

Recognition of Very Low-Resolution Characters from Motion Images Captured by a Portable Digital Camera

Shinsuke Yanadume¹, Yoshito Mekada², Ichiro Ide¹, and Hiroshi Murase¹

¹ Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8603 Japan
yanadume@murase.nuie.nagoya-u.ac.jp, {ide,murase}@is.nagoya-u.ac.jp

² Life System Science and Technology, Chukyo University
101 Tokodachi, Kaizu, Toyota, Aichi, 470-0393 Japan
y-mekada@life.chukyo-u.ac.jp

Abstract. Many kinds of digital devices can easily take motion images such as digital video cameras or camera-equipped cellular phones. If an image is taken with such devices under everyday situations, the resolution is not always high; moreover, hand vibration can cause blurring, making accurate recognition of characters from such poor images difficult. This paper presents a new character recognition algorithm for very low-resolution video data. The proposed method uses multi-frame images to integrate information from each image based on a subspace method. Experimental results using a DV camera and a phone camera show that our method improves recognition accuracy.

1 Introduction

Recently, opportunities for taking videos with such portable equipments as digital video cameras (DV camera) or camera-equipped cellular phones (phone camera) continue to increase. If a system could automatically recognize the characters from such video data, it could become a key piece of technology for the next generation of human-machine interfaces. For example, in the future, we will easily be able to scan and input URLs from magazines by phone cameras or send text by e-mail by recognizing characters from the images of captured notes. Many character recognition methods have already been proposed[1]. However, such methods generally assume that the image quality of characters is quite high. On the other hand, the quality of characters captured by portable digital cameras is often not sufficient to apply these methods; the image of the characters might be too small when a full document is captured in a single shot. Moreover, hand movement or poor lens quality might blur the image. It is difficult to recognize such low-quality characters from a single image. Eims et al.[9] proposed a method to recognize low-quality images from a image scanner, but this method is not sufficient in our case when the resolution of a single image is not enough. When we capture a character on video, we obtain a variety of character images

as a sequence of motion images. If we properly use such information, recognition of very low-resolution characters may become possible, even if we cannot recognize them from a single image.

In this paper, we propose a method that recognizes characters from poor quality video images. Cheeseman et al.[2] generated an image with higher resolution from multi-frame low-resolution images. Thus restoration of low-resolution images is one solution[10]. On the other hand, we take information directly from multi-frame images at the recognition step and integrate that information with a subspace method[3,4,5,6,7]. We generate subspaces that approximate a set of a large number of training images and compute the degree of similarity of the subspace. Finally, we use the multi-frame images input for recognition.

The proposed method consists of the following three parts: gathering training data, constructing subspaces, and recognizing characters from input images. In the training step, our method uses many variations of characters that are segmented from sequences of videos at various resolutions. The recognition step does not need to estimate camera movements to recognize a character, unlike a previously proposed method by Sawaguchi et al.[8].

We describe the characteristics of the characters in the video data in section 2, propose the algorithm in section 3, and show the experimental results in section 4.

2 Characters in Video Data

2.1 Portable Digital Cameras

Figure 1(c) shows a typical example of a character captured by a portable digital camera that is obviously difficult to recognize from the single image shown. When we photograph a full document with a common portable camera, as shown in (Fig.1 (a)), each character is in low-resolution. Our aim is to recognize such poor quality characters as shown in Fig.1(c) by using the information from multi-frame images.

2.2 Characteristics of Videodata

When we take a video using a portable handy camera, hand movement slightly shifts and rotates the camera, making it difficult to fix the camera position difficult. Therefore, a large variation generally exists in a sequence of video images, even for the same character. If we can properly integrate the information from these images, recognition of a very low-resolution character may become possible, even if we cannot recognize it from a single image. Figure 2 shows character "A" obtained from two frames captured by a digital video camera. Typical character recognition algorithms might not be able to recognize these characters from a single image. However, the subtle difference between these two images provides a clue to improving the recognition accuracy.

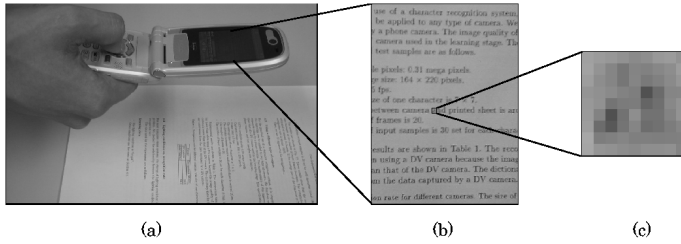


Fig. 1. Taking document image with a phone camera. (a): Taking an image with a phone camera. (b): Captured document image. (c): Segmented image of character “a”.

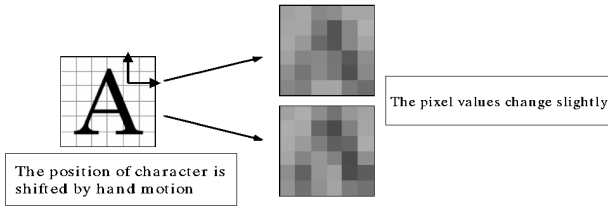


Fig. 2. Changes in pixel values due to hand motion.

3 Recognition of Characters from Motion Images

The proposed method consists of the following three parts: gathering training data, constructing subspaces, and recognizing characters from input images. We used character images captured by a portable camera as training data that helps achieve a high recognition rate and includes various cases of characters to be recognized. Eigenvectors were computed from the training data to be used to recognize input characters.

Training data and input data were generated from character images captured by a portable camera. We printed characters on a sheet with a fixed print pitch and segmented each character by this pitch information. The size of the segmented characters was normalized for use as training data and input data.

3.1 Creating Training Data

The target characters for recognition are:

- Printed characters.
- Upper and lower cases of the alphabet and the Arabic numerals.
- Characters whose images are bigger than 6×6 pixels.

The training data consisted of printed characters captured by a portable camera. We used multi-frame images from a sequence of motion images for training data because they contain many variations of the same character. Since the size of characters was unknown beforehand, we prepared training data captured at

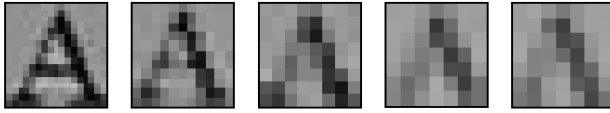


Fig. 3. Excerpt from the training data “A”.

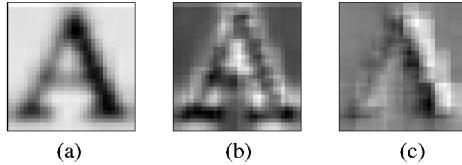


Fig. 4. Picturized eigenvector of “A”. (a): the first eigenvector. (b): the second eigenvector. (c): the third eigenvector.

various resolutions by changing the distance between the camera and the sheet. Figure 3 is an excerpt from the training data.

3.2 Construction of the Subspace from the Training Data

First, our method found the orthogonal bases of the training data for each category. Each i -th learning data image was converted to a unit vector whose average was 0 (normalization). The normalized vector is represented by

$$\mathbf{x}_i = [x_1, x_2, \dots, x_N]^T,$$

where N is the number of pixels. Next, matrix \mathbf{X} is defined as

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k],$$

where k is the amount of learning data for this category. Then, we calculated an autocorrelation matrix \mathbf{Q} for the category using matrix \mathbf{X} :

$$\mathbf{Q} = \mathbf{X}\mathbf{X}^T.$$

We constructed the subspace for each category using R eigenvectors that corresponded to the largest R eigenvalues. A set of eigenvectors was represented by,

$$\{\mathbf{e}_1^{(c)}, \mathbf{e}_2^{(c)}, \dots, \mathbf{e}_R^{(c)}\},$$

Figure 4 shows an example of eigenvectors that were computed and picturized implying that the blurred characters at several resolutions are included.

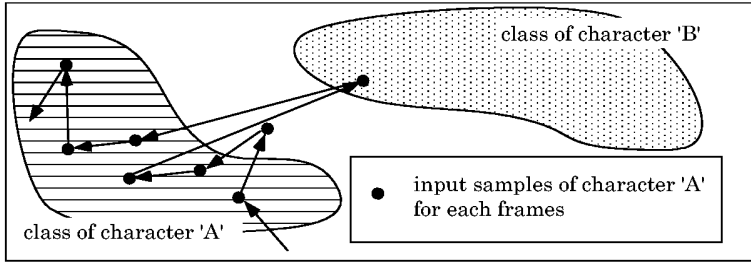


Fig. 5. Advantage of multi-frame input. This figure shows input multi-frame samples projected in subspace.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
			0	1	2	3	4	5	6	7	8	9													

Fig. 6. Set of target characters (Font: Century).

3.3 Recognition

Each character is segmented from an input video and normalized. A set of vectors for the character was constructed from multi-frame images and represented by

$$\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M\},$$

where M is the number of input frames. The similarity between category c and the input images is defined as

$$L^{(c)}(\mathbf{a}) = \frac{1}{M} \sum_{m=1}^M \sum_{r=1}^R (\mathbf{a}_m, \mathbf{e}_r^{(c)})^2$$

where (\mathbf{x}, \mathbf{y}) denotes an inner product. Then the category of input images was determined to maximize the above equation. If one sample is input which was closer to an incorrect class than a correct class, integration of multi-frame samples should enable correct category output (see Fig. 5).

4 Experiments

We verified the capability of this method experimentally by capturing a sequence of printed characters with either a portable digital camera or a phone camera. An alphanumeric ‘‘Century’’ font was used in the experiments on a total of characters in 62 categories, as shown in Fig. 6.

4.1 Recognition Rate Versus Number of Input Frames

To verify the performance of our method when applied to very low-resolution characters, we evaluated recognition rates by changing the number of frames and the size of characters. The data used for this experiment are as follows:

Training Data

- Captured with a DV camera.
- Character size controlled by changing the distance between the camera and the sheet on which the characters were printed.
 - Distance: less than 70cm.
 - Average character size : 16×16 , 11×11 , 8×8 , 7×7 , or 6×6 pixels.
- Multi-frames of each character for a total of 50 frames per character.

Dictionary Data

- Ten eigenvectors that corresponded to the ten largest eigenvalues.

Test Data

- Captured with the same DV camera
- Different from the training data.
- Two test sample sets for character size controlled by changing the distance between the camera and a sheet.
 - **small:** 70cm (approximately 6×6 pixels).
 - **medium:** 60cm (approximately 7×7 pixels).
- Number of input samples: 30 sets for each character for a total of 1,860 sets.

The results are shown in Fig. 7. Recognition rates increased as the number of input frames increased until reaching a saturation point at around 15 frames. In the *medium* size, the recognition rate almost reached 100%, indicating that our method improves recognition accuracy by inputting multi-frame images.

4.2 Lighting Conditions Versus Recognition Rate

Since changes in light conditions are a serious problem for most computer vision systems, we checked the relationship between light conditions and recognition rate.

Training data were captured in “bright” light conditions. The remaining conditions of the training data and the dictionary data are identical to Section 4.1.

Test Data

- Captured with a DV camera.
- Character size: *small*.
- Lighting conditions: “bright”, “middle”, or “dark”.
- Number of frames: 20.
- Number of input samples: 30 sets for each character for a total of 1,860.

The results in Table 1 show that our method is generally independent of light conditions. We also found that normalization (in Section 3.2) was effective.

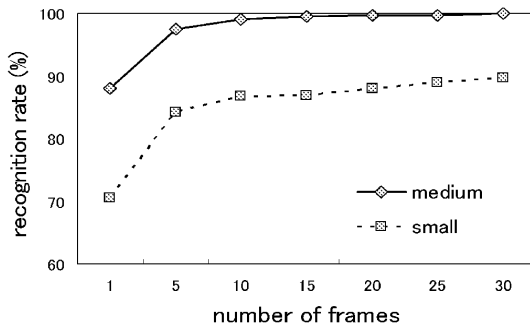


Fig. 7. Recognition rate vs. number of input frames.

Table 1. Recognition rates for change in light conditions. The character size is *small*.

Light condition	Recognition rate(%)
bright	88.1
middle	82.8
dark	85.4

4.3 Using Different Types of Cameras

For a character recognition system to be practical for use, its algorithm must be applicable to any type of camera. Therefore, we tried using image sequences taken by a phone camera in addition to the previous experiment. The image quality of this camera was worse than the DV camera used in the training stage. The specifications of this phone camera and test samples are as follows:

- Actual number of pixels of the CCD: 0.31 mega pixels.
- Captured image size: 164×220 pixels.
- Frame rate: 7.5 fps.
- Character size: *medium*.
- Distance between the camera and the printed sheet: approximately 20cm.
- Number of frames: 20.
- Number of input samples: 30 sets for each character for a total of 1,860 sets.

In the experimental results, the 92.0% recognition rate when using a phone camera is slightly lower than the 99.9% when using a DV camera because the image quality of the phone camera is inferior to the DV camera. The dictionary data for both cases were constructed from the data captured by a DV camera.

5 Conclusion

In this paper, we proposed a new framework based on a subspace method for recognizing low-quality, especially low-resolution, characters. We used various

resolutions of image sequences to construct the subspace in the training step. Experimental results show that a recognition rate of 99.9% is obtained for low-resolution alphanumeric characters about 7×7 pixels in size. Our method performs well even when devices or light conditions are changed. We conclude that our method is useful in recognizing very low-resolution characters captured by a portable digital camera. Although slight shift and rotation of camera are absorbed in the training data set, the method can not cope with a large tilt or rotation.

Future work includes adding such figures as Japanese characters and different fonts. When we recognize a character from document images, it is difficult to segment characters from low-resolution sentence images. Since we used printed, pre-segmented character images in this research, in the future we must apply this algorithm to words and sentences and explore the ramifications.

Acknowledgments. The authors thank their colleagues for useful suggestions and discussion. 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. S. Mori, K. Yamamoto, and M. Yasuda, "Research on machine recognition of hand-printed characters," *Trans. PAMI*, vol.PAMI-6, no.4, pp.386-405, July 1984
2. P. Cheeseman, B. Kanefsky, R. Hanson, and J. Stutz, "Super-resolved surface reconstruction from multiple images," Technical Report FA-94-12, NASA Ames Research Center, Artificial Intelligence Branch, October 1994
3. E. Oja, "Subspace methods of pattern recognition," Hertfordshire, UK: Research Studies, 1983.
4. 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," *IECE Trans.*, vol.J64-D, no.3, pp.276-283, March 1981
5. H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *International Journal of Computer Vision*, vol.14, pp.5-24, 1995
6. S. Omachi, and H. Aso, "A qualitative adaptation of subspace method for character recognition," *IEICE Trans.*, vol.J82-D-II, no.11, pp.1930-1939, November 1999
7. S. Uchida and H. Sakoe, "Handwritten character recognition using elastic matching based on a class-dependent deformation model," *Proc. ICDAR*, vol.1 of 2, pp.163-167, August 2003
8. M. Sawaguchi, K. Yamamoto, and K. Kato, "A proposal of character recognition method for low resolution images by using cellular phone," Technical Report of IEICE, PRMU2002-247, March 2003
9. A.J.Elms, S.Procter, and J. Illingworth, "The advantage of using an HMM-based approach for faxed word recognition," *IJDAR*, vol.1, no.1, pp.18-36, 1998
10. Paul D. Thouin and Chein-I Chang, "A method for restoration of low-resolution document images," *IJDAR*, vol.2, no.4, pp.200-210, 2000