RECOGNITION OF CAMERA-CAPTURED CHARACTERS IN BLURRED IMAGE USING MOTION-BLUR PARAMETERS

H. Ishida[1,2], T. Takahashi[1,2], I. Ide[1], Y. Mekada[3], H. Murase[1]

[1] Nagoya University, {hishi, ttakahashi, ide, murase}@murase.m.is.nagoya-u.ac.jp
[2] Japan Society for the Promotion of Science
[3] Chukyo University, y-mekada@life.chukyo-u.ac.jp

We propose a recognition method of camera-captured characters using motion-blur parameters. One of the most challenging problems in recognizing characters with hand-held cameras is motion-blur effect. In order to cope with this problem, a generation-based learning method is introduced in the training step to simulate blurred images. In the recognition step, motion-blur parameters are estimated and used for the matching to the artificial blurred images. We have experimentally proved that the effective use of the parameters improves the recognition accuracy of camera-captured characters.

Introduction

Camera-based character recognition has gained attention with the growing use of camera-equipped portable devices [1]. However, even with the improvement of the devices, the quality of captured images is still not sufficient for recognizing characters in many practical cases. One of the main approaches to cope with such degradation is image restoration. Various attempts have been made for image restoration [2, 3]. In practical applications, however, restoring an image is not always effective for character recognition because small characters are difficult to de-blur. This paper proposes a recognition method that does not need any restoration. It instead, copes with the degradations by learning artificially degraded images and using estimated motion-blur parameters.

Figure 1 illustrates the flow of the proposed method. The training step is based on the generative learning method [4, 5], where training data undergoing various speeds and orientations of motion blur are generated. The recognition method consists of two steps. The first employs the subspace method [6]. However, the subspace method constructs a single subspace from the training data with various speeds and orientations of blur, which often yields the misclassification among structurally similar characters. The eigenspace method [7] can be more effective for such characters, since the similarity to each training datum is evaluated. A reclassification based on the eigenspace method is then introduced as the second step to improve the recognition accuracy of such characters. This second step reclassifies characters by effective use of the motion blur. For this purpose, motion-blur parameters are estimated from camera motion; the similarity between the characters and training data simulated with the motion-blur parameters is evaluated in the recognition step.

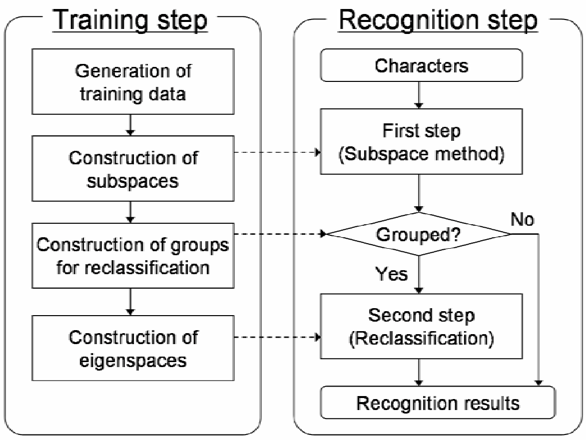


Fig. 1. Flow of proposed method.

Motion-blur model

A motion-blurred image $g(x, y)$ is generated from an original image $f(x, y)$ from a blur extent parameter b and a blur angle parameter θ as,

$$g(x,y)=\int_{-1/2}^{1/2} f\left(x-bt\cos\theta,\, y-bt\sin\theta\right)dt\,. \qquad (1)$$

This operation can also be simplified in the form of a convolution with a motion-blur PSF $h_{(b,\theta)}(x,y)$. The two-dimensional Fourier transformation is used to separate the blur component from the term $f(x,y)$ as,

$$H_{(b,\theta)}(u,v)=\frac{\sin[\pi b(u\cos\theta+v\sin\theta)]}{\pi b(u\cos\theta+v\sin\theta)}, \qquad (2)$$

with $h_{(b,\theta)}(x,y)$ obtained by inverting $H_{(b,\theta)}(u,v)$. Consequently, Eq. (1) is represented as

$$g(x,y)=f(x,y)*h_{(b,\theta)}(x,y)\,. \qquad (3)$$

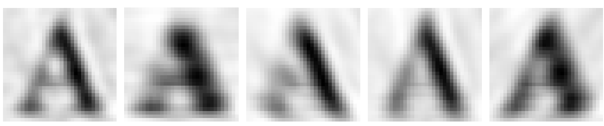
Figure 2 shows an example of $h_{(b,\theta)}(x,y)$.

(a) $H_{(b,\theta)}(u,v)$ (b) $h_{(b,\theta)}(x,y)$

Fig. 2. Example of motion-blur PSF ($b=5,\ \theta=\pi/4$).

Generation of training data

The generative learning method [4, 5] is used to generate synthetic degraded patterns by simulating actual degradation. Various training data are generated from original character images by applying motion-blur PSF $h_{(b,\theta)}(x,y)$. Examples of the generated training data are shown with corresponding parameters in Fig. 3.



$b=0$ | $b=20$, $\theta=0$ | $b=20$, $\theta=\pi/4$ | $b=20$, $\theta=\pi/2$ | $b=20$, $\theta=3\pi/4$

Fig. 3 Examples of generated images for category "A".

First step of the recognition (Subspace method)

The first step of the recognition is based on the subspace method [6]. In the training step, a subspace is constructed from generated training data for each category. The constructed subspaces have ability to classify low-quality characters robustly, except for some structurally similar categories.

[Training step] Let $\boldsymbol{x}_{(b,\theta)}^{(c)}$ be a vector constructed from the pixel values of category c's image generated with parameters b and θ. An auto-correlation matrix $\boldsymbol{Q}_1^{(c)}$ is calculated as,

$$\boldsymbol{Q}_1^{(c)}=\boldsymbol{X}_1^{(c)}\left(\boldsymbol{X}_1^{(c)}\right)^{\mathrm{T}}, \qquad (4)$$

$$\boldsymbol{X}_1^{(c)}=\left[\ \boldsymbol{x}_{(b,\theta)}^{(c)}\ \right],\quad (0\le b\le 20,\ 0\le\theta<2\pi)\,. \qquad (5)$$

The eigenvalues and corresponding eigenvectors of this matrix $\boldsymbol{Q}_1^{(c)}$ are then calculated. Examples of the eigenvectors are illustrated in Fig. 4.

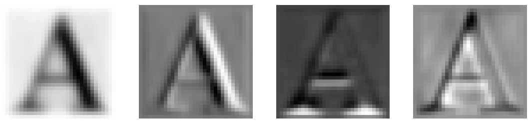


Fig. 4 Top four eigenvectors of category "A".

[Recognition step] Using the constructed subspaces $\{\boldsymbol{e}_1^{(c)}\cdots\boldsymbol{e}_R^{(c)}\}$, a given image $\boldsymbol{z}$ is classified to c_1 as,

$$c_1=\arg\max_{c}\sum_{r=1}^{R}(\boldsymbol{e}_r^{(c)}\cdot\boldsymbol{z})^2\,. \qquad (6)$$

Second step of the recognition (Eigenspace method)

The recognition results obtained in the first step tend to involve misclassification within certain groups of structurally similar categories. The second step attempts to reclassify such dubious results to the correct category using the eigenspace method [7]. The blur parameters estimated from camera motion are used for the matching of characters in this step. This attempt is grounded on the idea that the blur parameters should supply supplementary information for differentiating structurally similar categories.

[Training step] For each category g, characters that are frequently misclassified to category g are grouped and described as $\mathcal{G}^{(g)}$. Such groups can be organized by applying the first step of recognition to a certain amount of samples. Let $\rho(g|c)$ denote the rate at which a character in category c is classified to category g in the first step. The category c is grouped if $\rho(g|c) \geq \tau$, where τ is a grouping threshold; and of course, g itself also needs to be a member of $\mathcal{G}^{(g)}$. In brief, $\mathcal{G}^{(g)}$ is organized as,

$$\mathcal{G}^{(g)} = \{\, c \mid \rho(g \mid c) \geq \tau \} \cap \{g\}. \qquad (7)$$

An eigenspace used for this second step of recognition is constructed in each group. Similar to the step in which the subspaces were constructed, an auto-correlation matrix $\boldsymbol{Q}_2^{(g)}$ of group g is calculated as

$$\boldsymbol{Q}_2^{(g)} = \boldsymbol{X}_2^{(g)} \left(\boldsymbol{X}_2^{(g)}\right)^{\mathsf{T}}, \qquad (8)$$

$$\boldsymbol{X}_2^{(g)} = \left[\, \boldsymbol{x}_{(b,\theta)}^{(c)} - \boldsymbol{\mu}^{(g)} \,\right], \quad \begin{pmatrix} c \in \mathcal{G}^{(g)} \\ 0 \leq b \leq 20 \\ 0 \leq \theta < 2\pi \end{pmatrix}. \qquad (9)$$

$$\boldsymbol{\mu}^{(g)} = \mathcal{E}\,[\, \boldsymbol{x}_{(b,\theta)}^{(c)} \,]. \qquad (10)$$

The eigenvalues and corresponding eigenvectors of this matrix $\boldsymbol{Q}_2^{(g)}$ are then calculated. Examples of the eigenvectors are illustrated in Fig. 5.

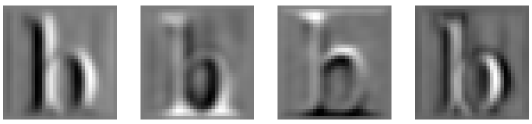


Fig. 5 Top four eigenvectors of group $\mathcal{G}^{(h)}$={“b”, “h”}.

[Recognition step] Given that a recognition result from the first step is g, if g is grouped, we dismiss the results from the first step once and compute the final result as described below.

As for blur information, a blur extension parameter $\hat{b}$ and a blur angle parameter $\hat{\theta}$ are estimated from camera motion as

$$\hat{b} = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad \text{and} \qquad (11)$$

$$\hat{\theta} = \tan^{-1}\left(\Delta y / \Delta x\right), \qquad (12)$$

where Δx and Δy are estimated from the motion vector of the target characters.

The given image $\boldsymbol{z}$ is projected onto the eigenspace $[\boldsymbol{e}_1^{(g)} \cdots \boldsymbol{e}_R^{(g)}]$ of group $\mathcal{G}^{(g)}$ and then classified as illustrated in Fig. 6 by

$$c_2 = \arg\min_{c \in \mathcal{G}^{(g)}} \left[\min_{\substack{\hat{b}-\Delta b \leq b \leq \hat{b}+\Delta b \\ \hat{\theta}-\Delta\theta \leq \theta \leq \hat{\theta}+\Delta\theta}} \left\| \boldsymbol{\zeta} - \boldsymbol{\xi}_{(b,\theta)}^{(c)} \right\| \right], \qquad (13)$$

$$\boldsymbol{\zeta} = [\boldsymbol{e}_1^{(g)} \cdots \boldsymbol{e}_R^{(g)}]^{\mathsf{T}} (\boldsymbol{z} - \boldsymbol{\mu}^{(g)}), \qquad (14)$$

$$\boldsymbol{\xi}_{(b,\theta)}^{(c)} = [\boldsymbol{e}_1^{(g)} \cdots \boldsymbol{e}_R^{(g)}]^{\mathsf{T}} (\boldsymbol{x}_{(b,\theta)}^{(c)} - \boldsymbol{\mu}^{(g)}). \qquad (15)$$

Here $\mathcal{B}$ in Fig. 6 is equivalent to the parameter range in Eq. (13). The classification is based on the nearest neighbor rule.

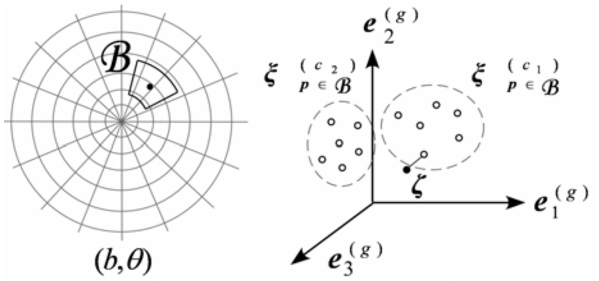


Fig. 6 Eigenpoints used for the classification are selected using estimated blur parameters.

Experiment

We evaluated the effectiveness of the proposed method with a digital camera (Panasonic DMC-FX9) that provides the ability to record video with a spatial resolution of 640×480 pixels and 30 frames per second. As test data, 62 characters (A-Z, a-z, 0-9: Century font) printed on a paper were captured. The distance to the paper was 30 cm, and the focal length of the camera was 5.8 cm; the average character size in the captured images was 11×11 pixels. The segmented area for each character was the minimum square that includes the whole character.

For the training, the blur extent parameter was changed by 11 steps $(b = 0, 2, ..., 20)$, and the blur angle parameter by 12 steps $(\theta = 0, \pi/12, ..., 11\pi/12)$. The original images were also in Century font. The size of the generated training data was 32×32 pixels. The number of eigenvectors was determined such that the cumulative contribution was over 95%. Next, the groups used for the reclassification were organized using some samples. As the recognition samples for the grouping, 300 image sequences composed of 10 successive frames each were taken in the same way as the test data. Table 1 shows the constructed groups under several grouping thresholds τ.

Table 1. Groups (g: $\mathcal{G}^{(g)}$) within which characters are reclassified.

τ = 0.01	τ = 0.02	τ = 0.05
L:{L, t} O:{O, o} R:{F, R} S:{S, 8} V:{V, v} h:{b, h} W: {W, w} l: {I, i, j, l, 1} 1: {i, 1}	S: {S, 8} V: {V, v} W:{W, w} h: {b, h} l:{I, i, l, 1} 1: {i, 1}	V: {V, v} W:{W, w} h: {b, h} l: {I, i, l, 1}

For testing, two photographic conditions were set for this experiment. Image sequences for the test data were taken under:

Condition A : Camera held as still as possible.
Condition B : Camera held by vibrating hand.

We used image sequences composed of ten successive frames for the tests. The number of the image sequences for Conditions A and B were 1,736 and 503, respectively. The image sequences were taken by six persons. The distribution of the estimated blur extent parameter $\hat{b}$ is given in Fig. 7, where we can see that $\hat{b}$ was not always zero even under Condition A. The parameters in Eq. (13) were set as $(\Delta b, \Delta\theta) = (5, \pi/12)$.

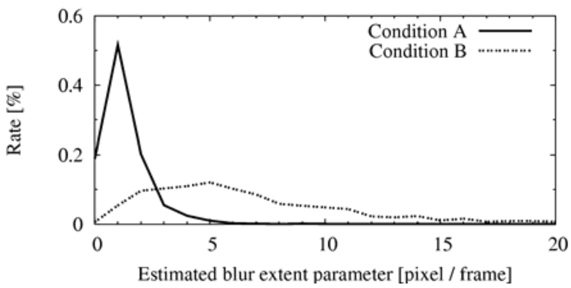


Fig. 7 Distribution of estimated blur extent $\hat{b}$ by two photographic conditions.

Recognition results under various grouping thresholds τ are presented in Fig. 7. The proposed method was compared also with the subspace method without reclassification. The results from both conditions indicate that the second step of the proposed method successfully reclassified some characters which were not correctly classified by the subspace method. On the other hand, grouping too many characters (τ = 0.01) was not effective. The determination of the optimum value of τ should be discussed in future work.

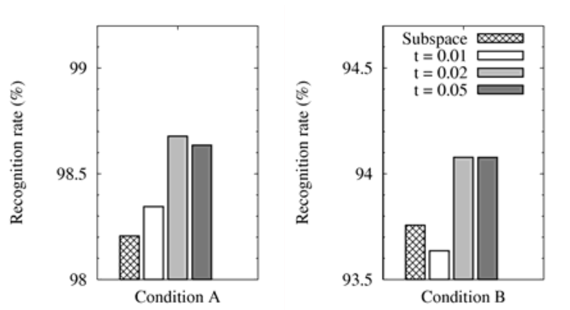


Fig. 7 Recognition rates under conditions A and B.

Conclusion

In this paper, we proposed a method for improving the recognition accuracy of camera-captured characters using motion-blur parameters. It was experimentally proved that the effective use of them can reduce classification errors. Evaluating the method's effectiveness under various conditions is a future work.

Acknowledgement

Parts of this research were supported by the Grants-In-Aid for Scientific Research (16300054 and 17650050).

References

1. D. Doermann, J. Liang, H. Li, Progress in camera-based document image analysis, Proc. 5th Int. Conf. on Document Analysis and Recognition, Edinburgh, Scotland. (2003) 606–616.
2. J. Hobby, H. Baird, Degraded character image restoration, Proc. 5th UNLV Symp. on Document Analysis and Information Retrieval, Las Vegas, USA. (1996) 177–189.
3. H. Li, D. Doermann, Text enhancement in digital video using multiple frame integration, Proc. 7th ACM Int. Conf. on Multimedia, Orlando, Florida. (1999) 19–22.
4. H. Murase, S. Nayar, Learning by a generation approach to appearance-based object recognition, Proc. 13th Int. Conf. on Pattern Recognition, Vienna, Austria. (1996) 24–29.
5. H. Ishida, S. Yanadume, T. Takahashi, I. Ide, Y. Mekada, 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) 45–51.
6. E. Oja, Subspace methods of pattern recognition, Research Studies, Hertfordshire, UK, 1983.
7. H. Murase, S. Nayar, Three-dimensional object recognition from appearance —parametric eigenspace method, Systems and Computers in Japan. 26 (8) (1995) 45–54.