

# Improved Linear Dynamical System for Unsupervised Time Series Recognition

**Ngoc Anh Nguyen Thi**

School of Electronics and Computer Engineering  
Chonnam National University, Gwangju 500-757, South Korea

**Hyung-Jeong Yang**

School of Electronics and Computer Engineering  
Chonnam National University, Gwangju 500-757, South Korea

**Soo-Hyung Kim**

School of Electronics and Computer Engineering  
Chonnam National University, Gwangju 500-757, South Korea

**Guee-Sang Lee**

School of Electronics and Computer Engineering  
Chonnam National University, Gwangju 500-757, South Korea

**Sun-Hee Kim**

School of Electronics and Computer Engineering  
Chonnam National University, Gwangju 500-757, South Korea

## ABSTRACT

*The paper considers the challenges involved in measuring the similarities between time series, such as time shifts and the mixture of frequencies. To improve recognition accuracy, we investigate an improved linear dynamical system for discovering prominent features by exploiting the evolving dynamics and correlations in a time series, as the quality of unsupervised pattern recognition relies strongly on the extracted features. The proposed approach yields a set of compact extracted features that boosts the accuracy and reliability of clustering for time series data. Experimental evaluations are carried out on time series applications from the scientific, socio-economic, and business domains. The results show that our method exhibits improved clustering performance compared to conventional methods. In addition, the computation time of the proposed approach increases linearly with the length of the time series.*

*Keywords: Time Series, Linear Dynamical System, Kalman Filter, Principal Component Analysis, Clustering, Feature Extraction*

## 1. INTRODUCTION

Unsupervised pattern recognition of time series has been utilized in numerous applications in fields such as bioinformatics [1-3], science [4], socio-economics [5] and business [6] and so on. Normally, the characteristics of time series are formed by high dimensionality and correlations among features [7] [8]. To have an efficiency of unsupervised

time series recognition, a very important processing step is to identify compact features extracted from co-evolving time series. This step can be used to not only convert a series of original values to more meaningful information, such as more understandable and interpretable information, but it also can lead time series to a lower dimensionality with the most relevant features. In other words, feature extraction can directly describe time series representation since suitable representation reduces feature space. This provides highly efficient features for knowledge discovery, which help so that they help to improve performance as well as the speed of the mining algorithm.

---

\* Corresponding author, Email: [hjyang@jnu.ac.kr](mailto:hjyang@jnu.ac.kr)  
Manuscript received Feb. 21, 2014; revised Mar. 07, 2014;  
accepted Mar. 17, 2014

Many feature extraction approaches have been proposed for time series recognition such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) [9]-[11], Discrete Fourier Transform (DFT) [12], [13], Linear Dynamical System (LDS), known as Kalman Filters (KF) [14]-[16], AutoRegression Moving Average (ARMA) [17]-[19]. However, these approaches do not enable compact features that are easy to interpret. Thus, applying clustering methodology with these feature extraction approaches show poor performance. Specifically, PCA and SVD are powerful tools to discover linear correlations across a time series. They effectively give the optimal low-rank approximation of the dataset by removing the variables with lower energy [11]. However, PCA and SVD ignore time-evolving patterns since their characteristics are not designed for tracking the ordering of the rows or the columns. Moreover, with PCA and SVD, the projection of the time series into low-dimensional principal components is not easy to interpret, and they do not clearly exhibit the characteristic of each pattern.

DFT is proposed to capture frequencies in a single time sequence. However, DFT lacks the dynamics, hence, clustering based on Fourier coefficients is unsuitable for data which include time-evolving patterns. The Linear Dynamical System, known as Kalman Filters, has been commonly used for time series analysis because of its simple implementation and extensibility [14]-[16]. Kalman Filter can capture correlations among time series and learn their evolving dynamics. However, the time shift effects, which are temporal gaps between two time series, cannot be handled. The representations of the extracted features by Kalman Filter are not also clear. The resulting model parameters thus lead to poor clustering performance.

Discrete Fourier Transform (DFT) can handle time shifts across the sequences. However, this approach ignores temporal dynamics. AutoRegression Moving Average (ARMA) is also a method for feature extraction. However, ARMA parameters do not provide a reliability since different sets of parameters can be obtained even from time series with similar structure.

In this paper, an improved linear dynamical system (LDS) for interpretability of prominent extracted features is applied in order to handle the challenges of time series, such as time shift effect and frequencies mixtures. Time shift effect or phase shift indicates temporal gaps between two time series. In other words, two time series have same frequency but different phase lags. The proposed method demonstrates the group of extracted features which obtains the same period with different lags and lightly amplitudes. Frequencies mixtures denote that time series often obey several periodicities which refer to combine different frequencies. The proposed method expects to identify periodic patterns and group them into one cluster.

The proposed feature extraction approach exploits two key characteristics of time series: correlation and dynamics. Correlation discovers the relationship between time series, while dynamic property discovers the temporal moving trend of time series by automatically identifying several hidden variables.

The contribution of the study is threefold. First, the proposed method can identify useful and understandable features that expected to have distinct characteristics of

different clusters. Second, the paper demonstrates the clustering accuracy and reliability of time series using k-means algorithm in a variety of domains that use real time series datasets, such as science, socio-economics, and business. Third, the execution time aspect is investigated since its computational complexity is significant with the linear scale-up in the duration of the time sequences.

The remainder of this article is organized as follows. Section 2 illustrates the proposed method setup for feature extraction. The experimental time series dataset is given in Section 3. In Section 4, empirical evaluations were conducted to assess the effectiveness of the proposed feature extraction method in clustering multiple time series.

## 2. PROPOSED METHOD

This section illustrates how to achieve the interpretability of extracted features using the proposed method from the time sequence in order to improve clustering quality. The given time series is denoted as a matrix  $X_{m \times T}$  by  $m$  sequences with duration  $T$ .

The proposed feature extraction method of the improved LDS includes four steps: (1) learning hidden variables using LDS, (2) calculating the Eigen-decomposition on transition matrix  $F$  to find the canonical form of hidden variables, which helps to find frequency combinations and determine the mixing weight of the frequency combinations, (3) measuring the magnitude of the frequency combination mixing matrix to eliminate phase shift, and (4) using SVD to combine the frequencies. The detail of the method is described as following:

### 2.1 Learning hidden variables based on Linear Dynamical System

Linear Dynamical System (LDS) has been used to model multi-dimensional time series data. LDS for multi-dimensional time sequence is modeled by following mathematical equations (1) – (3) with the parameters  $\theta = \{F, G, z_0, Q_0, Q, R\}$ .

$$z_1 = z_0 + \omega_0 \quad (1)$$

$$z_{n+1} = Fz_n + \omega_n \quad (2)$$

$$x_n = Gz_n + \varepsilon_n \quad (3)$$

Where vector  $x_n, z_n$  imply observed data sequences and hidden variable at time  $n$ , respectively. The vector  $z_0$  is initial state of hidden variable for the whole model.  $F$  is a transition matrix which relates to the transition of the state from the current time tick to the next time tick with noise  $\{\omega_n\}$ . While  $G$  is the observation projection with noise  $\varepsilon_n$ . All noise  $\omega_0, \omega_i$  and  $\varepsilon_i (i = 1 \dots T)$  are multivariate Gaussian noises with the following distribution with covariance matrix  $Q_0, Q$  and  $R$ , respectively. In LDS model, the expectation maximization (EM) algorithm is utilized to learn the component parameter models and estimate hidden variables.

The goal of EM is to maximize the expected log-likelihood of the given observation sequences. Specifically, the **E** step, it estimates the posterior distribution of the hidden variables conditioned on the data sequence with fixed model parameters; in the **M** step, it then updates the model parameters by maximizing the likelihood using some sufficient statistics from the posterior distribution [11], [20]-[24].

In this step, LDS learns the hidden dynamics, capturing the series of hidden variables that are evolving according to the linear transformation  $F$ . The hidden variables are linearly transformed to the observed numerical sequences [11], [21]. This implies that the series of hidden variables,  $z_n$ , is evolving over time ticks with linear transition matrix  $F$ . Also, the observed data sequences,  $x_n$ , are generated from these series of hidden variables with a linear output projection matrix  $G$  [21]. Each row of output matrix  $G$  corresponds to one sequence and therefore, can be used as feature in clustering step.

## 2.2 Finding canonical forms of hidden variables

The canonical form of hidden variables, which are achieved from the LDS system, is identified in this step. Each row of output transition matrix  $F$  of the LDS does not present distinct characteristics of the corresponding series, the rows need to be made compact and uniquely identified. We examine this further expecting to discover deeper hidden patterns for each sequence of LDS by determining the straight forward transition matrix  $F$  and output projection matrix  $G$ . Equation (2) represents the hidden variables depending on the eigenvalues of the transition matrix  $F$  [14]. These eigenvalues can capture the amplitude and frequencies of the underlying signals of hidden variables. As a result, normalizing the transition matrix  $F$  can directly reveal the amplitude, frequency, and mixture of the given set of sequences.

In this paper we applied an improved LDS which uses eigen-decomposition on transition matrix  $F$  in order to find frequency combinations as well as the mixing weight of frequency combinations. The eigen decomposition corresponds to the diagonal matrix  $\Lambda$  of eigen-dynamics and eigenvector ( $V$ ) of  $F$  as follows;

$$F = V\Lambda V^* \quad (4)$$

The output projection matrix  $G$  of LDS illustrates how the hidden variables are translated into observation sequences with linear combinations. Therefore,  $G$  matrix needs to be compensated to achieve the frequency combination mixing matrix  $G_h$  in order to obtain the same observation sequences from eigen-dynamics  $\Lambda$  as those of the transition matrix.

$$G_h = G \cdot V \quad (5)$$

The canonical hidden variables will be:

$$z_0^{new} = V^* \cdot \mu_0 \quad (6)$$

$$z_n^{new} = V^* \cdot z_n \quad (7)$$

$V$  contains conjugate pairs of columns corresponding to the conjugate pairs of eigenvalues in  $\Lambda$ . Therefore, the frequency combination mixing matrix  $G_h$  must contains conjugate pair of columns corresponding to the conjugate pairs of the eigenvalues in  $\Lambda$ .

$$\begin{aligned} z_n^{new} &= V^* \cdot z_n = V^* \cdot F \cdot z_{n-1} + noise \\ &= V^* \cdot V \cdot \Lambda \cdot V^* \cdot z_{n-1} + noise \end{aligned} \quad (8)$$

$$\begin{aligned} &= \Lambda^{n-1} \cdot \mu_0^{new} + noise \\ y_n &= G_h \cdot \Lambda^{n-1} \cdot \mu_0^{new} + noise \end{aligned} \quad (9)$$

From the equations (8) and (9), we obtain all hidden variables  $z_n$ , the canonical hidden variables  $z_n^{new}$  and observation  $x_n$ . These are mixtures of a set of scaling such as growing, shrinking or stationary sinusoid signals of data-dependent frequencies [14]. Their characteristics of frequencies and amplitudes are completely defined by the eigenvalues of the transition matrix  $F$ .

## 2.3 Eliminating phase shift effects

The frequencies mixing matrix  $G_h$  is calculated in the second step. From here, the contribution of each frequency mixture in the resulting observation sequences is found. This means that each row of  $G_h$  represents the characteristic of each sequence in the domain of the frequency combinations and thus, it can be used to cluster sequences. However, the frequency mixing matrix fails to group similar sequences with phase shifts or time shifts because it not only describes the strength of each eigen dynamic but also encodes the required phases for different sequences.

To eliminate the phase/lag effect by taking the magnitude of the frequency mixing matrix  $G_h$ , the same column for the conjugate column of  $G_h$  is obtained. By omitting these duplicated columns, we obtain the frequency magnitude matrix  $G_m$ , which tells how strongly each frequency combination base participates in the observation time sequences and also solves the lag challenge. [14].

## 2.4 Obtaining interpreted features

In order to obtain the interpreted features for each sequence, we apply the dimension reduction approach with SVD on the frequency magnitude matrix,  $\tilde{G}_m \approx U_k \cdot S_k \cdot V_k^T$ , where  $\tilde{G}_m$  is the column centered from  $\tilde{G}_m$ ,  $U_k$  and  $V_k$  are orthogonal matrices with  $k$  columns, and  $S_k$  is a diagonal matrix. The interpretability of the prominent features can be obtained as  $U_k \cdot S_k$ .

## 3. DATASET ACQUISITION

We conducted experiments on four different real datasets of multiple applications of time series as described below.

The first dataset is Cylinder-Bell-Funnel (CBF). This is the publicly available at the UCR time series data mining archive: <http://www.cs.ucr.edu/> [25]. CBF has three clusters of time series: Cylinder (C), Bell (B) and Funnel (F).

The second dataset was derived from a science application [26]. The temperature dataset is clustered, which is a collection of time series of the daily temperature in the year 2000 at various places in Florida, Tennessee, and Cuba[27]. The dataset includes temperature recordings from 9 locations in Tennessee, 4 locations in northern Florida, 8 locations in southern Florida, and 8 locations in Cuba. Based on geographical proximity and similarity in temperature, the time series from Tennessee and northern Florida form the first group, and those from southern Florida and Cuba form the second group.

The two remaining datasets are derived from socio-economics and business applications, such as the population dataset and personal income data. The population dataset contains a collection of 20 time series representing the population estimates from 1900-1999 in 20 states of the USA <http://eire.census.gov/popest/-archives/state/ststts.php> [8]. Eleven time series in the first group have an exponentially increasing trend, and the remaining time series in the second group have a stabilizing trend. The personal income dataset is a collection of time series representing the per capita personal income from 1929-1999 in 25 states of the USA <http://www.bea.gov/beat/regional/spi/> [5]. The 25 states are divided into groups based on their personal income growth rate in which the east coast states form the first group with a high growth rate and the mid-west states form the second group with a low growth rate.

#### 4. EXPERIMENTAL EVALUATION

To verify the successful identification of a few vital features, the clustering quality and scalability of the proposed method are considered. In particular, a comparative study was performed to assess the effectiveness of the proposed method against conventional feature extraction approaches such as PCA, DFT, and original Kalman Filter. In the experiment, we use the first two coefficients and clustered them by k-means algorithm with Euclidean distance. To evaluate the quality of the clustering results, we used the Cluster Similarity Measure (CSM) since we know the ground truth labels of each sequence. The formula equations can be defined as follows,

$$CSM(A, P) = \frac{1}{C} \sum_{1 \leq j \leq M} \max Sim(A_i, P_j) \quad (10)$$

$$Sim(A_i, P_j) = \frac{2|A_i \cap P_j|}{|A_i| + |P_j|} \quad (11)$$

where  $A = A_1, A_2, \dots, A_C$  is the actual cluster which is considered as a ground truth from the supervised data, while,  $P = P_1, P_2, \dots, P_C$  is a predicted cluster which obtained from a clustering method.  $|A_i|$  is the number of series in cluster

$A_i$ .  $|P_j|$  is the number of time series in cluster  $P_j$ . The similarity between clustering evaluation criteria has values ranging from 0 to 1, where 1 corresponds to the case when A and P are identical. This means that as the P criteria value obtained increase, the more similarity between A and P is achieved [23]. Note that this similarity measure is asymmetric.

The proposed method exploits deeper hidden patterns that can successfully capture correlation and temporal dynamics. It provides the group of distinct frequency mixtures that helps to manage the presence of the time shift effect with small shifts in frequency. These frequencies mixture groups represent good resulting features that lead to good clustering as well as visualization.

In order to assess the effectiveness of the proposed algorithm, the first contribution of the study is examined to show how much extracted features contributes to time series recognition. To do this, the first extracted feature of the CBF dataset is visualized as a specific example. In addition, comparisons are carried out to evaluate how much qualitative the extracted features of the proposed method compare with the original Kalman Filter, DFT and PCA. The visualization is shown in Figure 1. Among the color maps of Figure 1, the color map of the proposed method shows the different characteristics between the three clusters of the CBF dataset, such as Cylinder, Bell and Funnel. In each color map of the figure, the rows represent the number of time series of the dataset and the column performs the first features extracted from each method. As can be seen in Fig. 1, only the proposed method illustrates clearly the characteristics of the three clusters. On the other hand, the three remaining methods do not show meaningful interpreted features.

The second contribution of the proposed method represents the performance of clustering results on different multi-dimensional time series based on several extracted features. The clustering accuracy of all approaches are recorded in Table 1. Table 1 clearly shows that the proposed method for feature extraction gives the best accurate clustering for the particular application of a time series dataset with the highest clustering with similar measurement obtained, 0.7969, 0.8824, 1.00 and 0.7778 on CBF, Income, Temperature (Temp) and Population (Popul), respectively. Specifically, compared with the previously used main stream feature extraction methods, in the experiment of the CBF dataset, the proposed method on four dataset demonstrates a significant performance, which shows nearly 20%, 23% and 16.33% improvement clustering compared to the original Kalman Filter, DFT and PCA, respectively. Similarly, in the Income dataset, the clustering performance improvement is 20%, 20% and 17% more than of the Kalman Filter, DFT and PCA, respectively. On population dataset, the performance improvement is 18%, 26% and 26% greater than that of the Kalman Filter, DFT and PCA. The temperature time series dataset performs distinctively characteristics between two clusters; the proposed method

(PPM), the DFT and PCA therefore perform successful recognition.

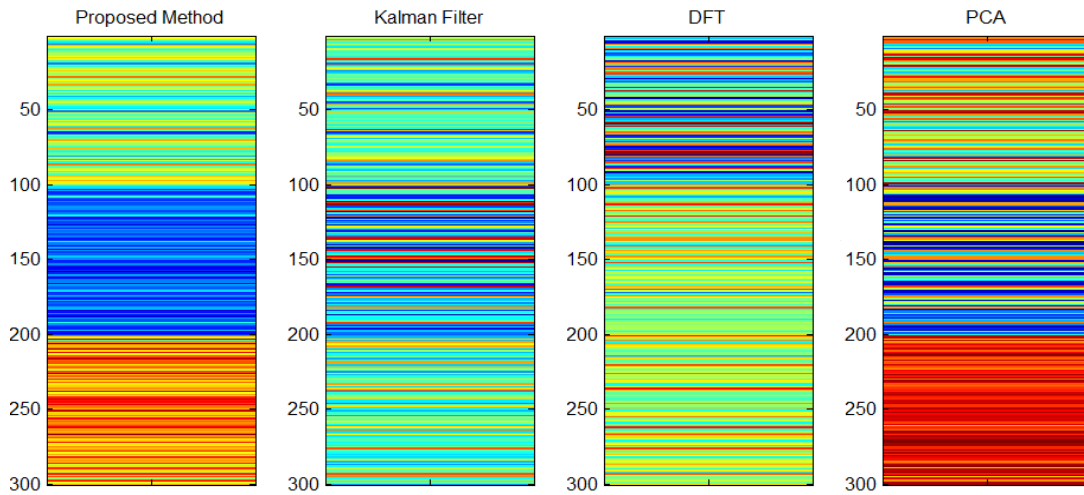


Fig. 1: Color map of the first extracted feature of four algorithm on CBF dataset

Table 1: Comparison of evaluation clustering criteria obtained from CSM of Proposed Method (PPM) for the three remaining approaches

	CBF	Income	Temp	Popul
PPM	0.7967	0.8824	1.000	0.7778
KF	0.6367	0.7022	0.800	0.6364
DFT	0.6100	0.7022	1.000	0.5758
PCA	0.6667	0.7316	1.000	0.5758

The execution time aspect is considered as the last contribution of the proposed method, shown in Fig. 2. It can be observed that the computational complexity is linear scale-up over the duration of the time series. On the other hand, the computation time of the proposed algorithm is scalable with linear complexity on the time duration T of sequences.

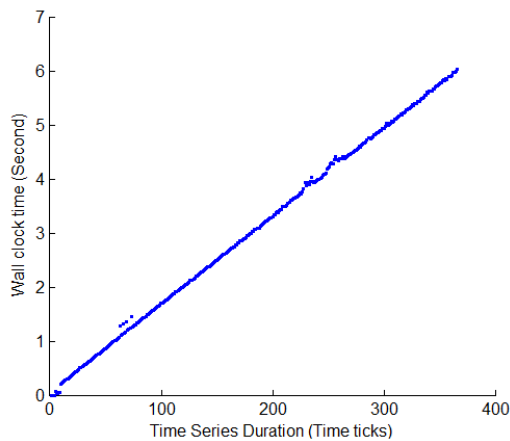


Fig. 2: Scalability of Execution Time

### 5. CONCLUSION

In summary, the paper presented an improved Linear Dynamical System by exploiting two distinct characteristics such as dynamics and correlation in time series for feature extraction. The proposed approach leads to a good set of extracted features which enables to improve time series

recognition compared with previous algorithms. Although, the paper demonstrated multi applications in time series domain, in the future work, we will extend on the theme of large time series recognition with the goal of developing time consumption and explore recognition accuracy. In addition, more challenges in time series mining will be investigated such as compression, forecasting, imputation, visualization and segmentation.

### ACKNOWLEDGEMENTS

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2013-H0301-13-3005). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST)(2013-056480).

### REFERENCES

- [1] O. Lange Wishmuller and D. R. Derch, et al, "Cluster analysis of biomedical image time-series," International Journal of Computer Vision, 46(2), 2002, pp.103-128.
- [2] C. S. Moller Levet, F. Klawonn, and O. Wolkenhauer, "Fuzzy clustering of short-time series and unevenly distributed sampling points," In lecture notes in computer science, vol. 2811, 2003, pp. 330-340.
- [3] S. Changdrakala and C. C. Sekhar, "A density based method for multivariate time series clustering in kernel feature space," In proceedings of IEEE international joint conference on neural networks, 2008, pp. 1885-1890.
- [4] K. Kalpakis, D. Gada, and V. Puttagunta, "Distance measure for effective clustering of ARIMA time series," In Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose CA, Nov. 2001, pp. 273-280.
- [5] H. S. Guan and Q. S. Jiang, "Cluster financial time series for portfolio," In proceedings of international conference

- on wavelet analysis and pattern recognition, 2007, pp. 851-856.
- [6] M. Gavrilov, D. Anguelov, P. Indyk, and R. Motwani, "Mining the stock market: which measure is best?," in: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, 20-23 August 2000, pp. 487-496.
- [7] X. Zhang, J. Liu, Y. Du, and T. Lv, "A novel clustering method on time series data," *Expert Systems with Applications*, vol. 38, issue 9, 2011, pp. 11891-11900.
- [8] T. Warren Liao, "Clustering of time series data- a survey," *The journal of the pattern recognition*, Elsevier, vol. 38, 2005, pp. 1857-1874.
- [9] Ding and X. He, "K-means clustering via principal component analysis," In Proceedings of the twenty-first international conference on Machine learning, 2004, pp. 29-32.
- [10] I. T. Jolliffe, *Principal Component Analysis*, Springer, 2002.
- [11] L. Li and B. Aditya Prakash, "Time Series Clustering: Complex is Simple!," In Proceedings of the 28th International Conference of Machine Learning, 2011.
- [12] M. Jahangiri, D. Sacharidis, and C. Shahabi, "Shift-split: I/O efficient maintenance of wavelet-transformed multidimensional data," In proceedings of the ACM SIGMOD international conference on Management of data, 2005, pp. 275- 286.
- [13] F. Morchen, "Time series feature extraction for data mining using dwt and dft," Technical Report no. 33, Philipps-University Marburg, 2003.
- [14] L. Li, B. Aditya Prakash, and C. Faloutsos, "Parsimonious Linear Fingerprint for Time Series," Proceedings of the VLDB Endowment, vol. 3, 2010, pp. 385-396.
- [15] L. Ralaivola and F. AlcheBuc, "Time series Filtering, Smoothing and Learning using the Kernel Kalman Filter," In Proceedings of International Joint Conference on Neural Network, Montreal, Canada, August 4, 2005.
- [16] A. Jain, E. Y. Chang, and Y.-F. Wang, "Adaptive stream resource management using Kalman Filter," In SIGMOD, 2004, pp. 11-22.
- [17] D. Gunopulos and G. Das, "Time series similarity measures and time series indexing," In SIGMOD Conference, Santa Barbara, CA, 2001.
- [18] K. Deng, A. Moore, M. Nechyba, "Learning to recognize time series: Combining arma models with memory-based learning," Proc. of the International Symposium on Computational Intelligence in Robotics and Automation, 1997, pp. 246-250.
- [19] X. Wang, K. Smith, and R. Hyndman, "Characteristic-Based Clustering for Time Series Data," *Data Mining and Knowledge Discovery*, vol. 13, 2006, pp. 335-364.
- [20] N. T. Ngoc Anh, H. J. Yang, S. H. Kim, and S. H. Kim, "Hidden Dynamic Learning for Long-Interval Consecutive Missing Values Reconstruction in EEG Time Series," In Proceedings of 2011 IEEE International Conference on Granular Computing, Taiwan, 2011, pp. 653-658.
- [21] L. Li, J. McCann, N. Pollard, and C. Faloutsos, "Dynammo: Mining and summarization of coevolving sequences with missing values," In KDD, ACM, New York, NY, USA, 2009.
- [22] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME C Journal of Basic Engineering*, 1960, pp. 25-45.
- [23] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, 1982, pp. 253-264.
- [24] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Technical Report CRG-TR-96-2, 1996.
- [25] E. Keogh and T. Folias, "The UCR time series data mining archive," 2002.
- [26] Y. Xiong and D. Y. Yeung, "Time series clustering with ARMA mixtures," *Pattern Recognition*, vol. 37, issue 8, August 2004, pp. 1675-1689.
- [27] <http://lwf.ncdc.noaa.gov/oa/climate/clima-tedata.html>



#### Ngoc Anh Nguyen Thi

She received the B.S. in Faculty Mathematics-Informatics from Da Nang Education University, Viet Nam in 2006, and M.S. at Dept. Electronics and Computer Engineering, Chonnam National University, Korea. She is currently a Ph.D. student at Dept. of Electronics and Computer Engineering, Chonnam National University, Korea. From 2006 to 2008, she worked as a lecturer and researcher at Faculty Mathematics-Informatics of Da Nang Education University, Viet Nam. Her research interests focus on the intelligent computing in many applications such as pattern recognitions, bioinformatics, data analysis of data mining and machine learning.



#### Hyung-Jeong Yang

She received her B.S., M.S. and Ph. D from Chonbuk National University, Korea. She is currently an associate professor at Dept. of Electronics and Computer Engineering, Chonnam National University, Korea. Her main research interests include multimedia data mining, pattern recognition, artificial intelligence, e-Learning, and e-Design.



#### Soo-Hyung Kim

He received the B.S. degree in Computer science from Seoul National University, Korea in 1986. He received the M.S. and Ph. D. degrees in Computer Science from KAIST, Korea in 1988 and 1993, respectively. He worked in Samsung Electronics as a senior researcher from 1990 to 1996 years. Since 1997, he has been a professor in the Electrical & Computer Engineering at Chonnam National University in Korea. His research interests include Artificial Intelligence, Pattern Recognition, Document Images Information Retrieval and Ubiquitous Computing.



**Guee-Sang Lee**

He received the B.S. and M.S. degrees in Electric Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received the Ph. D. degrees in Computer Science from University of Pennsylvania, in 1991. He has been a professor in the Electrical & Computer Engineering at Chonnam National University in Korea in 1994. His research interests include Multimedia Communication, Image Processing and Computer vision.



**Sun-Hee Kim**

She received M.S in Dongguk University, Korea. She received Ph. D degree at Dept. Electronics and Computer Engineering, Chonnam National University, Korea. She is currently a Post-doc researcher at School of Computer Science, Carnegie Mellon University, USA. Her research interests focus on data mining, sensor mining and stream mining.