Correction for Misrecognition of Korean Texts in Signboard Images using Improved Levenshtein Metric

Myung-Hun Lee[1], Soo-Hyung Kim[2], Guee-Sang Lee[2], Sun-Hee Kim[3] and Hyung-Jeong Yang[2]
[1] 1Media Service Group, Konan Technology Co. LTD, Korea
[e-mail: kh2821@naver.com]
[2]Department of Computer Science, Chonnam National University
Gwangju 500-757, South Korea
[e-mail: {shkim, gslee, hjyang}@jnu.ac.kr]
[3]Department of Computer Science, Carnegie Mellon University
Pittsburgh, 15213-3891, USA
[e-mail: wkdal749@cs.cmu.edu]
*Corresponding author: Hyung-Jeong Yang

\

Received September 5, 2011; revised November 14, 2011; accepted December 15, 2011; published February 28, 2012

Abstract

Recently various studies on various applications using images taken by mobile phone cameras have been actively conducted. This study proposes a correction method for misrecognition of Korean Texts in signboard images using improved Levenshtein metric. The proposed method calculates distances of five recognized candidates and detects the best match texts from signboard text database. For verifying the efficiency of the proposed method, a database dictionary is built using 1.3 million words of nationwide signboard through removing duplicated words. We compared the proposed method to Levenshtein Metric which is one of representative text string comparison algorithms. As a result, the proposed method based on improved Levenshtein metric represents an improvement in recognition rates 31.5% on average compared to that of conventional methods.

Keywords: Character recognition, sign image, Levenshtein Distance, post-processing

"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-2012-C1090-1211-0008), "This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technologies"(2011-0029429).

DOI: 10.3837/tiis.2012.02.016

1. Introduction

Different types of applications have been introduced according to the use of smart phones in our daily lives. Most of smart phones have no keyboards and include touch based interface to boost the efficiency of input. One of the most actively studied input methods is an automatic recognition system using a camera. Various applications that recognize barcodes using a camera [1] and provide book information by recognizing book covers [2] have been appeared. The common points in these applications are to provide specific information to users by recognizing the visual information.

Signs are a type of media for transferring different types of information. The feature of signs includes lots of intuitive visual information. Thus, it is possible to provide various types of information by recognizing such signs using high resolution cameras installed in smart phones or PDAs. For instance, as foreign tourists need to make some decisions through Korean signs, a system that provides proper information by translating the characters presented in signs that are obtained by taking pictures using their mobile phones will help their decision making. Therefore, a character recognition technology is an essential factor for implementing such a system.

The field of character recognition has been studied for a long time and largely used in various fields. The character recognition has been extended to the recognition of characters in natural scene images, which have complicate backgrounds, beyond the stage that recognizes printed letters in simple documents [3][4]. However, it is difficult to expect a high recognition level as same as that of the recognition in formalized types like documents. As human beings recognize characters, individual characters are to be first recognized and then the characters are to be verified whether these are fitted to the context of entire characters. Therefore, it is possible to accurately recognize such characters because this recognition does not depend on the type of character only. A recognizer outputs the results of its recognition through recognizing individual characters obtained from pre-processing phase. Thus, these individual character recognition approaches have a limitation in implementing a flexible system as same as the character recognition performed by human beings. Therefore, an effective method that can reduce misrecognition rates occurred in a character recognition process is required in order to overcome such a limitation.

In this study, a misrecognition correction post-processing method for improving individual character recognition in Korean sign images used in sign recognition and translation systems is proposed. improved Levenshtein Metric(*iLM*) is utilized in this paper to take advantages of ranks in candidate characters which are not counted in the previous approaches. The distance between candidate characters are calculated by means of weighted operations in terms of ranks. Candidate characters in top 5 are considered and 1.3 million of shop names are used to be matched. The proposed method can improve recognition rates compared to conventional methods by correcting characters misrecognized by a character recognizer from sign images taken by using mobile phone cameras. The proposed approach shows 31.5% of improvements compared to other approaches in recognition accuracy. Improving recognition rates through correcting misrecognized characters not only can reduce troubles in causing some confusions due to wrong information provided to users but also would provide more accurate information to users.

This study consists of five sections. Section 2 investigates background studies on character recognition and corrections of misrecognition. Section 3 describes the improved Levenshtein

Metric(*iLM*) method proposed in this study. Section 4 represents the results of the experiment that applies the proposed algorithm. Section 5 shows the conclusion of this study and future works.

2. Related Work

Most of conventional character recognition methods focus on its recognition by extracting characters from scanned images obtained from formalized types of information including decuments, name cards, and etc. However, recently studies on extracting characters from natural scene images have been conducted actively. The character recognition in natural scene images represents some difficulties because such scene images have various backgrounds differed from formalized documents. Thus, studies on recognizing characters through extracting them from natural scene images and reducing recognition error rates have been variously conducted. Haritaoglu performed a study on extracting and transforming characters from scene images obtained by using personal mobile devices such as camera installed smart phones and PDAs [5]. Yang developed a system that extracts, recognizes, and transforms semantic information and tested the system in various environments [6]. In addition, in [4], a study that translates English characters from sign images to Spanish characters was performed.

Wick applied the Viterbi algorithm (VA), which is based on the probabilistic express of context knowledge in optical character recognition (OCR), for minimizing misrecognition errors [7]. In [8] and [9], a string matching algorithm was applied to perform multifont OCR. In [10], a study on processing subordinate text strings was performed. However, these misrecognition correction methods show generally low correction rates less than 70% and require lots of costs for carrying correction processes[7][9]. Bansal corrected misrecognized words using a method that minimizes the post-probability based on the Markov assumption by limiting dictionary searching for solving previously menthioned problems [11].

Takahashi proposed a post-processing method that uses combinations in characters, which compose words, and candidates in individual sentences [12]. The combination in individual characters considers a case that can compose sentences using less than four documents, which have low frequencies, based on the investigation of application frequencies of characters in words. Also, the method measures their similarity and selects candidate words. The measurement of the similarity was carried out using the Levenshtein Metric through considering the rank of candidates. As the Levenshtein algorithm [13] is able to perform calculations as two text strings represent different lengths, it has been largely used in various fields including spell checkers, OCR systems, natural language translation systems, and etc. However, Levenshtein algorithm does not consider the rank of recognized candidates by determining the cost of each operation such as inserting, deleting, and replacing as 1. Therefore, a method to tackle this problem for increasing the accuracy rates of recognition.

In this study, an Improved Levenshtein Metric post-processing method that corrects the recognition results by means of condidates of recognized characters, which include misrecognition, is proposed to solve such misrecognition in sign images taken by using mobile phone cameras.

3. Correction of Misrecognized Characters using iLM

In this study, an improved Levenshtein Metric (*iLM*) that improves Levenshtein Metric (*LM*) algorithm, which has been most largely used among character correction algorithms, is proposed to correct misrecognition.

3.1 Sign database

A dictionary database is essential to obtain distances between input characters and candidates. In this study, a nationwide shop name database in Korea was established to recognize texts in a limited area of the signs. In the shop name database including entire shops which are about 4 millions, a shop name database with about 1.3 million words was established by excluding duplicated and English character included in signs. Korean is composed by 19 initial phonemes, 21 medial vowels, and 28 final consonants.

All possible combinations in Korean represent 11,172 characters and 2,667 characters are actually used by excluding practically unused characters. Thus, recognizing Korean is relatively more difficult than that of English characters due to its large number of characters in Korean. English texts are classified and recognized as 26 characters with the order of A to Z. In the results of the analysis of the nationwide shop name database, the practically used characters are 1,920. However, 808 Korean characters, which show a total frequency of 98%, are actually used in this study by excluding the characters that are not frequently used to improve recognition rate and time consumption.

3.2 Correction of Misrecognized Characters

A sign image character recognition system using mobile phones consists of three stages as shown in **Fig. 1**. In the first stage, text areas are to be searched in an input sign image and character and background areas in text areas are to be binarized. In the second stage, the binarized character image is segmented into each character image in which the character image distored due to the photographing angle. Then, the corrected image is inputted to an individual character recognizer. Then, the individual character recognizer performs the recognition of input characters one by one. In the third stage, the recognized characters are provided to users after applying post-processing.

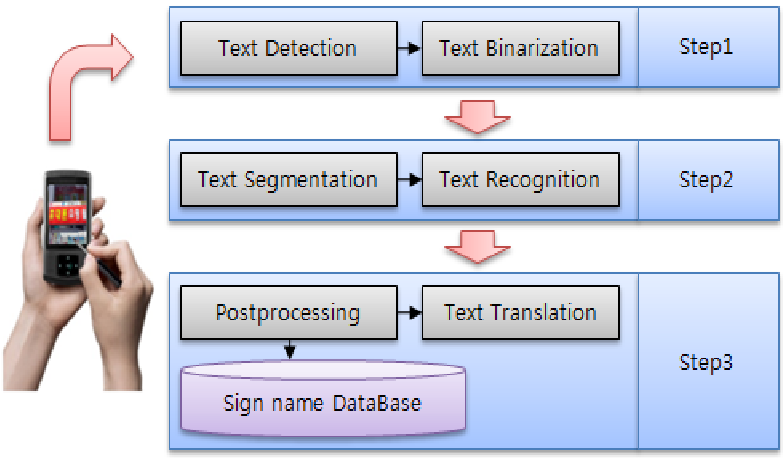

Fig. 1. Sign image character recognition system

Fig. 2. Input data from the recognizer

In this paper we focus on the corrections of misrecognized characters as a post-processing phase in the third stage to leverage the recognition accuracy. **Fig. 2** represents the input values from an individual character recognizer.

The input data has recognition candidates for each recognized character. The input data, R, is defined as Eq. (1).

$$R = \{R_1, \ldots, R_m\}, \tag{1}$$

where $R_1, R_2, \ldots, R_m$ show shop names with m syllables and each syllable has its own recognition candidate syllables. Thus, the extended input data, R', is defined as Eq. (2) by considering such candidates.

$$R' = \{r_{11}, \ldots, r_{mn}\} \tag{2}$$

For instance, if a shop name has m syllables with n candidate syllables for the recognition, the extended input data, R', can obtain a matrix as presented in Eq. (3).

$$R' = \begin{bmatrix} r_{11} & . & . & . & r_{1n} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ r_{m1} & . & . & . & r_{mn} \end{bmatrix} \tag{3}$$

In this paper D is defined as a set of searched shop names which is used to calculate the distance with the input string R to fine the best matched string. The distance between the search results, D and the input string, R is measured to find the most similar string to the input string. In the search of the database, the search is to be performed for the length of the input string, R, ±2. If the search is performed for the string that represents the same length as the input string, R, only, it will cause a problem that does not include a correct string in the search result, D, for the failure occurred in a character segmentation process and the over segmented string. The distance values between the input strings, R , are to be arranged for the obtained search result, D .

Levenshtein Metric (LM) is a representative scale for calculating distances among strings. The LM distance scale measures the calculation cost required in transforming a string to another string as two strings are given. The distance calculation in LM is performed using Eq. (4).

$$LM(R,D) = \min \begin{cases} d(R_{m-1}, D_{k-1}), & // R_m = D_k \\ d(R_{m-1}, D_k) + \alpha, & // \mathit{Insertion} \\ d(R_m, D_{k-1}) + \beta, & // \mathit{Deletion} \\ d(R_{m-1}, D_{k-1}) + \gamma, & // \mathit{Substitution} \end{cases} \tag{4}$$

where α , β , and γ are the calculation costs for insert, delete, and replace operations, respectively. These values are given as $\alpha = \beta = \gamma = 1$ in LM. Thus, the distances of the strings recognized from sign images in **Fig. 3** are calculated as presented in **Fig. 4**.

LM uses the first ranked syllable that shows the highest reliability in the candidates of R. Also, in the calculation for the correction of input characters, calculation costs for each operation of inserting, deleting, and replacing are determined as 1.

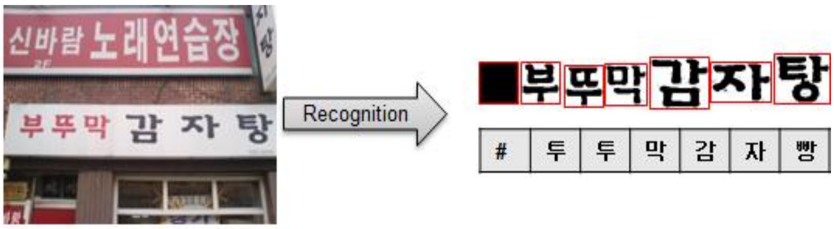

Fig. 3. Recognition of the characters in signs

		#	투	투	막	감	자	빵
	0	1	2	3	4	5	6	7
부	1	1	2	3	4	5	6	7
뚜	2	2	2	3	4	5	6	7
막	3	3	3	3	3	4	5	6
감	4	4	4	4	4	3	4	5
자	5	5	5	5	5	4	3	4
탕	6	6	6	6	6	5	4	4

Inserting(1) Deleting(1) Replacing(1) Stationary(0)

Fig. 4. Calculation process of the Levenshtein Ditstance

In this study, an improved Levenshtein Metric (*iLM*) that is able to consider the rank of candidates for complementing an issue that cannot consider the recognition candidates in *LM* is proposed. In *iLM*, calculation costs are controlled by considering the recognized candidates for replacing operation. The calculation costs of inserting and deleting are determined as 1, while as a weight for the calculation cost of replacing is controlled. That is, if there are some syllables, which agree with the candidate syllables in a replacement process, the calculation will be performed by applying the value of γ in which the weight value is applied. Here, the value of γ is presented in Eq. (5).

$$\gamma = \begin{cases} d(r_{m1} = D_k) + 0 \\ d(r_{m2} = D_k) + \frac{1}{n} \\ d(r_{m3} = D_k) + \frac{2}{n} \\ d(r_{mn} = D_k) + \frac{n-1}{n} \end{cases} \quad (5)$$

The value of γ is given as a weight value determined by between 0~1 according to the rank of considered candidates in which 0 and 1 represent agreement and disagreement, respectively, for each other. A replacing operation is to replace corresponding syllables to other characters. Thus, all characters except one character have the value of 1. The ranks of candidates represent its reliability. As the first rank shows 100% reliability, the continuing ranks represent decreases in its reliability. **Table 1** shows the weight for the replacing operation by considering candidates up to the fifth rank. As syllables that agree with the candidate syllables in considering such candidate syllables exist, calculation costs are to be controlled according to the rank of these candidate syllables.

A more decrease in the rank of candidates requires more calculations costs. If the candidate does not exist, the value of 1 that is the maximum value of the calculation cost will be required. Thus, the higher rank of candidates in recognition results represents shorter *iLM*. **Fig. 5** shows the calculation process of *iLM*.

Table 1. Replacing calculation costs for the candidate strings in *iLM*

input	**Candidate syllables**					
	1	**2**	**3**	**4**	**5**	**N/A**
R_1	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	
R_2	r_{21}	r_{22}	r_{23}	r_{24}	r_{25}	
...	...	...	...	...	...	
R_m	r_{m1}	r_{m2}	r_{m3}	r_{m4}	r_{m5}	
W	W_1=0	W_2=0.2	W_3=0.4	W_4=0.6	W_5=0.8	W=1

인식후보	5	#	브	추	악	김	저	황
	4	#	루	푸	박	갑	지	탑
	3	#	푸	뚜	떡	경	치	탈
	2	#	부	루	먹	강	자	탕
	1	#	투	투	막	감	차	빵
	0	1	2	3	4	5	6	7
부	1	1	1.2	2.2	3.2	4.2	5.2	6.2
뚜	2	2	2	1.6	2.6	3.6	4.6	5.6
막	3	3	3	2.6	1.6	2.6	3.6	4.6
감	4	4	4	3.6	2.6	1.6	2.6	3.6
자	5	5	5	4.6	3.6	2.6	1.8	2.8
탕	6	6	6	5.6	4.6	3.6	2.8	2

Inserting(1) Deleting(1) Replacing(1) Stationary(0)

Fig. 5. *iLM* calculation proces

4. Results of the Experiment

In this study, corrections for misrecognitions were performed using the proposed method after detecting and recognizing text areas in sign images taken by using mobile phone cameras. The recognition results used in this experiment were obtained using three different recognizers. Also, *LM* and *iLM* were compared to verify the efficiency of the proposed method. In this study, the recognition results were obtained using three different recognizers in order to verify the performance of misrecognition correction algorithms according to the performance of these recognizers. Also, for evaluating the performance of the proposed method, the correction rate that represents the level of correction for such misrecognitions and the recognition rate that indicates the improvement in recognition rates after correcting misrecognitions were compared. The correction rate was obtained using Eq. (6).

$$Correction\ \ rates = \frac{\#\,of\ \ right\ correction - \#\,of\ \ wrong\ correction}{\#\,of\ \ misrecognized\ characters} \times 100 \tag{6}$$

The recognition rate was calculated using Eq. (7).

$$recognition\ rates = \frac{\#\,of\ \ recognized\ signs}{\#\,of\ \ total\ signs} \times 100 \tag{7}$$

Table 2 shows the recognition and correction rates after correcting the misrecognitions obtained using three recognizers through applying four algorithms, *LM* and *iLM*. In the results of this experiment, *LM* that uses the recognition results of the first ranked candidate only without considering recognition candidates showed the recognition rate of 52.7% on average and that represents an increase in the recognition rate of 9.2% on average compared to that of before correcting the misrecognitions.

Table 2. Correction results of the misrecognitions

Methods	Recognized	misrecognition	Rate of recognition	# of success	# of failure	Recognition rate after correction	improvement	Rate of correction
LM	A	176	12%	34	142	29%	17%	19.32%
	B	99	50.5%	14	85	57.5%	7%	14.14%
	C	64	68%	7	57	71.5%	3.5%	10.93%
	Average	113	43.5%	18.3	94.7	52.7%	9.2%	14.79%
iLM	A	176	12%	88	88	56%	44%	50%
	B	99	50.5%	92	7	96.5%	46%	92.93%
	C	64	68%	61	3	98.5%	30.5%	95.31%
	Average	113	43.5%	80.3	32.7	83.6%	40.7%	79.41%

Table 3. Top five ranks in correcting misrecognitions

Original	Recognitions result	Ranks	*LM*	*iLM*
고흥식당	코 고 교 꼬 호 몸 봄 흥 찜 용 삭 식 허 석 직 담 탈 당 탑 탕	1	남교식당	**고흥식당**
		2	남흥식당	코너식당
		3	내흥식당	코보식당
		4	늘봄식당	코스식당
		5	담비식당	코아식당
삼화페인트	상 삼 심 싱 살 회 합 할 화 희 페 패 때 깨 메 언 인 연 안 얀 토 트 돌 들 르	1	상남페인트	삼회페인트
		2	상명페인트	**삼화페인트**
		3	상미페인트	상남페인트
		4	상아페인트	상명페인트
		5	상아탑페인트	상미페인트
광주상사	광 평 찜 망 낌 추 주 푸 루 투 싱 상 삼 심 성 샤 사 서 시 치	1	나주샤시	**광주상사**
		2	남광상사	백광주상사
		3	남주상사	광경상사
		4	낭주상사	광근상사
		5	내광상사	광남상사
만민장의사	깜 깔 낌 관 만 흰 던 믿 든 민 잠 장 꿈 찰 공	1	나동장의사	**만민장의사**
		2	나사렛의자	농민장의사
		3	나주장의사	만경장의사

	외 의 와 억 먹 사 자 서 샤 차	4	낙동장의사	만궁장의사
		5	낙영장의사	만물장의사
강동오케익	강 간 갑 긴 경 동 옹 등 청 둥 모 오 으 호 효 배 페 베 깨 메 이 대 취 미 질	1	낙동강오리	강동오토바이
		2	네모페이스	강동꽃배달
		3	네오페이스	강동배관
		4	대강이앤비	강동오치과
		5	대구바이오	**강동오케익**
월드문구	밀 월 필 활 말 드 트 도 두 돈 문 둔 운 분 윤 쿠 구 꾸 분 무	1	남도문구	**월드문구**
		2	능도문구	영월드문구
		3	다운문구	월드문구점
		4	다윤월드	랜드문구
		5	닥트월드	월드문고

iLM that considers recognition candidates showed the recognition rate of 83.6% on average and that represents an increase in the recognition rate of 40.7% on average compared to that of *LM* as much as 31.5%. **Table 3** shows the results of the correction using these algorithms. In *iLM*, the strings that include characters appeared in the rank of candidates were usually presented in recognition results candidates.

5. Conclusion

In this study, the distances between the candidates employed in sign firm name database were calculated by considering the recognition candidate rank based on the recognition results of a mobile phone camera based sign image recognition system. Also, a method that corrects the obtained data as the most similar firm name was proposed. For establishing a database for candidate groups, a nationwide firm name database with about 1.3 million firms was established. In addition, by investigating Korean characters used in this database, the characters used in the recognizers were decreased from 2,667 to 808. For evaluating the performance of the proposed system, post-processing works using *LM* and *iLM*, which is proposed in this study, were applied to the results recognized by using three different recognizers. Then, the recognition and correction rates were measured after applying these post-processing works.

The proposed *iLM* showed increases in the recognition and correction rates of 31.5% and 64.62%, respectively, compared to that of the existing *LM*. In general, *iLM* indicated better performances in correcting misrecognitions than that of *LM*. However, the processing speed of *iLM* showed a lower level than *LM* because it required lots of memory spaces and calculations in its calculation process. Since the processing speed is the salient points in the low computational devices such as a smart phone, the future works are required to solve the problem of processing speed by reducing comparison steps and memory usage.

In addition, we want to extend this study to other languages such as English, Japanese, Chinese etc. since this model doesn't dependent on any language structures. We will also continue our research on various images such as label images of products for more accurate text recognition using mobile cameras.

References

[1] A. Wojciechowski and K. Siek, "Barcode scanning from mobile phone camera photos delivered via MMS: Case Study," in *Proc. of 27th Int. Conf. on Conceptual Modeling*, pp.218-227, Oct.2008. Article (CrossRef Link)

[2] D.M. Chen, S.S. Tsai, R.Vedantham, R. Grzeszczuk and B. Girod, "Streaming mobile augmented reality on mobile phones," in *Proc. of 8th IEEE International Symposium on Mixed and Augmented Reality*, pp.181-182, Oct.2009. Article (CrossRef Link)

[3] C. Mancas-Thillou and B. Gosselin, "Natural scene text understanding," *Vision Systems: Segmentation and Pattern Recognition*, pp.307-333, Jun.2007. Article (CrossRef Link)

[4] A. Canedo-Rodriguez, S.H. Kim, J.H. Kim and Y. Blanco-Fernandez, "English to Spanish translation of signboard images from mobile phone camera," in *Proc. of IEEE SoutheastCon*, pp.356-361, Mar.2009. Article (CrossRef Link)

[5] I. Haritaoglu, "Scene text extraction and translation for handheld devices," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.408-413, Dec. 2001. Article (CrossRef Link)

[6] J. Yang, X. Chen, J. Zhang, Y. Zhang and A. Waibel, "Automatic detection and translation of text from natural scenes," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp.2101-2104, May.2002. Article (CrossRef Link)

[7] M.L.Wick, M.G. Ross and E.G. Learned-Miller, "Context-Sensitive error correction: Using Topic Models to Improve OCR," in *Proc. Of 9th Int. Conf. on Document Analysis and Recognition*, pp.1168-1172, September, 2007. Article (CrossRef Link)

[8] W.S. Rosenbaum and J.J. Hilliard, "Multifont OCR Postprocessing System," *IBM Journal of Research and Development*, vol.19, pp.398-421, Jul.1975. Article (CrossRef Link)

[9] S. Dobrisek, J. Zibert, N. Pavesic and F. Mihelic, "An edit distance model for the approximate matching of timed strings," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol.31, pp.736-741, Apr. 2009. Article (CrossRef Link)

[10] R.S. Boyer and J.S. Moore, "A fast string searching algorithm," *Communications of the ACM*, vol.20, pp.762-772, Oct.1977. Article (CrossRef Link)

[11] V. Bansal and R.M.K. Sinha, "Integrating knowledge sources in Devanagari text recognition system," *IEEE Transactions on Systems, Man, and Cybernetics,* vol.30, pp.500-505, Jul.2000. Article (CrossRef Link)

[12] H. Takashi, N.I. Amano and A. Yamashita, "A spelling correction method and its application to and OCR system," *Pattern Recognition*, vol.23, pp.363-377, May.1990. Article (CrossRef Link)

[13] T. Okuda, E. Tanaka and T. Kasai, "A method for the correction of Garbled Words based on the Levenshtein Metric," *IEEE Transactions on Computers*, vol.C-25, pp.172-178, Feb.1976. Article (CrossRef Link)

Myung-Hun Lee received B.S. in Computer Science from Gwangju University and his M.S. degree in Electrical & Computer Engineering from Chonnam National University. His recent works are to develop media service in Konan Technology. His main interests are Data Mining, Information Retrieval based on Smart Phone.

Soo-Hyung Kim 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 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 received the B.S in Multimedia from Korean Educational Development Institute in 2004 and the M.S. degree in Computer Science from Dongguk University, Korea in 2006. She received the Ph. D. degrees in Computer Science from Chonnam National in 2011. She recently works in Carnegie Mellon University as a researcher. Her research interests include Data Mining, Sensor Mining and Bioinformatics.

Hyung-Jeong Yang received the B.S. and M.S. degrees in Computer science from Chonbuk National University, Korea in 1991 and 1993, respectively. She received the Ph. D. degree in Computer Science from Chonbuk National University, Korea in 1998. She worked in Carnegie Mellon University, USA as a researcher from 2003 to 2005. Since 2005, she has been a professor in the Electrical & Computer Engineering at Chonnam National University in Korea. Her research interests include e-Design, Data Mining and e-Learning.