Recognition of low-resolution characters by a generative learning method

Hiroyuki Ishida, Shinsuke Yanadume, Tomokazu Takahashi,
Ichiro Ide, Yoshito Mekada and Hiroshi Murase
Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan
{hishi, yanadume, ttakahashi, ide, mekada, murase}@murase.m.is.nagoya-u.ac.jp

Abstract

Using appropriate training data is necessary to robustly recognize low-resolution characters by the subspace method. Former learning methods used characters actually captured by a camera, which required the collection of characters of all categories in various conditions. In this paper, we propose a new learning method that generates training data by a point spread function estimated beforehand by captured images. This method is efficient, since it eliminates collection of the training data. We confirmed its usefulness by experiments.

1. Introduction

Technologies to recognize low-resolution characters have gained attention in recent years for their potential applications to portable digital cameras. Recognizing documents with camera-equipped cellular phones is especially of practical concern [1]. Various attempts have been carried out on the recognition of printed characters on documents [2], [3], [4]. However, few studies have focused on very low-resolution characters captured by portable digital cameras. Even with the improvements of digital cameras, the characters in a captured image are still often small, when the target documents contain many characters.

The purpose of this work is to devise a new method for the efficient recognition of low-resolution characters using the subspace method [5], [6]. Generally, training data are collected from actual captured images. Since it is difficult and unrealistic to collect character images of all categories under various conditions, we propose a generative learning method in which training data are generated automatically from original character images. Since our goal is to recognize low-resolution characters, training data should be generated in accordance with actual degradation. The outline of this generative learning method is as follows: First, we estimate a point spread function (PSF) [7] of a camera from captured images. Next, we generate degraded training data by applying PSF to the original character images. This method is efficient, since it eliminates the collection of training data of all categories from captured images. This method is also suitable for the recognition of low-resolution characters compared with a simple learning method that only uses original character images without taking the actual degradation into consideration.

This paper is organized as follows: Section 2 describes the degradation process of the captured characters and our approach to simulate the degradation process. Section 3 describes the main idea of the generative learning method. Section 4 describes the recognition steps using the subspace method. Section 5 demonstrates the experimental results. Section 6 concludes the paper.

2. Degradation process

Understanding the actual degradation process of a captured character image is crucial for a generative learning method. A target character image is small and in low quality, which makes it difficult to recognize.

We divide the causes of character degradation into two factors: (1) Optical blur is caused by a process where the characters on a target document are projected onto CCD sensors through a camera lens, and (2) resolution decline is caused through the sampling process, where the character image is turned into a low-resolution digital image. Besides, captured character images vary with every frame even if captured from the same original image, since the light quantities of each CCD sensor shift due to slight camera vibration. Even such slight changes cannot be ignored for low-resolution character recognition.

We propose two generation models based on these factors: optical blur and resolution decline. The first generation model copes only with the low-resolution factor. In this model we generate training data by reducing the resolution of the original character images. Font data on a computer is used as the original image, based on the assumption

(a) original image (b) optically blurred image (c) low-resolution image

Figure 1. Degradation factors.

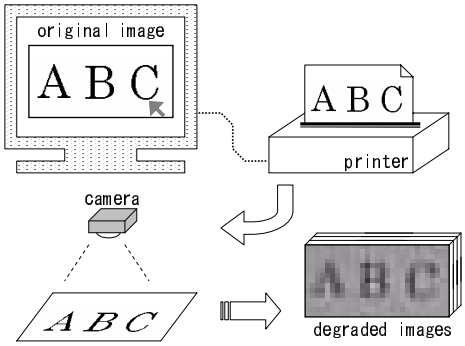


Figure 2. Flow of preparation for PSF estimation.

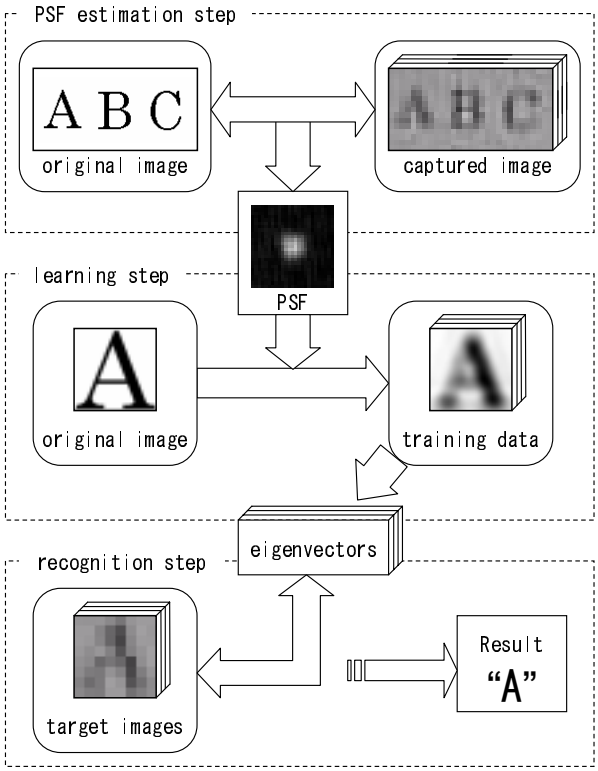


Figure 3. Flow of generative learning method based on PSF estimation.

that the characters to be recognized are printed. The second generation model copes with both degradation factors (low-resolution and optical blur). We generate training data for this second model by applying PSF estimated from actual captured images (Section 3). We use the same camera for both PSF estimation and recognition, and that enables us to simulate degradation features peculiar to the camera. In later sections, we discuss the effectiveness of these two generation models by experiments, and also compare them with a non-generative learning method in which none of the degraded factors are included.

3. Generative learning

In this paper, we propose a learning method that learns from artificially degraded training data. In the following sections, we compare two models for generative learning. The first generates training data only by reducing the resolution of the original images. For convenience, we call this method "Generative learning method (type-A)". The second generates training data by applying estimated PSF together with the reduction of the resolution. Similarly, we call this "Generative learning method (type-B)", whose flow is shown in Figure 2 and 3. In this section we focus on generative learning method (type-B).

3.1. Preparation for the PSF estimation

First we need to obtain images for PSF estimation. For the original image, we printed some characters and obtained degraded images for PSF estimation by capturing the printed figure with a camera (Figure 2). We need to capture a certain quantity of images for noise reduction.

Although PSF itself is independent from the original image, an appropriate PSF cannot be acquired if it is composed only of monotonous structural elements; an image composed of various structural elements is more appropriate as the original image for the PSF estimation.

3.2. PSF estimation

PSF is estimated from the original image and a number of degraded images captured by a camera. Image degradation is represented as

$$g(x,y) = f(x,y) * h(x,y) + n(x,y), \qquad (1)$$

where f is the original image, g is a degraded image, h is PSF, and n is the noise function. This equation indicates that a degraded image is generated by the convolution

of the original image and PSF [8]. To obtain h, we apply a two-dimensional fourier transformation to Equation 1 as follows:

$$H(u,v) = \frac{G(u,v)}{F(u,v)} - \frac{N(u,v)}{F(u,v)}, \tag{2}$$

where noise component N is unknown. Since the captured images contain a lot of noise, it is not possible to obtain an appropriate PSF from a single image. Thus, this method averages $H(u,v)$ calculated from variously degraded images to restrain noise [9]. Assuming that we use k degraded images $G_i(u,v)$ $(i = 1, \cdots, k)$, averaged $\hat{H}(u,v)$ can be calculated from multiple $H_i(u,v)$ as

$$\begin{aligned} \hat{H}(u,v) &= \frac{1}{k}\sum_{i=1}^{k} H_i(u,v) \\ &= \frac{1}{k}\sum_{i=1}^{k} \frac{G_i(u,v)}{F(u,v)} - \frac{1}{k}\sum_{i=1}^{k} \frac{N_i(u,v)}{F(u,v)}. \end{aligned} \tag{3}$$

Since we consider that no relation exists among the noise components of each image, consequently,

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \frac{N_i(u,v)}{F(u,v)} = 0. \tag{4}$$

Thus, when k is large enough, $\hat{H}(u,v)$ can be approximated as

$$\hat{H}(u,v) = \frac{1}{F(u,v)}\frac{1}{k}\sum_{i=1}^{k} G_i(u,v). \tag{5}$$

PSF is obtained by the inverse Fourier transformation of $\hat{H}(u,v)$.

Figures 4, 5, and 6 show the estimated PSF of each camera (digital video camera, digital camera, and camera equipped on cellular phone).

3.3. Generation of training data

Training data is generated with estimated PSF with various degrees of degradation. Here we introduce degradation parameter d to designate the degree of degradation. When $d = 0$, the generated image is equivalent to the original image, and when $d = 1$, the generated image is equivalent to the convolution of the original image and estimated PSF. We apply a PSF filter the size of $(R_h + 1) \times (R_h + 1)$ (See Figure 7) that transforms the original image into a degraded image. Each pixel of the training data is calculated using the PSF filter expanded in proportion to degradation parameter d as:

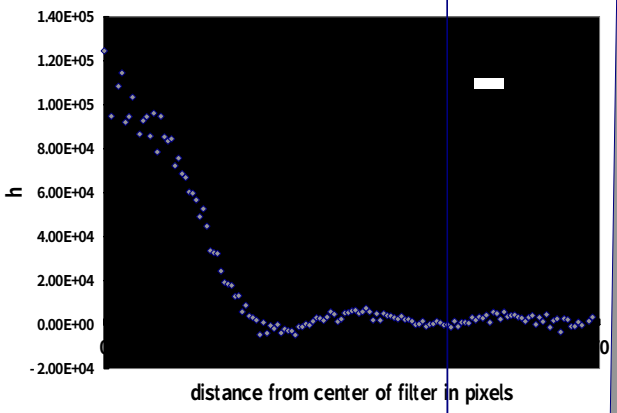


Figure 4. PSF of digital video camera.

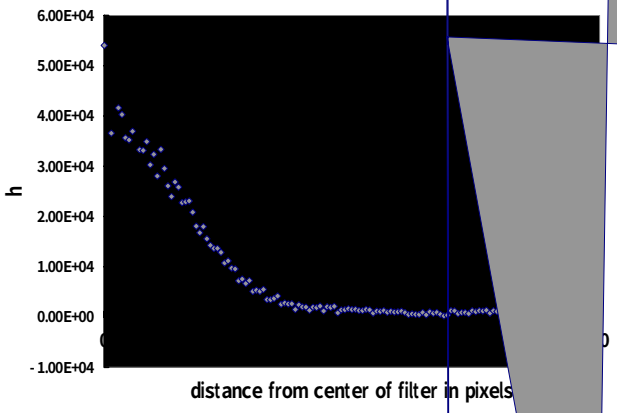
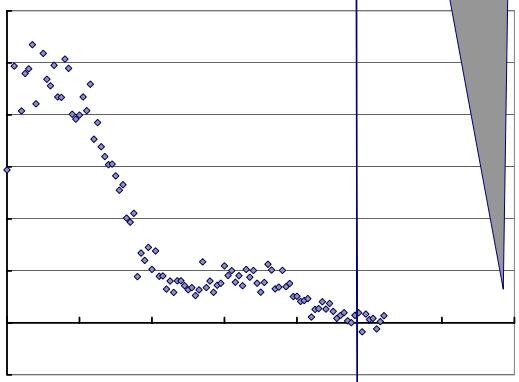


Figure 5. PSF of digital camera.



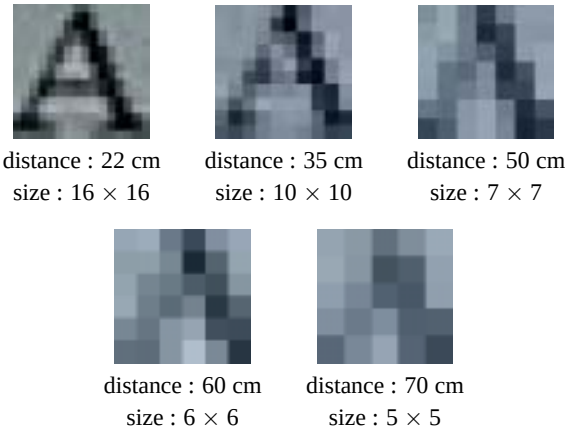


Figure 10. Test data captured with DV camera.

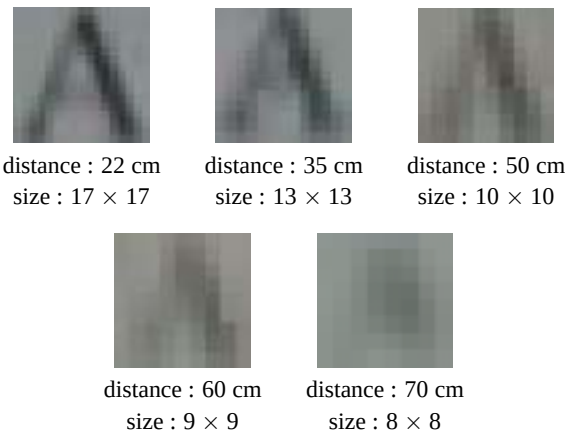


Figure 11. Test data captured with DC.

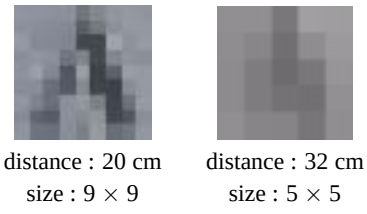


Figure 12. Test data captured with phone camera.

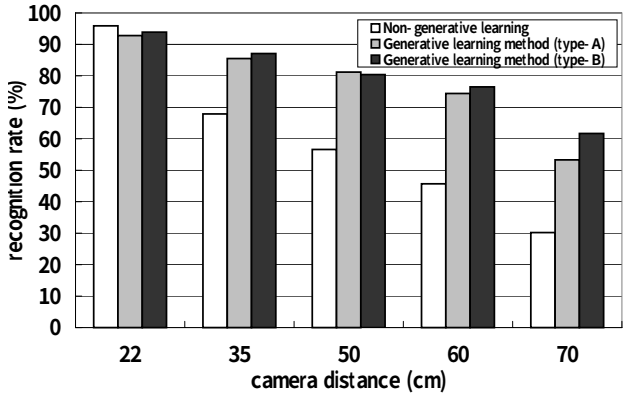


Figure 13. Recognition results (DV camera).

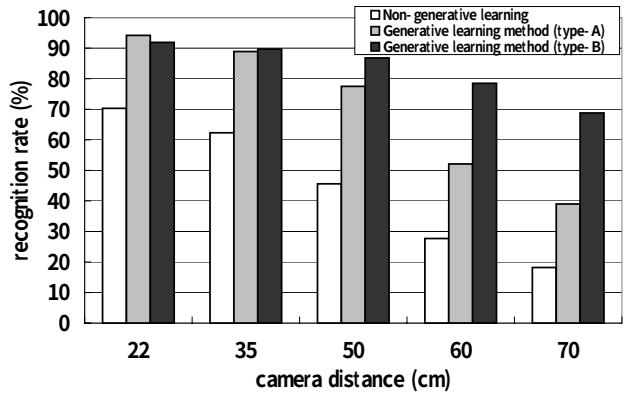


Figure 14. Recognition results (DC).

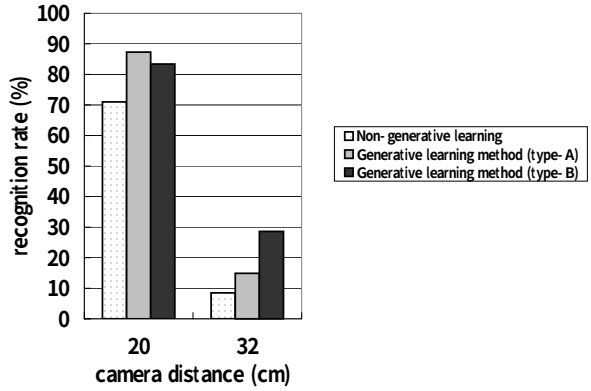


Figure 15. Recognition results (Phone camera).

The recognition rates of generative learning method types A and B were almost comparable where the size of the target characters were over 10×10 pixels. But the generative learning method (type-B) marked higher recognition rates for low-resolution characters. Results of experiments using a digital camera indicated the comparative robustness of generative learning method (type-B) while the recognition rates of other methods suddenly dropped in proportion to camera distance, indicating that generative learning method (type-B) is robust to the influence of optical blur. The shape of PSF estimated by a digital camera (Fig. 5) shows from its smooth waveshape that images captured by a digital camera are affected by optical blur. These experimental results showed that a generative learning method using estimated PSF is suitable for severely degraded and low-resolution characters.

6. Conclusion

In this paper, we proposed a learning method for the efficient recognition of low-quality characters. We proposed a generative learning method that artificially generates training data instead of collecting them from actual captured images. We applied a generative learning method with PSF estimated from captured images and examined the effectiveness of the method by experiments with three types of cameras. A generative learning method based on PSF proved to be efficient for our purposes.

Acknowledgment

Parts of this research were supported by the Grant-In-Aid for Scientific Research (16300054) and the 21st century COE program from the Ministry of Education, Culture, Sports, Science and Technology.

References

[1] D. Doermann, J. Liang and H. Li, Progress in Camera-Based Document Image Analysis. *Proc., 5th International Conference on Document Analysis and Recognition* : 606–616, August, 2003.

[2] V. Govindan and A. Shivaprasad, Character Recognition – A Review. *Pattern Recognition*, 23(7): 671–683, July 1990.

[3] S. Mori, K. Yamamoto, and M. Yasuda, Research on Machine Recognition of Handprinted Characters. *IEEE Trans. PAMI*, 6(4): 386–405, July 1984.

[4] S. Kahan, T. Pavilidis, and H. Baird, On the Recognition of Printed Characters of Any Font and Size. *IEEE Trans. PAMI*, 9(2): 274–288, March 1987.

[5] E. Oja, *Subspace Methods of Pattern Recognition.* Research Studies, Hertfordshire, UK, 1983.

[6] H. Murase, H. Kimura, M. Yoshimura, and Y. Miyake, An Improvement of the Auto-correlation Matrix in the Pattern Matching Method and its Application to Handprinted "HIRAGANA" Recognition (in Japanese). *IEICE Trans.*, J64-D(3): 276–283, March 1981.

[7] H. Andrew and B. Hunt, *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, N.J., 1977.

[8] S. Hashimoto and H. Saito, Restoration of Shift Variant Blurred Image Estimating the Parameter Distribution of Point Spread Function. *Systems and Computers in Japan*, 26(1): 62–72, January 1995.

[9] N. Tsunashima and M. Nakajima, Estimation of Point Spread Function Using Compound Method and Restoration of Blurred Images (in Japanese). *IEICE Trans.*, J81-D-II(11): 2688–2692, November 1998.

[10] J. Hobby and H. Baird, Degraded Character Image Restoration. *Proc., 5th UNLV Symp. on Document Analysis & Information Retrieval*, Las Vegas (USA), April 1996.

[11] H. Li and D. Doermann, Text Enhancement in Digital Video using Multiple Frame Integration. *Proc. 7th ACM International Conference on Multimedia* : 19–22, November, 1999.

[12] S. Yanadume, Y. Mekada, I. Ide, and H. Murase, Recognition of Very Low-resolution Characters from Motion Images. *Proc. PCM2004, Lecture Notes on Computer Science Springer-Verlag*, 3331: 247–254, December 2004.