Automatic Acquisition of Context-based Images Templates for Degraded Character Recognition in Scene Images

Minako Sawaki, Hiroshi Murase and Norihiro Hagita
{minako, murase, hagita}@apollo3.brl.ntt.co.jp
NTT Communication Science Laboratories
3-1, Morinosato-Wakamiya, Atsugi, Kanagawa, 243-0198, Japan

Abstract

This paper proposes a method for adaptively acquiring templates for degraded characters in scene images. Characters in scene images are often degraded because of poor printing and viewing conditions. To cope with the degradation problem, we proposed the idea of "context-based image templates" which include neighboring characters or parts thereof and so represent more contextual information than single-letter templates. However, our previous method manually selects the learning samples to make the context-based image templates and is time-consuming. Therefore, we attempt to make the context-based image templates automatically from single-letter templates and learning text-line images. The context-based image templates are iteratively created using the k-nearest neighbor rule. Experiments with 3,467 alpha-numeric characters in nine bookshelf images show that the high recognition rates for test samples possible with this method asymptotically approach those achieved with manual selection.

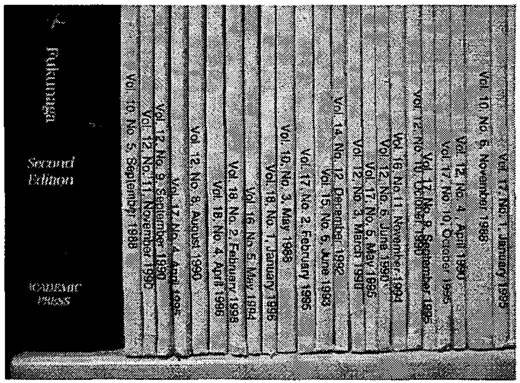

Fukunaga
Second Edition
Vol. 10 No. 5 September 1988
Vol. 12 No. 11 November 1990
Vol. 12 No. 9 September 1990
Vol. 17 No. 4 April 1995
Vol. 12 No. 8 August 1990
Vol. 18 No. 4 April 1996
Vol. 18 No. 2 February 1996
Vol. 16 No. 5 May 1994
Vol. 18 No. 1 January 1996
Vol. 10 No. 3 May 1988
Vol. 17 No. 2 February 1995
Vol. 14 No. 12 December 1992
Vol. 15 No. 6 June 1993
Vol. 12 No. 3 March 1990
Vol. 17 No. 5 May 1995
Vol. 12 No. 6 June 1990
Vol. 16 No. 11 November 1994
Vol. 17 No. 9 September 1995
Vol. 17 No. 10 October 1995
Vol. 12 No. 4 April 1990
Vol. 10 No. 6 November 1988
Vol. 17 No. 1 January 1995

Fig. 1 Example of a bookshelf image.

1. Introduction

A digital camera is currently one of the most inexpensive and simple tools for gathering image data. It may also become a convenient tool for image-to-text code conversion. As one application, we attempt to recognize characters on the spines of technical journals on bookshelves as recorded in digital camera images. The goal is to construct a personal library/filing database of the journals in the user's vicinity.

Characters in bookshelf images (Fig. 1) are often degraded because of poor printing and viewing conditions. This degradation problem involves problems with computer vision as well as conventional character recognition.

To cope with the degradation problem, conventional methods for character recognition try to construct document-specific character templates using ground truth data or lexical information. Kopec *et al.* extract templates by aligning text-lines and those ground truth data [1]. Nagy *et al.* isolate character templates from word images by word shift algorithm with truth data [2]. Ho [3] recognizes frequently occurring words and extracts templates based on the method in [2].

The authors, on the other hand, proposed templates that include neighboring characters or parts thereof to cope with degraded characters [4, 5]. The templates are referred to as context-based image templates, since they offer more contextual information than single-letter templates. Though they are robust for degraded characters, our previous method manually selects learning samples and so is too heavy a burden.

In this paper, we automatically create the context-based image templates iteratively from single-letter templates in an initial dictionary and learning text-line images. The learning text-line images are recognized with the initial dictionary. The recognized areas in the text-lines are used as learning patterns for context-based image templates. The learning patterns are labeled by the character category decided by the recognition process. Appropriate learning patterns are then automatically selected as context-based image templates using the *k-nearest neighbor* rule. This procedure can also be used to update an existing dictionary. Experiments show the recognition rates with the proposed method asymptotically approach those achieved with manual selection.

0-7695-0750-6/00 $10.00 © 2000 IEEE

2. Recognition process

An input color image captured by a digital camera is converted into a gray-scale image and then binarized. Skew detection is needed since journals usually slump on the bookshelf. Journal boundaries are detected using the dark lines caused by shadow. Next, text-line regions between the boundaries are extracted by the transition frequency of pixel colors (white-to-black and black-to-white) during vertical scanning, and the skew of each text-line is then corrected and the size is normalized.

Characters in the text-line region are recognized by displacement matching [4, 5]. An observation window ***F*** moves along the text-line pixel-by-pixel, and ***F*** is matched against stored templates. ***F*** is also shifted along the horizontal-axis to cover text-line misalignment. The complementary similarity measure, which is robust against noise, is employed for matching [4]. If the maximum similarity value exceeds a threshold, the category is selected as the recognition result.

3. Automatic template acquisition

3.1 Approach

To recognize low-quality characters, we proposed the complementary similarity measure S_C [4] and context-based image templates [5]. The complementary similarity measure is robust against additive/deletion noise, however, it is not robust enough against the deformation in scene images arising from poor printing and viewing conditions.

To overcome this problem, we introduced the context-based image templates, which include neighboring characters or parts thereof [5]. Fig. 2 shows examples of learning patterns in a text-line and the context-based image templates. Our previous method [5] used manual operations to screen the learning patterns, label the correct category name, and use them to make the contest-based image templates. However, positioning the characters and typing the correct categories by hand are very tiring activities.

In this paper, we provide a method that automatically acquires context-based image templates from learning images without manual extraction of learning patterns or the transcription process such as ground-truth information. This approach reduces the tiresome work of typing. Also, it enables to a dictionary to be adapted to the input conditions automatically and rapidly without keyboard work. This feature is convenient for a portable recognition system with a digital camera.

Since the adverse effect of mis-classified templates caused by automatic acquisition is a problem, we utilize the *k-nearest neighbor* rule within the learning patterns to eliminate such mis-classified templates. Even with this technique, some mis-classified patterns may remain. However, we assume that when the ratio of the number of mis-classified templates to the number of correctly classified templates is low, the mis-classified templates do not decrease the recognition rates significantly.

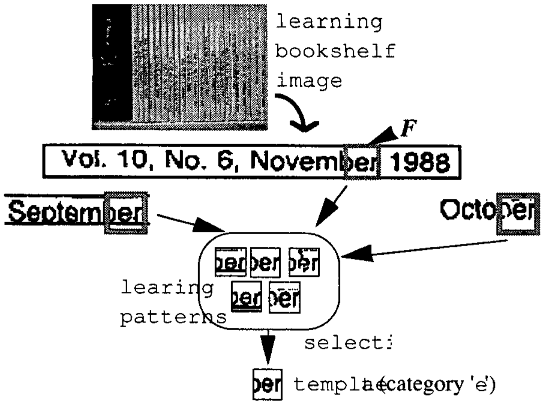


Fig. 2 Context-based image templates.

3.2 Template acquisition for low-quality characters

This section explains the detailed process of the proposed method. The flowchart is shown in Fig. 3.

Step 1) Making the initial dictionary

Initial dictionary is made by manually drawing one single letter image per character category in learning text-line images and registering them as single letter templates. This is the only manual step with the method.

Step 2) Recognition of learning text-lines with the (t-1)th dictionary

The learning text-lines are recognized by the (t-1)th dictionary by displacement matching in the t-th iteration. When the similarity exceeds the threshold, the category is determined as a recognized category. The recognized areas are extracted as learning patterns.

In the first iteration, the initial dictionary is used for recognition. In this case, the dictionary is divided into two in size, and large character group and small character group are recognized alternately one after the other. As each observation window usually includes multiple characters, the recognition with single-letter templates may not achieve high recognition accuracy. In order to avoid mis-classified by the existence of neighboring characters in the observation window, large characters which are relatively robust even with neighboring characters are recognized first and these regions are then whitened out. The small characters are then recognized. Later on, large characters and small characters are recognized alternately one after the other for several times. This recognition method is relatively time-consuming but it's effective in obtaining higher recognition rates with the initial dictionary than using the whole initial dictionary at once.

Step 3) Pattern selection using *k-nearest neighbor* rule

Since the extracted learning patterns may include misclassified patterns, uncertain patterns are eliminated from the dictionary using the *k-nearest neighbor* rule. Patterns whose *k-nearest neighbors* belong to many different categories are assumed to be uncertain templates and are not registered in the updated dictionary.

k-nearest neighbors of a learning pattern are obtained from the learning patterns. When a category achieves a majority, the pattern is registered in the updated dictionary. If the pattern number of the most significant category of *k-nearest neighbor* patterns does not exceed 50% of *k*, the learning pattern is not registered in the updated dictionary.

Step 4) Stop criterion

When the number of patterns removed becomes small, the iteration of learning is stopped. Otherwise, Steps 2 and 3 are repeated using the updated dictionary.

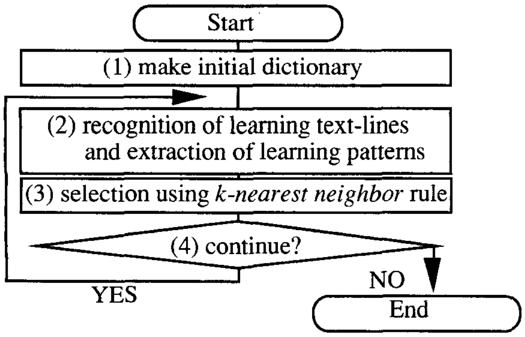

Start
(1) make initial dictionary
(2) recognition of learning text-lines and extraction of learning patterns
(3) selection using *k-nearest neighbor* rule
(4) continue?
YES
NO
End

Fig. 3 Template acquisition flow.

4. Recognition experiments

4.1 Experimental conditions

Bookshelf images were captured by using a digital camera (832 x 608 pixels) from a straight view. Journals on the bookshelf were the same kind with different publication dates (journals : *IEEE Trans. Pattern Anal. Machine Intell.*, 1988-1996). Threshold value for binarization was 128 out of 256 gray levels.Thirty seven categories were used consisting of numbers (0-9) and alphabetical categories (*A,D,F,J,M-O,S,V,a-c,e,g-i,l-p,r-v,y*) which are enough for recognizing *No*, *Vol*, *month* and *year*, all of which are often printed on journal spines. Dots (.) and commas (,) were not recognized. 3,860 patterns in 10 bookshelf images and 3,467 patterns in 9 bookshelf images were used as learning and test patterns, respectively (both had 14 words and numbers). The size of ***F*** was n = 24 x 24 pixels. In the experiments, the result was regarded as correct when the correct category was determined at the correct position. k of the *k-nearest neighbor* rule was determined as 10 (including own category) from preliminary experiments.

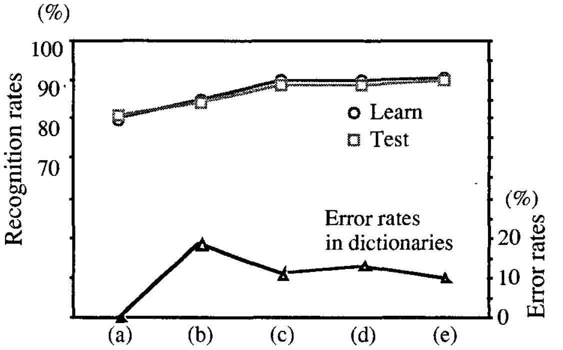


Fig. 4 Recognition rates for bookshelf images and error rates in the dictionaries. (a)initial dictionary, (b)all learning patterns (t=1 iteration), (c)selected dictionary (t=1 iteration), (d) all learning patterns (t=2 iteration), (e) selected dictionary (t=2 iteration).

4.2 Experimental results

The recognition rates are shown in Fig 4. The plots correspond to recognition rates with the initial dictionary, all learning patterns (1st iteration), selected dictionary (1st iteration), all learning patterns (2nd iteration), selected dictionary (2nd iteration). We stopped learning after two iterations as the removed pattern number basically saturated at this point. The template numbers in the dictionaries were 37, 3840, 3428, 3906, 3736, respectively. Figure 5 shows examples of templates in the initial and the last dictionary. The recognition rate with the last dictionary was 89.5% for the test patterns. Our previous method achieved 96.3%. Fig. 4 shows that the recognition rates of the proposed method increased with iteration number and approached that achieved with manual template selection.

The error rates, percentage of templates with erroneous category name, were 0.0%, 18.3%, 10.9%, 13.0%, 10.0% for the five dictionaries, respectively (Fig. 4). Error rates in the dictionaries fluctuated during these iteration steps. This is because the error rates increase with pattern extraction and decrease with selection. The values for (c) and (e) compared to (b) and (d) show that error rates were decreased by *k-nearest neighbor* based selection. Throughout learning, the error rates decrease gradually with iteration number. Therefore, our method is effective in automatically acquiring context-based image templates.

All eliminated templates of category '*a*' in the 1st selection step are shown in Fig. 6. Nine patterns of the 13 eliminated patterns (69%) were correctly eliminated.

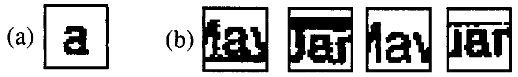


Fig. 5 Examples of templates (category '*a*') (a)initial dictionary, (b) last dictionary.

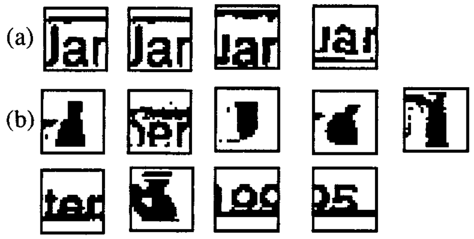


Fig. 6 Templates eliminated by *k-nearest neighbor* rule in the 1st iteration (category '*a*') (a) mis-eliminated patterns, (b) correctly eliminated patterns.

For comparizon, the iteration was continued till t=6 where the recognition rate (test data) started to decrease. The maximum recognition rate (test data) of 90.4% was obtained with the selected dictionary at t=5 while the error rate in the dictionary was 29.7%.

5. Journal volume retrieval system

As an application of the proposed method, a journal volume retrieval system was constructed. This system enables us to locate the desired volume within similar magazines on the bookshelf or to determine if some volumes are missing.

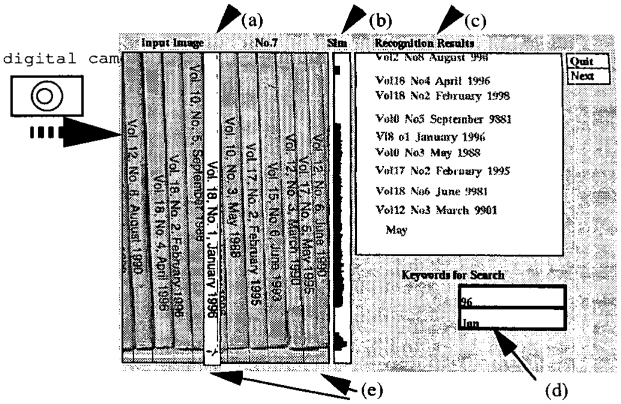
This system consists of a digital camera for image capture and a computer for recognition and retrieval, both commercial products. It recognizes characters in bookshelf images, and search strings are matched against the recognition results. When the desired journal volumes exist, they are displayed on the screen as binary images. Two search strings can be typed at one time, and the system retrieves the volumes using AND search.

The system image is shown in Fig. 7. The recognition results from top to the bottom correspond to the magazines from left to right. In Fig. 7, one journal is retrieved using the search strings of "*96*" and "*Jan*".

6. Conclusions

This paper proposed an automatic method for acquiring templates for degraded character recognition in scene images. To cope with the degradation, we used our context-based image templates, which include neighboring characters or parts thereof. The method creates context-based image templates automatically from single-letter templates and learning text-lines. Learning patterns in the learning text-lines are extracted by the recognition results using an initial dictionary. The learning patterns are then selected by the *k-nearest neighbor* rule for inclusion in an updated version of the original dictionary. These steps are iterated to refine the dictionary. Experiments with 3,467 characters in nine bookshelf images show that the recognition performance of the proposed method asymptotically approached to that achieved with manual extraction.

Our future works include designing an good initial dictionay and applying this method to various kinds of books and observation conditions.



```
Vol2 No8 August 998
Vol18 No4 April 1996
Vol18 No2 February 1998
Vol0 No5 September 9881
Vl8 o1 January 1996
Vol0 No3 May 1988
Vol17 No2 February 1995
Vol18 No6 June 9981
Vol12 No3 March 9901
  May
```

96
Jan

Fig. 7 Journal volume retrieval system
(a) input color image, (b) maximum similarity values, (c) recognition results, (d) search strings, (e) retrieved volumes.

Acknowledgements

We wish to express our gratitude to Dr. K. Ishii for several fruitful discussions. We also thank Mr. T. Iwata for his cooperation.

References

[1]G.E.Kopec and M.Lomelin, "Document-Specific Character Template Estimation", *Proc. of SPIE*, Vol. 2660, pp.14-26 (1996).

[2]G.Nagy and Y.Xu, "Automatic Prototype Extraction for Adaptive OCR", *Proc of ICDAR'97*, pp.278-282 (1997).

[3] T.K.Ho, "Bootstrapping Text Recognition from Stop Words", *Proc. of ICPR'98*, pp.605-609 (1998).

[4] M. Sawaki and N. Hagita, "Text-line Extraction and Character Recognition of Document Headlines with Graphical Designs using Complementary Similarity Measure", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 20, No. 20, pp. 1103 - 1109 (1998).

[5] M. Sawaki, H. Murase and N. Hagita, "Character Recognition in Bookshelf Images using Context-based Image Templates", *Proc. of ICDAR99*, pp.79 - 82 (1999).