2009 10th International Conference on Document Analysis and Recognition

Conspicuous Character Patterns

Seiichi Uchida
Kyushu Univ., Japan

Ryoji Hattori
Kyushu Univ., Japan

Masakazu Iwamura
Osaka Pref. Univ., Japan

Shinichiro Omachi
Tohoku Univ., Japan

Koichi Kise
Osaka Pref. Univ., Japan

Abstract

Detection of characters in scenery images is often a very difficult problem. Although many researchers have tackled this difficult problem and achieved a good performance, it is still difficult to suppress many false alarms andalthough missings. This paper investigates a conspicuous character pattern, which is a special pattern designed for easier detection. In order to have an example of the conspicuous character pattern, we select a character font with a larger distance from a non-character pattern distribution and, simultaneously, with a smaller distance from a character pattern distribution. Experimental results showed that the character font selected by this method is actually more conspicuous (i.e., detected more easily) than other fonts.

1. Introduction

Character detection, or character localization, in scenery images is a well-known difficult problem. Generally, past attempts at this problem (e.g., [1, 2, 3, 4]) have utilized several heuristics. For example, they used contrast, edge complexity, texture features, pattern regularity, etc. Detection performance, however, is still not satisfactory due to inevitable trade-off of the problem; if we want to detect all characters, we often have many false alarms around complex edge regions (such as tree regions). Conversely, if we want to suppress those false alarms, we may miss most characters. Actually, it will be very difficult to relax this trade-off in the strategy of the past attempts.

In this paper, we try to escape the trade-off by employing a new approach on this problem — the new approach is to use character patterns designed for easier detection. In other words, instead of trying to detect all character patterns, we try to detect some special character patterns which are designed to appeal their existence strongly to detectors.

This approach may seem strange from the viewpoint of the past attempts — it, however, is not so strange. In fact, we can find a similar approach in OCR/MICR fonts, which are specially designed to be recognized correctly by machines. In [5, 6], we can find updated versions of OCR fonts for camera-based character recognition. Two-dimensional bar codes, such as QR-code, are designed to be detected easily by cameras. Characters on signboards are often designed to be detected easily by humans !! (So, they are meant to catch one's eye.)

According to this approach, we can set our research purpose as the design of *conspicuous character patterns*. If the conspicuous character patterns are perfectly conspicuous in any scenery image, it is possible to detect all of them without any false alarm. Thus, we can escape from the above trade-off.

A naive example of conspicuous character patterns is red-colored characters; we can detect them in a given scenery image only by seeking red areas in the image. Another naive example is Times-Roman characters; we can detect them through a template matching-based search technique using original Times-Roman font templates. Although these naive examples are useful to understand the idea of the conspicuous character patterns, they are, clearly, not the best. It is hard to guarantee that there is no red area other than the characters. It is also hard to guarantee that Times-Roman is most distinguishable from non-character areas. We must take a more reasonable and systematic method for designing conspicuous character patterns.

These naive examples suggest an important fact that conspicuous character patterns should be designed together with the detection scheme. In other words, we must design the characters to be detected most easily for the detection scheme assumed. In fact, the red-colored characters are not conspicuous for the detector based on Times-Roman template matching. Similarly, the Times-Roman fonts are not conspicuous for the color-based detector.

Generally, each character detection scheme is defined by the following four factors and thus we must find the most conspicuous character pattern while considering those factors of the assumed detection scheme:

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.196

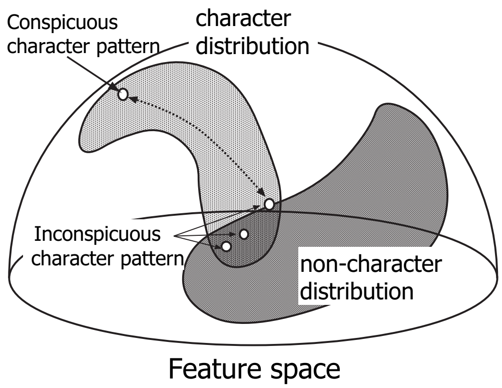

Figure 1. Character detection scheme assumed in this paper.

(i) How do we determine whether the target area is a character area or not?

(ii) How do we assume the general property of character patterns?

(iii) How do we assume the general property of non-character patterns?

(iv) What features do we use for representing patterns?

In the above example of the red-colored font, we will (i) determine a character area if the color of the area is red enough, (ii) assume that characters are red-colored, (iii) assume that non-characters are not red-colored, and (iv) use RGB color feature. It is clear that the most conspicuous character pattern for this detection scheme is to be designed as vividly red-colored characters. Every past attempt at character detection also can be described by these four factors and its conspicuous pattern can be defined from them.

In this paper, we will show an example of a reasonable and systematic design method of a conspicuous character pattern for a certain detection scheme. The designing method is rather general and possible to be modified for another detection scheme. The conspicuousness of the designed characters is verified through several experiments.

Note that we will not create a conspicuous character pattern from scratch. (That is, we will not synthesize it.) Instead, we will select the most conspicuous character pattern from various fonts. This is just because the selection is far simpler than the creation as the designing procedure. The creation will be considered as our important future work.

2. Assumed detection scheme

In this paper, we assume the following detection scheme as an example:

(i) The target area is determined as the character area when its distance to a character distribution is small and its distance to a non-character distribution is large.

(ii) The general property of characters is empirically defined by preparing a large set of font images. In other words, the property is defined as the actual distribution of font images.

(iii) The general property of non-characters is also empirically defined by preparing a large set of scenery image patches.

(iv) A normalized shape feature is used for representing patterns.

Figure 1 illustrates this detection scheme. Since any character or non-character pattern $\mathbf{p}$ is normalized so that $\|\mathbf{p}\| = 1$, the feature space becomes a hyper-sphere. Distributions of character patterns and non-character patterns are on the sphere surface. Then the target area is detected as a character area if it is far to the non-character distribution and near to the character distribution. The former condition is necessary to differentiate the character area from non-character patterns. The latter condition is introduced to guarantee human readability.

Note that this detection scheme is far more reasonable than the naive examples in Section 1. This is because the properties of characters and non-characters are defined empirically by using actual patterns. There is no strange assumption such that each character is red-colored.

3. Designing a conspicuous character pattern

3.1. Conspicuousness

Now, our task is to find the pattern which is the most easily detected by the assumed detection scheme of Section 2. Formally speaking, we must optimize the pattern $\mathbf{p}$ according to the criterion of maximizing its distance from the non-character distribution and, simultaneously, minimizing its distance from the character distribution.

According to the above criterion, we can define the *conspicuousness* of $\mathbf{p}$ by using its distances from the character and non-character distributions as follows. Let $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_i, \ldots, \boldsymbol{\phi}_I\}$ denote the principal I eigen-vectors of the covariance matrix of a distribution. Then the distance of a normalized pattern $\mathbf{p}$ from the distribution can be measured by using the I-dimensional subspace whose bases are $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_i, \ldots, \boldsymbol{\phi}_I\}$. Specifically, we can have the distance of $\mathbf{p}$ as the canonical angle $\theta_{\mathbf{p}}$ given by

$$\cos^2 \theta_{\mathbf{p}} = \sum_{i=1}^{I} \frac{\langle \mathbf{p}, \boldsymbol{\phi}_i \rangle^2}{\|\mathbf{p}\|^2 \|\boldsymbol{\phi}_i\|^2} = \sum_{i=1}^{I} \langle \mathbf{p}, \boldsymbol{\phi}_i \rangle^2. \qquad (1)$$

Using the canonical angle, the conspicuousness of $\mathbf{p}$ is defined as

$$C_{\mathbf{p}} = \begin{cases} \theta_{\mathbf{p}}^N / (\theta_{\mathbf{p}}^N + \theta_{\mathbf{p}}^C) & \text{if } \theta_{\mathbf{p}}^N \geq \epsilon, \\ 0 & \text{otherwise,} \end{cases} \qquad (2)$$

where $\theta^N_{\mathbf{p}}$ is the canonical angle between $\mathbf{p}$ and the non-character subspace and $\theta^C_{\mathbf{p}}$ is the canonical angle between $\mathbf{p}$ and the character subspace. Both canonical angles are derived by (1). The value $\theta^N_{\mathbf{p}}/(\theta^N_{\mathbf{p}} + \theta^C_{\mathbf{p}})$ becomes large if $\theta^N_{\mathbf{p}}$ becomes large and/or $\theta^C_{\mathbf{p}}$ becomes small. The maximum and the minimum values of $C_{\mathbf{p}}$ are 1 (the most conspicuous) and 0 (the least conspicuous), respectively. The threshold ϵ (> 0) is introduced so that a small $\theta^N_{\mathbf{p}}$ does not give a large $C_{\mathbf{p}}$ with a very small $\theta^C_{\mathbf{p}}$.

The subspaces of the character and non-character distributions are determined by using empirical covariance matrices of many font images and many scenery image patches, respectively. Note that the dimensions I of the two subspaces can be different. Also note that it is possible to use any image feature for representing those images, while we will use a shape feature, called direction code histogram [7], instead of the bitmap feature. The aim of using the shape feature is to minimize the effect of color and brightness. The detail of the shape feature will be given in Section 4.2.

3.2. Selection of most conspicuous character pattern

As noted in Section 1, we will not synthesize a conspicuous character pattern which globally maximizes $C_{\mathbf{p}}$. Instead, we will select the most conspicuous character pattern among various fonts. Specifically, we calculate the conspicuousness $C_{\mathbf{p}}$ for every font image $\mathbf{p}$, and then select the font image $\mathbf{p}$ with the largest $C_{\mathbf{p}}$ as the most conspicuous character pattern.

4. Experimental result

Using the above method with font images and scenery image patches, the selection and the evaluation of the most conspicuous character pattern were done experimentally. The following is the detail of this experiment.

4.1. Experimental setup

About 8,000 English capital letter images of 308 fonts were collected (26×308∼ 8,000). Those fonts were selected using their PANOSE [8]. PANOSE is comprised of 10 elements and the first element represents the font family, such as “Latin Text”, “Latin Handwritten”, “Latin Decorative”, etc. The selected 308 fonts were of “Latin Text”, which is the font family suitable for general documents[1].

As non-character images, about 80,000 image patches (of 64×64 pixel size) were collected from scenery images of the CalTech’s background database[9], which is available as a public image database.

[1]The authors found many “decorative” fonts with “Latin Text” PANOSE. The authors, however, did not remove them, because there is no objective criterion other than PANOSE for separating Latin Text fonts from decorative fonts.

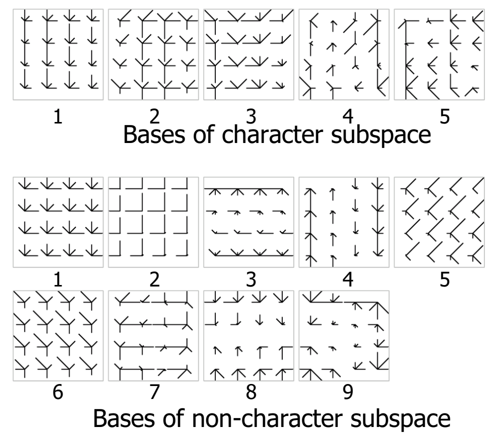

Figure 2. Bases of character distribution and non-character distribution.

4.2. Feature space

Each image was represented as the direction code histogram [7]. First, each 64×64 image is converted into an edge map and divided into 16 blocks of 16×16 pixel size. The edge pixels in each block were quantized into four directions and then the number of edge pixels of each direction was counted within a block. Thus the dimension of the direction code histogram is 64 = 16(blocks)× 4(quantized directions). Since this feature is totally based on the edge direction, it is affected by neither color nor intensity.

Then a 64×64 covariance matrix was prepared for each of the character set and the non-character set and its I principle eigen-vectors were calculated as $\{\phi_1, \ldots, \phi_i, \ldots, \phi_I\}$. By observing the cumulative ratio, the number I was fixed at 5 for characters and 9 for non-characters. Figure 2 visualizes those eigen-vectors (i.e., the bases of the subspace) of characters and non-characters, respectively. Each eigen-vector is visualized by 64 short line segments. Each line segment represents the direction and the value of an element of the vector [2].

4.3. Selected patterns

According to the procedure in Sections 3.1 and 3.2, the most conspicuous character pattern maximizing $C_{\mathbf{p}}$ was selected for each of 26 letter categories (“A” to “Z”). During the selection, the threshold ϵ was fixed at 0.4, which was determined experimentally.

Figure 3 (a) shows the most conspicuous character pattern of all categories. Round-shaped characters were selected as the conspicuous character patterns. This is because the direction code histogram of a round-shaped character has almost zero-value elements around the center area and

[2]In Fig. 2, each element is visualized with its polarity; for example, if an element of 45° has positive (negative) value, it will be visualized as a 45° ($225^\circ = 45^\circ + 180^\circ$) line segment.

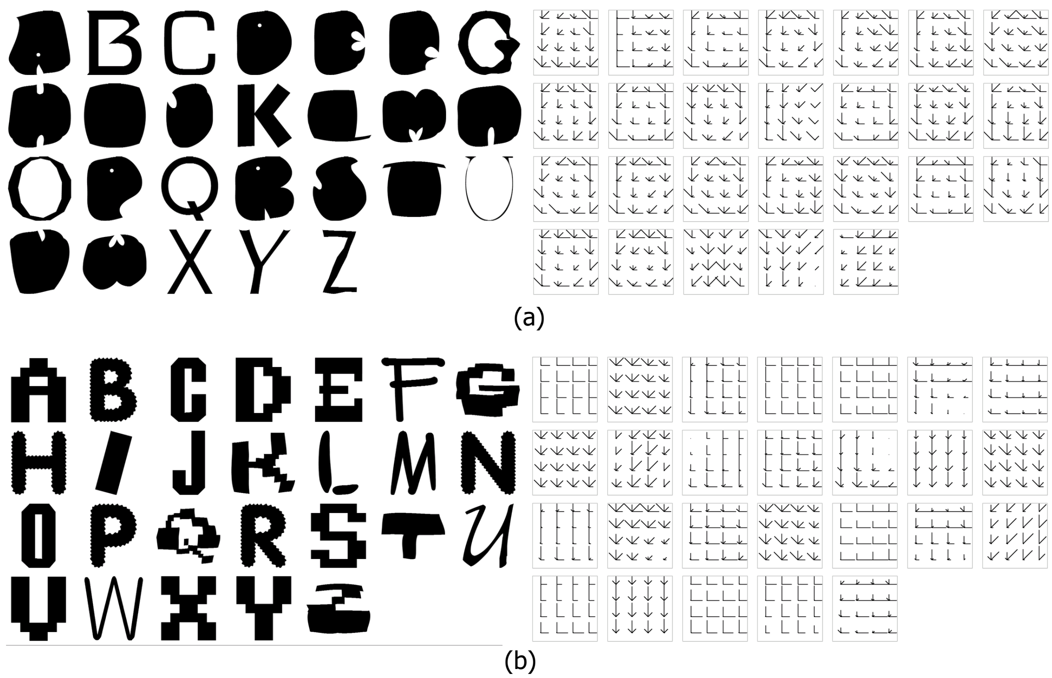

Figure 3. (a) Example of the most conspicuous pattern and (b) the least conspicuous pattern. Left: original font image. Right: direction code histogram.

thus is different from, especially, the bases ϕ_1 and ϕ_2 of non-character subspace. Consequently, $\theta_{\mathbf{p}}^N$ becomes large and thus $C_{\mathbf{p}}$ becomes large for the round-shaped characters.

It will be important to note that the conspicuous pattern is not limited to the pattern of Fig. 3 (a). Actually, the pattern of second-place was a similar round-shaped font. Therefore, we can use various round-shaped fonts as conspicuous patterns.

Figure 3 (b) shows the least conspicuous character pattern, which has the minimum $C_{\mathbf{p}}$ value by a small $\theta_{\mathbf{p}}^N$. Those characters often have a direction code histogram similar to one of the basis of non-character subspace. Thus $\cos^2 \theta_{\mathbf{p}}^N$ becomes large and, equivalently, $\theta_{\mathbf{p}}^N$ becomes small.

4.4. Detection experiment

Actual easiness of detecting the selected conspicuous patters was examined. As shown in the leftmost figures in Fig. 4, the selected most conspicuous patterns were printed on a paper sheet and put in scene. Similarly, the selected least conspicuous patterns were also printed and put in the scene. The detection was performed by exhaustive search; the conspicuousness $C_{\mathbf{w}}$ was evaluated at all possible square areas, $\{\mathbf{w}\}$. Note that the size of square areas was also varied during the search. If we have a large $C_{\mathbf{w}}$ on a square area $\mathbf{w}$, we can consider that a conspicuous character pattern exists within $\mathbf{w}$.

Figure 4 (a) verifies that the most conspicuous patterns are really conspicuous; almost all the conspicuous patterns were detected successfully. (Only "B" was not detected here.) In addition, there are very few false alarms. This promising result indicates the importance of designing conspicuous patterns for the detection problem. Another important fact is that none of the least conspicuous patterns was detected. This fact is also promising because it proves the conspicuousness is very stable even in real scene.

Figures 4 (b) and (c) still verify the usefulness of the conspicuous patterns in very complex backgrounds. In (b), the conspicuous patterns were successfully detected regardless of the sizes. In (c), the conspicuous pattern were printed in different colors. Since the assumed feature (the direction code histogram) is affected by neither color nor intensity, the colored conspicuous patterns were detected without obvious degradation. It is interesting to note that most characters in the backgrounds were not detected successfully. This also indicates that the characters designed to maximize their conspicuousness are very privileged patterns in the detection problem.

5. Conclusion

Like OCR/MICR fonts for easier character recognition, *conspicuous* character patterns were newly introduced for easier character detection in scenery images. This paper has discussed the approach of the conspicuous character pattern and detailed a systematic methodology of selecting the most

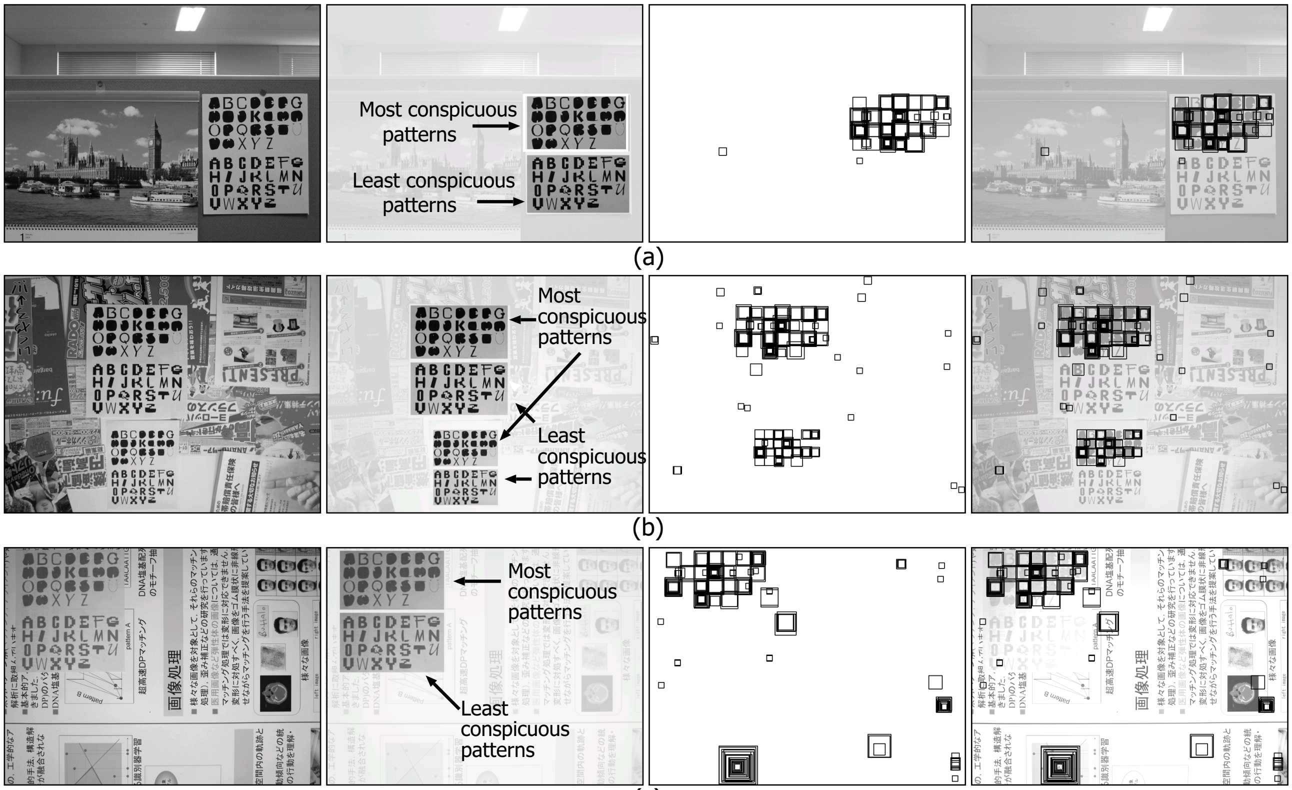

Figure 4. Character detection results on three different scenery images. From left to right: original scenery image with the most conspicuous and the least conspicuous character patterns (of Fig. 3), the locations of those patterns, detected areas, and the detected areas superimposed on the scenery image.

conspicuous character font among various candidate fonts for a certain detection scheme. The actual conspicuousness of the selected pattern was confirmed experimentally; their detection was very easier than the detection of other characters.

This paper is the first attempt at developing conspicuous character patterns and therefore there are many future works. In this paper, we first assume a detection scheme and then design the conspicuous pattern for the scheme. In future, simultaneous design of the detection scheme and the conspicuous pattern should be investigated. Another future work is the systematic creation (synthesis) of the most conspicuous pattern instead of selection. For this future work, some parametric font-shape models will be necessary.

References

[1] J. Keechul, K. I. Kwang, and A. K. Jain, “Text information extraction in images and video: a survey, Pattern Recognition, vol. 37, no. 5, pp. 977-997, 2004.

[2] R. W. Lienhart and F. Stuber, “Automatic text recognition in digital videos,” Proceedings of SPIE, vol. 2666 Image and Video Processing IV, pp. 180-188, 1996.

[3] X. Chen and A. L. Yuille, “Detecting and reading text in natural scenes,” CVPR 2004, vol. 2, pp. 366-373, 2004.

[4] V. Wu, R. Manmatha, and E. M. Riseman, “An automatic system to detect and recognize text in images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 11, pp. 1224-1229, 1999.

[5] S. Uchida, M. Sakai, M. Iwamura, S. Omachi, K. Kise “Extraction of embedded class information from universal character pattern,” Proc. ICDAR, vol. 1 of 2, pp.437-441, 2007.

[6] S. Omachi, M. Iwamura, S. Uchida, and K. Kise, “Affine invariant information embedment for accurate camera-based character recognition,” Proc. ICPR, vol. 2 of 4, pp. 1098-1101, 2006.

[7] F. Kimura, T. Wakabayashi, S. Tsuruoka, and Y. Miyake, “Improvement of handwritten Japanese character recognition using weighted direction code histogram,” Pattern Recognition, vol. 30, no. 8, pp. 1329-1337, 1997.

[8] `http://www.w3.org/Printing/stevahn.html`

[9] `http://www.vision.caltech.edu/html-files/archive.html`