ISSN:0975-9646

Survey of Pattern Recognition Approaches in Japanese Character Recognition

Soumendu Das, Sreeparna Banerjee

School of Engineering Technology
West Bengal University of Technology, India

***Abstract*— Optical Character Recognition (OCR) in Japanese, both handwritten and printed, is difficult to perform, owing to several reasons. Firstly, the Japanese language is comprised of over 3000 characters which can be classified as syllabic characters, or Kana, and ideographic characters, called Kanji. Secondly, Japanese text does not have delimiters like spaces, separating different words. Thirdly, several characters in the Japanese alphabet could be homo-morphic, i.e. have similar shape definition which could add to the complexity of the recognition process. In this paper, we present a survey that has been conducted of some of the approaches that have attempted to address these issues and implement schemes for Japanese character recognition in texts.**

***Keywords*— Japanese character recognition hiragana, katakana, kanji, handwritten, printed document analysis.**

I. INTRODUCTION

The storage, retrieval and analysis of Japanese documents require a conversion of the documents to electronic media. For the purpose of conversion to electronic (digitized) media, optical character readers are required. Due to the complexity of Japanese handwritten as well as printed text, this process is error prone. Furthermore, there exists a necessity of efficiently combining optical character recognition along with document retrieval techniques. In Japanese, this process is complicated by the fact that the text comprises of both the syllabic/ phonetic characters (Kana) as well as the ideographic characters (Kanji) and similar shape definition of several Japanese characters. Secondly, Japanese text is not separated by delimiters such as spaces. Thirdly, homomorphism or similar shape definition for different Japanese characters also poses problems especially in sans serif fonts. Hence, current research on OCR for Japanese attempts to address these issues. This survey describes several approaches described in the literature, for Japanese text recognition and retrieval.

II. JAPANESE LANGUAGE MODEL

Japanese text is written as a combination of syllabic/phonetic characters known as kata as well as ideographic characters known as Kanji. These are described in the following paragraphs.

A. Japanese Text

Japanese text is written using more than 3000 characters, many of which have complex and similar shapes, and the text is not separated by delimiters such as spaces:

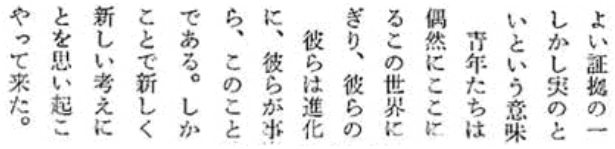

Fig. 1 Sample Japanese Text

B. Japanese Character Sets

Japanese writing system has three different characters sets, namely, Hiragana, Katakana and Kanji. For Japanese words, Hiragana (see Fig2a) is used, mostly for grammatical morphemes. Katakanas (see Fig2b) are used for transcribing foreign words, mostly western, borrowing and non-standard areas. In addition, diacritic signs like dakuten and handakuten are used (see Fig3 and 4)

あ	い	う	え	お
a	i	u	e	o
か	き	く	け	こ
ka	ki	ku	ke	ko

Fig. 2a Hiragana Script

ア	イ	ウ	エ	オ
a	i	u	e	o
カ	キ	ク	ケ	コ
ka	ki	ku	ke	ko

Fig. 2b Katakana script

Dakuten are used for syllables with a voiced consonant phoneme. The dakuten glyph (゛) resembles a quotation mark and is directly attached to a character (Foljanty 1984).

が	ぎ	ぐ	げ	ご
ga	gi	gu	ge	go
ガ	ギ	グ	ゲ	ゴ
ga	gi	gu	ge	go

Fig. 3 Hiragana and Katakana Dakuten Alphabets

Handakuten are used for syllables with a /p/ morpheme. The glyph for a 'maru' is a little circle (゜) that is directly attached to a character (Foljanty 1984).

ぱ	ぴ	ぷ	ぺ	ぽ
pa	pi	pu	pe	po
パ	ピ	プ	ペ	ポ
pa	pi	pu	pe	po

Fig. 3 Hiragana and Katakana Handakuten Alphabets

Kanji are content bearing morphemes. In Japanese text Kanji are written according to building principles like Pictograms (graphically simplified images of real artefacts), ideograms (combinations of two or more pictographically

characters) and phonograms (combinations of two Kanji characters).

C. *Inter-Word Spacing in Japanese*

Japanese words are not separated by delimiters like spaces, thus making character recognition difficult. Although Japanese is a word based language, segmenting text into word is not as clear cut as in languages using word spacing as a rule. Spacing is incorporated as in at least two ways. The first way is by adding spaces not only between their grammatical modifiers and post positions. The second way is to consider the modifiers and post positions as a part of the modified word. Based on the study conducted by Saino et al. [1] using 16 subjects in Japanese reading, 60 word texts from excepts of newspapers and internet columns, it was concluded that in pure Hiragana text, inter-word spacing is an effective segmentation method, in contrast to Kanji-Hiragana text, since visually silent kanji characters serve as effective segmentation uses by themselves.

D. *Character Features Vectors Identification*

Every character has its own features and identities. By identifying features we can recognize characters from a textual image document. By feature extraction the critical characteristics of characters gets isolated, and that reduces the complexities of the pattern. After classification it compares with known patterns and then matched with the character that has the same characteristics. The characters can be further subdivided into segments and the strokes in each of these segments exhibit certain characteristics in terms of shape, angle of inclination. In addition, presence of dakuten and handakuten also changes the character. All these aspects need to be taken into account in devising feature vectors for identification.

III. EARLIER ATTEMPTS

Three approaches have been attempted to overcome the problems due to imperfect recognition, as described below.

A. *Error Correction*

Recognition output is made clear by the error conversion. This can be done using a spelling checker which is capable of integrating characteristic patterns of recognition errors which differ from normal typing errors. A second method is to use linguistic knowledge [2] which includes knowledge about the content of documents [3] in addition to syntactic and lexical knowledge. A third method is the category utilizes vocabulary derived from similar documents in order to improve the word recognition rate [4].

B. *Document Search Without Optical Character Recognition*

The document processing system "Transmedia Machine" is used for this purpose [5]. Character images of scanned documents are encoded into two binary features for each character succeeded by a "string matching" based on incomplete codes. Word-level encoding [6] has been proposed as a more reliable alternative. Searching for text passages in document image database and subsequent pattern matching using a number of feature descriptors has also pattern been proposed [7].

C. *Error Tolerant Search System*

In order to make the search system tolerant of recognition error, multiple candidates have been used in the search process [8]. The optical character recognition keeps multiple candidates for ambiguous recognition and outputs them as a result text. Segmentation ambiguities [9] can also be included, with multiple hypotheses in both character segmentation and recognition represented as a network of hypotheses.

IV. DOCUMENT RETRIEVAL TOLERATING CHARACTER RECOGNITION ERRORS

Marukawa et al. [10] have proposed two methods of combining character recognition for retrieving Japanese documents. In their recognition process they used a multi-template based on directional features. The segmented character pattern is normalized and the contour derived from this geometrically normalized binary pattern is represented by eight directional codes. Each directional code is mapped into one of the four feature patterns. Each of them corresponds to the horizontal, vertical, right-up, and left-up directions. These features are blurred into a "directional feature" pattern.

Their first method [10] is the Extended Query Term Method using (Method I) confusion matrix, which uses two steps. In the first step characters similar to a character in the query terms by using a confusion matrix, and strings combining the "similar characters" are expanded as new query terms. In the next step new query terms are created by supplemented similar characters. In their second method, (Method II) [9] non-deterministic text, which keeps multiple candidates in a text file, is used. Searching is done by using clean query terms.

V. BI-GRAM AND ITS APPLICATION TO ONLINE CHARACTER RECOGNITION

Because Japanese language has a huge character set including characters with different entropies it is difficult to apply conventional methodologies based on n-gram to post-processing in Japanese character recognition. Itoh [11] proposed a method to overcome these two problems using a clustering scheme based on different parts of speech of Kanji and also by homogenizing the entropies of different Kana and Kanji characters. A bi-gram approach was used, based on these two techniques to Japanese language model. Experiments resolved the imbalance between Kana and Kanji characters, and reduced the perplexity of Japanese to less than 100, when Japanese newspaper texts were used. A post-processing technique was proposed using the model for on-line character recognition and about half of all substitution errors were obtained when the correct characters were among the candidates.

This approach needs to be extended to Katakana characters. Among the different parts of speech, verbs and post positions were considered extensively, but more minute classification of nouns were required. Finally the language model was applied as an online optical character recognition post processing method, caused failures in cases where the correct character is not included in the candidates. Integration of the language model into the recognition methodology should be attempted.

VI. Offline Character Recognition Using Virtual Example Synthesis

In character recognition, both for printed and handwritten character recognition, the performance of classifiers strongly depend on quality of naming samples. A very large database containing a sufficiently large number of good examples are required for classifiers to perform well, particularly in the case of hand written character samples. This is costly and time consuming. Miyao and Maruyama [12] attempt to overcome this difficulty by synthesizing virtual examples from a small number of real samples. Their approach is implemented in two steps.

Their results indicated that with an appropriate number of eigenvectors and base samples, the recognition rates are higher than or equal to those without PCA based pattern segmentation and the classification time is faster as the support vector of support vector machine is further reduced for recognition of handwritten Hiragana characters due to (I) determination of cumulative recognition rate for improvement in effectiveness and (II) designing a decision directed acyclic graph based on support vector machines (SVM).

VII. Stricr-Fb Algorithm

Barners and Manic [13] proposed an algorithm that contributes an original approach to constructing feature vectors; their proposed methodology creates a neural network by design and not by training. They showed that the centre of gravity remains consistent even a character is rotated. The centre of gravity will move proportionally as the characters change in size, translation, or rotation. The dakuon and handakuon characters will also be accurately identified due to the presence of the dakuten and handakuten markers. The centre of gravity moves proportionally with the additional pixels and produces a set of unique feature characteristics.

The Size-Translation-Rotation-Invariant Character Recognition and Feature vector Based STRICR-FB algorithm is based on the Kohonen Winner Take All [13, 14], type of unsupervised learning. The algorithm comprises of two phases; Construction of Character Unique Feature Vectors which calculates distance between characters and in an expanded form of the Euclidean distance defined in (15). The next phase is passing character unique feature vectors through a neural network for character recognition.

They conducted three list sets to validate the algorithm. In general, a training set is used to create an artificial neural network (artificial neural network). A test set of random characters is then used to determine the effectiveness of this artificial neural network. The experiment by random characters produces three sets of results; among which rotation set produced 96.2% accuracy rate and that of random character set is 93%.

Using this technique [14] Das and Banerjee have attempted to perform character recogntion on Japanese Hiragan characters with an accuracy of 92.2%.

VIII. Performance Improvement Strategies On Template Matching

Handwritten text differs due to differences in writing styles, and hence, handwritten character recognition suffers from absorbing variations of the same characters among different writing styles. Also, resolution of the graphical similarity of different characters in Japanese text is another consideration to be taken into account. To overcome the problem, an offline effective algorithm for large scale character recognition for large set characters like Korean and Chinese was proposed by Kim [17]. The algorithm was developed based on template matching and improvement strategies; First, Multi-stage pre-classification that reduces the processing time of the template matching by cutting off a number of recognition target classes [18] is done. It is desirable to cut off as many classes as possible with little or no degradation of recognition accuracy. Second, the pair wise reordering is done to enhance the recognition accuracy by performing a fine detail classification on the recognition candidates generated from the template matching [19].

The resulting algorithm consists of three processing stages of multi-stage pre-classification, template matching and pair wise reordering. The algorithm showed its effectiveness by an experiment where handwritten Korean character came up with 86.0% of recognition accuracy and 15 characters per second from PE92 [20] handwritten Korean character database.

IX. Recent Attempts

Most of the recent Japanese character recognition approaches, both for handwritten and printed text, either use soft computing based approaches for classification, or, image shape/morphology characteristics for classification. These approaches are outlined in the following.

A. Soft Computing approaches.

Among the soft computing approaches, Neural Networks (NN) are most widely used, though Genetic algorithms and support vector machines have also been explored. In the following paragraphs are some of the approaches from [19], listed in Table I.

Printed multi-font Kanji characters from books Table I [Sl.1] have been directtly classified using Support Vector Machines (SVM). The Peripheral Direction Contributivity (PDC) of given Kanji character images for learning and recognition are used with SVM. The recognition rate reported was 92%.

Simple recurrent NN have been used in [Sl.2] to recognize offline handwritte character based on template matchnig have been proposed. A template pattern and an unknown

pattern are divided into some partial regions and each partial region is input into simple recurrent networks. Simple recurrent networks calculate similarities between partial regions of two patterns and recognize the unknown pattern. In [Sl. no.4] multi-layered neural networks have been applied to handwritten character recognition with directional element feature learning and noise injection put into inputs by a theoretical analysis and simulation. In [Sl. no. 5] a 3 layer Back Propagation NN along with a copy learning model is used. The copy-learning model is a neural network which learns standard characters by using a part of the 3-layer BP net, copies obtained weights and biases on an unused part of the net based upon a rule, and recognizes the standard characters from the superimposed characters by using all parts of the net. For character recognition problem including deformation characters, such as ancient documents[[SL.no.7], the recognition system using neural network is considered to work effectively. The recognition accuracy of the system is demonstrated by comparing with a statistical method. [Sl. no.8]. A peripheral localoutline vector (P-LOVE) is used in conjunction with an unified NN in [Sl. no. 9].P-LOVE extracts a feature at aperipheral point on a character which is scanning from 8 directionsaround the character. The NN based matching of handwritten characters is done on this basis.

Genetic algorithms (GA) have also been used as classifiers. In [Sl. no. 3], a GA based method has been proposed to classify diversiform handwritten characters. Generated characters have been compared with those existing in an ETL database. The recognition efficiency was claimed to be very high (> 90%). A post processing character recognition using GA was also reported in [sl. no. 10]. After a character recognition system selects suitable candidates, a post-processing system combines them and uses again for word or sentence recognition. First the post-processing system by GA combines a number of sentences from candidates randomly. It evaluates each sentence by the approach of the natural language. The low evaluated sentences are excluded by selection. Suitable sentences are combined by the crossover. Chances for correct sentence combination are low if random crossover is used. In our approach the sentences crossover by utilizing three kinds of the crossover masks effectively.

Hidden Markov Models (HMM) have also been used for Japanese character recognition. In [s-no 6], off-line character recognition has been performed using HMM.They proposed a directional HMM based character recognition method that integrates 4 simply structured 1D HMMs all of which are based on directional feature extraction using linear filters.The results of our evaluation experiment using the Hand-Printed Character Database (ETL6) . The average recognition rate of 98.5% was claimed with the cumulative recognition rate of the top three was 99.3%.

B. Image Shape based approaches.

Image shape based approaches characteristically perform a segmentation of the candidate character and compare with a template of the prototype character template. The size, shape, extent and angular inclinations of the different strokes have also been taken into account in many instances. Additionally, the images of the characters (both candidate and prototype) have been divided into segments and the contours of the strokes pertaining to each of the segments have been studied. These processing prove to be helpful, especially for handwritten character texts where there are variations. Some of the major efforts are summarized below. A listing of the approaches obtained from [19] is given in Table II.

Proposal of character segmentation using the convergence of the shortest path has been described in [Sl. no.2] where boundaries between single characters are detected. The proposed method is to detect the starting point using more than one shortest path that has a different starting point with the end beingf the same. The shortest route from nature, is to concentrate on the boundary between the patterns, in the proposed scheme, to cut out and get into contact.

An aggregating visual confirmation method for character recognition result by OCR to reduce the cost of document scanning and capture has been mentioned in [sl. no.1]. First, this method has a clustering algorithm which is based on the recognition result and on the shape feature of the clustering algorithm which is based on the recognition result and on the shape feature of the string image. Second, it creates the composite image by superimposing the images which are classified in the same cluster. The composite image enables operators to determine whether it consists of the same string image. Therefore, the operators can confirm and correct the recognition results in the same clusters at a time by comparing them.

A Fast Deformable template matching for character recognition in scene images has been described in [sl. no. 4]]. They perform a new fast character segmentation method of characters including the isolated and the touching ones with robustness for perpective distortion and character deformations. By applying the proposed method to scene images,it is confirmed that the method is robust against perspective distortion and character deformations,and the method is superior in computational time against a simple template matching method.

[sl. no 5], an on-line Cuneiform modeled handwritten Japanese character recognition method has been suggested. The features of the algorithm are: i) newly developed cuneiform model which divides continuous handwritten strokes into segments of lines (cuneiform) and tries to match the cuneiform model to the dictionary; ii) calculation of the distance taking both the number of cuneiform and increase in the number of cuneiform segments. An 83.9% accuracy has been reported.

[sl.no.6] proposes the recognition of handwritten character string using a complete correspondence method. A segmentation candidate lattice is used for the character string in Japanese address. Segmentation processing divides a characters string of handwritten address pattern into small basic segments by labeling processing, and this processing creates a segmentation candidate lattice by combining basic segments.The proposed technique is using a complete correspondence methodas a role of post-processing. This method limits a recognition objecto the address of the same number of characters as the number of rectangles of segmentation candidate. A 97.57 % recognition rate has been reported.

[sl. no.7] describes an off-line handwritten character recognition method using a relaxation process based on a

character shape model. The probability of unknown character pattern is calculated from a skeleton pattern model which consists of a writing sequence of feature points, and the recognition process is formalized as a likelihood estimation problem of the probability.

A adaptive online handwriting character recognition system which combines adaptive classifier with adaptive context processing has been developed in [sl. no 8]. The adaptive context processing automatically collects

misrecognized patterns inputted by a user so that the system increases recognition accuracy for repeatedly input characters. The adaptive context processing automatically collects the strings inputted by a user so that the system reduces misrecognition of repeatedly input strings. The adaptive recognition system has a mechanism which prevents adaptive context processing from causing over adaptation even if the adaptive context processing is combined with the adaptive classifier. In our experiments, while 11 subjects of all 14 subjects answered that they were satisfied with the recognition accuracy of the adaptive system, 6 subjects answered that they were satisfied with that of the non-adaptive system.

An On-line character recognition algorithm using Reparametrized Angle Variations (RAV) has been proposed in [sl. no. 9]. The recognition rate claimed was was 91.2% while cumulative top three recognition rates was 96.2%. The proposed algorithm RAV has several important features; (i) Raw data consisting of pen position trajectory is transformed into angle variation and resampled in a simple but very effective manner, (ii) A special distance function is proposed to evaluate distance between two characters taking into account the angle variations as well as pen up/down variations, and (iii) An automatic dictionary generation scheme is proposed in order to copy with stroke order variations.

Another adaptive context processing processing algorithm for handwritten character recognition has been presented in [sl. no.10]. This system automatically collects the strings inputted by a user and regards the collected strings as the context of the user. Based on the context, it chooses the string among the candidate strings returned by the pattern recognition module. They concluded that this method is effective for the following two reasons: it improves recognition accuracy for each user; and it reduces the stresses from the same misrecognitions being repeated and enhances user's satisfaction.

A novel OCR error correction method for languages without word delimiters and with a large Japanese character set is proposed in [sl. no. 11] . The usefulness of this method lies in the fact that it is an error correction method for languages without word delimiters such as Japanese. It consists of a statistical OCR model considering character shape similarity, a statistical language model considering unknown words, an approximate word matching method that can handle short words, and an N-best word segmentation algorithm. The method is accurate and robust enough to handle a wide range of Japanese sentences and character recognition accuracy. When the baseline character recognition accuracy was 70%, 80%, and 90%, it achieved 88.3%, 94.1%, and 97.4% character recognition accuracy, respectively.

C. *Named Entities in Japanese text*

Since the Japanese text does not have word boundary markers, a character recognizer which can extract named entitites (NE) is very useful. One such effort is presented in Table II [sl. no. 3]. This paper proposes two methods for improving the performance of Japanese Named Entity (NE) extraction. The first one is the combination of word unit and character-unit extraction. Most Japanese NE extractors use words segmented by a Japanese morphological analyzer because Japanese language has no word boundary marker. However, word unit is not always consistent with NE unit. To solve this problem, a method is proposed to combine word-unit and character-unit extraction. The other is feature argumentation techniques by using extraction results of NE from not-labeled data with NE extractors. Their method collects the candidate NE classes of each word and the NE classes of its surrounding words from not-labeled data. They use these collected information of each word as features.

The recent approaches, obtained from a website [s. give link] are listed in tables 1 (for soft computing approaches) and 2 (for image shape model based approaches.

X. CONCLUSION

From the preceeding sections it can be seen that numerous researchers are involved in developing and improving Japanese character recognition systems. Initial efforts focussed on error corection. Subsequently, bi-gram based approaches and virtual example synthesis have been attempted. A novel algorithm STRICR-FB for Korean alphabets, which the authors have adopted for Japanese Hiragana characters has also been presented. Template based algorithms have gradually been introduced in Japanese text. This led to the advent of more ecent techniques using soft computing and image shape models described in the later sections.

It is hoped that, with the proliferation of various character recognition techniques and advancement of software methodologies, Japanese character recognition will be error free in the near future, thus enabling electronic documentation in large databases.

REFERENCES

[1] Miia Sainio, Kazuo Bingushi, Raymond Bertram; Vision Research (2007); "The role of interword spacing in reading Japanese : An eye movement study"; Volume: 47, Issue: 20, Pages: 2577-2586

[2] D.A. Dahl, L. M. Norton and S. L. Taylor, Improving OCR accuracy with linguistic knowledge, Proc. Second Ann. Symp. Document Analysis and Information Retrieval, pp. 169-177 (1993).

[3] N. Niwa, K. Kayashima and Y. Shimeki, Postprocessing for character recognition using keyword information, IAPRWorkshop Machine Vision Applications, pp. 519-522 (1992).

[4] J. J. Hull and Y. Li, Word recognition result interpretation using the vector space model for information retrieval, Proc. Second Ann. Syrup. Document Analysis and Information Retrieval, pp. 147-155 (1993).

[5] Y. Tanaka and H. Torii, Transmedia machine and its keyword search over image texts, Proc. RlAO'88, pp. 248-258 (1988).

[6] J. M. Trenkle and R. C. Vogt, Word recognition forinformation retrieval in the image domain, Proc. Second Ann. Symp. Document Analysis and Information Retrieval, pp. 105-122 (1993).

[7] J. J. Hull, Document image matching and retrieval with multiple distortion--invariant descriptors, Proc. IAPR Workshop on Document Analysis Systems, pp. 383-399 (1994).

[8] H. Fujisawa, A. Hatakeyama, Y. Nakano, J. Higashino and T. Hananoi, Document storage and retrieval system, U.S. Patent 4985863 (1986).

[9] S. Senda, M. Minoh and K. lkeda, Document image retrieval system using character candidates generated by character recognition process, Proc. Second Int. Conf. Document Analysis and Recognition, pp. 541-546 (1993).

[10] Katsumi Marukawa, Tao Hu, Hiromichi Fujisawa and Yoshihiro Shima; "Document retrieval tolerating character recognition errors—evaluation and application"; Pattern Recognition, Volume 30, Issue 8, August 1997, Pages 1361-1371, Oriental Character Recognition.

[11] Itoh, N.[Nobuyasu], "Japanese language model based on bigrams and its application to on-line character recognition,"; PR(28), No. 2, February 1995, pages: 135-141.

[12] Maruyama, K.-I.; Maruyama, M.; Miyao, H.; Nakano, Y.; "Handprinted Hiragana recognition using upport vector machines"; Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop Digital Object Identifier: 10.1109/IWFHR.2002.1030884; 2002, Pages: 55 – 60.

[13] Barnes, D. Manic, M. Univ. of Idaho, Moscow, ID, USA "STRICR-FB, a Novel Size-Translation- Rotation-Invariant Character Recognition Method" Human System Interactions (HSI), 2010 3rd Conference on May 2010, pp.163 – 168.

[14] Das S. and Banerjee S. A Pattern Recognition Approach for handwritten Hiragana character identification, d.o.i.: 978-1-4673-1344-5, Proceedings IEEE-ICCIC 2012, Chennai, India.

[15] S.H. Kim, "Performance Improvement Strategies on Template Matching for Large Set Character Recognition," Proc. 17th International Conference on Computer Processing of Oriental Languages, pp. 250-253, Hong Kong, April 1997.

[16] C.H. Tung, H.J. Lee and J.Y. Tsai, “Multi-stage pre-candidate selection in handwritten Chinese character recognition systems”, PatternRecognition, vol. 27, no.8, pp.1093-1102, 1994.

[17] H. Takahashi and T.D. Griffin, “Recognition enhancement by linear tournament verification”, Proc. 2nd ICDAR, Tsukuba, Japan, 1993, pp.585-588.

[8] D.H. Kim, Y.S. Hwang, S.T. Park, E.J. Kim, S.H. Paek and S.Y. Bang, “Handwritten Korean character image database PE92”, Proc. 2nd ICDAR, Tsukuba, Japan, 1993, pp.470-473.

[19] http://www.ipsj.or.jp/english/index.html

TABLE I SOFT COMPUTING APPROACHES

[1]	Evaluation of the SVM based Multi-Fonts Kanji Character Recognition Method for Early-Modern Japanese Printed Books	Yurie Enomoto , Masami Takata , Tsukasa Kimesawa , Kazuki Joe	研究報告数理モデル化と問題解決（MPS ）,2011-MPS-82(14),1-6 (2011-02-28)
[2] 2008	Neural Networks which Consist of Simple Recurrent Networks for Character Recognition by Template Matching	Yoshihisa Shinozawa Faculty of Science and Technology, Keio University	Journal of Information Processing Society, 49 (10),3703 -3714 (2008-10-15), 1882-7764
[3] 2003	A Deformed Character Generation by a Computational Handwriting Model and a Genetic Algorithm	Hiroyuki Kasuga , Keiti Sumida , Kei Ohkawa , Yasuhiro Wada	情報処理学会論文誌,44(9),2363-2373 (2003-09-15) , 1882-7764
[4] 2003	Handwritten Character Recognition Using Directional Element Feature and Noise Injection into Neural Network	Terumitsu Ohyama , Hideo Kuroda , Sueharu Miyahara , Osamu Shiku , Hideaki Takahira	情報処理学会論文誌,44(5),1368-1371 (2003-05-15) , 1882-7764
[5] 2003	A Copy-learning Model for Recognizing Superimposed Character	Kenichi Suzaki , Mituhiro Inoue , Taketoshi Momohara , Shinji Araya	情報処理学会論文誌,44(4),1173-1176 (2003-04-15) , 1882-7764
[6] 2002	Off-line Hand-printed Character Recognition System Using Directional HMM based on Features of Connected Pixels	Hiromitsu Nishimura , Masayoshi Tsutsumi , Minoru Maruyama , Hidetoshi Miyao , Yasuaki Nakano	情報処理学会論文誌,43(12),4051-4058 (2002-12-15) , 1882-7764 Journal of Information Processing Society, 43 (12),4051 -4058 (2002-12-15), 1882-7764
[7] 2001	The Neural Networks for Offline Handwritten CharacterRecognition by Template Matching	Yoshihisa Shinozawa , Seiichi Okoma	情報処理学会論文誌,42(1),16-25 (2001-01-15) , 1882-7764
[8]	A Study for Character Recognition of Ancient Documents using Neural Network	Yuji Waizumi , Nei Kato , Yoshiaki Nemoto , Shoji Yamada , Mamoto Shibayama , Hiroshi Kawaguchi	Computers and the Humanities IPSJ SIG Technical Reports (CH), 2000 (8 (1999-CH-045)) ,9-15 (2000-01-21)
[9] 1999	An Unified Neural Network for Handwritten Similar Characters Recognition by Using Peripheral Local Outline Vector	Keitaro Hori , Hiroki Sugawara , Akiyoshi Itoh	情報処理学会論文誌,40(12),4239-4247 (1999-12-15) , 1882-7764
[10] 1999	The Post processing of Character Recognition by Genetic Algorithms	Yoshihisa Shinozawa , Seiichi Okoma	Journal of Information Processing Society, 40 (3),1106 -1116 (1999-03-15), 1882-7764

TABLE II IMAGE SHAPE MODEL BASED APPROACHES

[1] 2012	Aggregating Visual Confirmation Method for Character Recognition Result	Yasuaki , Iwata Mitsuharu , Oba Toshiko , Matsumoto Takashi , Onoyama Hitachi Solutions, Ltd.	研究報告コンピュータビジョンとイメージメディア（CVIM）Volume: 2012-CVIM-180, No:4, date : 2012-01-12, Page - 1 - 6
[2] japanese	Proposal of character segmentation using the convergence of the shortest path	Eiichi Tanaka Fuji Xerox Co., Ltd.	Meeting on Image Recognition and Understanding Proceedings (MIRU2011) ,2011,331-336 (2011-07-20)
[3] 2008	Japanese Named Entity Extraction by Augmenting Features with Unlabeled Data	Tomoya , Iwakura Seishi , Okamoto Fujitsu Laboratories Ltd.	Computer Vision and Image Media IPSJ SIG Technical Reports (CVIM), 2008 (115 (2008-CVIM-165)) ,125-130 (2008-11-20)
[4] 2006	Fast Deformable Template Matching for Character Recognition in Scene Images	Kazuya , NEGISHI Masakazu , IWAMURA Shinichiro , OMACHI Hirotomo , ASO	情報処理学会研究報告コンピュータビジョンとイメージメディア（CVIM） ,2006(25(2006-CVIM-153)),101-106 (2006-03-16)
[5] 2003	An On-line Cuneiform Modeled Handwritten Japanese Character Recognition Method Free from Both the Number and Order of Character Strokes	Toshimi Yokota , Soushiro Kuzunuki , Keiko Gunji , Kouyou Katsura , Nagaharu Hamada , Yasushi Fukunaga	情報処理学会論文誌,44(3),980-990 (2003-03-15) , 1882-7764
[6] 2003	Recognition of Handwritten Characters String of Japanese Address Using a Complete Correspondence Method	Michio Umeda , Daisuke Honjo	情報処理学会論文誌,44(1),114-123 (2003-01-15) , 1882-7764
[7] 2001	An Off - line Handwritten Character Recognition Method using a Character Shape Model	Takeshi Nagasaki , Masaki Nakagawa	情報処理学会研究報告数理モデル化と問題解決（MPS）,2001(110(2001-MPS-037)),5-8 (2001-11-19)
[8] 2001	Online handwriting character recognition combining adaptive classification with adaptive context processing	Naomi Iwayama , Katsuhiko Akiyama , Kazushi Ishigaki	情報処理学会研究報告モバイルコンピューティングとユビキタス通信（MBL） ,2001(108(2001-MBL-019)),55-62 (2001-11-15)
[9] 2000	An On-line Character Recognition Algorithm RAV (Reparameterized Angle Variations)	Mitsuru Kobayashi , Shinya Masaki , Osamu Miyamoto , Youitch Nakagawa , Yoshimitsu Komiya , Takashi Matsumoto	情報処理学会論文誌,41(9),2536-2544 (2000-09-15) , 1882-7764 Journal of Information Processing Society, 41 (9),2536 -2544 (2000-09-15), 1882-7764
[10] 1999	A Handwriting Character Recognition System with Adaptive Context Processing	Naomi Iwayama	IPSJ Technical Report, Information and Media (IM), 1999 (69 (1999-IM-035)) ,85-90 (1999-08-20)
[11] 1998	A Japanese OCR Error Correction Method using Character Shape Similarity and Statistical Language Model	Masaaki , Nagata Masaaki , Nagata NTT Information and Communication Systems Laboratories	情報処理学会研究報告自然言語処理（NL） 1998