

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Figure-Ground Discrimination and Distortion-Tolerant Recognition of Color Characters in Scene Images

Toru Wakahara

*Faculty of Computer and Information Sciences, Hosei University
3-7-2 Kajino-cho Koganei-shi
Tokyo 184-8584 JAPAN*

1. Introduction

Recently, recognition of web documents and characters in natural scenes has emerged as a hot, demanding research field (Doermann et al., 2003). In particular, recognition of characters in scene images with a wide variety of image degradations and complex backgrounds poses the following two key problems.

The first problem is figure-ground discrimination (Herault & Horaud, 1993) or correct binarization of color characters in scene images as a crucial step to the success of subsequent recognition. Most of the binarization methods are based on global, local/adaptive or multi-stage selection of threshold (Trier & Jain, 1995; Wolf et al., 2002; Wu & Amin, 2003). However, color-based binarization has not yet been fully addressed (Miene et al., 2001).

The second problem is distortion-tolerant character recognition under the condition of a small sample size because there is only a limited quantity of data against a wide variety of fonts and image degradations. Hence, we cannot make good use of statistical pattern recognition techniques, including sophisticated discriminant functions, neural networks, support vector machines or kernel methods.

Regarding the first problem we propose three promising approaches. The first approach is application of genetic algorithms (GA) to a combinatorial problem of determining an optimal filter sequence that correctly binarizes an input image (Kohmura & Wakahara, 2006). The filter bank contains a number of typical image processing filters as applied to one of the RGB color planes and logical/arithmetic operations between two color planes. The second approach is selection of a maximum separability axis in the RGB color space and an appropriate threshold on the axis for binarizing an input image as the two-category classification problem (Yokobayashi & Wakahara, 2006). Here, the key idea for solving this problem is application of the Otsu's criterion (Otsu, 1979) to the distribution of color pixels of the input image projected onto every possible axis in the RGB color space. The third approach is application of K-means clustering in the HSI color space to color pixels of the input image, generation of temporally binarized images via every dichotomization of K clusters, and their classification into two categories: character and non-character (Kato &

Wakahara, 2009). Here, a character vs. non-character classification is effectively implemented by support vector machines (SVM).

Regarding the second problem we try to make use of elastic image matching techniques (Uchida & Sakoe, 2005). Here, we apply two kinds of distortion-tolerant template matching based on the deterministic character deformation models. The first one is our global affine transformation (GAT) correlation technique (Wakahara et al., 2001). The GAT correlation absorbs distortion expressible by affine transformation by determining optimal affine parameters that maximize a normalized cross-correlation value between an affine-transformed input image and a template. In particular, image matching by means of normalized cross-correlation was shown to be robust against image blurring and additive random noise (Sato, 2000). The second one is the well-known tangent distance (TD) (Simard et al., 1993). The tangent distance absorbs distortion expressible by a linear combination of predefined geometric and topographical transformations as applied to both an input image and each template.

We show experimental results made on the public ICDAR 2003 robust OCR dataset (ICDAR Datasets, 2003) containing a wide variety of single-character images in natural scenes.

In Section 2, we explain ICDAR 2003 robust OCR dataset. Section 3 proposes three kinds of techniques for figure-ground discrimination or correct binarization of color characters in scene images. In Section 4, we describe two competing techniques of distortion-tolerant image matching for recognizing binarized characters. Section 5 shows experimental results. Section 6 is devoted to discussion and future work.

2. ICDAR 2003 robust OCR dataset

Several datasets used in ICDAR 2003 robust reading competitions (Lucas et al., 2003) are available for download from the website (ICDAR Datasets, 2003). We use the robust OCR dataset containing JPEG single-character images in natural scenes. In particular, we select a total of 698 images from “Sample” subset.

Figure 1 shows examples of images with a variety of image degradations and complex backgrounds.

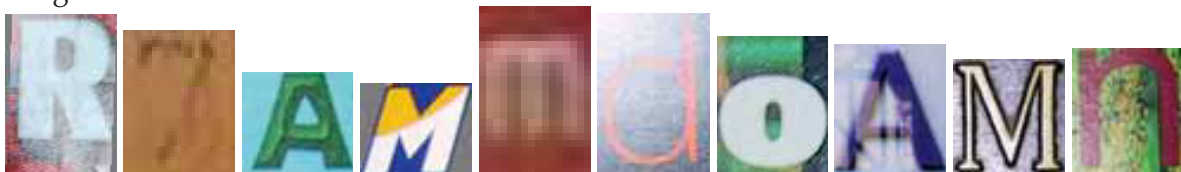


Fig. 1. Examples of images used in our experiments.

3. Figure-ground discrimination of color characters in scene images

In this section, we propose three kinds of techniques for figure-ground discrimination or correct binarization of color characters: determination of an optimal sequence of filters for binarization using GA, binarization using a maximum separability axis in a color space, and K-means clustering in a color space and figure-ground discrimination by SVM.

3.1 Determination of an optimal sequence of filters for binarization using GA

This technique is for binarization of color characters in scene images using genetic algorithms (GA) to search for an optimal sequence of filters through a filter bank. The filter bank contains simple image processing filters as applied to one of the RGB color planes and logical/arithmetic operations between two color planes. First, we classify images extracted from the ICDAR 2003 robust OCR dataset into several groups according to degradation categories. Then, in the training stage, by selecting training samples from each degradation category we apply GA to the combinatorial optimization problem of determining a filter sequence that maximizes the average fitness value calculated between the filtered training samples and their respective target images ideally binarized by humans. Finally, in the testing stage, we apply the optimal filter sequence to binarization of remaining test samples.

3.1.1 Grouping of character images according to degradation categories

By carefully examining a total of 698 images from “Sample” subset we classified them into six groups according to degradation categories: clear, background with pattern, character with pattern, character with rims, blurring, and nonuniform lighting. The criterion upon how to classify degradation categories is rather subjective just to show a wide variety of binarization problems. In practical application degradation categories should be selected automatically, and, also, it is necessary to automatically decide which degradation category a given input image belongs to.

Figure 2 shows examples of images in six degradation categories.

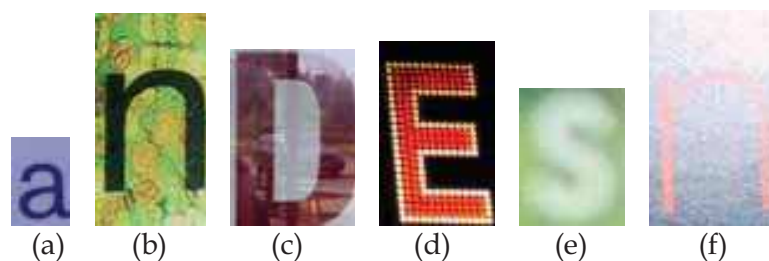


Fig. 2. Examples of images in six degradation categories. (a) Clear. (b) Background with pattern. (c) Character with pattern. (d) Character with rims. (e) Blurring. (f) Nonuniform lighting.

3.1.2 Image transformation by a sequence of filters and filter bank

Figure 3 shows a total flow of image transformation using a sequence of filters as applied to an original image so that a filtered image approximates its target image ideally binarized by humans as closely as possible.

We use GA in search of an optimal sequence of filters, equivalent to the image transformation L^* , while L specifies the ideal binarization. The degree of approximation of L^* to L is evaluated in terms of the fitness value calculated between target and filtered images.

Table 1 shows a list of filters in our filter bank. These filters are not sophisticated but rather primitive ones (Gonzalez & Woods, 2000).

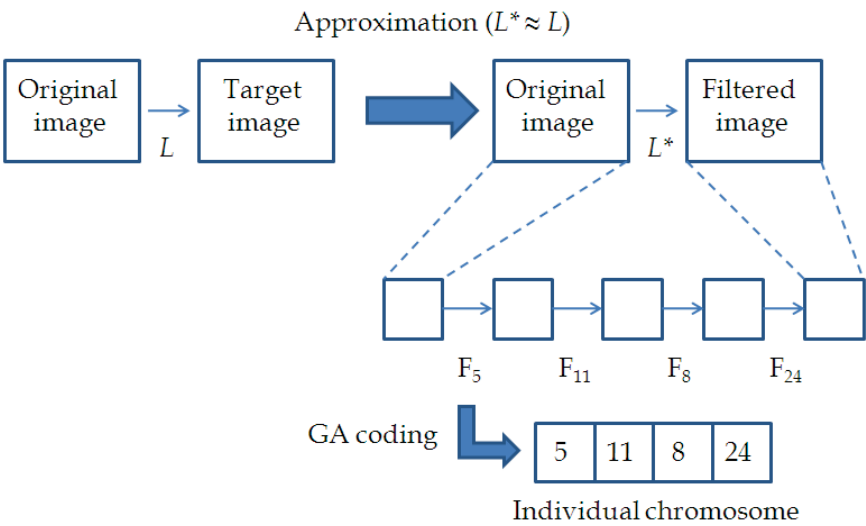


Fig. 3. Total flow of image transformation by a sequence of filters.

No.	Filter name	Function
1	Mean	local mean in a 3×3 window
2	Min	local min in a 3×3 window
3	Max	local max in a 3×3 window
4	Sobel (1)	horizontal differential
5	Sobel (2)	vertical differential
6	Sobel (3)	the norm of differential
7	LightEdge	Laplacian
8	DarkEdge	Laplacian + 255
9	Erosion	morphological erosion
10	Dilation	morphological dilation
11	Inversion	255 - g; g = pixel value
12	Logical sum	max of two color planes
13	Logical product	min of two color planes
14	Algebraic sum	sum of two color planes - their product / 255
15	Algebraic product	product of two color planes / 255
16	Bounded sum	g = sum of two color planes; if g > 255, g = 255
17	Bounded product	product of two color planes - 255; if g < 0, g = 0

Table 1. List of filters in a filter bank.

Each filtering operation is specified in either of the following two ways. One way is to select one of one-operand filters from no. 1 to no. 11 and one of the RGB color planes to which the selected filter is applied. The other way is to select one of two-operand filters from no. 12 to no. 17 and two of the RGB color planes to which the selected filter is applied, where the filtering result is overwritten onto either of the two color planes. Hence, the total number of filtering operations is 3 × 11 plus 6 × 6, and equals sixty nine.

3.1.3 Gene encoding and specifications of GA

We use GA (Goldberg, 1989) to search for an optimal filter sequence that transforms an original image so as to yield the maximum fitness value against its target image ideally binarized by humans. Here, the fitness value serves as a similarity measure.

As described in Section 3.1.2, the total number of filtering operations equals sixty nine. We specify each filtering operation by an ID number selected from one to sixty nine. Hence, a filter sequence or a chromosome is encoded as a string of 8-bit integers. Also, we set the maximum number of constituent filters in a chromosome at 80.

The initial population of 300 is randomly generated. We adopt the roulette selection rule based on the fitness values in each generation. We use the modified one-point crossover method that exchanges respective tails with the rate of 80%. Mutation also exchanges every constituent ID number within a chromosome for a different one with the rate of 0.1%.

Finally, we stop the GA process when the maximal fitness value of an elite chromosome exceeds the threshold value of 0.9 or when the number of generations arrives at the predetermined number of 800.

Here, by denoting target and filtered images by $T = \{ T_k(x, y) \}$ and $F = \{ F_k(x, y) \}$ ($k = R, G, B$), respectively, we calculate a fitness value, $f(T, F)$, between target and filtered images by

$$f(T, F) = 1 - \frac{\sum_{k=R,G,B} \sum_{x=1}^{W_x} \sum_{y=1}^{W_y} |T_k(x, y) - F_k(x, y)|}{3 \times W_x \times W_y \times 255}, \quad (1)$$

where W_x and W_y specify width and height of the image, respectively.

Figure 4 shows examples of binarization of training samples belonging to the degradation category “nonuniform lighting” using an optimal sequence of filters determined via GA.



Fig.4. Examples of binarization of training samples belonging to the degradation category “nonuniform lighting” using an optimal sequence of filters determined via GA. (a) Input images. (b) Binarized images.

It is to be noted that this technique provides us with an optimal sequence of filters for binarization of color characters in each of predetermined degradation categories. In other words, this technique cannot generate a single, all-purpose filter sequence to deal with a wide variety of image degradations and complex backgrounds. In this sense, we can say that this approach is very powerful when we know in advance that all of input images being considered belong to a particular kind of degradation category.

3.2 Binarization using a maximum separability axis in a color space

This technique is for binarization of color characters in scene images following two steps. The first step is temporary binarization by selecting one optimal projection axis with a maximum two-class separability in the RGB color space and an appropriate threshold on the

axis. Here, we apply Otsu's criterion to a two-class classification problem. The second step is figure-ground determination based on the figure-to-ground ratio on the image periphery and common characteristics that a character pattern should have.

3.2.1 Temporary binarization via Otsu's criterion in the RGB color space

First, color points of all pixels in an input image are projected onto an arbitrarily chosen axis in the RGB color space. Here, we adopt spherical polar coordinates, (r, θ, φ) , in 3D color space, and try all axes with angles, (θ, φ) , selected at intervals of one degree, respectively. Namely, a total of 180×180 axes in 3D color space are considered.

Second, for each point distribution on a chosen axis we calculate maximum between-class separability by setting an optimal threshold according to the Otsu's binarization technique (Otsu, 1979). We know that this idea is also based on the well-known Fisher criterion (Bishop, 2006) as applied to a two-class classification problem. Namely, the between-class separability, S , is defined as the difference of two means normalized by the averaged variance on the chosen axis according to

$$S = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2} \rightarrow \max \text{ for a bisection on the axis.} \quad (2)$$

Finally, we select the axis that gives the largest between-class separability and the corresponding threshold for temporary binarization of the input image. Here, from the viewpoint of figure-ground discrimination it is clear that this binarization result is only temporary because there are two possibilities of either class being a character.

Figure 5 shows projection of pixels onto a chosen axis in the RGB color space.

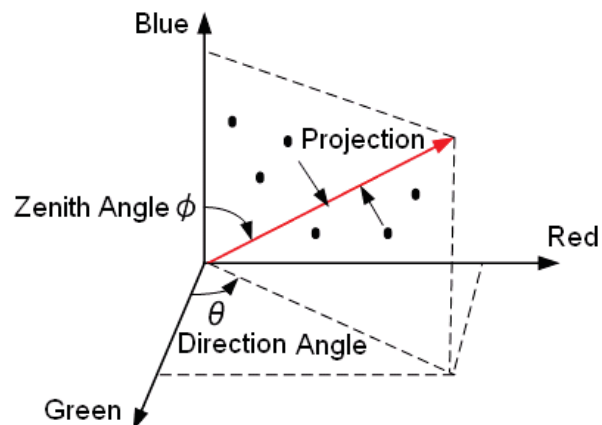


Fig. 5. Projection of pixels onto a chosen axis in the RGB color space.

3.2.2 Figure-ground determination using common characteristics of characters

We assume that an input image contains only one character and a character belongs to alphanumeric characters as shown in Fig. 1.

Granting this assumption, we can enumerate common characteristics that such single-character images should have as follows.

- (1) The majority of pixels on the image periphery belong not to a character but to a background.
- (2) The number of connected components in a character is one except "i" and "j."
- (3) The width of a character is narrower than that of a background.

Based on these common characteristics we propose a procedure for figure-ground determination written in a pseudo-code as shown below.

If the figure-to-ground ratio on the image periphery is less than a threshold value of Th , then consider the present binarized image as the correct one and goto END.

Else if the figure-to-ground ratio on the image periphery is more than the inverse of Th , then consider the reversed image as the correctly binarized one and goto END.

Else if the width of a figure is narrower than that of a ground, then consider the present binarized image as the correct one and goto END. Here, we define the width of a figure or a ground in the image as twice the number of erosion operations (Gonzalez & Woods, 2000) applied to the corresponding region until it vanishes.

Else consider the reversed image as the correctly binarized one and goto END.

END: Select and save only the maximum connected component of the figure and output the resultant image as the final result of figure-ground discrimination.

Figure 6 shows examples of binarization using a maximum separability axis in a color space.



Fig. 6. Examples of binarization using a maximum separability axis in a color space. (a) Input images. (b) Binarized images.

It is to be noted that this technique assumes that a character in an input image is made up of color pixels with similar values in the RGB color space and, hence, binarization is handled correctly as a two-class classification problem using only color information. Therefore, this technique is not well suited to deal with multi-color characters and/or characters with nonuniform backgrounds.

3.3 K-means clustering in a color space and figure-ground discrimination by SVM

This technique is for binarization of color characters in scene images following three steps. The first step applies K-means clustering in the HSI color space to points in an input image, and, then, generates a set of tentatively binarized images by every possible dichotomization of a total of K clusters or subimages. The second step calculates the degree of character-likeness of each tentatively binarized image by SVM in an appropriately chosen feature space. In advance, SVM is trained to determine whether and to what degree each binarized image represents a character or non-character. The third step outputs the binarized image with the maximum degree of character-likeness as an optimal binarization result.

3.3.1 K-means clustering in the HSI color space

First, values of R , G , and B in the RGB color space are converted to values of H , S , and I in the HSI color space, where H , S , and I represent hue, saturation, and intensity, respectively (Gonzalez & Woods, 2000). In particular, we scale each value of H , S , and I to range from 0 to 255 as follows.

$$\begin{aligned}
 I &= \max(R, G, B), \quad m = \min(R, G, B), \\
 \text{if } I = 0 \text{ or } I = m &\text{ then } S = 0, \quad H = \text{indefinite}, \\
 \text{else } S &= \frac{I - m}{I} \times 255, \\
 r &= \frac{I - R}{I - m}, \quad g = \frac{I - G}{I - m}, \quad b = \frac{I - B}{I - m}, \\
 \text{if } R = I &\text{ then } h = \frac{\pi}{3}(b - g), \quad \text{if } G = I &\text{ then } h = \frac{\pi}{3}(2 + r - b), \\
 \text{if } B = I &\text{ then } h = \frac{\pi}{3}(4 + g - r), \quad \text{if } h < 0 &\text{ then } h = h + 2\pi, \\
 H &= h \times 255.
 \end{aligned} \tag{3}$$

When an input image of size $W_x \times W_y$ is given, a total of $W_x \times W_y$ points corresponding to those pixels are scattered in the HSI color space.

Second, K-means clustering is applied to a total of $W_x \times W_y$ points in the HSI color space to generate K clusters, where a number of clusters, K , is determined in advance. The K-means clustering algorithm or nearest mean reclassification algorithm (Bishop, 2006) is as follows.

Step 1: Select K points at random from a total of $W_x \times W_y$ points scattered in the HSI color space as initial cluster centers, $\{\mu_k^{(\tau=0)}\}$, ($k = 1, \dots, K$). τ specifies an iteration number.

Then, assign each of $W_x \times W_y$ points to its nearest cluster center among $\{\mu_k^{(\tau=0)}\}$, ($k = 1, \dots, K$), and a set of points assigned to the same cluster center forms one cluster.

Step 2: Compute a mean vector of each cluster and set the mean vector as an update on its cluster center. Then, $\tau = \tau + 1$, and cluster centers thus updated are denoted by $\{\mu_k^{(\tau)}\}$, ($k = 1, \dots, K$).

Step 3: Each point is re-assigned to a new set according to which is the nearest cluster center among $\{\mu_k^{(\tau)}\}$, ($k = 1, \dots, K$), and each new set of points corresponds to a cluster.

If there is no further change in the grouping of the data points, output the present K clusters as the clustering result and stop. Otherwise, go to Step 2.

By inverse projection of a set of points forming each cluster in the HSI color space onto a 2D image plane, respectively, we obtain a total of K subimages the sum of which is equivalent to the input image.

3.3.2 Generation of tentatively binarized images by dichotomization of K subimages

We dichotomize K subimages into two groups, and set values of pixels belonging to the one group at 0 (black) and the other group at 255 (white). As a result, we obtain one binarized image, where black pixels represent figure and white pixels represent background.

By considering every possible dichotomization of K subimages we can generate multiple tentatively binarized images the total number of which, N_{binary} , is given by

$$N_{binary} = \sum_{i=1}^{K-1} {}_K C_i = 2^K - 2, \quad (4)$$

where ${}_K C_i$ denotes a binomial coefficient.

Figure 7 shows one example of generation of tentatively binarized images from an input image.

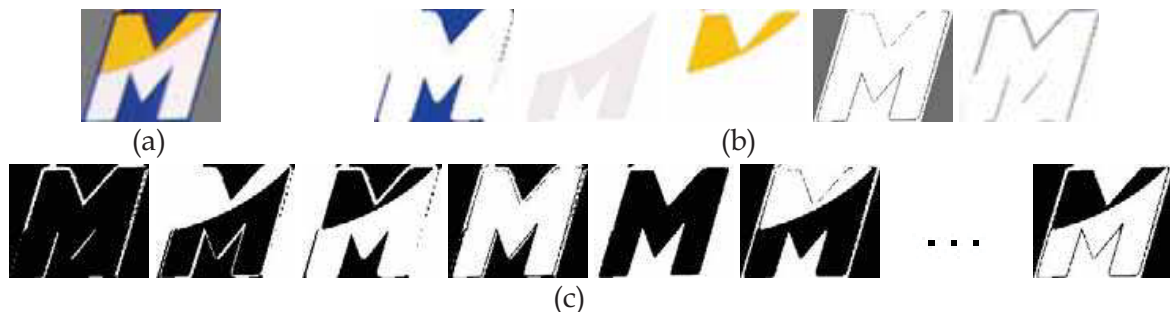


Fig. 7. One example of generation of tentatively binarized images from an input image. (a) An input image. (b) K subimages obtained by K-means clustering ($K = 5$). (c) $(2^K - 2)$ tentatively binarized images.

From Fig.7, it is seen that a correctly binarized image is included in a set of tentatively binarized images even when a character is represented by multiple colors in the original input image.

It is to be noted that this technique has the possibility of correctly binarizing multi-color characters and/or characters with complex backgrounds. However, it is necessary to devise a means of selecting a correctly binarized image from a set of tentatively binarized images. Also, the total number of clusters, K , should be large enough to just guarantee that a correctly binarized image will be included in a set of $(2^K - 2)$ tentatively binarized images.

3.3.3 Feature extraction from a binary image for estimating character-likeness

We extract a feature vector from a binary image so that a feature vector should represent a kind of character-likeness as much as possible. Selection of a good feature vector is a clue to the success of SVM that determines whether and to what degree each binary image represents a character or non-character in the given feature space.

As preprocessing, position and size normalization is applied to each binary image by using moments (Casey, 1970). Namely, the center of gravity of black pixels is shifted to the center of the image, and the second moment around the center of gravity is set at the predetermined value. Here, we set a size of a preprocessed binary image at 80×120 pixels.

Then, we extract three kinds of feature vectors all of which are well-known in the field of character recognition: mesh feature, direction code histogram feature, and weighted direction code histogram feature.

Mesh feature:

We divide the input binary image into a total number of $8 \times 12 (= 96)$ square blocks each of which has a size of 10×10 pixels and, then, calculate the percentage of black pixels in each of blocks. Finally, those measurements together form the 96-dimensional mesh feature vector.

Direction code histogram feature:

One of 4-directional codes, i.e., H (horizontal), R (right-diagonal), V (vertical), and L (left-diagonal), is assigned to every contour pixel of black regions. Then, we divide the input binary image into a total number of $4 \times 6 (= 24)$ square blocks each of which has a size of 20×20 pixels. Finally, in each block we count the number of contour pixels assigned to H , R , V ,

and L , respectively, and their measurements together form the 96-dimensional direction code histogram feature vector.

Weighted direction code histogram feature:

In order to improve robustness against shape distortion we introduce a locally weighted sum of the direction code histogram feature (Kimura et al., 1997). First, we divide the input binary image into a total number of $8 \times 12 (= 96)$ square blocks each of which has a size of 10×10 pixels. Hence, we obtain the 384-dimensional direction code histogram feature vector. Then, using a locally weighted sum around each block taken at intervals of two blocks, both horizontally and vertically, the dimension of the feature vector is reduced from 384 to 96. As a result, we obtain the 96-dimensional weighted direction code histogram feature vector.

Figure 8 shows a Gaussian mask for generating the weighted direction code histogram feature.

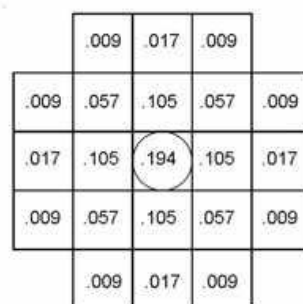


Fig. 8. A Gaussian mask for generating the weighted direction code histogram feature. A circle denotes the loci of a block around which a locally weighted sum is calculated.

3.3.4 Discrimination between character and non-character via SVM

The support vector machines, SVM, (Vapnik, 2000) map the input feature vectors, x , into a high-dimensional feature space, $\Phi(x)$, through some nonlinear mapping, chosen *a priori*. In this space, an optimal separating hyperplane that maximizes the margin is constructed.

The training data set comprises N input feature vectors $x_1, \dots, x_N$, with corresponding target values $y_1, \dots, y_N$ where $y_i \in \{-1, +1\}$, and new data points x are classified according to the sign of $f(x)$ given by

$$\begin{aligned}
 f(x) &= \sum_{i=1}^N \alpha_i y_i (\Phi(x_i) \cdot \Phi(x)) - b \\
 &= \sum_{i=1}^N \alpha_i y_i K(x_i, x) - b,
 \end{aligned}
 \tag{5}$$

where $(\Phi(x) \cdot \Phi(y))$ is an inner product in the high-dimensional feature space, and is replaced with the kernel function $K(x, y)$ by making use of the kernel trick.

Non negative coefficients $\{\alpha_i\}$ that maximize the margin are determined by solving a convex quadratic programming problem. The data points $\{x_k\}$ for which coefficients $\{\alpha_k\}$ are nonzero are called support vectors because they correspond to points that lie on the maximum margin hyperplanes in the high-dimensional feature space.

We implemented SVM via SVM^{light} (Joachims, 1998), and made use of the following three kinds of the kernel functions: linear, polynomial, and radial basis functions.

$$\begin{aligned} K_1(\mathbf{x}, \mathbf{y}) &= (\mathbf{x} \cdot \mathbf{y}), \quad K_2(\mathbf{x}, \mathbf{y}) = (s \times (\mathbf{x} \cdot \mathbf{y}) + c)^d, \\ K_3(\mathbf{x}, \mathbf{y}) &= \exp \left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2} \right), \end{aligned} \quad (6)$$

where parameter values are set at default ones: $s = c = 1.0$, $d = 3$ and $2\sigma^2 = 1.0$.

Training data were prepared for the training phase of SVM to discriminate between two classes of character and non-character as follows.

Training data for the *character* class:

First, we selected correctly binarized images from a total of $(2^K - 2)$ tentatively binarized images obtained for each of training samples. Secondly, we added a total of 136 available font sets to the training data.

Training data for the *non-character* class:

We selected incorrectly binarized images from a total of $(2^K - 2)$ tentatively binarized images obtained for each of training samples.

Figure 9 shows examples of training data for the character and non-character classes.

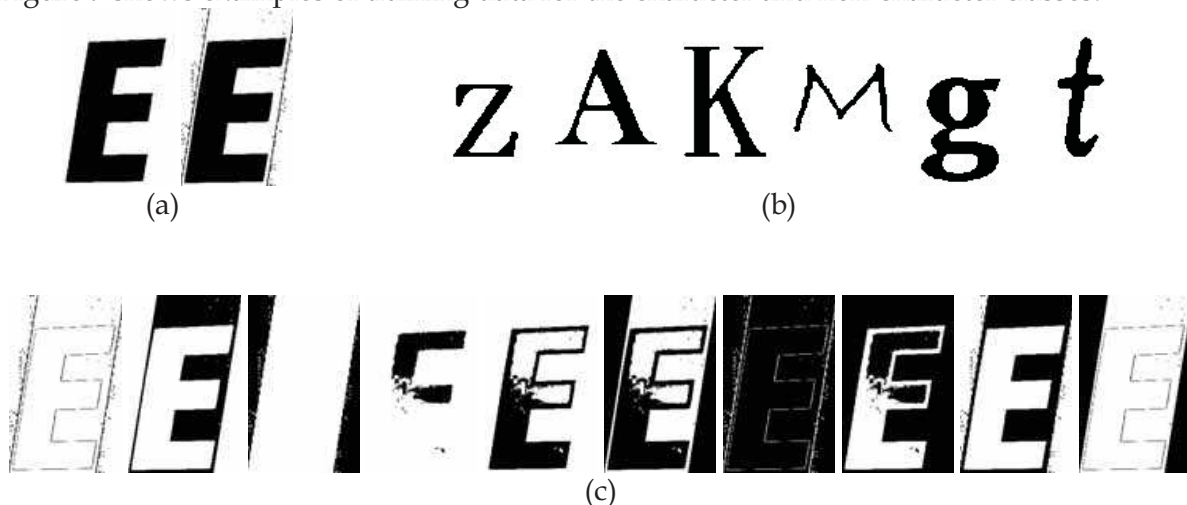


Fig. 9. Examples of training data. (a) Character class (correctly binarized images). (b) Character class (available fonts). (c) Non-character class (incorrectly binarized images).

3.3.5 Selection of correctly binarized image via SVM

For a given color character image a total of $(2^K - 2)$ tentatively binarized images are generated by K-means clustering. Then, feature vectors are extracted from each tentatively binarized image. Those feature vectors are fed into the trained SVM. SVM outputs the values of $f(\mathbf{x})$ of Eq. (5), where positive and negative values of $f(\mathbf{x})$ indicate character and non-character classes, respectively.

Here, we regard the value of $f(x)$ as estimating the degree of character-likeness, and, also, assume that the larger the value of $f(x)$ is the more its character-likeness is.

Then, we select a single tentatively binarized image with the maximum value of $f(x)$ among those of $(2^K - 2)$ candidates as an optimal binarization result.

It is to be noted that this technique tackles the problem of how to discriminate between character and non-character using SVM in the high-dimensional feature space based not on deterministic but on probabilistic means. In particular, this technique has a possibility for correctly binarizing both multi-color characters and/or characters with nonuniform backgrounds. Of course, K-means clustering in the HSI color space should generate a sufficient number of subimages for obtaining a successful dichotomization that corresponds to a correctly binarized image.

4. Distortion-tolerant character recognition as elastic template matching

In this section, we compare two competing techniques of distortion-tolerant template matching or elastic image matching. The first one is our global affine transformation (GAT) correlation technique (Wakahara et al., 2001). GAT correlation absorbs distortion expressible by affine transformation by determining optimal affine parameters that maximize a normalized cross-correlation value between an affine-transformed input image and a template. The second one is the well-known tangent distance (TD) (Simard et al., 1993). The tangent distance absorbs distortion expressible by a linear combination of predefined geometric and topographical transformations as applied to both an input image and each template.

First of all, considering that there is only a limited quantity of data against a wide variety of fonts and image degradations we dare to take the position that only a single template is provided for each character category.

Here, we use the "HGP Gothic E" font set for 62 alphanumeric characters as templates. As preprocessing, position and size normalization together with blurring operation is applied to each template. We set a size of each preprocessed gray-scale template at 28×28 pixels.

Figure 10 shows examples of templates.

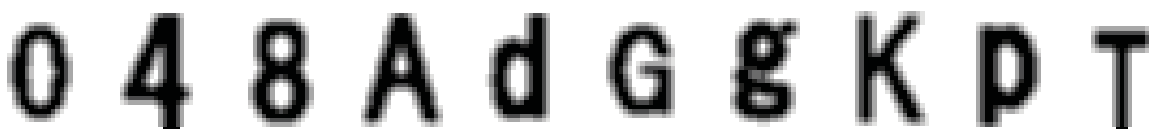


Fig. 10. Examples of templates.

4.1 GAT correlation

This technique provides a computational model for determining optimal affine parameters that deform an original input image, f , so as to yield the maximum correlation value against a template image, g .

First, both an input image, $f = \{f(r)\}$, and each template, $g = \{g(r)\}$, are linearly transformed to take the zero mean and the unit variance. As a result, a normalized cross-correlation value is

made equal to an inner product (f, g) . It is to be noted that image matching by means of normalized cross-correlation was shown to be robust against image blurring and additive random noise (Sato, 2000).

We denote the GAT-superimposed input image by $Affine[f]$. Here, $Affine[\bullet]$ stands for the operation of affine transformation in the 2D space, defined by a 2×2 matrix, A , representing rotation, scale-change, and shearing, and a 2D translation vector, b :

$$A = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \quad b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}. \quad (7)$$

The objective function Φ to maximize the value of $(Affine[f], g)$ is given by

$$\begin{aligned} \Phi &\equiv (Affine[f], g) = \sum_r Affine[f](r) \times g(r) \\ &= \sum_r f(r)g(\tilde{r}) \rightarrow \max \text{ for } A \text{ and } b, \\ \text{where } \tilde{r} &= Ar + b. \end{aligned} \quad (8)$$

Then, to avoid an exhaustive search for optimal A and b , we employ another objective function Ψ with Gaussian kernels given by

$$\begin{aligned} \Psi &\equiv \sum_r \sum_{r'} \gamma f(r)g(r')G(A, b, r, r') \\ &\rightarrow \max \text{ for } A \text{ and } b, \\ \text{where } \gamma &: \text{a function of } \nabla f \text{ and } \nabla g, \\ G(A, b, r, r') &= \exp\left(-\frac{\|r' - \tilde{r}\|^2}{2D^2}\right), \quad \tilde{r} = Ar + b, \end{aligned} \quad (9)$$

where the weight function γ serves as matching constraints. Also, D controls the spread of the Gaussian kernel.

Here, we explain how to practically design the values of γ and D .

First, γ of Eq. (9) is a function of ∇f and ∇g as matching constraints with the aim of promoting matching between pixels with the similar gradients. Here, we propose the concrete form of γ given by

$$\gamma(\nabla f, \nabla g) = \begin{cases} 1 & \text{if } \nabla f = \nabla g, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

where gradients ∇f and ∇g are quantized into eight directions at intervals of $\pi/4$. The introduction of γ into Ψ has an effect that optimal affine transformation forces matched pixels to have the same gradient direction.

Second, the parameter D of Eq. (9) controls the spread of the Gaussian kernel or the radius of search area for matching pixels by affine transformation. Hence, a suitable selection of D is the key to stabilizing the whole matching process. We propose to adaptively determine

the value of D prior to GAT application according to the disparity of input and template images in a gradient space as follows.

$$D = \frac{1}{2} \left\{ Av \left[\min_{r'} \left(\| \mathbf{r} - \mathbf{r}' \|; \gamma(\nabla f, \nabla g) = 1 \right) \right] + Av \left[\min_r \left(\| \mathbf{r} - \mathbf{r}' \|; \gamma(\nabla f, \nabla g) = 1 \right) \right] \right\}, \quad (11)$$

where Av stands for an averaging operation over either input or template images. Namely, D is the average minimum distance between two points, one in f and the other in g , with the same gradient direction.

Now, by setting the derivatives of Ψ with respect to each of six unknown parameters, a_{00} , a_{01} , a_{10} , a_{11} , b_0 , and b_1 , equal to zero, respectively, we obtain a set of nonlinear equations. Next, by using the 0th order approximation that sets $A = I$ and $\mathbf{b} = \mathbf{0}$ in the Gaussian kernel, we have a set of simultaneous linear equations. Finally, we solve these simultaneous linear equations by conventional techniques and obtain a sub-optimal solution of A and $\mathbf{b}$.

In order to obtain the true optimal GAT of Eq. (8), we use the successive iteration method by iteratively updating the input gray-scale image by sub-optimal affine parameters of Eq. (9) until the value of Φ arrives at a maximum.

4.2 Tangent distance

This technique encourages invariance of distance-based methods to a set of predefined transformations, which realizes distortion-tolerant template matching.

Concretely, by using a set of predefined geometric or topographical transformations applicable to an input image, f , and each template, g , we generate a tangent vector corresponding to each geometric or topographical transformation. Here, it is to be noted that all elements of both input/template images and tangent vectors are gray-scale values in the image plane.

The tangent distance, $D_T(f, g)$, is calculated as the minimum distance between two hyperplanes expanded by a set of tangent vectors around input and template images given by

$$D_T(f, g) = \min_{\alpha_f, \alpha_g} \| \tilde{\mathbf{f}} - \tilde{\mathbf{g}} \|, \quad (12)$$

$$\tilde{\mathbf{f}} = \mathbf{f} + T_f \alpha_f, \quad \tilde{\mathbf{g}} = \mathbf{g} + T_g \alpha_g,$$

where matrices T_f and T_g have their corresponding tangent vectors as column vectors. Also, α_f and α_g represent expansion coefficient vectors.

Tangent vectors are obtained via convolution between input/template images and Gaussian filters operated in advance by corresponding geometric or topographical transformations. Here, 2D Gaussian filters are given by

$$G_{\sigma}(\mathbf{r}) = \exp\left(-\frac{\|\mathbf{r}\|^2}{2\sigma^2}\right), \quad (13)$$

where the value of σ was set at $\sigma = 0.7$, and the size of a convolution mask was 19×19 . We deal with seven kinds of geometric or topographical transformations: X-translation, Y-translation, rotation, scaling, parallel hyperbolic transformation, diagonal hyperbolic transformation, and thickening (Simard et al., 1993).

Figure 11 shows examples of tangent vectors.

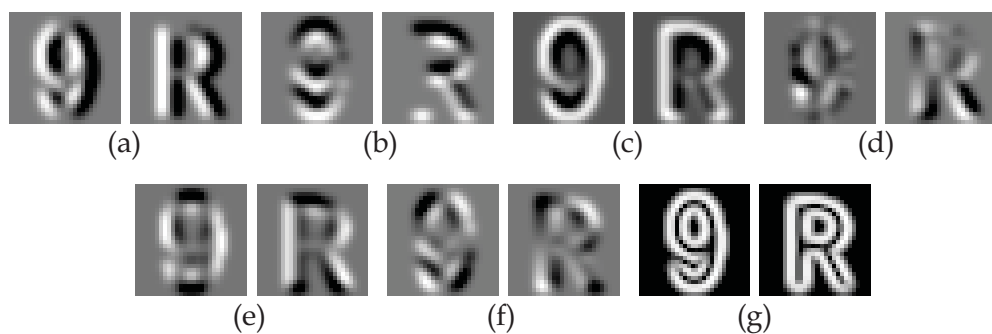


Fig. 11. Examples of tangent vectors. (a) X-translation. (b) Y-translation. (c) Rotation. (d) Scaling. (e) Parallel hyperbolic transformation. (f) Diagonal hyperbolic transformation. (g) Thickening.

5. Experimental results

In this section, we show two kinds of experimental results using ICDAR 2003 robust OCR dataset: figure-ground discrimination of color characters in scene images and distortion-tolerant character recognition as elastic template matching.

5.1 Abilities of figure-ground discrimination of color characters in scene images

Determination of an optimal sequence of filters for binarization using GA:

Figure 12 shows examples of binarization results obtained for both training and test samples in all of six degradation categories.

From Fig. 12, it is found that binarization of test samples is remarkably successful even if embedded characters in training and test samples are totally different in shape.

Moreover, In order to evaluate the ability of binarization in a more quantitative manner, we calculated a normalized cross-correlation value between optimally filtered images and their respective target images ideally binarized by humans.

Figure 13 shows relations between average correlation values and image degradation categories obtained from both training and test samples against their target images.

From Fig. 13, it is found that optimal sequences of filters determined by GA have the marked ability to achieve a fairly high correlation value, more than 0.9, between filtered and target images against most of all image degradation categories.

These results show clearly that we can select the optimal filter sequence for binarization of a given image if its degradation category is automatically determined. In other words, when we deal with the case where the cause of degradation is found to be unique and specific, this technique for binarization using the optimal filter sequence is expected to be very powerful.

Group	Training samples	Test samples
(a)		
(b)		
(c)		
(d)		
(e)		
(f)		

Fig. 12. Examples of binarization results. (a) Clear. (b) Background with pattern. (c) Character with pattern. (d) Character with rims. (e) Blurring. (f) Nonuniform lighting.

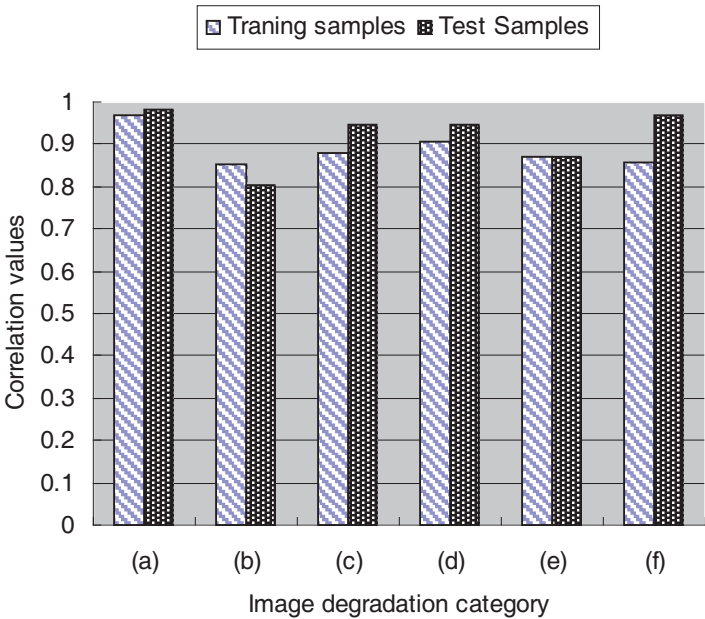


Fig. 13. Relations between correlation values and image degradation categories. (a) Clear. (b) Background with pattern. (c) Character with pattern. (d) Character with rims. (e) Blurring. (f) Nonuniform lighting.

Binarization using a maximum separability axis in a color space:

Table 2 shows rates of successful and unsuccessful binarization. Figure 14 shows examples of unsuccessful figure-ground discrimination. From Table 2 and Fig. 14, it is found that the task of temporary binarization poses a more serious problem than that of figure-ground determination does.

Results	Rates
Successful binarization	75.3%
Unsuccessful temporary binarization	17.5%
Unsuccessful figure-ground determination	7.2%

Table 2. Rates of successful and unsuccessful binarization.



Fig. 14. Examples of unsuccessful figure-ground discrimination. (a) Unsuccessful temporary binarization. (b) Unsuccessful figure-ground determination.

K-means clustering in a color space and figure-ground discrimination by SVM:

The number of clusters, K , in the K-means clustering was set at 5, and, hence, a total number of tentatively binarized images was 30 ($= 2^5 - 2$).

First, we evaluated the ability of discrimination between character and non-character via SVM using three kinds of feature vectors extracted from tentatively binarized images. Here, we adopted the technique of S -fold cross-variation (Bishop, 2006), which allows a proportion $(S-1)/S$ of the available data to be used for training while making use of all of the data to assess performance. We set at the value of S at 10.

Based on evaluation of false reject/acceptance rates (FRR, FAR) according to the sign of $f(x)$ of Eq. (5), we found that the radial basis function (RBF) as a kernel function of SVM achieved the minimum sum of FRR and FAR.

Figure 15 shows the distribution of SVM outputs for test samples using the RBF kernel and the weighted direction code histogram feature.

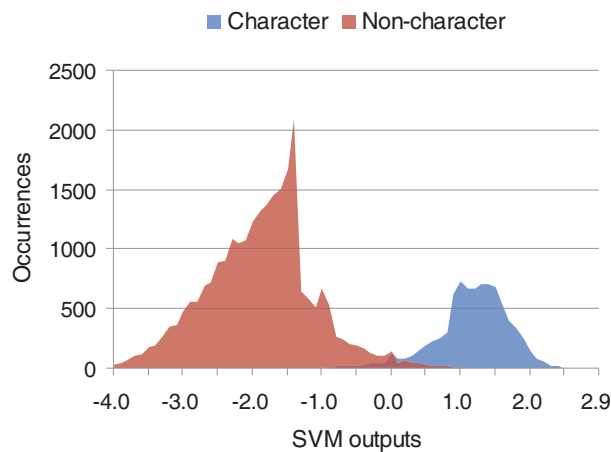


Fig.15. Distribution of SVM outputs for test samples using the RBF kernel and the weighted direction code histogram feature.

Figure 16 shows ROC (Receiver Operating Characteristic) curves obtained by moving a threshold for discrimination between character and non-character on the SVM output.

From Fig. 16, it is found that SVM fed with the weighted direction code histogram feature is the top of the three feature vectors and achieved the minimum equal error rate, EER, of 6.2%. Next, we investigated the ability of selecting a correctly binarized image from a total of 30 tentatively binarized images based on the values of SVM outputs. Here, we selected the binarized image with the maximum value of SVM outputs as an optimal binarization result. Namely, a total of 30 candidate binary images were arranged in the decreasing order of SVM outputs, and the top one was selected as a correctly binarized image.

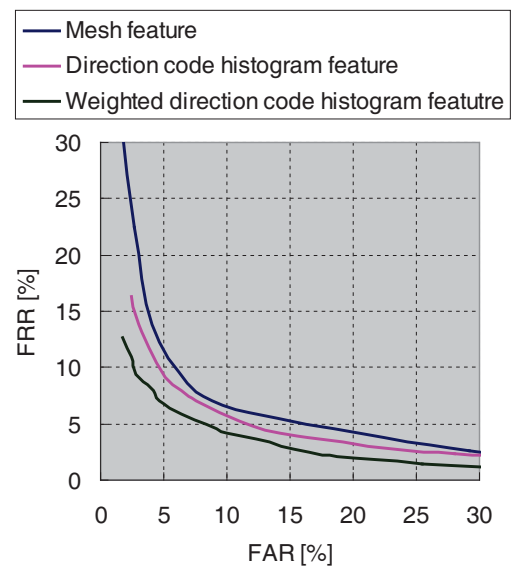


Fig.16. ROC curves obtained for three kinds of feature vectors via SVM with the RBF kernel.

Figure 17 shows cumulative binarization rates via SVM. The k th cumulative binarization rate is an average rate at which the top k candidate binary images contain a correctly binarized image. From Fig. 17, it is found that the correct binarization rate or the 1st cumulative binarization rate is 92.2%, and the 7th cumulative binarization rate is over 99.0%.

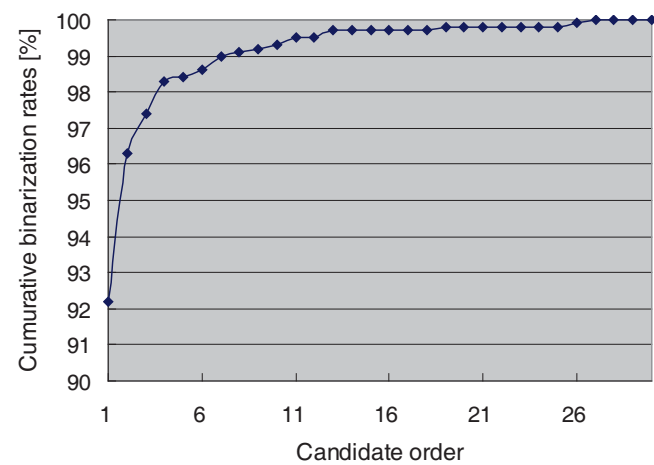


Fig.17. Cumulative binarization rates via SVM.

5.2 Abilities of distortion-tolerant character recognition as elastic template matching

In this subsection, we show results of distortion-tolerant recognition of correlctly binarized characters by the GAT correlation and the tangent distance (Wakahara, 2008).

Input images were normarized with respect to position and size, and were set at a size of 28×28 pixels. The matching measure of the GAT correlation is a normalized cross-correlation value, while the matching measure of the tangent distance is a pixelwise distance in gray-scale values, as described in Section 4. The dimension of a feature vector is 28×28 . It is to be noted that $f(r)$ and $g(r)$ in the GAT correlation can be any features extracted from images as far as they are a function of 2D loci vectors, r . On the other hand, the tangent distance can use no features besides gray-scale values.

Table 3 shows recognition rates for correctly binarized characters. The matching measure of simple correlation is a normalized cross-correlation value calculated between an input image and each template. Moreover, in the GAT correlation, we tried the well-known gradient features for correlation matching. Here, the dimenstion of a gradient feature vector is $28 \times 28 \times 8$, where an original 2D gray-scale image is decomposed into eight gradient images calculated along the direction at intervals of $\pi/8$.

Methods	Recognition rates (%)
Simple correlation (gray-scale values)	80.4
GAT correlation (gray-scale values)	90.3
GAT correlation (gradient values)	94.1
Tangent distance (gray-scale values)	91.6

Table 3. Recognition rates for correctly binarized characters.

From Table 3, it is first found that both GAT correlation and tangent distance reduced the error rate of the simple correlation more than by half. Secondly, it is found that the use of gradient features in GAT correlation improved the recognition accuracy markedly. Figure 18 shows examples of correctly recognized and misrecognized images by both of GAT correlation and tangent distance.



Fig. 18. Examples of correctly recognized and misrecognized images by both of GAT correlation and tangent distance. (a) Correctly recognized. (b) Misrecognized.

Furthermore, in order to evaluate the robustness of GAT correlation and tangent distance against rotation which cannot be compensated by position and size normalization, we fed each of recognizers with artificially rotated templates to be matched against upright templates.

Figure 19 shows relations between rotation angles and mean of normalized cross-correlation values, where elastic image matching was performed between each upright template and their artificially rotated templates from -45 degrees to +45 degrees at intervals of 5 degrees. From Fig. 19, it is clear that the GAT correlation is superior to the tangent distance in robustness against rotation at an angle of more than 20 degrees.

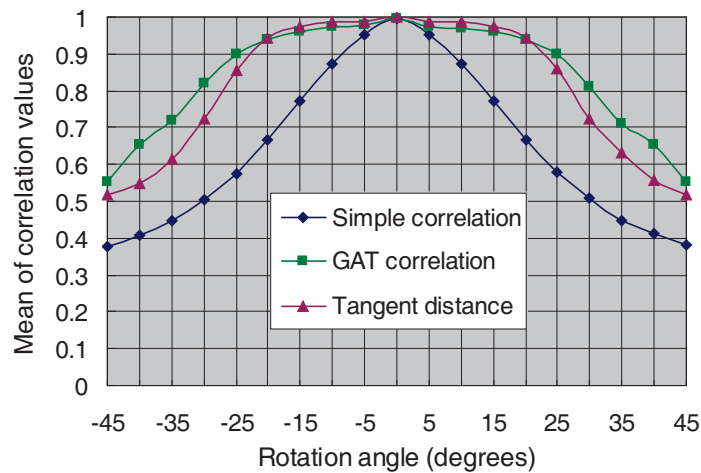


Fig. 19. Relations between rotation angles and mean of normalized cross-correlation values.

6. Discussion and future work

We tackled two challenging problems: figure-ground discrimination or correct binarization of color characters in scene images as a crucial step to the success of subsequent recognition, and distortion-tolerant character recognition under the condition of a small sample size.

Regarding the first problem, we proposed three kinds of techniques. Although each of three techniques showed promising preliminary results, we dare to enumerate their weak points, respectively, as follows.

The first technique of generating an optimal sequence of filters for binarization using GA had the following two disadvantages: not automatic but manual selection of degradation categories and the limited ability against a wide variety of complex backgrounds even if the degradation category is specified.

The second technique of using a maximum separability axis in a color space based on Otsu's criterion had also one major disadvantage: the insufficient adaptability to multi-color characters and/or characters with nonuniform backgrounds.

The third technique of using K-means clustering in a color space and figure-ground discrimination by SVM showed the most promising preliminary results mainly because of its potential ability to deal with multi-color characters and/or characters with nonuniform, complex backgrounds.

Hence, we enumerate several issues concerning the third technique of using K-means clustering in a color space and figure-ground discrimination by SVM still need to be addressed.

- (1) Adaptive and stable determination of the optimal number of clusters in K-means clustering,
- (2) Selection of more efficient feature vectors for evaluating character-likeness, and
- (3) Systematic expansion of training data in SVM using a kind of degradation or deformation models.

Regarding the second problem, we compared two competing techniques as elastic template matching: GAT correlation and tangent distance. Although both of them achieved recognition rates of more than 90% for correctly binarized characters, the recognition accuracy still needs to be much improved to meet the practical demands of the market. From this viewpoint, the following issues remain to be solved.

- (1) Appropriate selection of multiple templates per category, and
- (2) Cooperation between distortion-tolerant template matching and statistical pattern recognition techniques.

Finally, it is necessary and interesting to extend and apply techniques of recognizing single-character images to recognition of character strings in scene images.

7. Acknowledgments

The author would like to thank his former excellent students, Ms. Hanako Kohmura, Mr. Minoru Yokobayashi, and Mr. Junpei Kato, for their fruitful discussions and long, careful experiments.

8. References

- Bishop, C. B. (2006). *Pattern Recognition and Machine Learning*, Springer, 2006
- Casey, R. G. (1970). Moment normalization of handprinted characters, *IBM J. Res. Develop.*, Vol. 14, No. 5, pp. 548-557, September 1970
- Doermann, D.; Liang, J. & Li, H. (2003). Progress in camera-based document image analysis, *Proceedings of 7th International Conference on Document Analysis and Recognition*, Vol. I, pp. 606-616, Edinburgh, August 2003
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, 1989
- Gonzalez, R. C. & Woods, R. E. (2000). *Digital Image Processing*, Second Edition, Prentice Hall, 2000
- Herauld, L. & Horaud, R. (1993). Figure-ground discrimination: a combinatorial optimization approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 899-914, September 1993
- ICDAR Datasets. (2003). <http://algoval.essex.ac.uk/icdar/Datasets.html>. 2003
- Joachims, T. (1998). Making large-scale SVM learning practical, In: *Advances in Kernel Methods: Support Vector Learning*, Schölkopf, B.; Burges, C., Smola, A., Chap. 11, MIT Press, 1998
- Kimura, F.; Wakabayashi, T., Tsuruoka, S., Miyake, Y. (1997). Improvement of handwritten Japanese character recognition using weighted direction code histogram, *Pattern Recognition*, Vol. 30, No. 8, pp. 1329-1337, August 1997
- Kato, J. & Wakahara, T. (2009). Binarization of color characters in scene images using K-means clustering and support vector machines, *Proceedings of 12th Meeting on Image Recognition and Understanding*, pp. 351-358, Matsue, July 2009 (in Japanese)

- Kohmura, H. & Wakahara, T. (2006). Determining optimal filters for binarization of degraded characters in color using genetic algorithms, *Proceedings of 18th International Conference on Pattern Recognition*, Vol. III, pp. 661-664, Hong Kong, August 2006
- Lucas, S. M.; Panaretos, A., Sosa, L., Tang, A., Wong, S. & Young, R. (2003). ICDAR 2003 robust reading competitions, *Proceedings of 7th International Conference on Document Analysis and Recognition*, Vol. I, pp. 682-687, Edinburgh, August 2003
- Miене, A.; Hermes, T. & Ioannidis, G. (2001). Extracting textual inserts from digital videos, *Proceedings of 6th International Conference on Document Analysis and Recognition*, pp. 1079-1083, Seattle, September 2001
- Otsu, N. (1979). A threshold selection method from gray-level histogram, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 9, No. 1, pp. 62-69, January 1979
- Sato, A. (2000). A learning method for definite canonicalization based on minimum classification error, *Proceedings of 15^h International Conference on Pattern Recognition*, Vol. II, pp. 199-202, Barcelona, September 2000
- Simard, P.; LeCun, Y. & Denker, J. (1993). Efficient pattern recognition using a new transformation distance, *Advances in Neural Information Processing Systems*, Vol. 5, [NIPS Conference], pp. 50-58, Morgan Kaufmann, 1993
- Trier, O. & Jain, A. K. (1995). Goal directed evaluation of binarization methods, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 12, pp. 1191-1201, December 1995
- Uchida, S. & Sakoe, H. (2005). A survey of elastic matching techniques for handwritten character recognition, *IEICE Transactions on Information and Systems*, Vol. E88-D, No. 8, pp. 1781-1798, August 2005
- Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*, Second Edition, Springer, 2000
- Wakahara, T.; Kimura, Y. & Tomono, A. (2001). Affine-invariant recognition of gray-scale characters using global affine transformation correlation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 4, pp. 384-395, April 2001
- Wakahara, T. (2008). Figure-ground discrimination and distortion-tolerant recognition of color characters in scene images, *Proceedings of 19th International Conference on Pattern Recognition*, Tampa, December 2008
- Wolf, C.; Jolion, J. & Chassaing, F. (2002). Text localization, enhancement and binarization in multimedia document, *Proceedings of 16th International Conference on Pattern Recognition*, Vol. 2, pp. 1037-1040, Quebec, August 2002
- Wu, S. & Amin, A. (2003). Automatic thresholding of gray-level using multi-stage approach, *Proceedings of 7th International Conference on Document Analysis and Recognition*, Vol. I, pp. 493-497, Edinburgh, August 2003
- Yokobayashi, M. & Wakahara, T. (2006). Binarization and recognition of degraded characters using a maximum separability axis in color space and GAT correlation, *Proceedings of 18th International Conference on Pattern Recognition*, Vol. II, pp. 885-888, Hong Kong, August 2006



Pattern Recognition

Edited by Peng-Yeng Yin

ISBN 978-953-307-014-8

Hard cover, 568 pages

Publisher InTech

Published online 01, October, 2009

Published in print edition October, 2009

For more than 40 years, pattern recognition approaches are continually improving and have been used in an increasing number of areas with great success. This book discloses recent advances and new ideas in approaches and applications for pattern recognition. The 30 chapters selected in this book cover the major topics in pattern recognition. These chapters propose state-of-the-art approaches and cutting-edge research results. I could not thank enough to the contributions of the authors. This book would not have been possible without their support.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Toru Wakahara (2009). Figure-Ground Discrimination and Distortion-Tolerant Recognition of Color Characters in Scene Images, Pattern Recognition, Peng-Yeng Yin (Ed.), ISBN: 978-953-307-014-8, InTech, Available from: <http://www.intechopen.com/books/pattern-recognition/figure-ground-discrimination-and-distortion-tolerant-recognition-of-color-characters-in-scene-images>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen