OCR Fonts Revisited for Camera-Based Character Recognition

Seiichi Uchida
Kyushu Univ., Japan
uchida@is.kyushu-u.ac.jp

Masakazu Iwamura
Osaka Pref. Univ., Japan
masa@cs.osakafu-u.ac.jp

Shinichiro Omachi
Tohoku Univ., Japan
machi@aso.ecei.tohoku.ac.jp

Koichi Kise
Osaka Pref. Univ., Japan
kise@cs.osakafu-u.ac.jp

Abstract

In order to realize accurate camera-based character recognition, machine-readable class information is embedded into each character image. Specifically, each character image is printed with a pattern which comprises five stripes and the cross ratio derived from the pattern represents class information. Since the cross ratio is a projective invariant, the class information is extracted correctly regardless of camera angle. The results of simulation experiments showed that recognition rates over 99% were obtained by the extracted cross ratio under heavy projective distortions.

1. Introduction

For accurate camera-based character recognition [1], an embedment of machine-readable class information into character image is examined. Specifically, each character image is printed with a horizontal stripe pattern, called a *cross ratio pattern*. Figure 1 (a) shows a character image "K" printed with a cross ratio pattern. The cross ratio derived from the pattern is related to the class of the character image. Since the cross ratio is a projective invariant, correct class information will be extracted even from character images captured from an arbitrary camera angle.

The character image with the cross ratio pattern can be considered as a new machine-readable font. Thus, it is closely related to so-called "OCR fonts" and "MICR (magnetic ink character recognition) fonts", which were proposed in the dawn of OCR/MICR research [2]. DataGlyph [3, 4] is a more recent trial. Those conventional fonts were designed for scanner-based recognition and therefore not suitable for camera-based recognition. For example, they are not robust to projective distortions.

The number of distinguishable cross ratio patterns is often fewer than the number of classes under a limited image resolution. In this case, the same cross ratio should be assigned to several different character classes and the character class cannot be determined uniquely by the extracted cross ratio. Thus, we employ a shape similarity between reference and input character images together with the cross ratio.

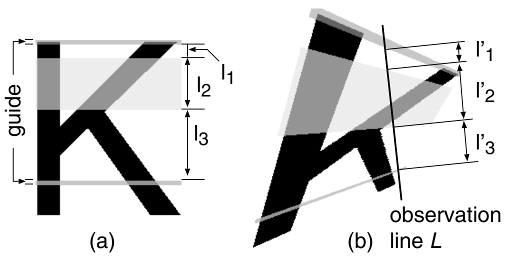

Figure 1. (a) A character image "K" printed with a cross ratio pattern. (b) Projective distortion and observation line $\mathcal{L}$.

2. Embedment of cross ratio pattern to character image

2.1 Cross ratio pattern

The cross ratio pattern is comprised of five horizontal stripes as shown in Fig 1 (a). The top and the bottom stripes are guides which have a fixed width and define the beginning and the end of the cross ratio pattern. The remaining three stripes have variable widths, l_1, l_2, and l_3 and provide a well-known projective invariant called *cross ratio*,

$$r = \frac{(l_1 + l_2)(l_2 + l_3)}{l_2(l_1 + l_2 + l_3)}. \tag{1}$$

The cross ratio pattern can provide class information by assigning a cross ratio r_k $(k = 1, 2, \ldots, K)$ to the class $c \in \boldsymbol{C}$. (This implies that the character of the class c is always printed with the cross ratio pattern which gives $r = r_k$.) The detail of the assignment will be discussed in Section 4. Figure 2 shows character images printed with different cross ratio patterns (i.e., $K = |\boldsymbol{C}|$).

2.2 Extraction of cross ratio

The cross ratio r_k can be extracted from a character image printed with a cross ratio pattern by drawing an *observation line* $\mathcal{L}$ which crosses two guides and then measuring the widths of the three stripes on $\mathcal{L}$ (Fig. 1(b)). Since the cross ratio is a projective invariant, we can obtain the original cross ratio r_k by using measured widths l'_1, l'_2, and l'_3

Figure 2. Character images printed with different cross ratio patterns (i.e., $K = 26$).

		recog res.				opt. assign.	bad assign.
		A	B	C	D		
input	A	0.6	0.2	0	0.2	k=1	1
	B	0	0.7	0.3	0	2	1
	C	0	0	1.0	0	1	2
	D	0.2	0	0	0.8	2	1

Figure 3. Confusion matrix and two assignments of cross ratios.

instead of l_1, l_2, and l_3 in (1), regardless of camera angle and the position and the slope of the observation line $\mathcal{L}$.

The extracted cross ratio may be incorrect due to insufficient camera resolution, blurring, lighting condition, etc. Thus, we use the following voting strategy for a robust estimation of r_k: (i) we draw the observation line $\mathcal{L}$ on the character image P times changing its position and slope randomly, (ii) obtain P cross ratio values, (iii) quantize each of those values into one of $\{r_k\}$, and (iv) choose the most frequent r_k as the cross ratio embedded.

3. Combining cross ratio with shape similarity

When the variations of the cross ratios are fewer than character classes (i.e., $K < |\boldsymbol{C}|$), the same cross ratio is assigned to several different character classes and we cannot determine the character class uniquely from the extracted cross ratio. Unfortunately, this case is not rare; for example, a Chinese character set has $|\boldsymbol{C}| > 1000$ classes, whereas K is bounded by image resolution. Let $\boldsymbol{C}_k \subset \boldsymbol{C}$ denote the set of classes to which the cross ratio r_k is assigned. We should select the most reliable class from $|\boldsymbol{C}_k|$ candidates when r_k is extracted.

For the selection, some shape similarity is helpful; that is, (i) extract the embedded cross ratio r_k, (ii) calculate the shape similarity between the input character image and the reference character image of each class in $\boldsymbol{C}_k$, and (iii) select finally the class with the highest shape similarity. Note that this selection procedure (i.e., the recognition procedure) totally relies on the extracted cross ratio r_k. If a wrong r_k is extracted, the correct class is never chosen by the procedure. Fortunately, the cross ratio r_k can be extracted with high accuracy thus good performance is expected.

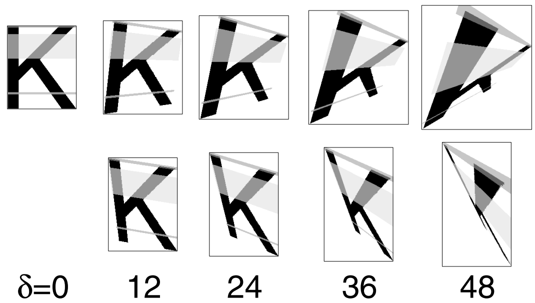

Figure 4. Test patterns, which undergo projective distortion.

4. Optimal assignment of cross ratios to classes

The assignment of K cross ratios to $|\boldsymbol{C}|$ classes, i.e., the partition of $\boldsymbol{C}$ into the disjoint subsets $\{\boldsymbol{C}_k\}$, is crucial. The recognition procedure of Section 3 provides correct recognition results if (i) the cross ratio r_k is correctly extracted and (ii) the correct class has the highest shape similarity among $\boldsymbol{C}_k$. The condition (ii) indicates that the subset $\boldsymbol{C}_k$ should be comprised of classes which are "less likely to confuse" for the shape similarity [5].

Figure 3 is an example of a confusion matrix ($|\boldsymbol{C}| = 4$) which represents the confusing classes of some shape similarity. Here, "A" and "B" are a pair of confusing classes. This is because when an input pattern is determined as "B" by the shape similarity, the correct class is "A" or "B". If the bad assignment in Fig. 3 is used, the set $\boldsymbol{C}_1 = \{$"A", "B", "D"$\}$ includes the confusing classes and thus we will suffer from the misrecognition between "A" and "B", even though their cross ratios are correctly extracted as r_1. In contrast, if the optimal assignment in Fig. 3 is used, "A" and "B" can be distinguished by their cross ratios and therefore correct recognition results will be provided.

As shown by this example, the assignment $\{\boldsymbol{C}_k\}$ should be optimized with a criterion that major confusing classes are assigned to different subsets. The detailed procedure of the assignment optimization can be found in [5].

5. Simulation experiment

A simulation experiment was conducted by using synthetic test samples for quantitative evaluations of the proposed framework.

5.1 Experimental setup

The 26 capital English letter images from the font-set called "Arial" were used as original character images. Their heights were around 200 pixels. After embedding cross ratio patterns into them, those images were used as not only reference patterns but also the source patterns for preparing test patterns.

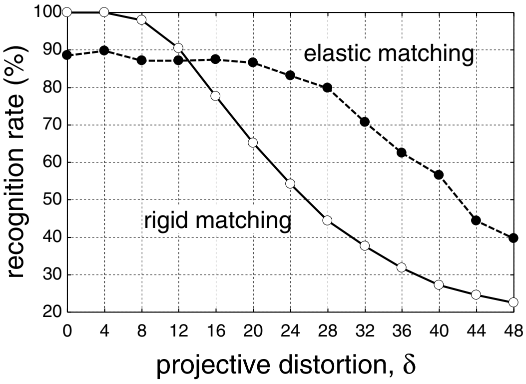

Figure 5. Recognition rates attained by using shape similarity alone.

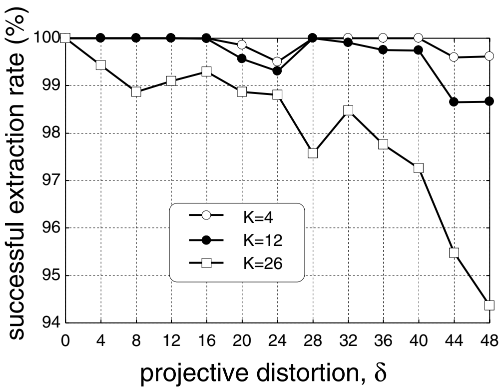

Figure 6. Extraction accuracy of cross ratios.

Test patterns were prepared by applying projective distortions on the original character images with the cross ratio patterns. The projective distortion was controlled by the (x, y)-displacement of four corners of a character image. For each corner, there were four variations of the (x, y)-displacement, $(\pm\delta, \pm\delta)$ and $(\pm\delta, \mp\delta)$, where δ denotes the displacement by the pixel. Thus, for a fixed δ, $4^4 = 256$ test patterns were created from a single original character image. Figure 4 shows several test patterns of "K".

5.2 Shape similarity

Two matching techniques, simple rigid matching (i.e., superimposing) and elastic matching [6], were employed for evaluating the shape similarity used in the recognition procedure of Section 3. The elastic matching provides a shape similarity after fitting the reference pattern to the input pattern nonlinearly.

Figure 5 shows recognition accuracy only by the shape similarities. The rigid matching was very sensitive to projective distortions and its recognition accuracy decreases drastically according to the increase of δ. On the other hand, the elastic matching is rather robust to the projective distortions. For more heavy distortions, however, its accuracy decreases like the rigid matching.

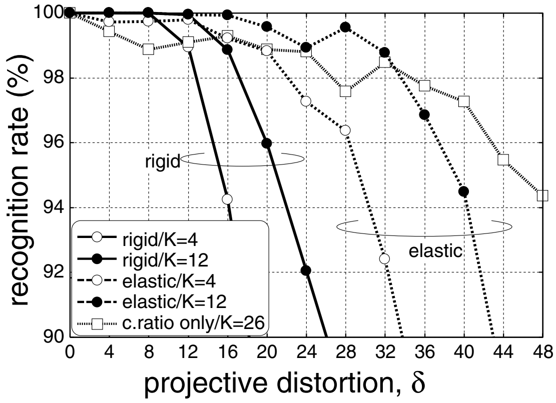

Figure 7. Recognition rate by the combination of shape similarity and cross ratio. Exceptionally, when $K = 26$, character class was determined only by extracted cross ratio.

5.3 Extraction accuracy of cross ratios

Figure 6 shows the extraction accuracy of the cross ratios as a function of δ. This graph indicates that the cross ratios can be extracted very accurately even under heavy distortions. By comparing Figs. 5 and 6, it is shown that this accuracy is far higher than the recognition rates by shape similarities. Thus, the cross ratio is more reliable information than the shape similarities for camera-based character recognition.

The graph at $K = 26$ in Fig. 6 means the recognition rate attained with the cross ratio only, i.e., without any character shape information. High recognition rates over 98% were attained for $\delta \leq 24$.

Extraction failures were mainly due to slight errors of l'_1, l'_2, l'_3 by insufficient resolution. In fact, at $K = 26$, 85% of extraction failures were "near-misses" that r_k was detected as $r_{k\pm1}$. More serious failures that r_k was detected as $r_{k\pm\Delta}$ $(\Delta \geq 2)$ were 10%. The remaining 5% were the failures in the detection of guides thinned by heavy projective distortion.

5.4 Recognition accuracy by using cross ratio and shape similarity

Figure 7 shows the recognition rates by using the cross ratios and the shape similarity according to the procedure of Section 3. The cross ratios ($K = 4$ or 12) were optimally assigned to $|\boldsymbol{C}| = 26$ classes according to the policy of Section 4. In Fig. 7, the recognition rate at $K = |\boldsymbol{C}| = 26$, i.e., the recognition rate by the cross ratio alone, is also plotted.

Figure 7 indicates that recognition rates were drastically improved from the rates of Fig. 5. At $\delta = 16$, for exam-

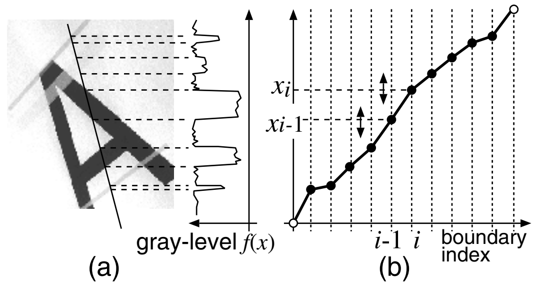

Figure 8. Segmentation of observation line.

ple, the recognition rate attained by the shape similarity by the elastic matching was 87.4% and improved to 99.2% and 99.9% by $K = 4$ and 12 cross ratios, respectively. This improvement was achieved by the help of class information provided by the cross ratio pattern. For example, the misrecognitions between "N" and "H", which were confusing classes for the elastic matching, were fairly reduced because different cross ratios were assigned to those classes and the wrong class was not included in the candidate class set $\boldsymbol{C}_k$.

Figure 7 also indicates that $K = |\boldsymbol{C}|$ cross ratios are not necessary if a shape similarity is available. In fact, only four cross ratios could attain satisfying recognition rates (around 99%) when projective distortion was not severe. This result suggests that the proposed technique can be applied to the recognition of characters with many classes.

6. Toward practical use

The experiments in the previous section were conducted on the synthetic test patterns free from any gray-level change. Thus, the measurement of the stripe widths (l_1, l_2, l_3) was easy. In contrast, if such patterns are placed in a scene and captured by a camera, the measurement is not easy because the patterns undergo gray-level change. Thus, we should perform the segmentation of $f(x)$ which represents the observed gray-scale value at x on the observation line $\mathcal{L}$ and then determine the stripe widths (Fig. 8 (a)).

As a promising segmentation strategy, we choose an optimization-based strategy here. A simple objective function J for the segmentation is defined as

$$J = \sum_{i=2}^{I} \int_{x_{i-1}}^{x_i} \| f(x) - \overline{f}_{x_{i-1},x_i} \| dx \to \min, \quad (2)$$

where I is the number of segment boundaries on $\mathcal{L}$ and x_i is the ith boundary and $\overline{f}_{x_{i-1},x_i}$ is the average of $\{f(x) \mid x_{i-1} < x < x_i\}$. The optimization of J with respect to $\{x_i\}$ can be considered as an optimal path problem (Fig. 8 (b)) and solved efficiently by a dynamic programming-based algorithm (whose details are omitted here).

For practical evaluation of this segmentation strategy, a limited experiment was conducted using 167 character images captured by a digital camera from various angles. The experimental result showed that 80.4% boundaries on $13,527$ $(= 167 \times 81$, i.e., $P = 81)$ observation lines were correctly detected if I was given. Note that we can expect improvement of this accuracy by incorporating a priori knowledge of the character patterns and the cross ratio patterns into (2). Also note that even though several boundaries are missed, we can expect a correct cross ratio r_k, because multiple observation lines and a voting strategy are used as discussed in Section 2.2.

7. Conclusion and future work

For accurate camera-based character recognition, the embedment of class information into each character pattern is investigated. Specifically, a character pattern is printed with a horizontal stripe pattern, called a cross ratio pattern, representing the class of the character. Since cross ratio is a projective invariant, the same class information can be extracted from character images captured at an arbitrary camera angle. Experimental results showed that the cross ratio can provide class information accurately under heavy projective distortions. In addition, it was also experimentally shown that conventional shape similarities are helpful when we cannot prepare enough variations of the cross ratio pattern to distinguish all character classes.

We have a lot of future work. First of all, practical experiments should be conducted thoroughly using camera-captured character patterns. In addition, the design of the cross ratio patterns should be revised. Other distortion invariants are also to be examined.

Acknowledgment: This work is partially supported by Microsoft Public Trust for Intellectual Property Research Support Fund.

References

[1] D. Doermann, J. Liang and H. Li. "Progress in camera-based document image analysis", Proc. Int. Conf. Doc. Anal. Recog., pp. 606–616, 2003.

[2] The British Computer Society. Character Recognition 1967, Unwin Brothers Limited, 1966.

[3] D. L. Hecht. "Printed embedded data graphical user interfaces", IEEE Computer, vol. 34, no. 3, pp. 47–55, 2001.

[4] K. L. C. Moravec. "A grayscale reader for camera images of Xerox DataGlyphs", Proc. British Machine Vision Conference, pp. 698–707, 2002.

[5] M. Iwamura, S. Uchida, S. Omachi, and K. Kise. "Recognition with supplementary information —how many bits are lacking for 100% recognition?—", Proc. First Int. Workshop Camera-Based Doc. Anal. Recog. (CBDAR2005), pp. 68–75, 2005.

[6] S. Uchida and H. Sakoe. "A survey of elastic matching techniques for handwritten character recognition", IEICE Trans. Info. & Syst., vol. E88-D, no. 8, pp. 1781–1790, 2005.