldhill	765	331	1868	271
Jet	756	333	1882	281
Peppers	740	372	1914	304
Average	740	342	1870	278

4. Conclusion

This paper proposed a hybrid watermarking method that used SVD on average coefficients of IHWT. The watermark image is embeds through dither quantization process on singular coefficie-

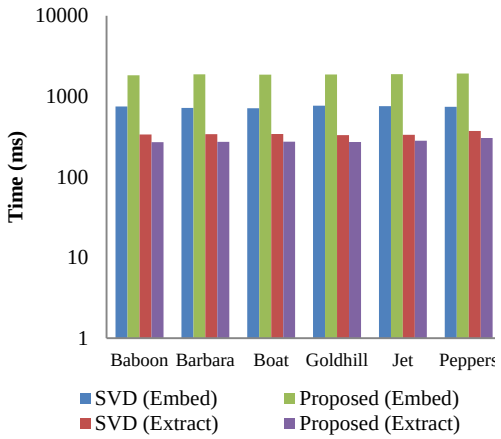


Figure 8. Robustness against Filtering

TABLE 5
EXTRACTED WATERMARK IMAGES

Methods	Attacks	Images					
		Baboon	Barbara	Boat	Goldhill	Jet	Peppers
SVD [16]	JPEG QF 50						
	JPEG QF 70						
	JPEG2000 CR 5						
	JPEG2000 CR 3						
	Average Filtering						
	Gaussian Filtering						
Proposed	JPEG QF 50						
	JPEG QF 70						
	JPEG2000 CR 5						
	JPEG2000 CR 3						
	Average Filtering						
	Gaussian Filtering						

nts value. The experiment results shows that the proposed method has proper watermarked images quality above 38dB. In term of robustness against compression and filtering attacks, the proposed method has outperform the previous method by Mohan and Kumar which perform SVD directly on image pixel value. IHWT is able to perform down sampling on image which in turn increase the robustness of watermark image and reduce the computational complexity in extraction of watermark. Moreover, the IHWT has finite integer value which can avoid reversibility problem in wavelet based watermarking system, making it suitable for further development of reversible watermarking system.

Acknowledgement

The authors would like to thank to Institute of Research and Community Services (LPPM), Universitas Dian Nuswantoro for providing financial support on this study by Research Grants Scheme (021/A.35-02/UDN.09/X/2015)

References

[1] M. Ali, C. W. Ahn, M. Pant, and P. Siarry, "An image watermarking scheme in wavelet domain with optimized compensation of singular value decomposition via artificial bee colony," *Inf. Sci. (Ny).*, vol. 301, pp. 44–60, 2015.

[2] N. Muhammad and N. Bibi, "Digital image watermarking using partial pivoting lower and upper triangular decomposition into the wavelet domain," *IET Image Process.*, vol. 9, no. 9, pp. 795–803, 2015.

[3] N. M. Makbol and B. E. Khoo, "A new robust and secure digital image watermarking scheme based on the integer wavelet transform and singular value decomposition," *Digit. Signal Process.* vol. 1, no. 134, pp. 1–14, 2014.

[4] V. Singh, R. Kumar, and A. Ojha, "Significant region based robust water-marking scheme in lifting wavelet transform domain," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8184–8197, 2015.

[5] C.-C. Lai and C.-C. Tsai, "Digital Image Watermarking Using Discrete Wavelet Transform and Singular Value Decom-position," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 11, pp. 3060–3063, 2010.

[6] S. Rao V, "A DWT-DCT-SVD Based Digital Image Watermarking Scheme Using Particle Swarm Optimization," in *2012 IEEE Students' Conference on Electrical, Electronics and Computer Science*, 2012, pp. 10–13.

[7] S. D. Lin, "A robust DCT-based watermarking for copyright protection," *IEEE Trans. Consum. Electron*, vol. 46, no. 3, pp. 415–421, 2000.

[8] J. C. Patra, J. E. Phua, and C. Bornand, "A novel DCT domain CRT-based water-marking scheme for image authentication surviving JPEG compression," *Digit. Signal Process*, vol. 20, no. 6, pp. 1597–1611, 2010.

[9] C. Li, Z. Zhang, Y. Wang, B. Ma, and D. Huang, "Dither modulation of significant amplitude difference for wavelet based robust watermarking," *Neurocomputing*, vol. 166, pp. 404–415, 2015.

[10] K. Loukhaoukha and J.-Y. Chouinard, "Hybrid watermarking algorithm based on SVD and lifting wavelet transform for ownership verification," in *2009 11th Canadian Workshop on Information Theory*, 2009, no. 1, pp. 177–182.

[11] M. Botta, D. Cavagnino, and V. Pomponiu, "A modular framework for color image watermarking," *Signal Processing*, vol. 119, pp. 102–114, 2016.

[12] N. Raftari and A. M. E. Moghadam, "Digital image steganography based on Integer Wavelet Transform and assignment algorithm," *Proc.- 6th Asia Int. Conf. Math. Model. Comput. Simulation, AMS 2012*, pp. 87–92, 2012.

[13] P. W. Adi, F. Z. Rahmanti, and N. A. Abu, "High Quality Image Steganography on Integer Haar Wavelet Transform using Modulus Function," in *2015 International Conference on Science in Information Technology (ICSITech)*, 2015, pp. 79–84.

[14] J. Xu, a. H. Sung, P. Shi, and Q. Liu, "JPEG compression immune steganography using wavelet transform," *Int. Conf. Inf. Technol. Coding Comput. 2004. Proceedings. ITCC 2004.* vol. 2, pp. 0–4, 2004.

[15] N. A. Abu, P. W. Adi, and O. Mohd, "Robust Digital Image Steganography within Coefficient Difference on Integer Haar Wavelet Transform," *lnternational J. Video Image Process. Netw. Secur.*, vol. 14, no. 02, pp. 1 – 8, 2014.

[16] B. C. Mohan and S. S. Kumar, "A Robust Image Watermarking Scheme using Singular Value Decomposition," *J. Multimed.*, vol. 3, no. 1, pp. 7–15, 2008.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[18] Y. P. Lee, J. C. Lee, W. K. Chen, K. C. Chang, I. J. Su, and C. P. Chang, "High-payload

image hiding with quality recovery using tri-way pixel-value differencing," *Inf. Sci. (Ny).*, vol. 191, pp. 214–225, 2012.

Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/1 (2016), 35-42
DOI: http://dx.doi.org/10.21609/jiki.v9i1.364

ELECTROCARDIOGRAM ARRHYTHMIA CLASSIFICATION SYSTEM USING SUPPORT VECTOR MACHINE BASED FUZZY LOGIC

Sugiyanto, Tutuk Indriyani, and Muhammad Heru Firmansyah

Department of Informatics Engineering, Faculty of Information Technology,
Adhi Tama Institute of Technology Surabaya
Arief Rachman Hakim 100, Surabaya, 60117, Indonesia

E-mail: sugianto@itats.ac.id, tutuk@itats.ac.id, herufirmansyah93@gmail.com

Abstract

Arrhythmia is a cardiovascular disease that can be diagnosed by doctors using an electrocardiogram (ECG). The information contained on the ECG is used by doctors to analyze the electrical activity of the heart and determine the type of arrhythmia suffered by the patient. In this study, ECG arrhythmia classification process was performed using Support Vector Machine based fuzzy logic. In the proposed method, fuzzy membership functions are used to cope with data that are not classifiable in the method of Support Vector Machine (SVM) one-against-one. An early stage of the data processing is the baseline wander removal process on the original ECG signal using Transformation Wavelet Discrete (TWD). Afterwards then the ECG signal is cleaned from the baseline wander segmented into units beat. The next stage is to look for six features of the beat. Every single beat is classified using SVM method based fuzzy logic. Results from this study show that ECG arrhythmia classification using proposed method (SVM based fuzzy logic) gives better results than original SVM method. ECG arrhythmia classification using SVM method based fuzzy logic forms an average value of accuracy level, sensitivity level, and specificity level of 93.5%, 93.5%, and 98.7% respectively. ECG arrhythmia classification using only SVM method forms an average value accuracy level, sensitivity level, and specificity level of 91.83%, 91.83%, and 98.36% respectively.

Keywords: *arrhythmia classification, ECG, fuzzy logic, heart rate, Support Vector Machine*

Abstrak

Aritmia adalah penyakit kardiovaskular yang dapat didiagnosis dokter menggunakan elektrokardiogram (EKG). Informasi yang terdapat di EKG digunakan oleh dokter untuk menganalisis aktivitas elektrik jantung dan menentukan jenis aritmia yang diderita oleh pasien. Dalam penelitian ini, proses klasifikasi aritmia EKG dilakukan dengan menggunakan Support Vector Machine berbasis fuzzy logic. Pada metode yang diusulkan, fungsi keanggotaan fuzzy digunakan untuk mengatasi dengan data yang tidak dapat diklasifikasikan dalam metode Support Vector Machine (SVM) satu-terhadap-satu. Tahap awal pengolahan data adalah proses *baseline wander removal* pada sinyal EKG asli menggunakan Transformasi Wavelet Diskrit (TWD), dan kemudian sinyal EKG bersih dari *baseline wander* tersegmentasi ke unit denyut. Tahap berikutnya adalah untuk mencari enam fitur dari denyut, dan setiap denyut tunggal diklasifikasikan menggunakan metode SVM berbasis fuzzy logic. Hasil dari penelitian menunjukkan bahwa klasifikasi aritmia EKG menggunakan metode yang diusulkan (SVM berdasarkan logika fuzzy) memberikan hasil yang lebih baik daripada metode SVM asli. Klasifikasi aritmia EKG menggunakan metode SVM berbasis logika fuzzy membentuk nilai rata-rata tingkat akurasi, tingkat sensitivitas, dan tingkat spesifisitas 93,5%, 93,5%, dan 98,7%. Klasifikasi aritmia EKG menggunakan metode SVM asli hanya membentuk tingkat rata-rata nilai akurasi, tingkat sensitivitas, dan tingkat spesifisitas 91,83%, 91,83%, dan 98,36%.

Kata Kunci: *klasifikasi aritmia, ECG, logika fuzzy, denyut jantung, Support Vector Machine*

1. Introduction

Arrhythmias are disorders of the heart in the form of interference on the frequency, irregularity, place of origin pulse or conduction of electrical impulses in the heart [1]. Arrhythmia is a dangerous disease, so patients need immediate treatment and therapy regularly to prevent a worsening condition [2]. In general, the diagnosis of arrhythmias can only be done by a cardiologist. Along with the development of science and technology, many researchers doing research on the diagnosis of arrhythmias to find a system that can classify arrhythmias more accurately.

Many researchers have previously raised the topic of arrhythmia classification. Srivastava et al. [3] create an arrhythmia classification system using Fuzzy Sugeno method. The proposed sys-tem is able to categorize an ECG wave to one of thirteen types of arrhythmia. Results from the classification system are consistent with the results of testing by the cardiologist as many as 91 of the 105 data.

Moavenian and Khorrami [4] conducted a comparison using the method of Support Vector Machine (SVM) and Artificial Neural Network (ANN) to classify the six types of arrhythmia among others Right Bundle Branch Block (RBBB), Left Bundle Branch Block (LBBB), Premature Ventricular Beat (PVB), Paced Beat (PB) Premature Atrial Beat (PAB), and Fusion of Paced and normal Beat (FB). The testing process using three assessment criteria includes performance training, performance testing and training time. The results show that SVM is superior in performance training and training time while ANN is superior in performance testing.

Castillo et al. [5] did a comparison between three methods of classification of arrhythmias include Fuzzy K-Nearest Neighbours, Multi-Layer Perceptron with Back Propagation Gradient Descent with momentum, and Multi-Layer Perceptron with Back Propagation Gradient Scale Conjugated. Each method resulted considerable accuracy. After the arrhythmia classification system combined with Fuzzy Mamdani method, it rise the accuracy up to 98%.

From several previous studies, the authors found a solution to make the arrhythmia classification system using SVM method that has proven better for classifying arrhythmias, and combined with fuzzy methods to address data that is not classified. One of the electrocardiogram signal noises that often arise is the Baseline Wander, a condition in which an electrocardiogram signal have movement at low frequencies irregular [6]. So that the system can perform accurate classification electrocardiogram signal, the signal must be free from noise. The solution resulted in higher levels of accuracy.

2. Methods

Electrocardiogram

Electrocardiogram (ECG) is a representation of a signal generated by the electrical activity of the heart muscle [7]. ECG signal is recorded using an electrocardiograph device. Electrocardiograph devices are medical devices used by patients to measure the electrical activity of the heart by measuring the difference biopotential from the outside into the body. In the medical field, electrocardiograph is used to diagnose some types of diseases related to the heart of which is a heart attack, disease/heart conditions, ischemia, hypertension (high blood pressure). One-piece ECG describes the condition of normal heartbeat consists of one P wave, one QRS complex, and one T wave. Figure 1 shows a schematic representation of normal ECG.

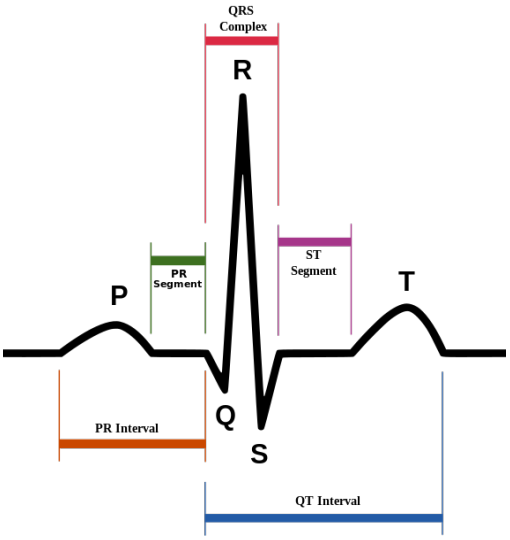


Figure 1. Schematic representation of normal ECG

Arrhythmia

Arrhythmia is a change in the frequency of heart rhythm caused by abnormal electrolyte conduction or automatically [1]. Heart rhythm disorder is not confined to the irregularity of the heart rate but also includes the rate and conduction disturbances. Heart rhythm disorder is caused by lack of oxygenated blood supply to the heart muscle, usually arises throbbing sensation that is too slow, too fast, or irregular throbbing pulsations [8]. Based on the type of beat, Association for the Advancement of Medical Instrumentation (AAMI) classifies the arrhythmia into 15 types of beat [9]. This study will classify the six types of disorder's arrhythmias selected from disorders that often appear in the dataset MIT-BIH [10] among other's Normal Beat (NB), Premature Ventricular Contraction (PVC), Paced Beat (PB), Left Bundle Branch Block Beat (LBBB), Right Bundle Branch Block Beat (RBBB), and Atrial Premature Beat (APB).

Support Vector Machine

Support Vector Machine (SVM) is a linear classifier method with the predetermined feature set [11]. SVM delivers maximum results while using fewer data training and there is no overlap bet-

ween the existing classes [12]. SVM looking for a hyperplane with the largest margin is called Maximum Marginal Hyperplane (MMH) to separate the existing class. $\boldsymbol{D}(\boldsymbol{x},\boldsymbol{w},\boldsymbol{b})$ is a decision function to determine the MMH can be expressed by equation(1).

$$D(\mathrm{x},\mathrm{w},\mathrm{b}) = W^T . X + b = \sum_{i=1}^{n} \mathrm{W}_i x_i + b \quad (1)$$

To optimize the equation(1) in order to obtain a maximum margin, quadratic programming optimization is performed using equation(2).

$$L_d(a) = -0.5a^2 Ha + f^T a \quad (2)$$

Equation(2) has a constraint such as equation(3), and equation(4).

$$y^T a = 0 \quad (3)$$

$$ai \geq 0, i = 1, \ldots . l, \quad (4)$$

where $\alpha = [\alpha_1,\alpha_2,\ldots,\alpha_l]$T, H is the Hessian matrix notation for $H_{ij} = y_i\, y_j\, x_i^T\, x_j$, f is a (l, l) vector, and f = [1 1 …. 1]T. With regard to equation (1), then we used equation(5) and (6) to obtain the value of w, and b are optimal.

$$W_o = \sum_{l=1}^{l} a_{oi}\, y_i x_i \quad (5)$$

$$b_0 = \frac{1}{N_{sv}} \left(\sum_{s=1}^{N_{sv}} \left(\frac{1}{y_s} - x_s{}^T w_o \right) \right) \quad (6)$$

where equation(6) is only used on data that has a support vector (α>0).

Support Vector Machine Based Fuzzy Logic

Support Vector Machine (SVM) based fuzzy logic is the development of Support Vector Machine methods to overcome the problems of multiclass. In each class, there will be defined a polyhedral pyramidal decision membership function using function obtained from SVM method for a class pair. SVM method will look for a hyperplane with the largest margin called Maximum Marginal Hyperplane (MMH). The distance between the hyperplane with a side of the margin is equal to the distance between the hyperplane with margin on the other side. If the classification is done in pairs, decision function for class i and j formulated in equation(7).

$$D_{ij}(x) = w_{ij}{}^T x + b_{ij} \quad (7)$$

SVM based fuzzy logic used membership functions to classify areas that cannot be classified by the decision function. Equation(8) shows the membership function m_{ij}.

$$m_{ij}(x) = \begin{cases} 1, for\ D_{ij}(x) \geq 1, \\ D_{ij}(x)\ for\ the\ other \end{cases} \quad (8)$$

By using $m_{ij}(x)$, the membership function x to the class i can be defined using the operator equation(9).

$$m_i(x) = \min_{\mathrm{j=1,\ldots..n}} m_{ij}(x) \quad (9)$$

Equation(8) and (9) are equivalent to equation (10).

$$m_i(x) = min(1, \min_{\mathrm{j\neq i,j=1,\ldots..n}} D_{ij}(x)) \quad (10)$$

To simplify the calculations, equation(10) can be written into the equation(11).

$$m_i(x) = \min_{\mathrm{j\neq i,j=1,\ldots..n}} D_{ij}(x) \quad (11)$$

Furthermore, the data is classified based on the highest membership value according to the equation (12).

$$Data\ class\ x = \max_{\mathrm{i=1,\ldots..n}} m_i(x) \quad (12)$$

ECG Dataset

The dataset used in this study is the MIT-BIH arrhythmia database published by PhysioNet [10]. The source of the ECG recording consists of 4000 Holter recordings originating from Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. ECG randomly selected each from 48 the data recording ECG signals with a duration of 30 minutes. Examples of ECG signal representation can be seen in Figure 2.

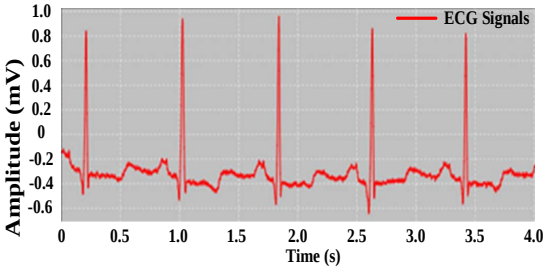


Figure 2. ECG signal representation

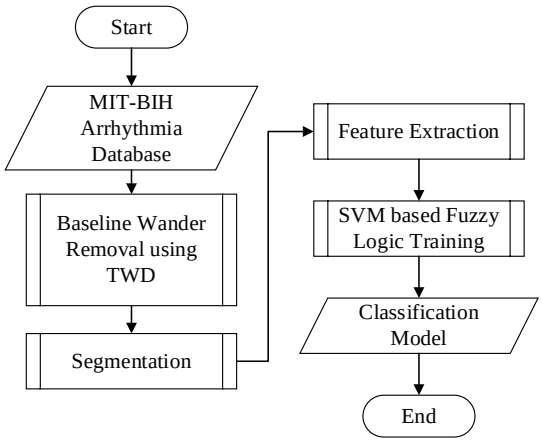


Figure 3. Training System Flowchart

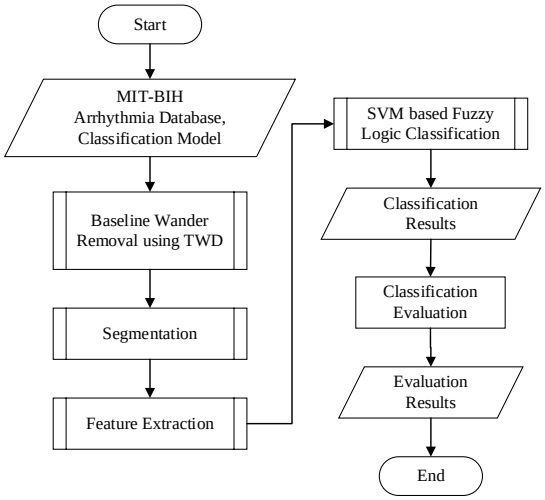


Figure 4. Testing System Flowchart

System Flowchart

System flowchart is divided into two parts, namely the training system flowchart, and testing system flowchart. Training system doing the learning from the training data to find a classi-fication model. The early stage of ECG data processing is the baseline wander removal process on the original ECG signal using Discrete Wavelet Transformation (TWD). The ECG signal is clean from the baseline wander will be segmented into units beat. The next stage is to find six features of the beat, and every single beat is classified using SVM method based fuzzy logic. Researchers added fuzzy membership function in SVM method to solve the problem of classified data when using Support Vector Machine (SVM) one-against-one. Figure 3 shows the training system flowchart.

The testing system created a prediction based on testing data processing to get the classification results, and continued by analyzing the classification results. Figure 4 shows the testing system flowchart.

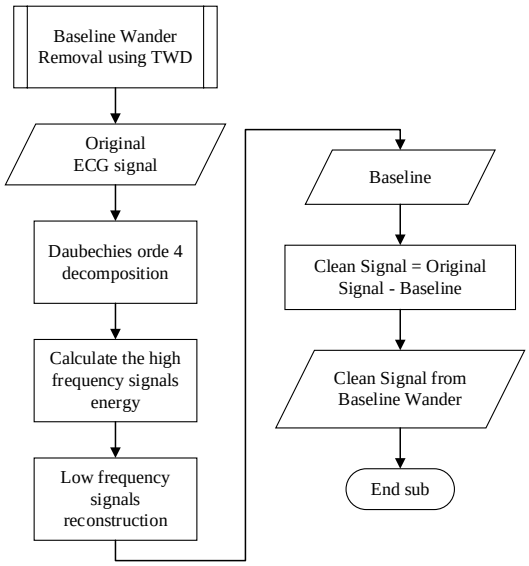


Figure 5. Baseline Wander Removal Flowchart

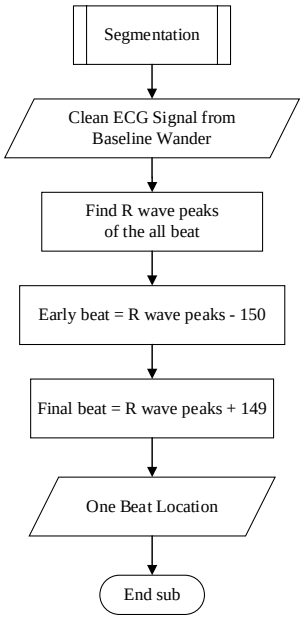


Figure 6. Beat Segmentation Flowchart

Baseline Wander Removal

Baseline Wander is a condition when the ECG signal is shifted up or down to the isoelectric line (line axis). To overcome the noise, the researchers conducted the Baseline Wander Removal using Transform Wavelet Discrete method [6].

The original signal is decomposed using the Transform Wavelet type Daubechies order 4. The system calculated the energy value at a high-frequency signal and found the conditions in which

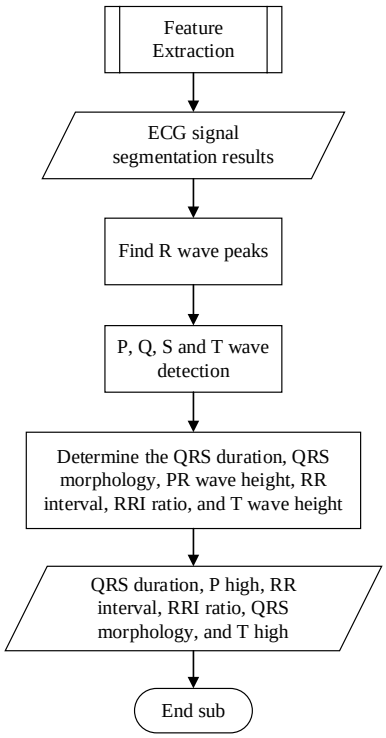


Figure 7. Feature Extraction Flowchart

the value of decomposition level is lower than the value of decomposition level before. Furthermore, the system performed approximation signal reconstruction of this level by removing the value of the high-frequency signal. Figure 5 shows the Baseline Wander Removal flowchart.

Beat Segmentation

In this stage, the ECG signals segmented into each beat. This process used additional annotation file to locate the position of the R-wave peaks. Assuming the width of each beat is done by positioning the peak of R as a pivot for each beat [8]. Early signals are sliced starting from the position of R-150 to R-149, so we get as many as 300 samples of beat data. Figure 6 shows the beat segmentation flowchart.

Feature Extraction

Feature extraction is the process of detection characteristics of the electrocardiogram that is used as an input variable in the classification process. The process consists of several steps as shown in Figure 7. This process begins by determining the R-wave peaks on the signal segmentation results, and then continued found for the wave that is P, Q, S and T. After all waves found, then continued to characterize the one beat of ECG signal. These include the QRS duration, QRS morphology, P wave height, RR interval, RRI ratio, and T wave height.

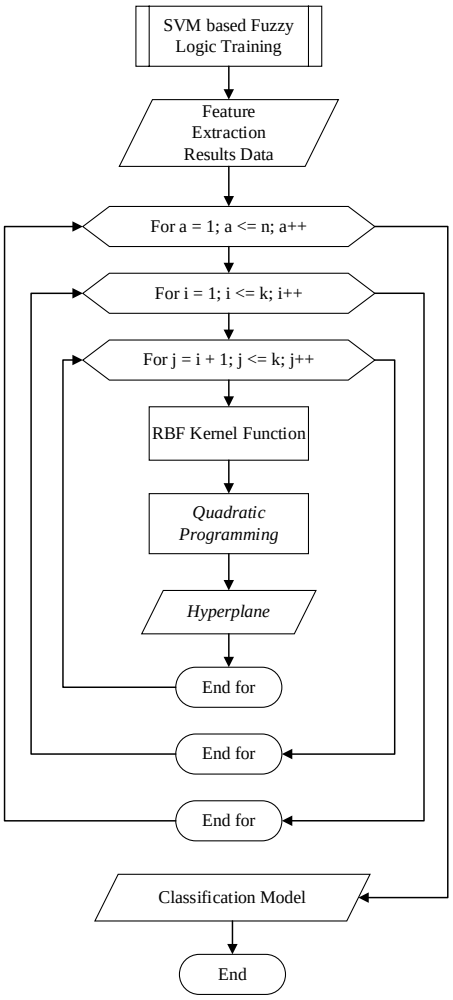


Figure 8. Training Process Flowchart

Training Process

SVM based fuzzy logic training process used the same training process with SVM one against one method, and build $k(k-1)/2$ parts of the binary classification model (k is the classes number). Each classification models is trained using data from two classes. For classroom training data to class i and class j, the system will solve the problems with the 6 classes by finding 15 separator functions as follows:

$D_{ij}(x) = D_{12}(x), D_{13}(x), D_{14}(x), D_{15}(x), D_{16}(x),$
$D_{23}(x), D_{24}(x), D_{25}(x), D_{26}(x), D_{34}(x), D_{35}(x),$
$D_{36}(x), D_{45}(x), D_{46}(x), D_{56}(x).$

So that the data can be separated linearly, this system uses Radial Basis Function (RBF) kernel $K(x,y) = exp(\frac{-1}{2\sigma^2}\|x-y\|^2)$. To obtain a solution or the separation of these two classes, this system using quadratic programming. Result of

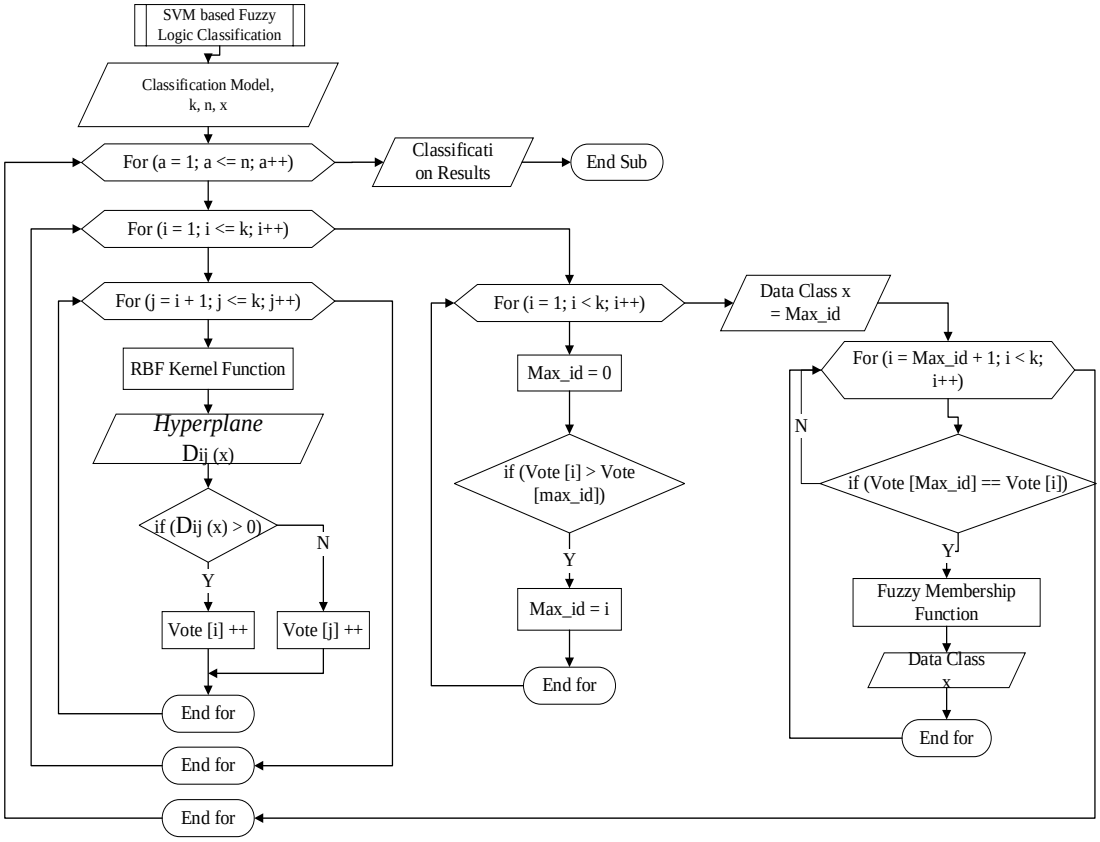


Figure 9. Classification Process Using SVM based fuzzy logic

quadratic programming are w, x, and B that used for the testing process. The flowchart for the stages of the training process can be seen in Figure 8.

Classification Process Using SVM based fuzzy logic

A common strategy in the testing process using one-against-one SVM method is max-voting [9]. Based on this strategy, if the representation of a data on the hyperplane $D_{ij}(x)$ is in class i, then voting added one for the class i. Otherwise, if the representation of a data on the hyperplane $D_{ij}(x)$ is in class j, then voting added one for the class j. These steps are repeated for all hyperplane. Then, system predicted x is in a class where, based on the value of the highest voting. In cases where there are two classes with the same voting value, then selected the smallest index value, the process is arguably less effective. To resolve with the data that is not classifiable, this study used the membership function according to the equation(11) using $m_i(x)$. The data x are classified based on the highest membership value according to the equation(12). Flowchart for the stages SVM based fuzzy logic method can be seen in Figure 9.

3. Results and Analysis

Evaluations of the classification results of the arrhythmia abnormalities consisted of two scenarios, among others, using the original SVM, and proposed method (SVM based fuzzy logic). This study used 600 testing data, which consisted of each 100 Normal Beat (NB) data, 100 Left Bundle Branch Block Beat (LBBB) data, 100 Right Bundle Branch Block Beat (RBBB) data, 100 Premature Ventricular Contraction (PVC) data, 100 Atrial Premature Beat (APB) data, and 100 Paced Beat (PB) data. Then the results are compared based on three measures among others accuracy, sensitivety, and specificity.

$$\text{Accuracy} = \frac{\text{number of data classified correctly}}{\text{number of tests performed}} \text{x} 100\% \quad (13)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{ x } 100\% \quad (14)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \text{ x } 100\% \quad (15)$$

where True Positive (TP) is arrhythmia correctly

TABLE 1
CLASSIFICATION RESULTS USING SVM METHOD

	Arrhythmia Type / Class					
	NB	LBBB	RBBB	PVC	PB	APB
TP	93	95	87	91	91	94
FN	7	5	13	9	9	6
FP	14	6	5	7	10	7
TN	486	494	495	493	490	493

identified, False Positive (FP) is arrhythmia wrongly identified, True Negative (TN) is arrhythmia correctly not identified, and False Negative (FN) is arrhythmia wrongly not identified.

Scenario 1: Classification using SVM Method [4]

Experimental results of classification using SVM method [4] can be seen in Table 1.

$$\text{Sensitivity of NB Class} = \frac{93}{93+7} \text{ x } 100\% = 93\%$$
$$\text{Sensitivity of LBBB Class} = \frac{95}{95+5} \text{ x } 100\% = 95\%$$
$$\text{Sensitivity of RBBB Class} = \frac{87}{87+13} \text{ x } 100\% = 87\%$$
$$\text{Sensitivity of PVC Class} = \frac{91}{91+9} \text{ x } 100\% = 91\%$$
$$\text{Sensitivity of PB Class} = \frac{91}{91+9} \text{ x } 100\% = 91\%$$
$$\text{Sensitivity of APB Class} = \frac{94}{94+6} \text{ x } 100\% = 94\%$$

Sensitivity values for each class are summed and divided by six, to obtain the average sensitivity value of 91.83%.

$$\text{Specificity of NB Class} = \frac{486}{14+486} \text{ x } 100\% = 97.2\%$$
$$\text{Specificity of LBBB Class} = \frac{494}{6+494} \text{ x } 100\% = 98.8\%$$
$$\text{Specificity of RBBB Class} = \frac{495}{5+495} \text{ x } 100\% = 99\%$$
$$\text{Specificity of PVC Class} = \frac{493}{7+493} \text{ x } 100\% = 98.6\%$$
$$\text{Specificity of PB Class} = \frac{490}{10+490} \text{ x } 100\% = 98\%$$
$$\text{Specificity of APB Class} = \frac{493}{7+493} \text{ x } 100\% = 98.6\%$$

The classification system using SVM method has average specificity value of 98.36%. From Table 1, we can calculate the overall accuracy value of 91.83%, the sensitivity average value of 91.83%, and the average specificity value of 98.36%.

Scenario 2: Classification using Proposed Method

Experimental results of classification using proposed method (Support Vector Machine based fuzzy logic method) can be seen in Table 2. Experiments scenario 2 uses the same dataset as that used in scenario 1.

TABLE 2
CLASSIFICATION RESULTS USING PROPOSED METHOD

	Arrhythmia Type / Class					
	NB	LBBB	RBBB	PVC	PB	APB
TP	95	95	93	91	93	94
FN	5	5	7	9	7	6
FP	10	6	5	7	4	7
TN	490	494	495	493	496	493

$$\text{Sensitivity of NB Class} = \frac{95}{95+5} \text{ x } 100\% = 95\%$$
$$\text{Sensitivity of LBBB Class} = \frac{95}{95+5} \text{ x } 100\% = 95\%$$
$$\text{Sensitivity of RBBB Class} = \frac{93}{93+7} \text{ x } 100\% = 93\%$$
$$\text{Sensitivity of PVC Class} = \frac{91}{91+9} \text{ x } 100\% = 91\%$$
$$\text{Sensitivity of PB Class} = \frac{93}{93+7} \text{ x } 100\% = 93\%$$
$$\text{Sensitivity of APB Class} = \frac{94}{96+4} \text{ x } 100\% = 94\%$$

Sensitivity values of classification using proposed method for each class are summed and divided by six, to obtain the average sensitivity value of 93.5%.

$$\text{Specificity of NB Class} = \frac{490}{10+490} \text{ x } 100\% = 98\%$$
$$\text{Specificity of LBBB Class} = \frac{494}{6+494} \text{ x } 100\% = 98.8\%$$
$$\text{Specificity of RBBB Class} = \frac{495}{5+495} \text{ x } 100\% = 99\%$$
$$\text{Specificity of PVC Class} = \frac{493}{7+493} \text{ x } 100\% = 98.6\%$$
$$\text{Specificity of PB Class} = \frac{496}{4+496} \text{ x } 100\% = 99.2\%$$
$$\text{Specificity of APB Class} = \frac{493}{7+493} \text{ x } 100\% = 98.6\%$$

The classification system of classification using proposed method has average specificity value of 98.7%. From Table 2, we can calculate the overall accuracy value of 93.5%, the average sensitivity value of 93.5%, and the average specificity value of 98.7%. These results indicate that the arrhythmia classification system uses proposed method deliver performance values of accuracy, sensitivity, and specificity higher than the arrhythmia classification system using the original SVM method. This is because the use of fuzzy membership functions can cope with the data that is not classifiable in one against one Support Vector Machine method.

4. Conclusion

From the results of experiments and calculations, it can be concluded that the use of support vector machine based on fuzzy logic methods for classification system abnormalities on ECG arrhythmia delivered performance with an accuracy level of 93.5%, the sensitivity of 93.5% and specificity of 98.7%. The result is higher when compared with the use of Support Vector Machine method with

an accuracy of 91.83%, sensitivity of 91.83%, and specificity of 98.36%. The use of fuzzy membership function can cope with data that are not classifiable in the one against one SVM method to deliver performance values of accuracy, sensitivity, and specificity higher than SVM method. This research is only using one lead of ECG signals. Some arrhythmia disorders common feature but is different leads, such as abnormalities in Left Bundle Branch Block Beat (LBBB), and Beat Right Bundle Branch Block (RBBB). Further research can be done using two leads to improve the accuracy of the arrhythmia classification system.

References

[1] A. Dallali, A. Kachouri, & M. Samet. “Fuzzy c-means clustering, Neural Network, WT, and HRV for classification of cardiac arrhythmia”, ARPN Journal of Engineering and Applied Sciences. Vol. 6(10). 2011.

[2] Arrhythmia Rules. Biomedical Telemetry. http://www.biotel.ws/ekgs/TheRules.htm accessed on 12 May 2014.

[3] P. Srivastava, N. Sharma, & C.S. Aparna. Fuzzy Soft System and Arrhythmia Classification. Hindawi. India. 2013.

[4] M. Moavenian, & H. Khorrami. “A qualitative comparison of Artificial Neural Networks and Support Vector Machines in ECG arrhythmias classification”. Elsevier. 3088-3093. 2010.

[5] O. Castilo, P. Melin, E. Ramirez, & J. Soria. “Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbours and neural networks combined with a fuzzy system”. Elsevier. 2947-2955. 2012.

[6] D. Sripathi. Efficient Implementations of Discrete Wavelet Transform Using FPGAs. Florida State University. Florida. 2003.

[7] A.K.M.F. Haque, H. Ali, M. A. Kiber, & Md. T. Hasan. “Detection of Small Variations of ECG Features Using Wavelet”, ARPN Journal of Engineering and Applied Sciences. 4(6):27-30. 2009.

[8] B. Anuradha, & V.C. Veera Reddy. “ANN for Classification of Cardiac Arrhythmias”, ARPN Journal of Engineering and Applied Sciences. 3(3):1-6. 2008.

[9] A. Dallali, A. Kachouri, & M. Samet. “Integration Of HRV, WT and Neural Networks for ECG Arrhythmia Classification”, ARPN Journal of Engineering and Applied Sciences. 06/2011; 6(5). 2011.

[10] "MIT-BIH Arrhythmia database," 2015. [Online]. Available:http://www.physionet.org/physiobank/database/mitdb/. Accessed on 29 May 2014.

[11] K. J. Reza, S. Khatun, M. F. Jamlos, M. M. Fakir, & S. S. Mostafa. “Performance evaluation of diversified SVM kernel functions for breast tumour early prognosis”, ARPN Journal of Engineering and Applied Sciences. 03/2014; 9(3):329-335. 2014.

[12] T. Kitamura, S. Abe, & K. Fukui. Subspace Based Least Squares Support Vector Machines For Pattern Classification. IEEE International Joint Conference on Neural Networks, 2009.

Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/1 (2016), 43-51
DOI: http://dx.doi.org/10.21609/jiki.v9i1.365

PERFORMANCE COMPARISON OF USART COMMUNICATION BETWEEN REAL TIME OPERATING SYSTEM AND NATIVE INTERRUPT

Novian Habibie[1], Machmud Roby Alhamidi[1], Dwi M J Purnomo[1], and Muhammad Febrian Rachmadi[2]

[1]Faculty of Computer Science, Universitas Indonesia, Kampus Baru UI, Depok, 16424, Indonesia
[2]School of Informatics, The University of Edinburgh, 11 Crichton Street, Edinburgh EH8 9LE, United Kingdom

E-mail: novian.habibie@ui.ac.id[1], s1467961@sms.ed.ac.uk[2]

Abstract

Comunication between microcontrollers is one of the crucial point in embedded sytems. On the other hand, embedded system must be able to run many parallel task simultaneously. To handle this, we need a reliabe system that can do a multitasking without decreasing every task's performance. The most widely used methods for multitasking in embedded systems are using Interrupt Service Routine (ISR) or using Real Time Operating System (RTOS). This research compared perfomance of USART communication on system with RTOS to a system that use interrupt. Experiments run on two identical development board XMega A3BU-Xplained which used intenal sensor (light and temperature) and used servo as external component. Perfomance comparison done by counting ping time (elapsing time to transmit data and get a reply as a mark that data has been received) and compare it. This experiments divided into two scenarios: (1) system loaded with many tasks, (2) system loaded with few tasks. Result of the experiments show that communication will be faster if system only loaded with few tasks. System with RTOS has won from interrupt in case (1), but lose to interrupt in case (2).

Keywords: *embedded system RTOS, interrupt, USART, performance analysis*

Abstrak

Komunikasi antar mikrokontroller adalah salah satu hal krusial dalam sebuah *embedded system*. Di sisi lain, *embedded system* juga harus dapat menangani beberapa *task/*pekerjaan dalam satu waktu. Untuk itu, diperlukan sebuah sistem yang dapat melaksanakan proses *multitasking* tanpa mengganggu performa dari masing-masing *task* yang ada. Ada dua metode *multitasking* yang populer digunakan pada *embedded system,* yaitu menggunakan *Interrupt Service Routine* (ISR) dan menggunakan *Real Time Operating System* (RTOS). Penelitian ini membandingkan performa komunikasi USART pada mikrokontroller dengan RTOS dengan yang hanya menggunakan *interrupt*. Uji coba dilakukan pada dua *development board* XMega A3BU-Xplained dengan sensor internal (cahaya dan temperatur) dan menjalankan sebuah servo. Uji performa dilakukan dengan menghitung waktu *ping*, yaitu waktu yang dibutuhkan untuk mengirim satu karakter data ke *board* tujuan dan menerima balasan satu karakter sebagai tanda bahwa data telah diterima oleh *board* tujuan. Skenario yang digunakan adalah (1) sistem memiliki banyak task, dan (2) saat sisem memiliki sedikit *task*. Berdasarkan eksperimen yang dilakukan, secara umum proses komunikasi akan berjalan lebih cepat jika sistem hanya mempunyai sedikit *task*. Sistem dengan RTOS akan memiliki waktu *ping* yang jauh lebih cepat dari yang menggunakan *interrupt* pada kasus (1), namun sistem dengan *interrupt* akan lebih cepat dari sistem dengan RTOS pada kasus (2).

Kata Kunci: *embedded system, RTOS, interrupt, USART, analisa performa*

1. Introduction

Nowadays the usage of embedded systems are widely spread in every aspects of our life. It is because embedded systems are the right solution to implant an automatic behaviour or responses into physical world which is small, low-powered, and specific to one dedicated purpose. Implementation of embedded system are everywhere, start from daily utensils like refrigerator, television, calculator, until many device that runs daily life like traffic light, automatic gate in the railstation, etc.

Although one embedded systems can be only dedi-cated to specific purposes, its purposes itself may contain some tasks. Because of that, one of a capability that embedded system must have is an ability to handle multiple tasks without fail. To do that, the system that can handle parallel computation in small and low computational ability is urgently needed. One of the solution by using Inter-

rupt Service Routine (ISR) to run tasks as a process that interrupting its main program simultaneusly. Another solution is Real Time Operating System (RTOS), a tiny operating system that can fit into embedded system's device and run processes as tasks that run in specific time slice. Either interrupt or RTOS have its own advantages and disadvantages, depends on needs of the system.

Another crucial point for embedded system is an ability to communicate and exchange data from one microcontroller to another. With communicating each other, the functionality of the system can be increased and can be made as a wider system that can coordinate each other. There is many methods to do a communication between device, one of them is communication using USART (Universal Synchronous Asynchronous Receiver Transmitter) method.

Communication between system usually followed with another tasks that run in parallel, e.g. sensor reading or move an actuator. However, sometimes comunication process disturbed by another task, so in result communication process can be slower than expected and/or occurred an error that cause data loss. To overcome that, the system that capable to run parallel processes without reduce any process's performance is needed.

Several research about embedded system and RTOS have been conducted in recent years. The latest one, Manju Nanda et al. [1] conducted research about qualifying RTOS for use in safety critical applications using formal methods due to effectiveness and preciseness. Manju Nanda et al [4] provides guidelines for development and implementation of formal approach to qualify a Commercially off the Shelf (COTS) RTOS as per the civil aerospace standard RTCA DO-178C.

Yonghyun Hwan et al. [2] present an accurate timed RTOS model within transaction level models (TLMs). There are two key features used in this research. The RTOS behavior model provides dynamic scheduling, IPC (inter-process communication), and external communication for timing annotated user applications. The RTOS overhead model has processor to specific pre-characterized overhead information to provide cycle approximate estimation. To demonstrate the model, Yonhyun Hwan et al [2] used a multicore platform executed a JPEG encoder and provide results that RTOS model present high accuracy.

Su-Lim TAN and Tran Nguyen Bao Anh [3] present a research about RTOS for small microcontroller. Su-Lim TAN and Tran Nguyen Bao Anh [3] used 16 bit microcontroller to perform RTOS multitasking. To demonstrate the ease of RTOS platform migration, the mTKernel RTOS is chosen for porting to the H8S/2377 16-bit microcontroller.

Ji Chan Maeng et al [4] present a research about produce an RTOS specific code using an automated tool and model-driven approach embedded software development. Generic RTOS APIs was defined to capture most of typical RTOS ser-vices and for describing application's RTOS related behaviour at an early stage. Generic RTOS APIs have been transformed into RTOS specific APIs using an automated transformation tool.

Fabiano Hessel [5] present a research about abstract RTOS model that allows refining the abstract model to an implementation model at lower abstraction levels. Fabiano Hessel [5] used C language with some extension to build the model and as a results fifty task with four priority levels shows the usefulness of this model.

Our previous research [6], do a comparison between RTOS and Interrupt using ultrasonic sensor and rack movement mechanism where it will move from distance 0 cm to 10 cm repeatedly. Defined threshold at a certain distance that is beginning, middle and end. Ultrasonic sensor will indicated "HIT" if distance between assigned threshold and rack movement is equal. Result shows that RTOS has higher accuracy performance than Interrupt butlesser precision.

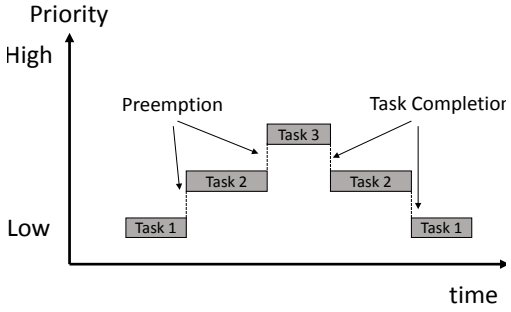


Figure 2. Schematic of RTOS scheduling.

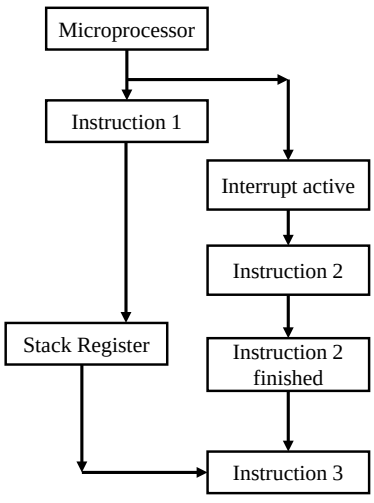


Figure 1. Schematic of interrupt

In this research, the peformance comparison between RTOS and native interrupt will be investigated in the case of serial communication between microcontroller. This research will test performance of USART communication while undergoing another some other tasks.

The rest of this paper is organized as follows. Section 2 describe the methodology used in this research. Section 3 presents results and analysis of experiment. Finally, section 4 presents conclusions of this paper.

2. Methods

This research focused on testing the performance of USART communication on various multitasking environment. Performance in this research measured by two aspects: (1) communication speed, and (2) communication reliability. Experiments conducted on two multitasking system which connected each other with USART communication. Each board have an identic environment and specifications, either hardware or software.

Communication speed can be measured by obtaining data of amount of elapsed communication time. This aspect tested by conduct "ping" process and count its elapsed time. Similar to ping in networking [7], ping is a process to check a reachability of destination device. Ping conducted by sending a packet of data to destination device and get a reply data as a sign that the data has been received. For experiment on this research, ping conducted as character sending and receiving process. Ping function transmit character 't' to destination board, and the destination board will reply with character 'r'. Elapsed time obtained by count the time differences between data sending and receiving process. For each experiment, ping conducted several times and the mean of ping time become the result. Detail of experiment process explained in sub-section 2.2, on experiment scenarios subsection.

Reliability of communication conducted to see how much error occurred when USART runs on various condition. To see the system's error, experiment still use the same process as ping does, but it now focused on the amount of data that transmiited and received. As explained above, each experiment conducted ping several times. To check reliability of the system, this experiment will count the differences between the amount of data which is received and retransmitted. This aspect obtained in each board separately. The amount of received ping accumulated on the destination board, and the amount of obtained reply data that comes from destination board accumulated in source board. The amount of differences between received ping and received reply become amount of packet that loss on ping process. From amount of loss packet, reliability of the system can be measured and analyzed. Detail of this aspect explained in section 4, subsection experiment scenarios.

RTOS and Native Interrupt

RTOS is a new approach as an alternative of interrupt in microcontroller world. Its capability of undertaking multitask performance better than native interrupt has become an attraction for many of researchers. RTOS eminences comprise flexibility of architecture can be empowered, reliable for many tasks, actively developed, simple, and many others [8].

The main difference between RTOS and native interrupt is illustrated in Figure 1 and Figure 2. Figure 1 shows the task management in RTOS. If there is task whose higher priority, the lower priority task will be suspended and the high priority task will be executed. Whereas in native interrupt, the task which is inside the interrupt cannot be interrupted by another task before the previous task finished (Figure 2).

The utilization of RTOS (in this research free-RTOS was employed) due to its capability of multi-tasking. Multitasking is highly related to setting the priorities of tasks. In the freeRTOS the priorities of tasks can be determined by utilizing FreeRTOSConfig.h. The aforementioned priorities are ranging from 0 to (`configMAX_PRIORITIES`-1) [9]. The value of (`configMAX_PRIORITIES`-1) could be defined freely, as long as it does not exceed the RAM capacity. However, if the chosen value is 1 (one) for `configUSE_PORT_OPTIMISED_TASK_SELECTION` in the `FreeRTOSConfig.h`, the value of `configMAX_PRIORITIES` is limited to 32. Whereas the task whose priority of 0 is called `tskIDLE_PRIORITY.`

In this research, native interrupt was also empowered to be compared with RTOS. The priorities of the above mentioned interrupt are listed in the microcontroller datasheet [10]. To cite an instance, `RESET` whose the highest priority. To assign the interrupt, Interrupt Service Routine (ISR) must be written in the source code. Further-more, to activete the global interrupt in order to make the interrupt executed, macro `sei()` was assigned.

Interrupt employed to the system by activate the timer interrupt feature. Timer interrupt will interrupt main program with specific task in a specific time slice. Timer interrupt will execute the task after the timer is overflow. The amount of tick required (TC) to make timer overflow can be determined by put a value to the Timer Counter Register (TCCR) from equation(1) as follows:

$$\boldsymbol{TC = \frac{t_{req}}{T_{CPU}} - 1} \qquad (1)$$

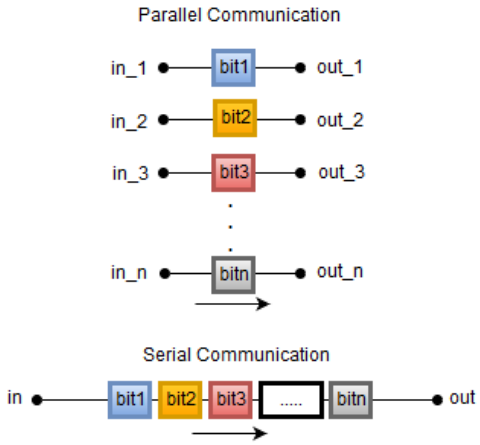


Figure 3. Illustration of comparison between parallel and serial communication

TABLE 1
SPECIFICATION OF ATXMEGA256A3BU-AU

Specification	Value
Flash	256KB + 8KB
EEPROM	4KB
SRAM	16KB
Max Speed	32 MHz
Power Supply	1.6-3.6 V

TABLE 2
PORT CONNECTION

Component/Feature	Chip Port / Board Port
USART Transmit Port (TX)	J1-PIN3 / PC3
USART Receive Port (RX)	J1-PIN2 / PC2
Servo Motor Data	J1-PIN0 / PC0
GPIO to Arduino	J1-PIN4-5 / PC4-5

Which t_{req} is actual amount of time required (in second), and $T_{CPU}$is clock time period of the microcontroller. $T_{CPU} = prescaler/f\_CPU$, where f_{CPU} is clock of microcontroller and prescaler is divider of real frequency depends on how long timer defined. Prescaler defined as a value of the power of two, e.g. 1,2,4,8, etc.

USART

Communication between devices in embedded systems have two different forms, parallel and serial communication. Parallel communication employs transmitting and receiving data via multiple GPIO port, which each port represent one bit of data. On the other hand, serial communication transmit or receive data in one port only, and data transmitted sequentially in a form of data packet. Each parallel or serial communication have its own advantages and disadvantages. Parallel communication can have a faster speed because it can transmit/receive multiple data at one time because it use multiple ports. However, the usage of multiple port itself became its disadvantages because it's too costly and give extra complexity to the system. Nowadays, serial communication is commonly used because it use less port and transmit data in a form of packet. Illustration of differences between parallel and serial communication can be seen on Figure 3.

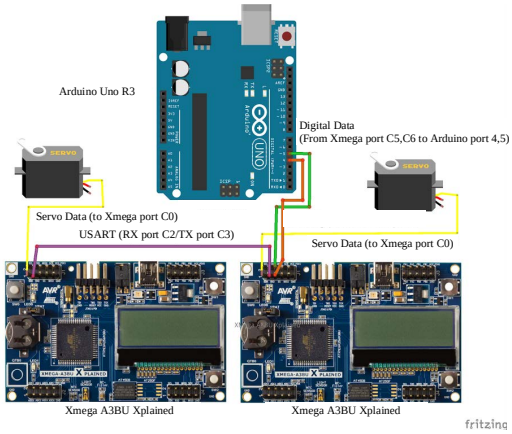


Figure 4. Schematic of system board images via atmel.com

TABLE 3
TASK DESCRIPTION

Task Name	Description	Executed Every
LCD	Print system's status and data from sensor	2 s
Servo Motor	Repeatly move servo's shaft left and right	1 s
Temperature Sensor	Read temperature from environment	300 ms
USART: Receive	Standby to receive data form another board	~1 ms
USART: Transmit	Repeatly send data to another board	~1 ms

Universal Synchronous Asynchronous Receiver Transmitter (USART) is one of a serial communication protocol. In general it uses two ports, one for transmitting data (TX port), and another one is for receiving data (RX port). If necessary, it can use one additional port as a clock for synchronous communication. USART can be activated via USART register in microcontroller. USART have three modes, asynchronous normal mode, asynchronous double speed, and synchronous mode.

Similar to interrupt, performance of USART depends on system clock and baud rate. System clock can be defined based on specification of microcontroller used on the system. Baud rate is a term of how many data/symbol can be transmitted in one time, which one symbol can contain more than one bit.

If N is an amount of bits in one symbol, required symbol to be sent is $S = 2^N$. Baud rate can

be converted as bit rate by counting $\boldsymbol{R = baud\ rate\ x\ log_2 S}$. Correlation of system clock and baud rate, known as BSEL, used as input value to UBBR register. For example, the value for UBBR in USART asynchronous normal mode can be obtained from equation(2).

$$\boldsymbol{BSEL = \frac{f_{PER}}{2^{BSCALE}.16 f_{BAUD}} - 1} \qquad (2)$$

Which $\boldsymbol{f_{PER}}$ is clock of microcontroller, $\boldsymbol{BSCALE}$ is parameter to tune $\boldsymbol{BSEL}$ to make it as close as its real value, and $\boldsymbol{f_{BAUD}}$ is a desired value of baud rate. Common used baud rate are 4800, 9600, 19200, and so on.

Experiment Environment

Hardware

Experiments run on development board Xmega A3BU-Xplained, which use microcontroller ATXmega A256A3BU manufactured by Atmel. Full specification of the board can be seen on Table 1. Internal components from the development board such as button, LED, monochrome LCD, and internal sensor (light sensor and temperature sensor) have been used as the components to simulate the multitasking system. Experiments use two identical development board powered with USB cable and connected each other with USART communication, which use port Rx (receiver) and Tx (transmitter). For external components, servo motor is connected to each board in port SDA to use PWM feature from the microcontroller. To count ping time from the system, one of the development board connected to Arduino Uno board via GPIO. Arduino will connected to the destkop computer via USB serial. Arduino Uno acts as a timer to count elapsed time for ping time
from XMega board. For further hardware details, schematic diagram is on Figure 3, and external components port connection table on X Mega board is on Table 2.

Software

Program that used in this research for XMega board was developed on Atmel Studio 7 IDE, MinGW C compiler, and firmware downloader FLIP from atmel. The program use Atmel Software Framewok (ASF) library as main library to use various features of development board Xmega A3BU-Xplained. For Arduino Uno board, the program developed on Arduino IDE with standard Arduino Library.

RTOS which is used in this research is FreeRTOS, an open source RTOS for various microcontroller. Raw FreeRTOS source code obtained from its official website, *http://freertos .org*. Raw source code has been configured to be compatible with Xmega board and only use a required features. However, current FreeRTOS version is not compatible to Xmega-family microcontroller yet. To overcome this, the additional configuration from [11] has been used in configur-ation file of Free-RTOS. Timer counter used as data logger in arduino was Timer1 library from Arduino Library[12]. Timer1 provide library for timer counter in mili-second up to two decimal places.

Multitasking Configurations

This reseach conduct two type of multi-tasking system: (1) system with primitive Interrupt Service Routine (ISR) and (2) system with RTOS. Both system loaded with five parallel tasks : LCD display, servo, temperature sensor, USART receive process, and USART transmit process. Each of them scheduled in specific time. Detail of parallel tasks described in Table III.

This experiment use interrupt library from Atmel Software Framework (ASF): Programmable Multi-level Interrupt Controller (PMIC) module, specifically use timer interrupt. Based on feature of XMegaA256ABU, microcontroller used for this experiment provide four timer/counter register: C, D,E,F with each of them have two channel, channel 0 and 1. For this experiment, timer counter used for parallel task are C1 (USART: transmit), D0 (LCD), D1 (Temperature Sensor), E0(servo), E1 (USART: receive).

Configuration used in this research followed the standard of FreeRTOS, which configuration of system is defined in `FreeRTOSConfig.h` file and when the task is created via the `xTask-Create()` function. For this research, every task is configured with the identical settings. Each task assigned to the same priority (priority 0) to assure they have the same amount of time slice. Moreover, each task have the same depth of stack, 500.

Software Interface Configurations

System used for experiment have sensor as input simulation and display and actuator as output simulation. For software driver, system use ASF modules to simpify implementation process. Internal temperature sensor from XMega board used to simulate input via Analog to Digital Converter (ADC) module, which connected to ADC register A. System also use GFX Monochrome module to print data from system. PWM used in this system run via direct register access on register C0.

USART Configurations

USART for this experiment also use ASF's implementation, which need some parameters to specify its feature. This experiment use two values for baud rate, 4800 and 9600, which will be explained in experiment scenario. As the system would transmit

and receive data in character form, capacity of data for each packet is 8-bit length. USART for this experiment not use any parity bit and stop bit.

Experiment Scenarios

The experiments are divided into two main scenarios: (1) count and compare elapsed time of USART communication between two board, and (2) compared USART's reliability by comparing transmited and received data in the destination board. Scenario (1) itself also have two sub-scenarios: count time when system is (1.a) heavy-loaded (run many tasks) and (1.b) light-loaded (only run fewer tasks than first scenario). Scenario (2) has two components: (2.a) check differences of on amount of sent data with amount of received data and (2.b) send and receive string as sequence of characters. Table 4 shows scenario conducted in this research. Table 5 shows experimental parameter used in the scenario.

In details, scenario (1) counts ping time. For each experiment, ping is performed 100 times. Ping time obtained from total 100 ping time divided with 100 as a mean time. For scenario (1), both systems with RTOS and interrupt use same process and same amount of tasks. For sub-scenario (1.a), system is loaded with 5 different tasks: LCD, button, light sensor, temperature sensor, and servo, and in subscenario (1.b), system is only loaded with button and LCD with minimum display. See Table III for detail of task descriptions.

Scenario (2) will test data transmission and reception realibility. As same as scenario (1), each sub-scenario, (2.a) and (2.b) tested with three configurations: heavy-loaded, light-loaded and hybrid (heavy-light) loaded. Scenario (2.a) counts and compares amount of data received on destination board respect to replied data received on sender board. Scenario (2.b) tests data consistency by sending string as sequence of characters. Received string on destination board will be compared to sent string on source board to find whether any error or not.

TABLE 4
EXPERIMENTAL SCENARIO

Scenario/ Parameter	Baud Rate	Task Load
Ping Time	V	V
Data Loss	V	V

TABLE 5
EXPERIMENTAL PARAMETER

Baud Rate / Loaded	Hybrid Loaded	Heavy Loaded	Light Loaded
4800	V	V	V
9600	V	V	V

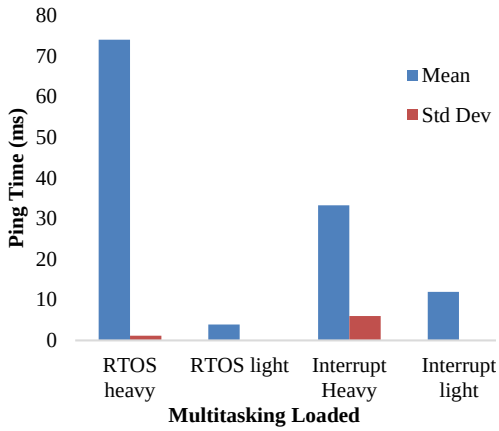


Figure 5. Ping time comparison between RTOS and Interrupt in Baud Rate: 4800

3. Results and Analysis

Figure 5 shows the ping time comparison between RTOS and Interrupt if multitasking is run in heavy loaded system or light loaded system in baud rate 4800. The RTOS in light loaded outperformed the Interrupt. On the contrary, the Inter-rupt shows better performance in heavy loaded task. In RTOS, multitasking will be done within the specified time, which means that each task will be done when the specified time arrives. On the other hand, at the Interrupt task will be done by interrupting main process.

Results from the experiments show that average ping time for the RTOS in heavy loaded is 74.011 ms and the Interrupt in the same configuration is 33.249 ms. In light loaded experiment, the average ping time for the RTOS is 3.912 ms and the Interrupt is 11.943 ms.

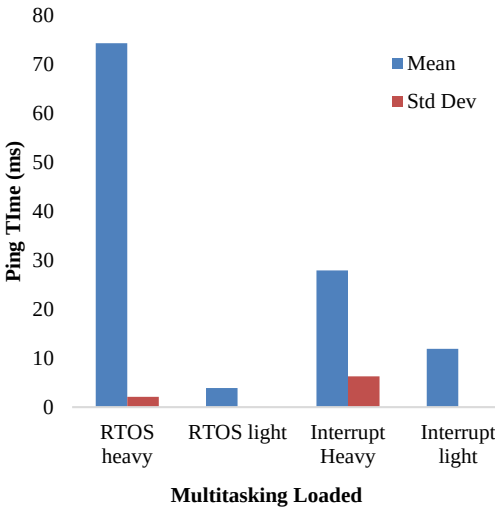


Figure 6. Ping time comparison between RTOS and Interrupt in Baud Rate: 9600

TABLE 6
PING TIME SCENARIO RESULTS FOR HYBRID LOADED

Baud Rate	RTOS				Interrupt			
	Light x Heavy		Heavy x Light		Light x Heavy		Heavy x Light	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
4800	68.045	1.7005	69.578	1.43418	17.008	3.23858	18.042	2.93538
9600	67.918	1.90209	68.645	1.77053	16.842	3.29598	16.654	3.40705

TABLE 7
DATA LOSS SCENARIO RESULTS FOR HYBRID LOADED

Baud Rate	RTOS				Interrupt			
	Light x Heavy		Heavy x Light		Light x Heavy		Heavy x Light	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
4800	0	0	1261.5	13.2686	3.6	0.96609	3.5	1.08012
9600	0	0	1254.2	11.2822	3.5	0.84984	3.7	0.67495

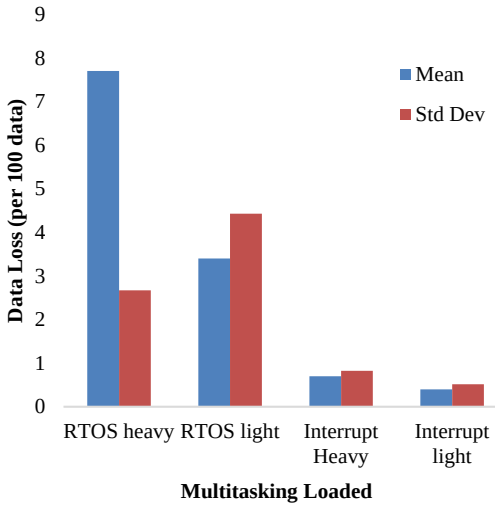


Figure 7. Data loss comparison between RTOS and Interrupt in Baud Rate: 4800

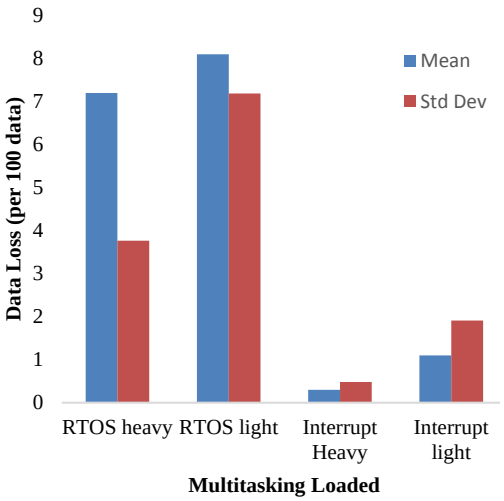


Figure 8. Data loss comparison between RTOS and Interrupt in Baud Rate: 9600

Figure 6 shows the ping time comparison between RTOS and Interrupt if multitasking is run in heavy loaded system or light loaded system in baud rate 9600. Similar to previous experiment which used baud rate 4800, RTOS gives slower communication than interrupt in heavy-loaded task but runs faster in light-loaded system (RTOS heavy-loaded = 74.357 ms; light-loaded = 3.884ms, Interrupt heavy-loaded = 27.923 ms; light-loaded = 11.91 ms).

From experiments above, we can infer that communication's performance in RTOS depends on how many tasks loaded into the system. Figure 7 shows data loss comparison between RTOS and Interrupt if multitasking is run in heavy loaded system or light loaded system in baud rate 4800. The Interrupt both in heavy and light loaded systems shows better performance than the RTOS. This is due to multitasking in RTOS which would be done within the specified time. It means that each task would be done when the specified time arrives. It could cause a loss of data when the task has not yet completed but had moved on to another task. On the other hand, a task will be done by interrupting another task and do not swich to another task until that task was completed in the Interrupt.

From the experiment, we show that average data loss for the RTOS in heavy loaded is 7.7% data whereas the Interrupt in the same loaded is only 0.7% data. In light loaded experiment, the average data loss for the RTOS is 3.4% data and the Interrupt is 0.4% data.

Figure 8 shows data loss comparison between RTOS and Interrupt if the multitasking is run in heavy-loaded system or light loaded syste, in baud rate 9600. The Interrupt in heavy or light loaded experiment shows better performance than RTOS. Results the experiments shows that average data

loss for the RTOS in heavy-loaded is 7.2% data whereas the Interrupt in the same loaded has data loss of 0.3% of data. In light loaded experiment, the average data loss for the RTOS is 8.1% data and the Interrupt is 1.1% data.

Table 6 shows experiment results of ping time scenario in hybrid loaded task. Ping time results show that light loaded task combined with heavy loaded task in interrupt method is the best in all baud rate. In details, if transmitter is light-loaded and the receiver is heavy-loaded then the ping time is faster than in reverse configuration. This result occurred because communication perf-ormance follows the heaviest side of the system, in this case is heavy-loaded side. Because of that, performance in hybrid configuration is almost similar for each configuration. However, just like the previous scenario, RTOS in heavy-loaded conf-iguration gives the worst result. From these results, we can say that both cases is not good for RTOS. However, RTOS in heavy loaded task combined with light loaded task gives the best standard deviation in all baud rate.

Table VII shows experiment results of data loss scenario in hybrid loaded task. The results show us that RTOS in light loaded combined with heavy loaded task gives the best performance with no loss of data. In contrary, RTOS in heavy loaded combined with light loaded task gives the worst performance with more than 1000 data loss. Received data from RTOS with heavy loaded task will be responded quickly in RTOS with light loaded task due to there is no other task interfere. Because of that, each response in RTOS with light-loaded configuration will be counted as data res-ponse and it will made data loss in the system. In contrary, transmitted data from RTOS with light loaded task will be responded slowly in RTOS with heavy loaded task due to there are many tasks pro-cessed in the system. So, RTOS with light loaded task will be waiting data response from RTOS with heavy loaded task and it made no loss of data. From Table 7, the Interrupt shows stable performan-ce with just 3.5 data loss in every baud rate and light-heavy load hybrid systems.

4. Conclusion

Inter-microcontroller communication is one of the crucial aspects in embedded systems. To do fast and reliable communication, the system have to manage its resources and do a simultaneus processes without interfering another tasks. Methods that can be used to manage the microcontroller's resources are Interrupt Service Routine (ISR) and Real Time Operating System. Interrupt and RTOS have different system workflows, so they have their own advantages and disadvantages.

From experiments conducted in this research, the results show that interrupt and RTOS give a competitive performance, either in communication speed and data reliability. In details, interrupt give better result in speed and data reliability than RTOS if loaded with many tasks. However, if the task load is minimum, RTOS give the best result in term of speed but still lose in data reliability. It is because each task in RTOS assigned its resource to the main CPU, so If the system contains combined task load (heavy-loaded board connected with light-loaded board), interrupt is the most stable system form speed and reliability. Specific in RTOS, data loss of the system which placed heavy system as transmitter will give the worst result, but in vice versa it give the best data accuracy.

In the end, speed and reliability of multi-tasking system to conduct inter-microcontroller communication depends on the task load of the system. Interrupt give better if the system want to focus on communication. However, interrupt only can handle a small amount of tasks because the limited amount of available timer counter register. If the main purpose of the system is to run a large amount of tasks, then RTOS is recommended.

Acknowledgement

This work was supported by Directorate Research of Universitas Indonesia funding in 2015. The title of the research is Laboratory Infra-structure. This grant number is 1831/UN2.R12/ HKP.05.00/2015.

References

[1] M. Nanda, S. Dhagem and J.Jayathi, "An Approach to Formally Qualify Commercial RTOS for Safety Application," in *International Conference on Computing for Sustainable Global Development*, 2015.

[2] Y. Hwang, G. Schirner, S. Abdi and D. G. Gajski, "Accurate Timed RTOS Model for Transaction Level Modeling," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2010.

[3] S.L. TAN and T. N. B. Anh, "Real-time operating system (RTOS) for small (16-bit) microcontroller," in *The 13th IEEE International Symposium on Consumer Electronics (ISCE 2009)*, 2009.

[4] J. C. Maeng, J.-H. Kim and M. Ryu, "An RTOS API Translator for Model-driven Embedded Software Development," in *the 12th IEEE International Conference on Embedded and Real-Time Computing System and Applications*, 2006.

[5] F. Hessel, V. M. d. Rosa, I. M. Reis, C. A. M.

Marcon and A. A. Susin, "Abstract RTOS Modelling for Embedded Systems," in *the 15th IEEE International Workshop on Rapid System Prototyping*, 2004.

[6] D. M. Purnomo, M. R. Alhamidi, G. Jati, N. Habibie, B. Hardjono and A. Wibisono, "Comparative Study of RTOS And Primitive Interrupt In Embedded System," *Jurnal Ilmu Komputer dan Informasi Volumne 8 Issue 1*, pp. 36-45, 2015.

[7] T. Inc, "Techopedia," 2016. [Online]. Available: https://www.techopedia.com/definition/2452/ping.

[8] R. Barry and R. T. E. Ltd, "Free RTOS," 2016. [Online]. Available: http://www.freertos .org.

[9] R. Barry and R. T. E. Ltd, "RTOS Task Priority," 2016. [Online]. Available: http:// www.freertos.org/RTOS-task-priority.html.

[10] A. Corp, "Atmel 8155D AVR Atmega 32A Datasheet," Atmel, 2014.

[11] Yuriykulikov, "FreeRTOS On XMEGA," 2012. [Online]. Available: https://github.com/yuriykulikov/FreeRTOS-on-XMEGA.

[12] Arduino, "Timer," 2016. [Online]. Available: http://playground.arduino.cc/Code/Timer1.

Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/1 (2016), 52-58
DOI: http://dx.doi.org/10.21609/jiki.v9i1.366

PARTICLE SWARM OPTIMIZATION (PSO) FOR TRAINING OPTIMIZATION ON CONVOLUTIONAL NEURAL NETWORK (CNN)

Arie Rachmad Syulistyo[1], **Dwi M J Purnomo**[1], **Muhammad Febrian Rachmadi**[2], **and Adi Wibowo**[3]

[1] Faculty of Computer Science, Universitas Indonesia, Kampus Baru UI, Depok, 16424, Indonesia
[2] School of Informatics, The University of Edinburgh, 11 Crichton Street, Edinburgh EH8 9LE, United Kingdom
[3] Department Micro-Nano System Engineering, Graduate School of Engineering, Nagoya University, 1 Furocho, Chickusa Ward, 464-8603 Japan

E-mail: arie.rachmad@ui.ac.id

Abstract

Neural network attracts plenty of researchers lately. Substantial number of renowned universities have developed neural network for various both academically and industrially applications. Neural network shows considerable performance on various purposes. Nevertheless, for complex applications, neural network's accuracy significantly deteriorates. To tackle the aforementioned drawback, lot of researches had been undertaken on the improvement of the standard neural network. One of the most promising modifications on standard neural network for complex applications is deep learning method. In this paper, we proposed the utilization of Particle Swarm Optimization (PSO) in Convolutional Neural Networks (CNNs), which is one of the basic methods in deep learning. The use of PSO on the training process aims to optimize the results of the solution vectors on CNN in order to improve the recognition accuracy. The data used in this research is handwritten digit from MNIST. The experiments exhibited that the accuracy can be attained in 4 epoch is 95.08%. This result was better than the conventional CNN and DBN. The execution time was also almost similar to the conventional CNN. Therefore, the proposed method was a promising method.

Keywords: *deep learning, convolutional neural network, particle swarm optimization, deep belief network*

Abstrak

Jaringan syaraf tiruan menarik banyak peneliti dewasa ini. Banyak universitas-universitas terkenal telah mengembangkan jaringan syaraf tiruan untuk berbagai aplikasi baik kademik maupun industri. Jaringan syaraf tiruan menunjukkan kinerja yang patut dipertimbangkan untuk berbagai tujuan. Meskipun begitu, kinerja dari jaringan syaraf tiruan merosot dengan signifikan untuk masalah-masalah yang kompleks. Untuk menyelesaikan masalah tersebut di atas, banyak penelitian yang dilakukan untuk meningkatkan kinerja dari jaringan syaraf tiruan standar. Salah satu pengembangan yang menjanjikan untuk jaringan syaraf tiruan pada kasus yang kompleks adalah metode *deep learning*. Pada penelitian ini, diusulkan penggunaan metode Particle Swarm Optimization (PSO) pada Convolutional Neural Networks (CNNs), yang merupakan salah satu metode dasar pada deep learning. Penggunaan PSO dalam proses pelatihan bertujuan untuk mengoptimalkan hasil vektor solusi pada CNN, sehingga dapat meningkatkan akurasi hasil pengenalan. Data yang digunakan dalam penelitian ini adalah data angka yang berasal dari MNIST. Dari percobaan yang dilakukan akurasi yang dicapai dengan 4 iterasi adalah 95,08%. Hasil ini lebih baik dari CNN konvensional dan DBN. Waktu eksekusinya juga mendekati CNN konvensional. Oleh karena itu, metode yang usulkan adalah metode yang menjanjikan.

Kata Kunci: *deep learning, convolutional neural network, particle swarm optimization, deep belief network*

1. Introduction

In recent years, many researchers conducted studies on machine learning with deep hierarchical architecture. The term deep hierarchical learning was introduced by Hinton et al. [1]. They proposed a method to transform high dimensional data into low dimension data. They employed multilayer neural network with a small middle layer to reconstruct the input vector. Today, machine learning with deep hierarchical learning is named as deep learning.

The concept of deep learning is derived from neural network research, therefore deep lear-

ning regarded as a "*new-generation of neural networks*" [2]. Deep learning is research intersection between many areas, such as neural networks, artificial intelligence, pattern recognition, signal processing, optimization, and graphical model. Feedforward neural networks or Multilayer Perceptron (MLPs) with many hidden layers, which is often called Deep Neural Networks (DNN), is a good example of a model with deep architecture. Deep learning demonstrated impressive results and has been applied in several fields, like object recognition, computer vision, voice search, conversational speech recognition, and language processing.

From a wide variety descriptions that exist, deep learning generally has two aspects [2]. The first is deep learning is a model consisting of multiple layers of nonlinear information process-ing. The second is deep learning is a supervised or an unsupervised learning method to represent the features of the bottom layer to the top layers.

Several variations of deep learning are continually being researched and many of them have been applied into some machine learning tasks. In many cases, deep learning exhibited significant improvement on results, compared to previous conventional methods. There are plenty of deep learning algorithms which have been developed such as, Deep Neural Networks (DNNs) [1], Deep Boltzmann Machines (DBMs) [3], Recurrent Neural Networks (RNNs) [4], and Deep Auto encoders [5].

From many deep learning methods, there are 3 basic models that underlie many of the deep learning methods, i.e. Deep Belief Networks (DBNs) [6], Convolutional Neural Networks (CNNs) [7], and Stacked Auto Encoder (SAE) [8]. These three models are the most prominent and become building blocks for many deep learning methods, such as Multiresolution DBN (MrDBN) [9], an extension of the DBN or Scale-Invariant Convolutional Neural Network (SiCNN) [10] developed based on the CNNs model.

In this research, we focus handwritten digit recognition problem based on data from MNIST. From three basic models, we empowered CNNs method as our basic algorithm. CNN was employed due to its high accuracy on MNIST datasets [11].

CNNs is a type of feed forward neural network inspired by the structure of visual system. CNN consists of many neurons that have weights and biases, where each neuron receives several inputs and perform dot products. In terms of architecture, CNNs composed of one or more convolutional layers with subsampling stages and one or more fully connected layers as found in a standard multi-layer neural networks.

Even though standard CNN has shown considerable accuracy, there are still a lot of space for improvement. To ameliorate performance of CNN in recognition task, we used PSO to optimize output vector from CNNs. The utilization of PSO is due to its powerful performance on the optimization problems.

PSO in an optimization method developed by Eberhart and Kennedy [12]. This method is inspired by social behavior of animals that do not have a leader in their group. PSO consists of a swarm of particles, where the particles represent a potential solution.

To assess our proposed method, we compare results obtained from proposed method with other existing algorithms results. The existing algorithms used for comparison are the original CNNs and Deep Belief Networks (DBNs). The performance criteria used in this research are error and accuracy.

The remaining of this paper is organized as follows: Section 2 describes the basic theory of DBN, CNN, and PSO. Experiment setup and results obtained in the comparison study presented in Section 3. The conclusions of this paper are given in Section 4.

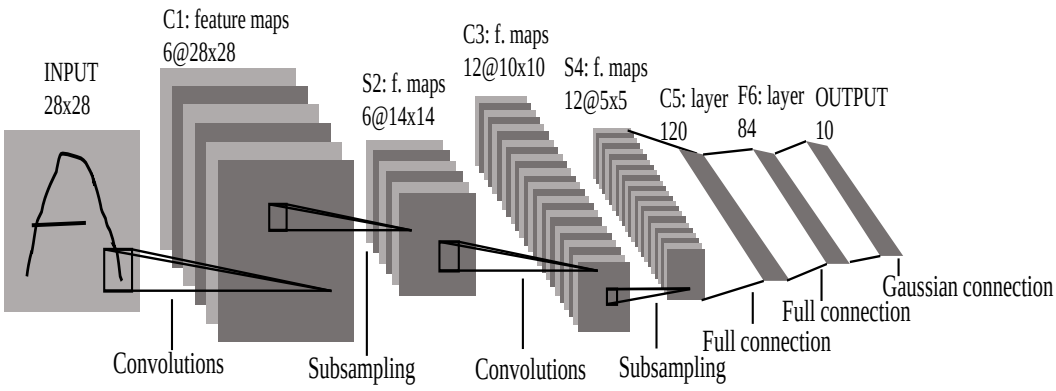


Figure 1. The original Convolutional Neural Networks (CNN) architecture [7]

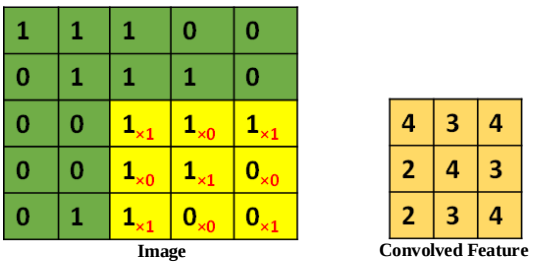


Figure 2. Convolution operation

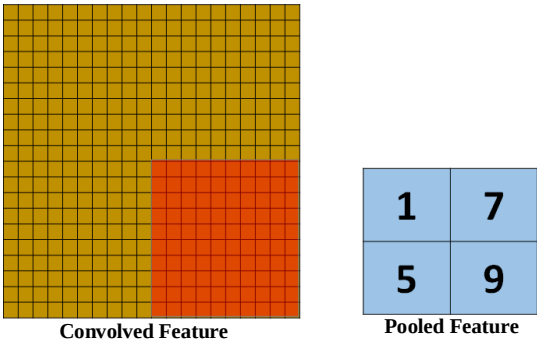


Figure 3. Illustration of pooling process

2. Methods

Convolution Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) is inspired by cat's visual cortex [12]. CNN generally consists of 3 layers which are convolutional layer, subsample/pooling layer, and fully connected layer. Convolution layer shares a lot of weights whereas pooling layer performs subsampling function to resulted output from convolution layer and reduce data rate of the layer below. The outputs from pooling layer are used as an input to several fully connected layers. Figure 1 shows the architecture of CNNs [7], and the convolution operation is illustrated on Figure 2. Convolved feature that is generally called feature maps is the result of convolving the filter/kernel on dataset.

The convolution process can be written by the following equation(1):

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \quad (1)$$

where I is an input image, K is kernel/filter used in convolution process, m is row of image, and n is column of image. The subsample or pooling is the process to reduce feature map. The concept of pooling process almost equal to convolution process that is convolving filter on input data. However, the differences pooling process on shifting filter that does not overlap on each filter compare with convolution process. The pooling illustration can be seen on Figure 2.

Deep Belief Networks (DBNs)

Deep Belief Networks (DBNs), which was introduced by Hinton, et al. [6], is a probabilistic graphical model consisting of multiple layers with hidden variables. DBNs are trained using greedy layer wise algorithm which can optimize the weight of DBNs at the time complexity that linear to the size and depth of the network. DBNs are trained to extract deep hierarchical representation on the input data. DBNs are composed of a number

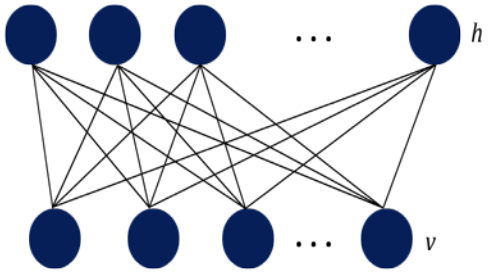


Figure 4. Restricted Boltzmann Machines (RBMs)

of Restricted Boltzmann Machines (RBMs), a special kind of Boltzmann Machines, that consisting a layer of visible unit and a layer of hidden units, with undirected and symmetrical connections between visible and hidden layer, but there is no connection among units in the same layer. Illustration of RBM models can be seen in Figure 4.

The lower layer v is the visible layer, and the top layer h is the hidden layer, where these two layers are stochastic binary variables. The weights between the visible layer and the hidden layer (W) are undirected. In addition each neuron has a bias.

The join distribution function p(v,h) of the visible units v and the hidden units are defined in equation(2) in the form of an energy function.

$$p(v,h) = \frac{\exp(-E(v,h))}{Z} \quad (2)$$

where Z is the partition function and given by summing all possible pairs of visible and hidden units as shown by equation(3).

$$Z = \sum_{v,h} \exp(-E(v,h)) \quad (3)$$

A joint configuration of visible and hidden units has an energy that given by equation(4).

$$E(v,h) = -\sum_{i=1} a_i v_i - \sum_{j=1} b_j h_j - \sum_{i,j} v_i h_j w_{ij} \quad (4)$$

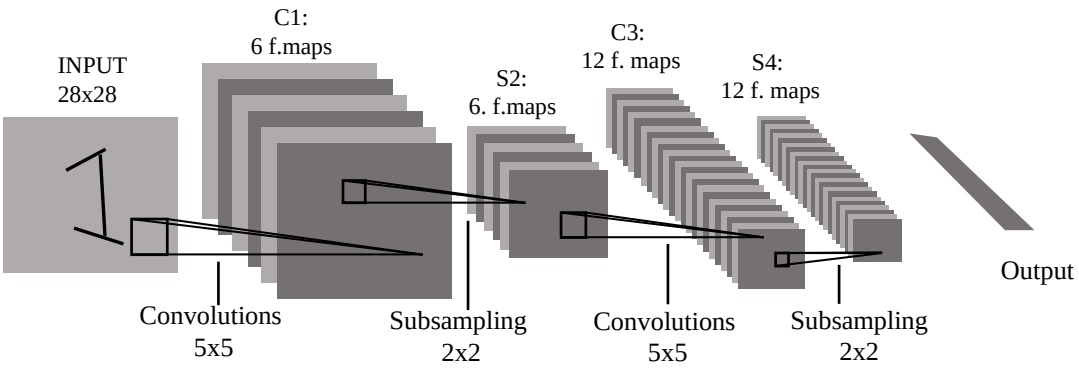


Figure 5. CNN Architecture in the Proposed Method

where v_i is binary state of visible unit i, h_j is binary state of hidden unit j, a_i is bias in visible unit, b_j is bias in hidden unit, and w_{ij} is the weight between visible and hidden units. The update rule for RBM weights are:

$$\Delta W_{ij} = E_d(v_i h_j) - E_m(v_i h_j) \quad (5)$$

where $E_d(v_i h_j)$ is the expectation in the training data and $E_m(v_i h_j)$ is the same expectation that defined by the model. RBMs are trained using *Contrastive Divergence* (CD) algorithm to approximate the expected value.

Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) algorithm is one of evolutionary algorithm which was firstly proposed in 1995 [13]. PSO has widely been employed in miscellaneous field, to cite an instance swarm robot for odour source localization purpose [14, 15].

PSO algorithm consist several consecutive steps. First of all, initialization which randomly select the particles as searching agents (x) as well as the velocities (v). Secondly the particles then inserted into cost function to find local bests (p_{best}) and global best (g_{best}). Local best is defined as the location on which the cost is the smallest for each particle. Meanwhile, global best is the location on which the cost is smallest among the local bests. Thirdly, the particles are updated by empowering equation(6) and equation(7).

$$v_{n+1} = v_n + c_1 r_1 (p_{best} - x_n) + c_2 r_2 (g_{best} - x_n) \quad (6)$$

$$x_{n+1} = x_n + v_{n+1} \quad (7)$$

Where c_1 and c_2 are the constants, r_1 and r_2 are random numbers, and n is iteration. The algorithm of PSO can be written as follows:

Algorithm 1: Particle Swarm Optimization

```
% Initialization
Number_Interation_PSO;
Number_PSO_Swarm;
Determine gbest from PSO_Swarm;
Determine pbest from PSO_Swarm;
calculate fitness_old;

for 1 to Number_Interation_PSO do
 for 1 to Number_of_Particle do

    %Update Velocity
    v(n)t = v(n)t+1 + c1*r1.*(pbest-
    x(n) +c2*r2.*(gbest-x(n))

    %Update position
    x(n+1) = x(n) + v(n)t

    %Evaluate the objective function
    fitness_new = f(x(n+1))
    if (fitness_new< fitness _old)
     fitness _old = fitness _new;
     x = x(n+1);
     else
     fitness_new = fitness _old;
     x(n+1) = x;
    end if
 end for;
 Index = min (fitness_new);
 pbest = x(index);
end for;
```

Proposed Method

Figure 5 shows the CNNs architecture used in the proposed method, where it is consist of an input image that will be processed using 6 convolution kernel with size 5x5 pixels, 6 sub-sampling kernel with size 2x2 pixels, 12 convolution kernel with size 5x5 pixels, 12 subsampling kernel with size 2x2 pixels and the last layer is the vector output of CNN. The proposed method process can be seen on Figure 6. In Figure 6 Y denotes the condition is met, whereas N represent the condition is not met.

PSO in this study would optimize the output vector. The output vector would be augmented by δx to acquire better value. The value of δx itself is the value which would be optimized by PSO. To

calculate the fitness function, the root mean square error between output vectors after augmented with δx and the true output would be employed.

Generally, the process of the proposed method consists of several steps as shown below: 1) the first step is initializing the learning rate of the CNNs with the value is 1 based on the experiment. Batch size of CNNs is 50, the number of CNNs epoch in the range of 1 to 4, PSO iteration is 10. The convergence status of PSO is used to check the convergences of PSO, if the error value has not changed for three iterations, then the PSO is considered as convergent; 2) after setting up the experiment, the next step is run CNNs training process, where the detail of the process can be seen in section 2.1.

The result of CNNs is vector output that will be optimizing using PSO algorithm. PSO optimization in this study serves to make the value of loss function on CNN becomes minimal; 3) the output vector will be update if the solution of swarm has less error compare with old vector output; 4) the PSO will run as long as the iteration number of PSO and the convergence solution have not fulfilled; 5) after the CNN Training, the model will be tested with testing data that consist of 10000 data; 6) the result of CNN test is accuracy of CNN, it represent how precise of the CNN model can predict the actual value of testing dataset.

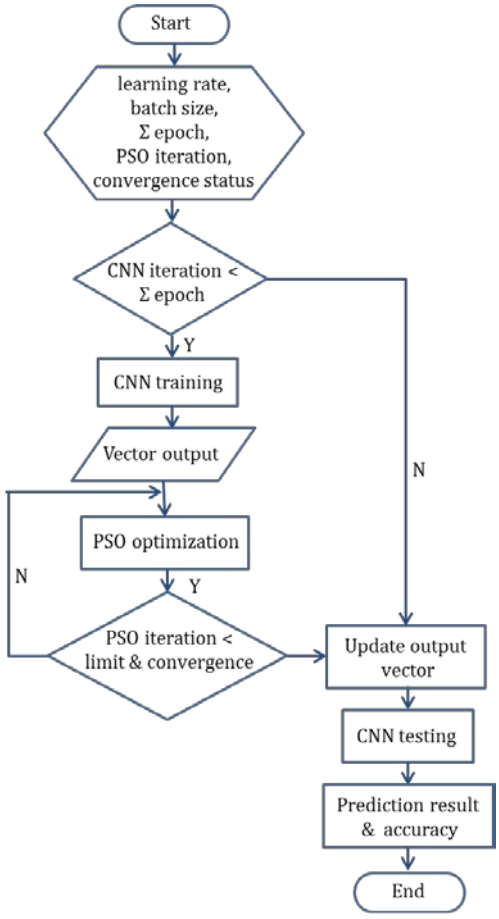


Figure 6. Flowchart of the proposed method

3. Results and Analysis

Dataset

In this study we use the handwritten digits data taken from MNIST database. This dataset has 28 x 28 pixel dimension and consist of 70000 data, in which 60000 data used for training and the rest data used for testing.

Experiment Result

In this chapter we will show the experiment result on deep learning method, such as Convolutional Neural Network, Deep Belief Network, and Convolutional Neural Network Optimize with Simulated Annealing (SA) [16] and also compare the performance with proposed method. The experiment use handwritten Digit dataset from MNIST. And the running time of all the deep learning method will compare each other to know how long the deep learning method can predict the test dataset.

Experiment on Handwritten Dataset

The overall experimental results on error and accuracy can be seen in Figure 7 and Table 1 respectively. The results of experiments showed the proposed method has a better accuracy than another deep learning method. Although the difference accuracy value is not high but it important be-cause the CNN does not need a lot of training epoch to get good accuracy. CNNPSO has better accuracy than CNNSA, it could be the advantages of PSO fast searching ability with minimum iteration of CNN. The error also exhibit the similar behavior. These due to the fact that PSO is a powerful algorithm for optimizing. Thus if com-pared to simulated annealing in particular, it has better performance.

As shown in Table 1, with only 4 epochs the accuracy of CNNPSO has reached 95.08%. This value was only slightly different from simulated annealing optimized CNN (95.19%). Meanwhile, CNN and DBN occupy the two last place with 94.81% and 91.14% respectively.

In execution time, DBN has fastest execution time, which architecture 50 hidden node and 25 hidden layer, because on training process using contrastive divergence that very fast [17].

CNNPSO consume longer time than CNN. This due to the addition iteration in the PSO algo-

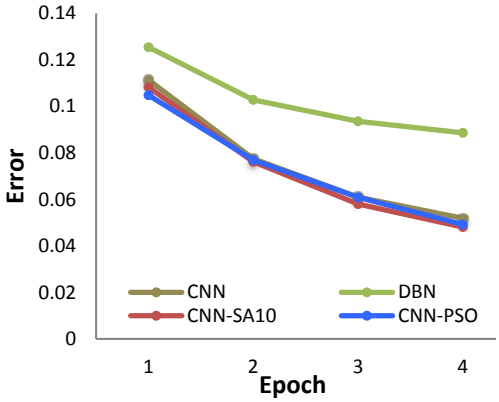


Figure 7. Error comparison of several algorithm

TABLE 1
ACCURACY PERFORMANCE ON ALL DEEP LEARNING METHOD

Method	Number of Epoch			
	1	2	3	4
CNN	88.87	92.25	93.9	94.81
DBN	87.46	89.72	90.64	91.14
CNN-SA10	89.18	**92.38**	**94.2**	**95.19**
CNN-PSO	**89.52**	92.31	93.91	95.08

rithm as well as the particles calculation. Nevertheless, if compared to CNNSA10, CNNPSO was still faster.

4. Conclusion

Based on the experiment result that have been conducted it can be conclude that the proposed method CNNPSO has good accuracy. The considerable accuracy (95.08%) was attained with only 4 epoch. Moreover the proposed method exhibited better performance than CNN, DBN. Even though its accuracy is lower than CNNSA, to obtain the accuracy nearby, CNNPSO consumed shorter time. If compared to the conventional CNN, CNNPSO consumed only slightly longer time. However, it has to be improved. The improvement can be focused on how to give restricted range on delta x so the proposed method get optimal vector output faster.

Acknowledgement

This work was supported by Directorate Research of Universitas Indonesia funding in 2015. The title of the research is Laboratory Infra-structure. This grant number is 1831/UN2.R12/ HKP.05.00/2015.

References

[1] G. E. Hinton and Ruslan R. Salakhutdinov, "Reducing the dimensionality of data with neural networks." *Science Journals* 313.5786 (2006): 504-507.

[2] L. Deng and D. Yu, "Deep Learning: Methods and Applications", Foundations and Trends® in Signal Processing, Vol. 7, Nos. 3–4 (2013) 197-387.

[3] R.R. Salakhutdinov and G. E. Hinton, "Deep Boltzmann Machines", *In Proceeding of International Conference of Artificial Intelligence and Statistics*, vol. 12, 2009.

[4] J. Martens and I. Sutskever, "Learning Recurrent Neural Networks with Hessian-free Optimization." *In Proceedings of International Conference on Machine Learning (ICML)*, 2011.

[5] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring Strategies for Training Deep Neural Networks." *J. Machine Learning Research*, vol. 10, pp. 1-40, 2009.

[6] G.E. Hinton, S. Osindero, and Y.W. Teh, "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, 2006.

[7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based Learning Applied to Document Recognition". *In Proceedings of the IEEE*, 86:2278–2324, 1998.

[8] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *in Advances in Neural Information Processing Systems 19* (NIPS'06), (B. Scholkopf, J. Platt, and T. Hoffman, eds.), pp. 153–160, MIT Press, 2007.

[9] Y. Tang and A.R Mohamed, "Multiresolution Deep Belief Networks", *In Proceeding of Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS) 2012*, La Palma, Canary Islands. Volume XX of JMLR: W&CP XX, 2012.

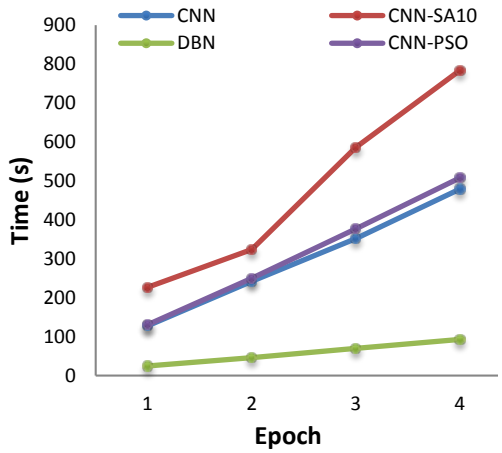


Figure 8. Time Comparison Based On Epoch Value

[10] Y. Xu, T. Xiao, and J. Zhang, Scale-Invariant Convolutional Neural Networks, *in arXiv*: 1411.6369v1 [cs.CV], 2014.

[11] Ciresan, Dan; Meier, Ueli; Schmidhuber, Jürgen, Multi-column deep neural networks for image classification, IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[12] Hubel, D. and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology (London), 195, 215–243.

[13] J. Kennedy and R. Eberhart, "Particle swarm optimization", *In Proceedings of the IEEE International Conference Neural Network (ICNN)*, Nov. 1995, vol. 4 pp. 1942-1948.

[14] W. Jatmiko, K. Sekiyama, and T. Fukuda, A mobile robots pso-based for odor source localization in dynamic advection-diffusion environment, in Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on. IEEE, 2006, pp. 4527–4532.

[15] W. Jatmiko, A. Nugraha, R. Effendi, W. Pambuko, R. Mardian, K. Sekiyama, and T. Fukuda, Localizing multiple odor sources in a dynamic environment based on modified niche particle swarm optimization with flow of wind, WSEAS Transactions on Systems, vol. 8, no. 11, pp. 1187–1196, 2009.

[16] L. M. R. Rere, M. I. Fanany, A. M. Arymurthy, Simulated Annealing Algorithm for Deep Learning, Procedia Computer Science, no. 72, pp. 137–144, 2015.

[17] Hinton, Geoffrey E. and Osindero, Simon and Teh, Yee-Whye, A Fast Learning Algorithm for Deep Belief Nets, Journal Neural Computing,no.7, pp 1527-1554, 2006.