Generation of Templates for Low-Resolution Text Recognition Using a Hypothesis Graph[1]

H. Ishida[a, b], T. Takahashi[a, b], I. Ide[a], and H. Murase[a]

[a] *Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464–8601, Japan*

[b] *Japan Society for the Promotion of Science, Japan*

e-mail: {hishi, ttakahashi, ide, murase}@murase.m.is.nagoya-u.ac.jp

Abstract—We propose a recognition method of character-string images captured by portable digital cameras. A challenging task in character-string recognition is the segmentation of characters. In the proposed method, a hypothesis graph is used for recognition-based segmentation of the character-string images. The hypothesis graph is constructed by the subspace method, using eigenvectors as conditionally elastic templates. To obtain these templates, a generation-based approach is introduced in the training stage. Various templates are generated to cope with low-resolution. We have experimentally proved that the proposed scheme achieves high recognition performance even for low-resolution character-string images.

DOI: 10.1134/S1054661808040159

1. INTRODUCTION

Text recognition technologies using portable digital cameras have gained attention in recent years [1]. A problem in camera-based text recognition is the segmentation of characters. The characters in camera-captured images tend to be small, which makes the segmentation task difficult. These characters should be segmented properly for the accurate recognition of character strings. As has been discussed in many studies [2], however, neither recognition nor segmentation of low-quality character-string images can be done independently. One solution is recognition-based segmentation, in which tasks of segmentation and recognition are performed jointly.

In recognition-based segmentation methods [3, 4], a hypothesis graph is employed to search for an optimal segmentation result. However, several practical problems of the conventional methods have been revealed by experiments with low-resolution images. Firstly, a plausible hypothesis is not obtained if characters contained in an image are even slightly different from their templates. To cope with this problem, templates for various conditions are required. The remaining question is how to obtain them. Secondly, solving the hypothesis graph with a large number of templates is computationally infeasible. In this paper, we introduce a generation-based approach [5] for the first problem. And for the second problem, we introduced the hypothesis graph based on the subspace method [7].

Since the method in [5] was proposed for individual character recognition, the problems in the segmentation task were not taken into account. The approach used in this paper is specialized for low-resolution text recognition. Various training images are generated in the training step. In the recognition stage, the subspace method is introduced. Using eigenvectors just as templates, the hypothesis graph is constructed.

This paper is organized as follows. Section 2 describes the generation stage of templates. Section 3 describes construction of subspaces from the generated templates. Section 4 describes the recognition stage of character-string images. Results are presented in Section 5.

2. GENERATION OF TEMPLATES

All templates of characters are generated from font images. Automatic generation of templates not only simplifies the training stage, but also makes it possible to control the degrees of resolution and degradation. Here we introduce segmentation and resolution parameters for simulating actual character images. In order to tolerate a certain degree of segmentation error, various templates are generated by changing the parameters.

Figure 1 illustrates the generation process of the templates. Characters shown in the top of the figure are examples of original font images. Horizontal dotted lines represent upper and lower boundaries of the characters. Likewise, vertical dotted lines represent left and right boundaries of character A. Let points $(x_0^{(c)}, y_0)$ and $(x_1^{(c)}, y_1)$ be coordinates at the top left and lower right of the original character region, respectively. A rectangular region to be extracted as a template is expressed using segmentation parameters (u_0, v_0, u_1, v_1) by

$$(x_0^{(c)} - u_0, y_0 - v_0) - (x_1^{(c)} + u_1, y_1 + v_1). \qquad (1)$$

The image in this region is transformed to a low-resolution image, its size being $w \times h$ pixels. This low-resolution image X_l is obtained from the original image X_o by

[1] The text was submitted by the authors in English.

Received March 3, 2008

$$X_l(p, q) = \frac{1}{|G_{(p,q)}|} \sum_{i, j \in G_{(p,q)}} X_o(i, j), \tag{2}$$

$$G_{(p,q)} = \left\{ i, j \left| \begin{array}{c} (x_1^{(c)} - x_0^{(c)})p/w \le i - x_0^{(c)} < (x_1^{(c)} - x_0^{(c)})(p+1)/w \\ (y_1 - y_0)q/h \le j - y_0 < (y_1 - y_0)(q+1)/h \end{array} \right. \right\}, \tag{3}$$

where $G_{(p,q)}$ is a set of pixels (i, j) integrated to form pixel (p, q) of low-resolution image X_l. Next, X_l is normalized to a template $X_n^{(c)}$ whose size is 32×32 pixels by linear interpolation. This normalization is necessary for the construction of a subspace.

3. CONSTRUCTION OF A SUBSPACE

A subspace is constructed from various templates. Suppose that we have N generated templates for each category. The principal component analysis (PCA) is applied to them to obtain R eigenvectors ($R < N$), which can also be used as templates. The recognition stage is based on the subspace method [7].

At first, each template $X_n^{(c)}$ is converted to a vector $x_n^{(c)}$ such that the mean of the elements becomes 0, and that the norm becomes 1,

$$x_n^{(c)} = \tilde{x}_n^{(c)} / (\tilde{x}_n^{(c)T} \tilde{x}_n^{(c)}), \tag{4}$$

$$\tilde{x}_n^{(c)} = [X_n^{(c)}(0, 0) - \bar{x}_n^{(c)} \ \dots \ X_n^{(c)}(31, 31) - \bar{x}_n^{(c)}]^T, \tag{5}$$

$$\bar{x}_n^{(c)} = \frac{1}{32 \times 32} \sum_{p=0}^{31} \sum_{q=0}^{31} X_n^{(c)}(p, q). \tag{6}$$

Next, an auto-correlation matrix is calculated from a list of N vectors $\mathbf{x}_n^{(c)}$ by

$$\mathbf{Q}_n^{(c)} = \frac{1}{N} [\mathbf{x}_1^{(c)} \ \dots \ \mathbf{x}_N^{(c)}][\mathbf{x}_1^{(c)} \ \dots \ \mathbf{x}_N^{(c)}]^T. \tag{7}$$

The eigenvalues and corresponding eigenvectors are calculated from this matrix $\mathbf{Q}_n^{(c)}$.

The eigenvectors $\mathbf{e}_r^{(c)}$ with the largest R eigenvalues are used for the recognition. Figure 2 shows examples of the eigenvectors.

4. RECOGNITION OF CHARACTER-STRING IMAGE

The hypothesis graph is introduced in the recognition stage. The task of character-string recognition is simplified to individual character recognition by using the hypothesis graph. The basic idea is similar to the candidate character lattice method [4], which is relatively simple but suitable for printed text in low-resolution. In the lattice method, individual characters are initially processed, and thereby the hypothesis graph with the lists of the candidate characters is constructed as illustrated in Fig. 3. Character-string recognition is performed by searching for the optimal path; it differs from the conventional lattice method in that the proposed method introduces a subspace-based approach to the hypothesis graph. A new representation of the hypothesis graph is proposed in this paper.

4.1. Recognition of Individual Characters

In the proposed method, eigenvectors of characters are used as elastic templates which are compared to subcomponents of a given character-string image. In the given image composed of W columns, a subcomponent from the mth column to the nth column is denoted as $Z_{(m,n)}$ ($1 \le m < n \le W$). Let it be normalized to a vector $\mathbf{z}_{(m,n)}$ so that the mean becomes 0 and the norm becomes 1. The similarity of $Z_{(m,n)}$ to category c is defined as a sum of squared inner product to the eigenvectors $\mathbf{e}_r^{(c)}$ by

$$s^{(c)} = \sum_{r=1}^{R} (\mathbf{e}_r^{(c)T} \mathbf{z}_{(m,n)})^2. \tag{8}$$

4.2. Construction of a Hypothesis Graph

A hypothesis graph is composed of candidate character-strings along with values of their plausibility. Here, a set of candidates for the subcomponent $Z_{(m,n)}$ are denoted as

$$\{(C_{(m,n)}, S_{(m,n)})\}, \tag{9}$$

where $C_{(m,n)}$ is a string of characters and $S_{(m,n)}$ is its plausibility. These candidates are calculated first for small subcomponents of the given character-string image, followed by larger subcomponents ($m_{\text{new}} \le m_{\text{old}} < n_{\text{old}} \le n_{\text{new}}$). Finally, the most plausible candidate for the entire image $Z_{(1,W)}$ is accepted as the recognition result.

The similarities calculated in Subsection 4.1 appears on the hypothesis graph as

$$C_{(m,n)} \leftarrow \text{character of category } c, \tag{10}$$

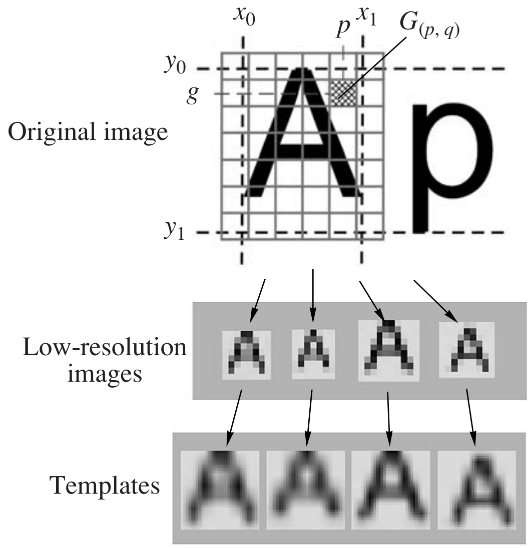


Fig. 1. Templates are generated from original font images. After converting them to low-resolution images, their size is normalized.

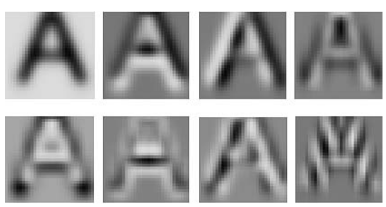


Fig. 2. Top eight eigenvectors of category A.

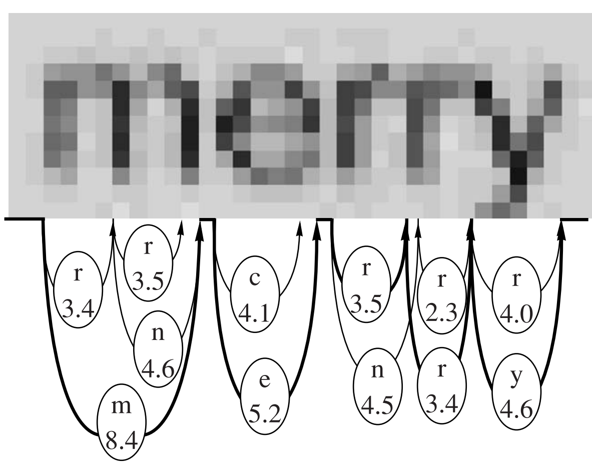


Fig. 3. Hypothesis graph constructed for character-string image "merry." Candidate characters are shown with their plausibility values. The path shown with bold lines maximizes the sum of the plausibility values.

$$S_{(m,n)} \longleftarrow \begin{cases} (n-m+1)s^{(c)} & \left(\dfrac{x_1^{(c)}-x_0^{(c)}}{y_1-y_0}h \right. \\ -t < n-m+1 < \dfrac{x_1^{(c)}-x_0^{(c)}}{y_1-y_0}h+t) & \\ 0 \quad \text{otherwise.} & \end{cases} \tag{11}$$

The weight $(n-m+1)$ acts as a normalization factor for the subcomponent $Z_{(m,n)}$. The height of the character-string image is denoted by h. Since we use the eigenvectors as conditionally elastic templates, the width of the subcomponents is restricted by a parameter t. For each subcomponent $Z_{(m,n)}$, candidate characters are sorted in order of the magnitude of plausibility. Several candidates with large values of plausibility are selected to make new candidates for larger subcomponents. If we select $(C_{(m,n)}, S_{(m,n)})$ and $(C_{(m',n')}, S_{(m',n')})$ as illustrated in Fig. 4 $(m < n \le m' < n')$, a new candidate $(C_{(m,n')}, S_{(m,n')})$ appears on the hypothesis graph. The new candidate is calculated by

$$C_{(m,n')} \longleftarrow \text{merge}\{C_{(m,n)}, C_{(m',n')}\}, \tag{12}$$

$$S_{(m,n)} \longleftarrow S_{(m,n)} + S_{(m',n')}. \tag{13}$$

The candidate character-string $C_{(1,W)}$ for the entire image $Z_{(1,W)}$ is then obtained as the recognition result.

5. EXPERIMENT

In this section, the effectiveness of the proposed method is evaluated. A digital camera (Panasonic DMC-FX9) was used to capture character-string images. 298 words [8] printed on paper were captured five times and, in all, 1490 character-string images were prepared as test samples. The number of categories was 62 (A–Z, a–z, 1—9: Ariel font). The average height of the character-strings was 12 [pixel]. In the process of extracting the character-string images, their height was initially estimated from the whole document image, and the extracted areas were then determined such that each of the contained character strings should be located at the center of the area.

In the training step, training images were generated from original images (Arial font) as described in Section 2. The parameters in Eq. (1) were determined as shown in table, where I is the width of the vertical stroke in the original font images, and $J = (y_1 - y_0)/24$. Two training sets were compared in order to evaluate the usefulness of generating the training images; by changing parameters as shown in table, $5 \times 5 \times 5 \times 5 \times 5 \times 2 = 6.250$ training images per category were generated for training set A. Similarly, $3 \times 3 \times 3 \times 3 \times 5 \times 2 = 810$ training images per category were generated for training set B.

Recognition results are presented in Fig. 5. A macro-averaged F_1 measure [6] was used for evalua-

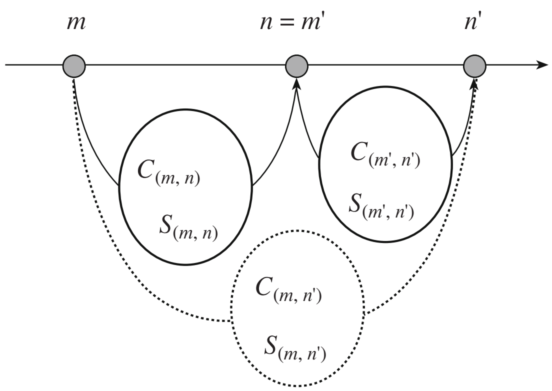


Fig. 4. Construction process of hypothesis graph. Candidate character-strings are calculated from two candidate character-strings of smaller subcomponents.

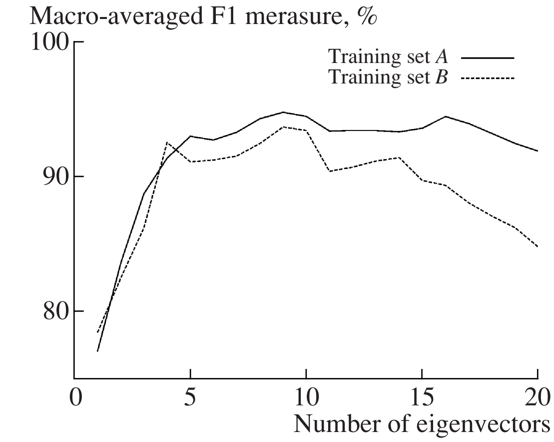


Fig. 5. Recognition rates for various numbers of eigenvectors used as templates.

tion, where F_1 is given for each test character-string by the formula $F_1 = 2pr/(p + r)$ with precision rate p and recall rate r. Letting A and B denote sets of characters in the correct string and in the recognized string, respectively, it follows that $p = |A \cap B|/|B|$ and $r = |A \cap B|/|A|$. In order to investigate the relationships between the performance and the number of eigenvectors, the rates were calculated by changing R. The parameter t in Eq. (11) was set to 3.

If R is 1, the methods are equivalent to a simple template-matching method that uses an averaged template. A better performance was obtained when multiple eigenvectors were used; the rates were raised to around 94% when R was in the range of 5–15. It is worth noting that the number of templates was reduced ($R \ll N$) by introducing the subspace method in the hypothesis graph. However, the performance was sensitive to the number of eigenvectors. The rate did not saturate but even decreased once $R > 16$. This result implies that some of the eigenvectors were not suitable for the construction of the hypothesis graph. For this reason, the value R should be determined experimentally. Table 2 shows some examples of the recognized character-strings. Some errors were eliminated by using multiple eigenvectors, although new errors can be yielded by setting R too high.

Parameters used for generating training sets

Parameter	Training set A	Training set B
u_0	$u_0 = I, \frac{5}{4}I, \frac{3}{2}I, \frac{7}{4}I, 2I$	$u_0 = \frac{5}{4}I, \frac{3}{2}I, \frac{7}{4}I$
u_1	$u_1 = I, \frac{5}{4}I, \frac{3}{2}I, \frac{7}{4}I, 2I$	$u_0 = \frac{5}{4}I, \frac{3}{2}I, \frac{7}{4}I$
ν_0	$\nu_0 = -2J, -J, 0, J, 2J$	$\nu_0 = -J, 0, J$
ν_1	$\nu_0 = -2J, -J, 0, J, 2J$	$\nu_1 = -J, 0, J$
h	$h = 8, 9, 10, 11, 12$	
w	$w = \left[\frac{x_1^{(c)} - x_0^{(c)}}{y_1 - y_0}h\right], \left[\frac{x_1^{(c)} - x_0^{(c)}}{y_1 - y_0}h\right] + 1$	

Figure 5 shows also the recognition rates from the training sets A and B. For almost all values of R, the recognition rates from the training set A were higher than those in the training set B. This result indicates that the recognition accuracy was improved by generating various templates.

6. CONCLUSIONS

In this paper, a recognition method for low-resolution character-string images is proposed. In order to cope with various levels of degradation, templates are generated by changing segmentation and resolution parameters. In the recognition stage, eigenvectors of the templates are used to construct a hypothesis graph from which the recognition result is obtained. Experimental result shows the effectiveness of the proposed method for character-string recognition from low-resolution images. Future work will introduce some new constraints to the hypothesis graph for further improvement of recognition accuracy.

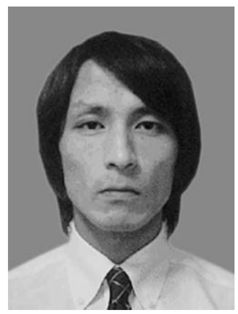
ACKNOWLEDGMENTS

Parts of this research were supported by the Grants-In-Aid for Scientific Research (16300054 and 17650050) and the 21st century COE program from the Ministry of Education, Culture, Sports, Science, and Technology. This work is implemented based on the MIST library (http://mist.suenaga.m.is.nagoya-u.ac.jp/).

REFERENCES

1. J. Liang, D. Doermann, and H. Li, "Camera-Based Analysis of Text and Documents: A Survey," Int. J. Document Analysis and Recognition **7** (2–3), 84–104 (2005).
2. S. Wachenfeld, H. Klein, and X. Jiang, "Recognition of Screen-Rendered Text," *Proc. 18th Int. Conf. on Pattern Recognition, Hong Kong, China, 2006*, pp. 1086–1089.
3. R. Casey and E. Lecolinet, "A Survey of Methods and Strategies in Character Segmentation," IEEE Trans. on Pattern Analysis and Machine Intelligence **18** (7), 690–706 (1996).
4. H. Murase, T. Wakahara, and M. Umeda, "Online Writing-Box Free Character String Recognition by Candidate Character Lattice Method," IEICE Trans. **J-68-D** (4), 765–772 (1985).
5. H. Ishida, S. Yanadume, T. Takahashi, I. Ide, Y. Mekada, and H. Murase, "Recognition of Low-Resolution Characters by a Generative Learning Method," *Proc. 1st Int. Workshop on Camera-Based Document Analysis and Recognition, Seoul, Korea, 2005*, pp. 45–51.
6. Y. Yang, "An Evaluation of Statistical Approaches to Text Categorization," J. Information Retrieval **1** (1–2), 69–90 (1999).
7. E. Oja, "Subspace Methods of Pattern Recognition," *Research Studies* (Hertfordshire, UK, 1983).
8. The constitution of Japan (in English), Prefaces, Nov. 1946; http://www.solon.org/Constitutions/Japan/English/english-Constitution.html.

Hiroyuki Ishida. Received his B.S. and M.S. degrees from the Department of Information Engineering and from the Graduate School of Information Science, respectively, at Nagoya University. He is currently pursuing a Ph.D. in Information Science at Nagoya University.



Tomokazu Takahashi. Received his B.S. degree from the Department of Information Engineering at Ibaraki University, and his M.S. and Ph.D. from the Graduate School of Science and Engineering at Ibaraki University. His research interests include computer graphics and image recognition.

Ichiro Ide. Received his B.S. degree from the Department of Electronic Engineering, his M.S. degree from the Department of Information Engineering, and his Ph.D. from the Department of Electrical Engineering at the University of Tokyo. He is currently an Associate Professor in the Graduate School of Information Science at Nagoya University.

Hiroshi Murase. Received his B.S., M.S., and Ph.D. degrees from the Graduate School of Electrical Engineering at Nagoya University.

He is currently a Professor in the Graduate School of Information Science at Nagoya University. He received the Ministry Award from the Ministry of Education, Culture, Sports, Science and Technology in Japan in 2003. He is a Fellow of the IEEE.