International Journal of Hybrid Information Technology
Vol. 9, No.9 (2016), pp. 283-292
http://dx.doi.org/10.14257/ijhit.2016.9.9.27

A Typeface Searching Technique Using Evaluation Functions for Shapes and Positions of Alphabets Used in Ancient Books for Image Searching

Jun-Ho Huh* and Kyungryong Seo*[1]

*Department of Computer Engineering, Pukyong National University at Daeyeon, Busan, Republic of Korea**
72networks@pukyong.ac.kr, krseo@pknu.ac.kr

Abstract

Various projects to digitize the document and book inventories around the world are proceeding recently. At the same time, utilizing the IT/ICT technologies for the paleographic works is becoming an interesting issue among the historical and academic circles especially in the Republic of Korea where both Chinese characters and native Korean alphabets have been used in their ancient articles. There are some subtle differences in the typefaces and Great Seals used in each dynasty or year depending on the writers or the popular writing style of the time. The typeface recognition technique proposed in this article includes an alphabet analysis DTW algorithm and an alphabet position analysis method to determine the periods and possible authors of old documents. The results of analysis showed that the former was more reliable compared to the latter.

Keywords*: Typeface Searching, ICT, Era Assessment System, DTW Algorithm, Java, Image Searching*

1. Motivation

The kingdoms in Korean history are distinguished by the names declared by the progenitor of each kingdom and the presiding kings usually belong to a certain clan. In most cases, distinctive Great seals and typefaces were used by the each king or high-ranking officials [*e.g.,* Shijung - Unified Silla Dynasty, Seungseon - Koryeo and Doseungji (head of Seungjeongwon) – Chosun Dynasty] of the time when preparing official statements or orders [1-12].

Accordingly, the written periods and authors could be identified by studying the seal types or typefaces typical of each generation. The algorithm described and proposed in this study will be able to achieve this aim.

2. Objective and Scope of Research

Since it is essential to prove effectiveness of the proposed algorithm and understand the unique forms of Korean alphabets which sometimes take an art form, the scope of research will be limited as follows: (a) The ancient Korean alphabets to be studied will be selected from those used in the ancient literatures and documents of Chosun Dynasty (A.D. 1392-1910), having distinctive types. (b) As for the calligraphic alphabets, those types (*e.g.,* Nakseongbiryong, Okwondungheeyeon, and Bongseo) used in the transcribed materials of the royal court and showing the penmanship that is unconventional compared to ancient Chinese are considered [7-19].

[1] Corresponding Author, Department of Computer Engineering, Pukyong National University at Daeyeon, Busan, Republic of Korea

ISSN: 1738-9968 IJHIT

As for the future task, the documents and literatures (*e.g.*, original Hunminjongum Haerebon, Yongbieocheonga, Seokbosangjeol and Wolin-seokbo) in the period of Hunminjongum invention whose typeface characteristics are similar or the same will be used. Also, those editions and printed books [*e.g.*, Hunminjeongeum eonhaebon (King Sejong's invention), Oryunhaengsildo, Songgang-gasa, Banggakbon Novels, king's and noblemen's writings (Jeong Cheol, Kim Jeong hee, Heo Mok and yang Sa-heon)] written or printed with a calligraphic penmanship and have a unique brush-stroke touch will be used.

3. Contribution and Performance Evaluation

3.1. Typeface Evaluation Function

Letters have a certain kind of shape displaying consistent characteristics and typeface is the subordinate symbols having the basic forms in principle to support the phonetic languages. These typefaces change over time displaying some visual distinctiveness. Such changes often seem to follow the individual taste of a writer or the popular writing style (tendency) of the age. Typefaces are usually affected by various elements of culture during the transformation period and show different distinctive features depending on the groups or statuses. The Typeface Search Evaluation Function proposed in this article performs both the 'Alphabet Forms Search' and the 'Alphabet Position' evaluation functions.

3.2. Alphabet Forms Search Evaluation Function

The dynamic time wrapping (DTW) algorithm was used to determine the level of similarity of two kinds of graph in which the time axis was designated as a reference axis. The DTW algorithm analyzes/compares these graphs and produces 2-dimentsional arrays while scanning them from left to right and bottom to the top along both the horizontal and the vertical axes. The Formula 1 is to calculate the 2-dimensional array D (I,j) and the calculated value of D indicates the level of similarity of the two graphs. If the value gets closer to 0, the similarity is higher and when the value becomes 0, the two graphs are considered to be the same ones.

$$D(i,j) = d(i,j) + minimum \begin{cases} D(i-1,j-1) \\ D(i-1,j) \\ D(i,j-1) \end{cases}$$

Formula 1

In analyzing a letterer '김' is written with two kinds of typefaces, the DTW algorithm has been applied, generating the graphs on which the pixels of each letter is projected downwards. Figure 1 shows typeface comparison without image resizing. Both letters would seem very similar but the algorithm calculated the value of D as 481.10472972972974, meaning that they are not so similar after all.

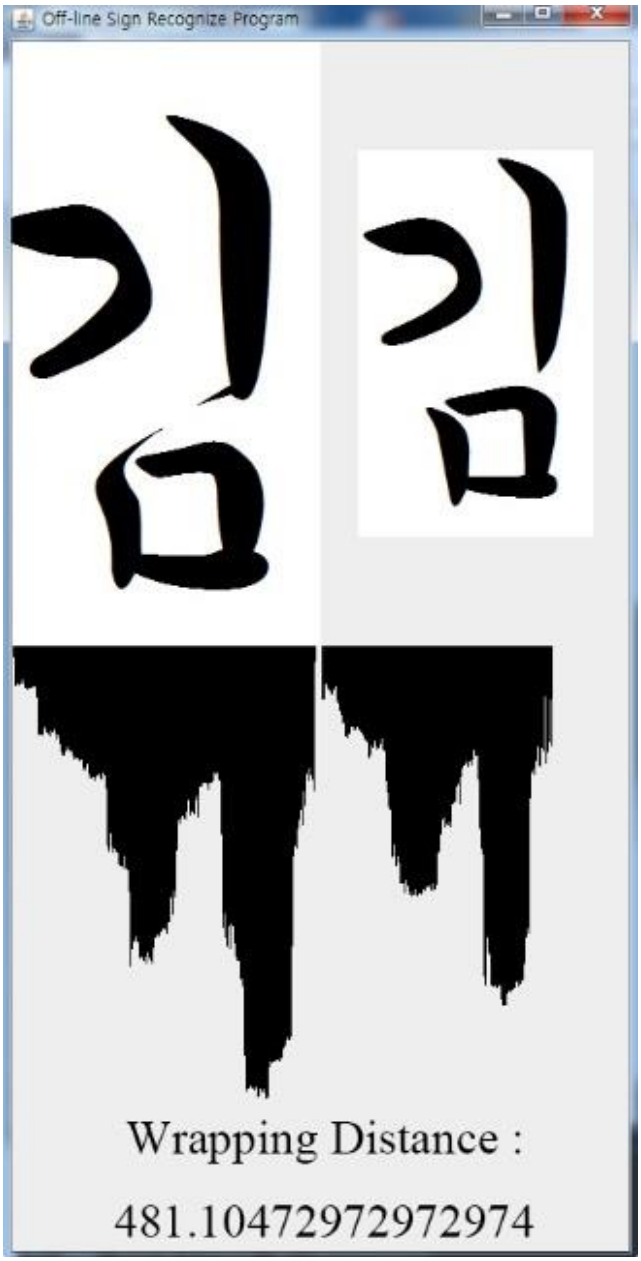

Figure 1. Typeface Comparison without Image Resizing

To increase the reliability of this assessment, another calculation was performed using Formula 2, which calculated the distance from the point where the letter begins to the point where it lasts along with the vertical axis, and resized the width of other typeface, matching the shorter length among the two.

$$resize\ rate = \frac{min\begin{cases} sample\ image\ width \\ compare\ image\ width \end{cases}}{max\begin{cases} sample\ image\ width \\ compare\ image\ width \end{cases}}$$

Formula 2

Then, after the adjustment, the calculation result was 64.692 showing increased similarity between the typefaces Figure 2.

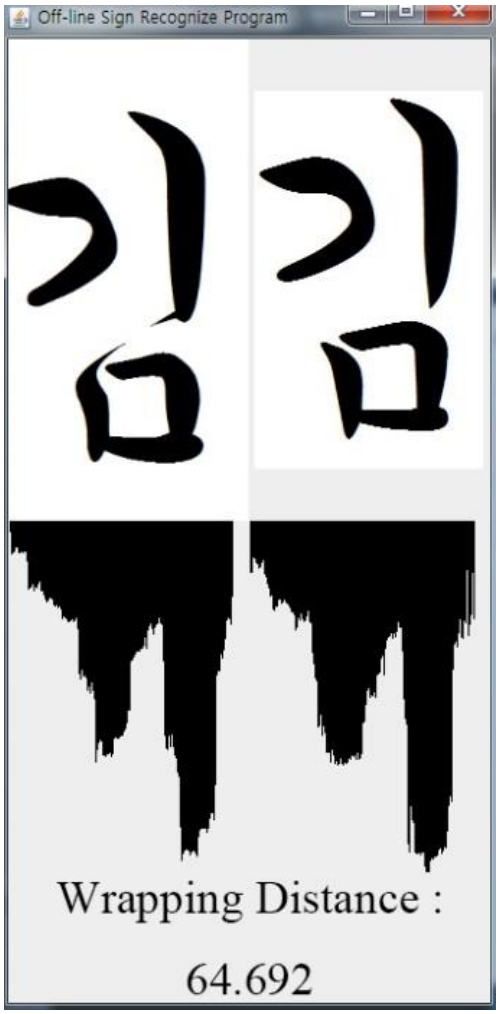

Figure 2. Typeface Comparison after Image Resizing

On the other hand, an additional factor needs to be considered when adopting this method. Considering that the Korean letters are formed from the combinations of consonants and vowels, comparison work must be carried out for each of them before making the calculations. Separation steps of alphabets are represented in Formula 3: 1) each point-value is taken from the position where the stroke starts. The 'point' refers to the pixel of the first alphabet that touches the horizontal axis. Clearly, setting just one point is not enough to provide the exact measurements to conduct a separation work. Thus, in order to obtain the measurements for the space availability, another point was extracted and the distance between the first and the second was measured. The resulting distance is then compared to other points which cover wider ranges (*i.e.*, points within a threshold value). Formula 4 represents this procedure.

$$
\begin{aligned}
&For\ each\ height\ pixel\ h \\
&\quad For\ each\ width\ pixel\ w \\
&\qquad \mathrm{If}(\mathrm{rgb}(\mathrm{w},\mathrm{h})\ !=\ white) \\
&\qquad\quad put\ point(w,h)\ to\ point\ list \\
&\qquad\quad break\ for\ loop \\
&return\ point\ list
\end{aligned}
$$

Formula 3

$$resize\ rate = \frac{min\begin{cases} sample\ image\ width \\ compare\ image\ width \end{cases}}{max\begin{cases} sample\ image\ width \\ compare\ image\ width \end{cases}}$$

Formula 4

Extract each consonants and vowel with a method of increasing width and length by 1 pixel unit starting from the points where each separation starts and move on to the next line when non-letter spaces appear. The separation result of this method is shown in Figure 3.

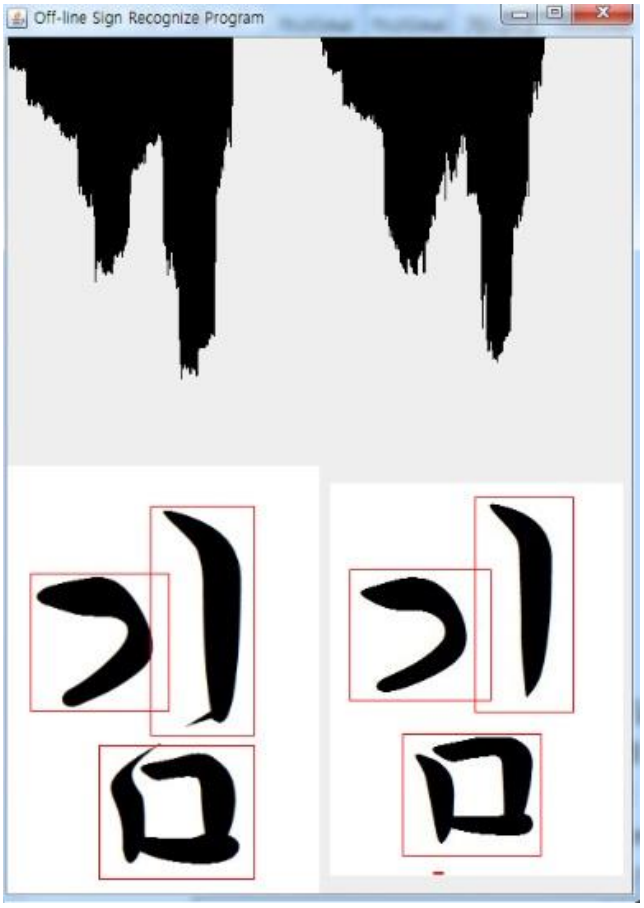

Figure 3. Typeface Alphabet (Consonants and Vowels) Separation

By applying DTW algorithm to the graph on which each separated alphabets have been projected, the authors obtained an output as in Figure 4.

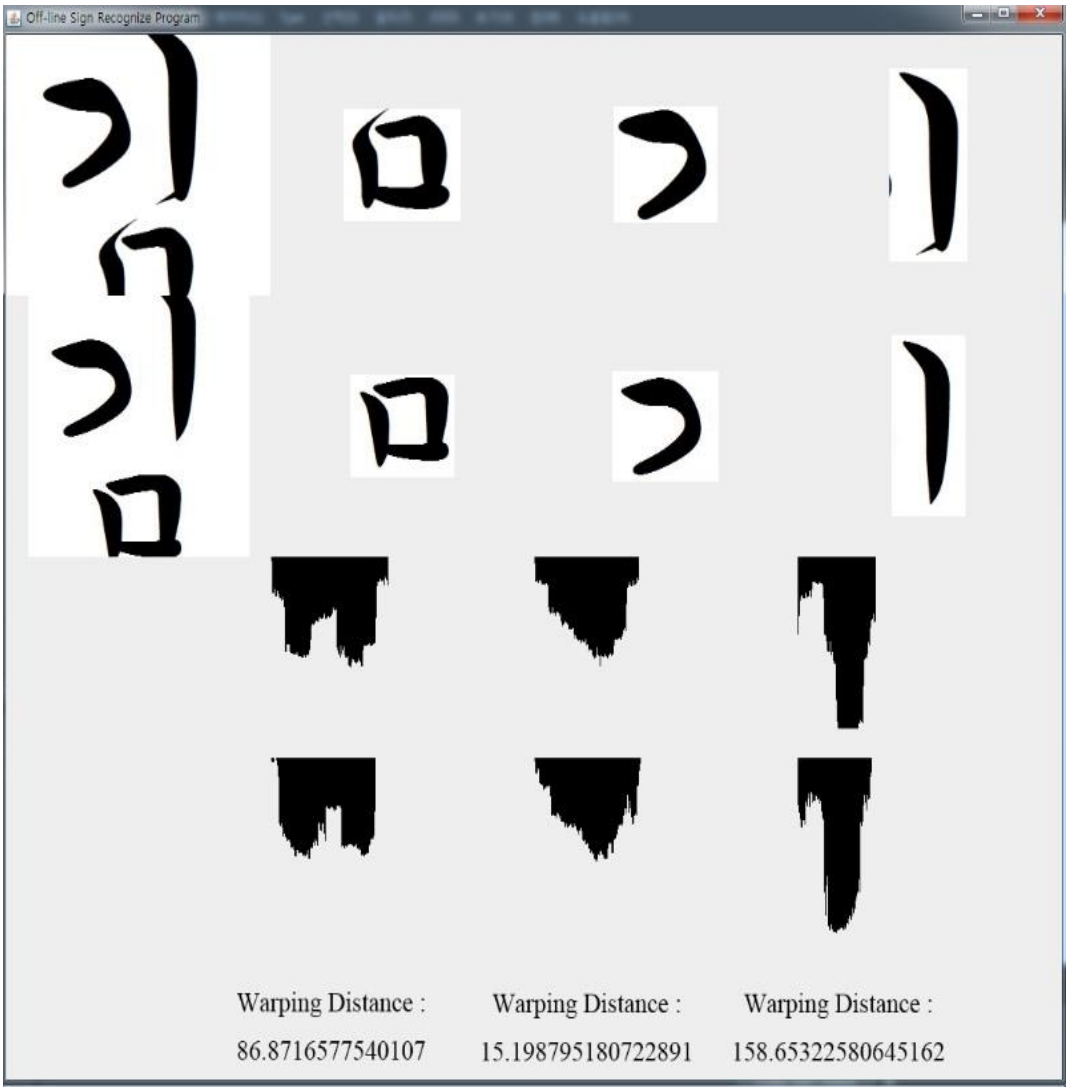

Figure 4. DTW Operation by Separating Alphabets

With the formula, the authors extracted a value of 86.90789291 as a value of the form evaluation function for final evaluation.

3.3. Alphabet Location Search Evaluation Function

The Korean letters are typically formed with an initial consonant and a vowel, frequently having an additional consonant(s) under the vowel. The position of each alphabet can be distinctive depending on the writer so that the authors used this to make comparisons.

Below Figure 5 indicates the characteristic points for two differing penmanship. Here, observing the letter '림' one can realize that the sizes of central spaces are different.

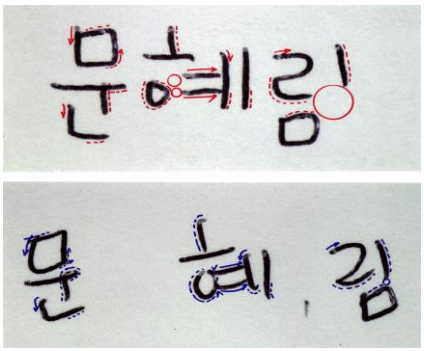

Figure 5. Positional Characteristic of Korean Writing (Source: Korea Document Authentication Institute)

Figure 6. shows the analysis for the central points. To apprehend positional characteristics of Korean writing, the authors first find the central points in each sound and compare the distances between the points. With this method, the authors can analyze how different the positional characteristics are between two letters.

As formula 5 shows below, the authors deduce an evaluation function of positional characteristics with an average value of distance differences for the sample alphabets and the template alphabets, all of which having initial, medial and final sounds.

$$f(letter) = \frac{1}{M}\sum_{i=1}^{M}|sample\_distance(i, i+1) - template\_distance(i, i+1)|$$

Formula 5

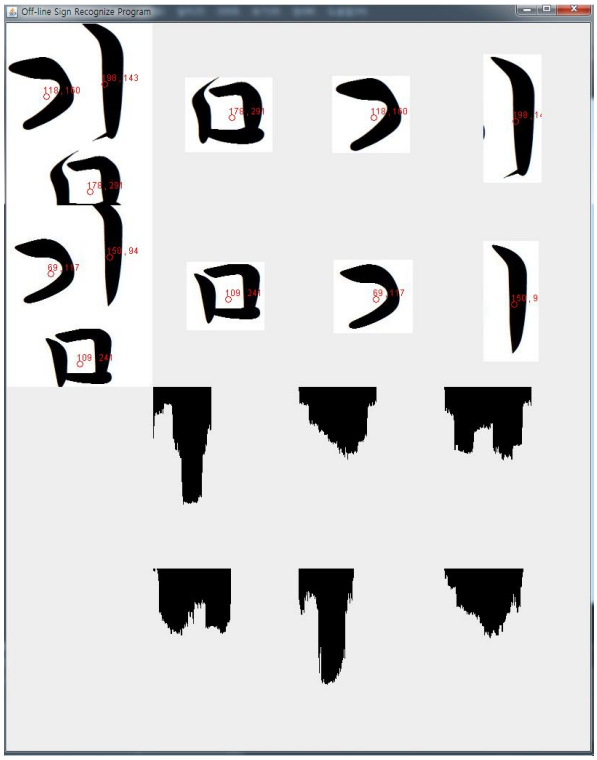

Figure 6. Analysis for Central Points

Table.1 shows the analysis of central points for Initial, medial and final consonants. The positional characteristics of the extracted central points are shown in the table below

and an evaluation value obtained by applying f (letter) to the positional characteristics will be 6.49194333.

Table 1. Analysis of Central Points for Initial, Medial and Final Consonants

	Upper“김”	Under“김”	Distance Difference
Distance(ㄱ, ㅣ)	81.7863069	84.2021377	2.41583
Distance(ㄱ,ㅁ)	149.3452376	152.6106156	3.2653
Distance(ㅣ,ㅁ)	144.0867794	130.2919798	13.7947

4. Conclusion and Future Work

To conduct accurate Korean typeface analyses/comparisons that are much different from English and other western languages, the alphabets forming a letter had to be separated prior to the analysis. The analyses were carried out with the evaluation functions which adopted the DTW algorithm. The comparison graphs were used to evaluate positional characteristics of the alphabets whose images were scanned from the top to the bottom.

Also, for the comparison, another evaluation function that employs the distance difference between the alphabets was used, but this method had produced rather unreliable results for the cursive scripts because the alphabets often overlapped or are not written in a formal form. Solving this problem with an advanced algorithm will be included in our future tasks. Meanwhile, the proposed algorithm can be used as a basic platform for the current paleographic works, or for assessment of authenticity. A simple application can be created with present algorithm for the use by the non-experts who wish to check the authenticity of ancient documents before purchasing them. Thus, commercialization of such application is feasible.

Acknowledgements

The first draft of this proceeding paper [16] was presented (Poster Session) in Spring Conference of Korea Multimedia Society, May, 28-30 (2015). I am grateful to 3 anonymous commentators who have contributed to the enhancement of the paper's completeness with their valuable suggestions at the Conference.

The second draft of this paper [1] was presented Oral Session in The 9th 2016 International Interdisciplinary Workshop Series, JEJU, April, 19-21 (2016). I am grateful to 3 anonymous commentators who have contributed to the enhancement of the paper's completeness with their valuable suggestions at the International Workshop.

References

[1] J.-H. Huh, N. Kim and K. Seo, “A Typeface Searching Technique Using Evaluation Functions for Shapes and Positions of Alphabets Used in Ancient Books”, SERSC, ASTL, vol.129, **(2016)**, pp.190-195
[2] S. G. Kim, “The Cause of Formation of Fonts Used in Hunminjeongeum”, IDAK Design Journal, Visual Information Design Association of Korea, vol.2, **(2002)** (In Korean).
[3] K. M Huh and I. T Kim, “Types and Designations of Korean Typefaces in Chosun Dynasty”, The Korean Language Society Journal, **(2007)** (In Korean).
[4] K. M. Huh, “Discussion for the calligraphic works of seven kinds of typefaces”, The 12the Korean Calligraphy Festival, Celebrating the 568th Hangul Day, Inviting the Major Works for Korean Calligraphy by Hannae ‘Huh, Kyeong-Moo’, a special member, Korean Language Society at Busan, Korean Typeface Research Association, **(2014)**, pp.4-17. (In Korean)
[5] K. M Huh, “The 12the Korean Calligraphy Festival”, Celebrating the 568th Hangul Day, Korean Language Society at Busan, Korean Typeface Research Association, **(2014)**, pp.70-94. (In Korean)
[6] Myers, Cory, L. Rabiner and A. E. Rosenberg, “Performance tradeoffs in dynamic time warping algorithms for isolated word recognition”, Acoustics, Speech and Signal Processing, IEEE Transactions, vol. 28, no. 6, **(1980)**, pp.623-635.

[7] Muda, Lindasalwa, M. Begam and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques", Journal of Computing, vol. 2, no. 3, **(2010)**, pp.138-143.
[8] Basil, Mohammad and B. Gawali, "Comparative Analysis of MSER and DTW for Offline Signature Recognition", International Journal of Computer Applications, vol.110, no.5, **(2015)**, pp.13-17.
[9] Rajesh, S. Avani, V. Bharadi and P. Jangid, "Performance improvement of complex plane based feature vector for online signature recognition using soft biometric features", Communication, Information & Computing Technology (ICCICT), 2015 International Conference on IEEE, **(2015)**, pp.1-7.
[10] Bharadi, A. Vinayak, R. R. Sedamkar nd P. S. Jangid, "Performance analysis of grid & texture based feature vector for dynamic signature recognition", Pervasive Computing (ICPC), 2015 International Conference on IEEE, **(2015)**, pp.1-6.
[11] J.-H. Huh, T. Koh and K. Seo, "NMEA2000 Ship Area Network (SAN) design and Test Bed using Power Line Communication (PLC) with the 3-Phase 3-Line Delta Connection Method", SERSC, ASTL, vol.94, **(2015)**, pp.57-63.
[12] J.-H. Huh and K. Seo, "Hybrid AMI Design for Smart Grid Using the Game Theory Model", SERSC, ASTL, vol.108, **(2015)**, pp.86-92.
[13] J.-H. Huh, N. Kim and K. Seo, "Design and Implementation of Mobile Push Service-Based Mobile Medication-Hour Notification System", Advanced Science and Technology Letters, vol.117, **(2015)**, pp.92-96.
[14] J.-H. Huh, T. Koh and K. Seo, "Design of a Shipboard Outside Communication Network and the Test Bed Using PLC: for the Workers' Safety Management During Ship-Building process", Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication, ACM IMCOM 2016, DaNang, vol.43, **(2016)**, pp.1-6.
[15]J.-H. Huh, S. Otgonchimeg and K. Seo, "Advanced metering infrastructure design and test bed experiment using intelligent agents: focusing on the PLC network base technology for Smart Grid system", The Journal of Supercomputing, Springer, USA, vol.72, no.5, **(2016)**, pp. 1862-1877.
[16] N. Kim, J.-H. Huh and K. Seo, "Design and Implementation of Foundation Technology for Authenticity and Creation Era Assessment System using Typeface Search Evaluation Functions", Proceedings of Spring Conference of the Korea Multimedia Society, Poster Session A, **(2015)**, pp. 42.
[17] J. Park, "Design of the Real-Time Mobile Push System for Implementation of the Shipboard Smart Working", Advances in Computer Science and Ubiquitous Computing, Lecture Notes in Electrical Engineering, Springer, Singapore, vol. 373, **(2015)**, pp. 541-548.
[18] C.-H. Lee and S.-H. Jung, "Fast thinning algorithm based on improved SOG (SOG*)", IECON 2004, 30th Annual Conference of IEEE, vol.3, **(2004)**, pp.2201-2206.
[19] J.-H. Huh, D. Lee and K. Seo, "Implementation of Graphic Based Network Intrusion Detection System for Server Operation", International Journal of Security and Its Applications, SERSC, vol.9, no.2, **(2015)**, pp.37-48.

Authors

Jun-Ho Huh, he finished the Cooperative Marine Science and Engineering Program, Texas A&M University at Galveston, United States of America in Aug. 2006.

Received B.S. in Science Degree from Department of Major of Applied Marine Sciences (Marine Aquaculture, Oceanography, Marine Life Sciences), B.S. in Engineering Degree (Double Major) from Department of Major of Computer Engineering from Jeju National University at Ara, Jeju, Republic of Korea in Aug. 2007. And completion of the Secondary School (Middle and High schools) Teacher Training Curriculum in accordance with Republic of Korea Secondary Education Act.

M.A. in Education Degree from Department of Major of Computer Science Education, Graduate School of Education, Pukyoug National University at Daeyeon, Busan, Republic of Korea in Aug. 2012. And completion of the Secondary School (Middle and High schools) Teacher Training Curriculum in accordance with Republic of Korea Secondary Education Act.

He received the Best Paper Award from Korea Multimedia Society four times (Nov. 2014, May. 2015, Nov. 2015, May 2016). And he

received Best Paper Award The 10th 2016 International Interdisciplinary Workshop Series from HSST (Aug. 2016).

Senior Research Engineer of SUNCOM Co., Busan, Republic of Korea (Aug. 2015- June. 2016).

Currently he is Research Professor of Dankook University at Jukjeon, Yongin, Republic of Korea (July. 2016-). And he is Clinical Assistant Professor of E-Green Remote Continuing Education (Mar. 2016-).

His research interests are Green IT, Smart Grid, Network Security, Curriculum of Computer, High Availability Computing.

Kyungryong Seo, he received B.S. in Engineering Degree from Department of Major of Electrical Machinery Engineering from Pusan National University, Busan, Republic of Korea in Feb. 1983. M.S. in Degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea in Feb. 1990.

Received the Ph.D. Degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea in Aug. 1995.

He received the Best Paper Award from Korea Multimedia Society five times (Nov. 2007, Nov. 2014, May. 2015, Nov. 2015, May 2016). And he received Best Paper Award The 10th 2016 International Interdisciplinary Workshop Series from HSST (Aug. 2016).

Currently, he is a Full Professor (Tenure) of Computer Engineering Departments, Pukyong National University at Daeyeon, Busan, Republic of Korea.

His research interests are High Speed Computer Network, Network Security, High Availability Computing.

