images are rotated. It is concluded that the HMM does capture the trend caused by rotation and provide effective information of dependence among the subbands.

3) The formation of the feature sequence is important. The feature sequences, starting from the lowest frequency subband, give similar recognition performances, which are far superior to those of the feature sequences starting from the highest frequency subband.

4) In the proposed scheme, only one parameter, the number of states in the HMM, needs to be specified. All other parameters are part of the design and training procedure.

5) The assumption of Gaussian density for the observation probability density function is only approximately true. Replacing the Gaussian observation density function by the mixture Gaussian density function, the recognition rate could be improved.

References

[1] F. Farrokhnia and A. K. Jain, "A multi-channel filtering approach to texture segmentation," in *Proc. CVPR-91*, Hawaii, pp. 364–370, June 1991.

[2] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," to appear in *Pattern Recognition*, vol. 24, pp. 1167–1186, 1991.

[3] J. M. Coggins and A. K. Jain, "A spatial filtering approach to texture analysis," *Pattern Recognition Lett.*, vol. 3, pp. 195–203, May 1985.

[4] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, no. 1, pp. 25–39, Jan. 1983.

[5] R. L. Kashyap and A. Khotanzad, "A model-based method for rotation invariant texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 472–481, July 1986.

[6] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 4, pp. 959–963, Aug. 1985.

[7] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[8] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 55–73, Jan. 1990.

[9] F. S. Cohen, Z. Fan, and M. A. Patel, "Classification of rotation and scaled textured images using Gaussian Markov random field models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 192–202, Feb. 1991.

[10] M. Unser and M. Eden, "Multiresolution feature extraction and selection for texture segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 717–728, July 1989.

[11] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 1, pp. 39–55, Jan. 1987.

[12] B. Julesz, "Visual pattern discrimination," *IRE Trans. Information Theory*, pp. 84–92, Feb. 1962.

[13] B. Julesz and J. R. Bergen, "Textons, the fundamental elements in preattentive vision and perception of textures," *Bell Syst. Tech. J.*, vol. 62, no. 6, pp. 1619–1645, July/Aug. 1983.

[14] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[15] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 4, pp. 269–285, Apr. 1976.

[16] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[17] C. W. Therrien, "An estimation-theoretic approach to terrain image segmentation," *Comput. Vision, Graph., Image Processing*, vol. 22, pp. 313–326, 1983.

[18] K. L. Laws, "Rapid texture identification," *Proc. SPIE*, vol. 238, pp. 376–380, 1980.

[19] F. Campbell and J. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol.*, vol. 197, pp. 551–566, 1968.

[20] J. Nachmais and A. Weber, "Discrimination of simple and complex gratings," *Vision Research*, vol. 15, pp. 217–223, 1975.

[21] M. Vetterli, "Multi-dimensional sub-band coding: Some theory and algorithms," *Signal Processing*, vol. 6, pp. 97–112, Apr. 1984.

[22] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 5, pp. 1278–1288, Oct. 1986.

[23] P. P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions and perfect reconstruction techniques," *IEEE ASSP Mag.*, vol. 4, pp. 4–20, July 1987.

[24] L. R. Rabiner and B-H. Juang, "An introduction to Hidden Markov Models," *IEEE ASSP Mag.*, pp. 4–16, June 1986.

[25] A. Ljolje and S. E. Levinson, "Development of an acoustic-phonetic hidden Markov model for continuous speech recognition," *IEEE Trans. Signal Processing*, vol. 39, pp. 29–39, Jan. 1991.

[26] A. Kundu, Y. He, and P. Bahl, "Recognition of handwritten word: First and second order hidden Markov model based approach," *Pattern Recognition*, vol. 22, pp. 283–297, 1989.

[27] Y. He and A. Kundu, "2D shape classification using HMM," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 11, pp. 1172–1184, Nov. 1991.

[28] P. Brodatz, *Textures—A Photographic Album for Artist and Designers.* New York: Dover, 1965.

[29] Anil K. Jain, *Fundamentals of digital image processing.* Englewood Cliffs, NJ: Prentice Hall, 1989, pp. 344–345.

Recognizing Characters in Scene Images

Jun Ohya, Akio Shio, and Shigeru Akamatsu

***Abstract*— An effective algorithm for character recognition in scene images is studied. Scene images are segmented into regions by an image segmentation method based on adaptive thresholding. Character candidate regions are detected by observing gray-level differences between adjacent regions. To ensure extraction of multisegment characters as well as single-segment characters, character pattern candidates are obtained by associating the detected regions according to their positions and gray levels. A character recognition process selects patterns with high similarities by calculating the similarities between character pattern candidates and the standard patterns in a dictionary and then comparing the similarities to the thresholds. A relaxational approach to determine character patterns updates the similarities by evaluating the interactions between categories of patterns, and finally character patterns and their recognition results are obtained. Highly promising experimental results have been obtained using the method on 100 images involving characters of different sizes and formats under uncontrolled lighting.**

***Index Terms*— Adaptive thresholding, character pattern extraction, character recognition, image processing, image segmentation, pattern recognition, relaxation, scene images.**

I. Introduction

In recent years, great progress has been made in optical character reader (OCR) technology. The range of recognizable characters has been extended from printed alphabetical letters, numbers, and kana

Manuscript received August 6, 1990; revised October 12, 1992. Recommended for acceptance by Editor-in-Chief A. K. Jain.

J. Ohya is with ATR Communication Systems Research Laboratories, Seika-cho, Soraku-gun, Kyoto 619-02, Japan.

A. Shio is with NTT Human Interface Laboratories, Yokosuka-shi 238-03, Japan.

S. Akamatsu is with ATR Human Information Processing Laboratories, Seika-cho, Soraku-gun, Kyoto 619-02, Japan.

IEEE Log Number 9208441.

(Japanese syllabic characters) to handprinted Chinese characters, and the character recognition rate has been greatly improved [1]. Most OCR's in current use, however, can only read characters printed on a sheet of paper according to rigid formatting restrictions and are mainly being applied to office automation systems such as document readers [2]. Nonetheless, if OCR's could read characters in scene images[1] acquired by a TV camera, they would be applied to more areas such as factory automation, automatic monitoring systems, and robotic systems.

Unlike characters in documents, characters in scene images can suffer from a variety of noise components. Characters in scene images originally exist in 3-D space, so they can be distorted by slant, tilt, and shape of objects on which the characters are printed. Other noise components include uncontrolled illuminating conditions. Moreover, these characters have sizes, pitches, positions, fonts, formats, and gray-levels that generally vary.

Conventional research work on character recognition in scene images includes automatic inspection processes in the manufacturing field [3] and vehicle license number plate readers [4], [5]. These systems, however, have the following limitations: positions of characters must be specified in advance [3], characters to be recognized must have simple shapes such as numerical and alphabetical letters [3], and characters must be uniformly illuminated [4], [5]. Automatic extraction of characters from images by estimating line pitches and character pitches was studied for document images [2], but this method is not useful for format-free situations in scene images. Character extraction from scene images was studied [6], [7], but involved unsatisfactory preprocesses for detecting character segments[2] by thresholding with a fixed threshold [6] and by image segmentation by merging neighboring pixels with small gray-level differences [7]. Image segmentation is an important preprocess to detect character segments accurately under uncontrolled illuminating conditions. Algorithms that could be useful for character segment detection have been studied [8]–[10], and among them there is a document image segmentation method that uses a training procedure for character segment detection [10].

The authors have proposed a new method that overcomes the problems in the conventional methods and achieves a very flexible character recognition system for scene images [11], [12]. The proposed method extracts and recognizes characters including multisegment characters such as Chinese characters from scene images under a variety of illuminating conditions on the assumption that the size, position, font, gray-level, and format of characters are unknown; this correspondence does not deal with the 3-D distortions described earlier. The strategy of the proposed method is: 1) before a character recognition process, character-like patterns corresponding to either characters or noncharacters are extracted; thereby, false dismissal of real characters is avoided; and 2) a character recognition process and a subsequent relaxational operation select and recognize character patterns. In this correspondence, characters in scene images are assumed to be almost upright (rotated slightly if any); to consist of character segments having the same gray-level; to not be texture-like characters; and to not be connected with other characters like a signature.

[1] In general 3-D scenes are acquired as 2-D images by a TV camera, and in this correspondence the acquired 2-D images are called scene images. Scene images include images of outdoor and indoor uncontrolled scenes.

[2] Characters consist of one or more strokes. Among the strokes there are not only isolated strokes, but also connected strokes, namely strokes that intersect other strokes. In this correspondence a character segment is defined as a connected stroke. A character with one character segment is called a single-segment character, and a character with more than one character segments is called a multisegment character.

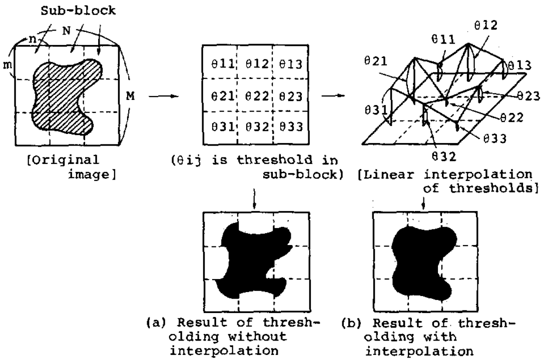


Fig. 1. Principle of local thresholding.

The concrete processes are as follows. Local thresholding [11], [13] which does not need a training process is utilized for image segmentation. In the image segmentation results, character candidate regions that may correspond to character segments are detected by evaluating gray-level differences between adjacent regions. Character pattern candidates are obtained by associating the detected regions that exist close to each other and have similar gray-levels. A character recognition process calculates the similarity between a character pattern candidate and each category in a dictionary. If the similarity value for a category is large, i.e., similar to a category, the character pattern candidate is selected as a high similarity pattern. A relaxational operation [12] removes the ambiguities and contradictions between high similarity patterns and achieves extraction and recognition of characters.

In the following, the details of the proposed algorithm are given. Section II describes the detection of character pattern candidates by image segmentation using local thresholding. Section III explains the selection of high similarity patterns by a character recognition process, and Section IV shows how to determine and recognize character patterns by a relaxational operation. Section V gives the experimental conditions, results, and discussions, and Section VI concludes this correspondence. Results of the experiment using 100 scene images confirm the potential of the algorithm for actual robotic systems.

II. Detecting Character Pattern Candidates

A. Image Segmentation Using Local Thresholding

Local thresholding [11], [13], which is a kind of adaptive thresholding, is utilized for image segmentation. The principle of local thresholding is shown in Fig. 1. An original image ($M \times N$ pixels) is divided into subblocks ($m \times n$ pixels, $m < M, n < N$), in each of which the gray-level threshold is determined. Here, if the original image is thresholded with the determined threshold in each subblock, there is discontinuity between subblocks as shown in Fig. 1(a). Therefore, the threshold of each pixel is determined by linearly interpolating thresholds between the central pixels of subblocks [12]. Using these interpolated thresholds, smooth bilevel (black/white) images can be obtained as shown in Fig. 1(b). Examples of results of local thresholding are shown in [11].

Generally, in scene images there are subblocks that do not include characters and whose gray-level is almost uniform. In such subblocks thresholding is meaningless by nature and yields small noise-like patterns. To deal with this matter, a bimodality test for the gray-level histograms in subblocks was introduced to judge whether the subblocks should be thresholded [13]. In this correspondence, to avoid losing real character segments, thresholding is done in all subblocks. Noise-like patterns are removed by the process described in Section II-B.

There are various ways of determining thresholds in subblocks. In this correspondence, the automatic and nonparametric threshold selection method using the discriminant criterion [15] is applied to each subblock.

Whether characters appear as black or white is unknown in bilevel images obtained by local thresholding, because gray-level information on characters and their background are not given in advance. Therefore, 8-adjacent pixels with the same level (black or white) are grouped into regions, and a label is given to each region. By the process in Section II-B each region is tested as to whether it could correspond to a character segment.

In local thresholding, fake regions tend to appear in relatively large areas where changes in the gray-level are small and smooth [11]. In such cases, image segmentation results do not conform to their original images. However, since character segments are narrow and form binary images locally, image segmentation results for character segments conform to their original character segments. In this sense local thresholding is suitable for character segment detection.

B. Detecting Character Candidate Regions

It is necessary to detect regions corresponding to character segments from the image segmentation results. The authors have studied features for detecting character segments [16]. The features are: 1) the gray-level contrast between character segments and their background is high, 2) the width of a character segment is uniform, 3) the gray-level of a character segment is uniform, and 4) the spatial frequency of character segments is high. As a result of a study using parameters representing the above features, 1) was found to be useful for removing noise-like small regions as described in Section II-A, but 2)–4) were not useful and tended to lead to false detection [16].

A parameter for evaluating the gray-level contrast is indicated in Fig. 2. Let L_C and L_B be the average gray-level in a region C and in an outer region B (the area in the rectangle that circumscribes C, excluding C). The gray-level difference ΔL which is the parameter for contrast is given by

$$\Delta L = |L_C - L_B|. \tag{1}$$

In each region, ΔL is obtained by (1). If ΔL satisfies

$$\Delta L \geqq T_L \tag{2}$$

then the region is detected as a character candidate region, where T_L is the threshold value. If not, it is judged as not corresponding to a character and is merged with the outer region. Here, regions touching the edges of images are judged as background and are not detected. By (1) and (2), both regions corresponding to characters and those not corresponding to characters are detected, but distinguishing characters from noncharacters using the low-level character segment features such as 2)–4) will lead to errors resulting in lost character segments.

C. Extracting Character Pattern Candidates

Whether or not the character candidate regions (in the following, regions) described in Section II-B correspond to characters is determined by a character recognition process. By this process, the loss of real character segments can be avoided, and character recognition results can be obtained at the same time. For the character recognition process, it is necessary to extract character patterns from images. However, extracting single regions is insufficient, because there are also multisegment characters, such as "i", "j", and some Chinese characters. To ensure the extraction of multisegment characters, character pattern candidates are obtained by associating the regions as explained below.

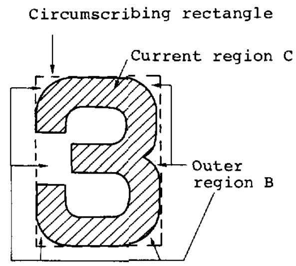


Fig. 2. Evaluation of gray-level difference.

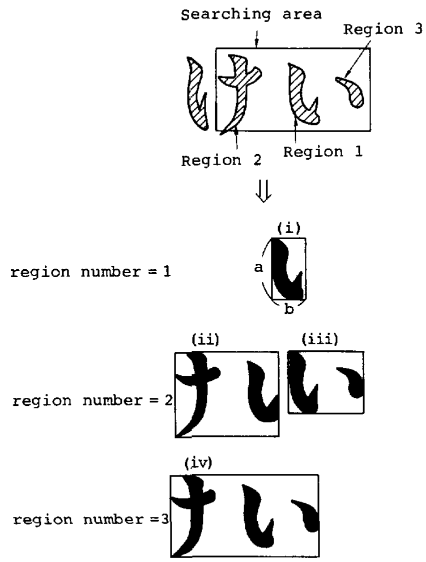


Fig. 3. Extraction of character pattern candidates.

An example of extracting character pattern candidates is shown in Fig. 3. Let L_i be the average gray-level, and B_i be the bilevel value (black/white) of the thresholded results of a region i. Let j be a region in the searching area near i, and T_c be the threshold value. If

$$B_j = B_i \tag{3}$$

or

$$|L_i - L_j| \leqq T_c \tag{4}$$

is true, then region j is listed as a region that may be part of the same character as region i, i.e., region i and j may be associated. Equations (3) and (4) are based on the idea that regions included in the same character have the same bilevel value and/or have almost equal gray-levels. If two or more regions satisfy (3) and/or (4), all the combinations are considered.

III. Selecting High Similarity Patterns

A character recognition process judges whether character pattern candidates extracted in Section II could correspond to characters and selects high similarity patterns. In this correspondence, the Mesh feature algorithm for multifont printed Chinese characters [17] is used for the character recognition process. This process divides a normalized character pattern into $64(=8\times 8)$ subareas and obtains mesh feature vector $Y=(y_{11},\cdots,y_{88})^T$, where y_{uv} is the number of black pixels in the (u,v) subarea and is normalized by the number of pixels in a subarea. The similarity $S(C_i)$ between a character pattern candidate and the standard pattern of category C_i in the dictionary is obtained from

$$\begin{aligned} S(C_i) &= (X(C_i),Y)/(|X(C_i)|\cdot|Y|) \\ i &= 1,2,\cdots,N \\ N &= \text{number of categories}, \qquad 0 \leqq S(C_i) \leqq 1 \end{aligned} \tag{5}$$

where $X(C_i)$ is the standard mesh feature vector for C_i, Y is the mesh feature vector for a character pattern candidate, and (,) denotes an inner product. If one or more $S(C_i)$ values satisfy

$$S(C_i) \geqq T(C_i), \qquad i=1,2,\cdots,N \tag{6}$$

where $T(C_i)$ is the similarity threshold value for each category, determined by the minimum similarity of the patterns to be recognized, then the character pattern candidate is extracted as a high similarity pattern. C_i's with $S(C_i)$ values satisfying (6) are extracted as candidate categories. If no $S(C_i)$ value satisfies (6), the character pattern candidate is assumed to be a noncharacter and rejected.

IV. Determining and Recognizing Character Patterns by a Relaxational Operation

The character recognition process calculates the similarities between character pattern candidates and each category in the dictionary. As described in Section III, the character recognition process is used not only to recognize characters but also to extract patterns with a high similarity to any categories (high similarity patterns), and it rejects character pattern candidates with a low similarity to any categories. Obviously, if by those processes only character pattern candidates corresponding to real characters were accurately extracted as high similarity patterns, there would be no problem. In reality, however, it is very difficult to achieve such a system because of the ambiguities inherent in the character recognition process.

Assume that with respect to regions and character patterns the following general rules exist:

1) One region corresponds to one or no character pattern.
2) If a region corresponds to a character pattern, its adjacent regions cannot correspond to any other character patterns.

However, high similarity patterns contradicting the above rules can be extracted, because character pattern candidates are obtained by associating regions as described in Section II. Moreover, simply comparing the similarities of each pattern may also lead to incorrect extraction, because the similarity values do not always predict the likelihood of the categories being correct.

This correspondence introduces a relaxational approach that effectively removes the ambiguities and contradictions between high similarity patterns and finally, achieves extraction and recognition of character patterns.[12]

The method proposed here is based on relaxation [18] and removes the ambiguities in the character recognition process by updating the similarities of character recognition results. This updating is done by evaluating interactions between candidate categories of high similarity patterns in order to select character patterns that comply with the above rules 1) and 2). Specific interactions $(A),(B)$, and (C) involve such primary category features as (A) whether or not the category is a multisegment character; (B) whether or not the category has holes; and (C) aspect ratio of circumscribing rectangle of the pattern.

The similarities of candidate categories of a high similarity pattern are updated by evaluating the specific interactions between the pattern

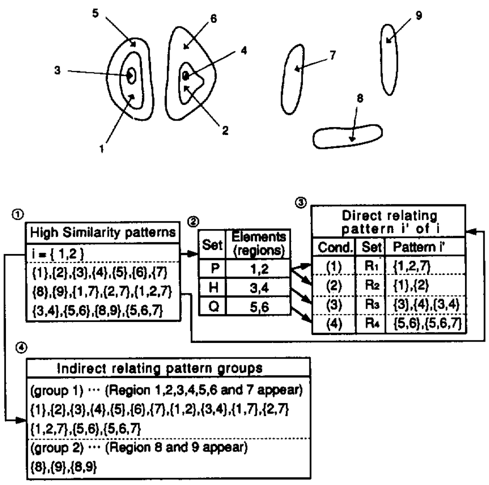


Fig. 4. Examples of direct relating pattern and indirect relating pattern group. $1\sim 9$ denote region numbers, { } denotes a high similarity pattern. In this figure, as a result of selecting high similarity patterns, the patterns denoted by { } in box ① are extracted. This figure shows how to get the direct relating patterns i's for $i=\{1,2\}$. The regions in sets P, H, and Q in Definition 1 are listed in box ②. According to Definition 1, patterns corresponding to R_1 and R_2 can be obtained from P by searching patterns except i in ①; similarly R_3 from H and R_4 from Q. Direct relating patterns i' of i are listed in box ③. Using the direct relating patterns of other patterns in ① (not shown), indirect relating pattern groups can be obtained according to Definition 2. In this figure, there are two groups shown in box ④.

and its direct relating patterns which cannot coexist with the pattern because they contradict the rules 1) and 2). The definition of a direct relating pattern is given as follows.

Definition 1: Direct relating pattern.

Direct relating pattern i' of a pattern i is defined as a pattern whose regions are the elements of either set R_1, R_2, R_3, or R_4 where R_1–R_4 satisfy the following conditions 1)–4), respectively:

1) $P \subset R_1$,
2) $R_2 \subset P$,
3) $R_3 \subseteq H$,
4) $Q \subseteq R_4$.

Here, P is a set of regions that belong to i, H is a set of holes surrounded by the elements of P, and Q is a set of regions surrounding the elements of P if all elements of P are the holes of other regions. Pattern i' derived from conditions 1) and 2) corresponds to interaction (A), and derived from 3) and 4) corresponds to (B). Direct relating patterns i' of i and examples of the sets described above are listed in Fig. 4. ∎

Now, if pattern i has m candidate categories $(C_1,\cdots,C_m)$ whose similarities are $(S_i^{(0)}(C_1),\cdots,S_i^{(0)}(C_m))$, the similarities of each candidate category are updated in terms of the interactions between i and the Ni' direct relating patterns i's. The similarities are updated using the following equations:

$$S_i^{(k+1)}(C_j)\begin{cases} = \{1-S_i^{(k)}(C_j)\}q_i + S_i^{(k)}C_j) \\ \quad \text{if } S_i^{(k)}(C_j) \geqq 0.5 \\ = S_i^{(k)}(C_j)q_i + S_i^{(k)}(C_j) \\ \quad \text{if } S_i^{(k)}(C_j) < 0.5 \end{cases}$$

$$j = 1,2,\cdots,m, \tag{7}$$

$$q_i = (1/N_{i'})\sum_{i'} d_{i'} \sum_{c'} r_{ii'}(C_j,C')S_i^{(k)}(C'). \tag{8}$$

TABLE I
EXAMPLES OF CATEGORY TABLE

Category	A	旧	1	B	—
Number of Segments	1	2	1	1	1
Number of Holes	1	2	0	2	0
Aspect Ratio QA^a	3	3	1	2	5

[a] Aspect ratio Ap is quantized to five levels.

$$
\begin{aligned}
QA &= 1 \text{ if } Ap \leqq 0.5 \\
&= 2 \text{ if } 0.5 < Ap \leqq 0.8 \\
&= 3 \text{ if } 0.8 < Ap \leqq 1.25 \\
&= 4 \text{ if } 1.25 < Ap \leqq 2.0 \\
&= 5 \text{ if } Ap > 2.0
\end{aligned}
$$

In (8), C' is a set of Nc' candidate categories of each i', and $d_{i'}$ is a coefficient defined as $d_{i'} = 1/N_{c'}$. Since $S_i^{(k)}(C_j)$ is in the range [0, 1] and the compatibility coefficient $r_{ii'}(C_j, C')$ is in the range $[-1, 1]$ as described below, q_i is in the range $[-1, 1]$. The compatibility coefficient $r_{ii'}(C_j, C')$ compares the compatibility of C_j to i and C' to i', and if C_j is more compatible than C', q_i approaches 1 and makes the similarity of C_j increase. In the opposite case, q_i approaches -1.

The compatibility coefficient $r_{ii'}(C_j, C')$ in (8) is derived in two steps, and the similarities are updated using (7) and (8) at each step. To perform these processes, the number of character segments, the number of holes and the quantized aspect ratio for each category are stored beforehand in a table, as shown in Table I.

In Step 1, $r_{ii'}(C_j, C')$ is obtained by evaluating interaction (A) between i and a direct relating pattern i' whose condition is 1) or 2) in Definition 1 and (B) for 3) and 4) in Definition 1. For example, for condition 1) in Definition 1, since i' corresponds to R_1, i' has two or more regions. If C' is a single-segment character, the likelihood of C' being correct is lower than that of C_j, and by making $r_{ii'}(C_j, C')$ positive, the similarity of C_j is increased. If C' is a multisegment character, $r_{ii'}(C_j, C')$ is made negative according to the difference between the number of regions in i' and segments in C'.

$$
r_{ii'}(C_j, C') \begin{cases} = e & (NR' = 1) \\ = -e\{1 - \min(1, \mid NR' - NI' \mid / NR')\} & (NR' \geqq 2) \end{cases} \tag{9}
$$

where e is a constant $0 < e < 1, NI'$ and NR' are the number of regions in i' and segments in C', respectively. Similar equations for conditions 2)–4) in Definition 1 are indicated in [12].

In Step 2, by comparing the aspect ratios of circumscribing rectangles for pairs of patterns, $r_{ii'}(C_j, C')$ is obtained. The aspect ratio of a circumscribing rectangle is quantized to five levels $(QA; QA = 1 - 5)$ as shown in Table I. Using these values,

$$
\begin{aligned}
r_{ii'}(C_j, C') = (e/4)(2 - |QAI - QAC|) \\
- (e/8)(|QAI - QAC| - |QAI' - QAC'|)
\end{aligned} \tag{10}
$$

where QAC, QAC', QAI, and QAI' are the QA value of C_j, C', i, and i', respectively. According to (10), if i and C_j have similar QA values and if i' and C' have different QA values, the similarity of C is increased by making $r_{ii'}(C_j, C')$ positive.

After the similarities of candidate categories of patterns are updated, the updated similarities are compared in each indirect relating pattern group so that the patterns with the highest updated similarities in each group are selected as character patterns. An indirect relating pattern group is defined as follows.

Definition 2: Indirect relating pattern group.

The patterns in an indirect relating pattern group are determined as follows: a pattern i_0 is put into a starting group. Next, the direct relating patterns i_1 of i_0 are put into the starting group. In the same way, the direct relating patterns i_{k+1} of each $i_k (k > 0)$, except the patterns already processed, are put into the group recursively. If there are no more patterns to be put into the group, another starting group is generated. This is repeated until all patterns are put into some group. Let R_p be a set of regions appearing in the patterns in a starting group; $R_{h'}$, a set of holes of any elements of R_p; and $R_{s'}$, a set of regions surrounding any elements of R_p. If any pattern in a different group has some elements of either $R_{p'}, R_h$, or $R_{s'}$, the starting groups are combined and form an indirect relating pattern group. All patterns are grouped into two indirect relating pattern groups in Fig. 4. ■

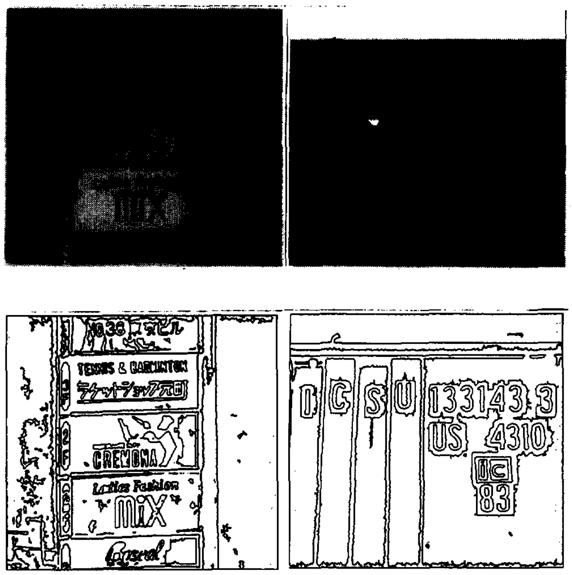
After k iterations of (7) and (8) for the two steps, the maximum updated similarity $S_{i,\max} = \max(S_i^{(k)}(C_1), \cdots, S_i^{(k)}(C_m))$ is obtained for every pattern i. Next, the pattern having the maximum $S_{i,\max}$ in an indirect relating pattern group is determined to correspond to a real character and extracted as a character pattern. Moreover, the candidate category providing the maximum $S_{i,\max}$ becomes the character recognition result. After character patterns are determined in the same way in all the indirect relating pattern groups, the patterns that include either 1) the regions appearing in the determined character patterns, 2) the holes of the appearing regions, or 3) the regions surrounding the appearing regions, are removed from the groups, because these patterns contradict the two rules mentioned in this section. For the pattern groups that become empty sets as a result of this removal, no more iterations are carried out. For the groups that are not empty, iterations continue. If all the groups become empty, the process stops. Note that our algorithms also works for scene images without characters unless character pattern candidates are similar to any characters by chance, because the character recognition process can reject all of the candidates in the images.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Image data for the experiments were acquired by taking photographic slides of scenes including characters of various sizes, gray-levels, and formats under uncontrolled lighting conditions, such as in cities or at train stations. These slides were then placed on a diffuser uniformly illuminated from behind by a fluorescent light, and acquired as images through a TV camera. These scenes were digitized into 256 gray-levels and 512×480 pixel images. For the following experiments, we used 100 images in which characters are printed on road signs, license number plates of automobiles, sign boards of shops, etc.

The authors did examine the influences of subblock sizes on local thresholding results for 20 scene images among the 100 images [11]. As the results of the experiments using the subblock size $4 \times 4, 8 \times 8, 16 \times 16$ and $32 \times 32, 8 \times 8$ (pixels) turned out to give the best results [11]. In this correspondence, a subblock size of 8×8 is used for local thresholding. For the thresholds of the gray-level difference between two adjacent regions in local thresholding results, $T_L = 3$ in (2) was used. Examples of detected regions are shown in Fig. 5, where the black lines indicate the borders of each region, and the white areas indicate the regions. As shown in the original image in Fig. 5(a), even if shadows are cast on part of the characters with various gray-levels, character segments can be correctly detected. In 95 images, all regions corresponding to character segments can be detected as regions although a few regions are in contact with other character segments due to lack of resolution. In the other five images, some characters merged with the background due to lack of contrast in the original images.

Experiments on extracting character pattern candidates were carried



(a) (b)

Fig. 5. Examples of detection of character candidate regions: (a) signboards, (b) freight train. Upper: original image. Lower: character candidate regions (white area: detected region, black line: border of region).

out using 50 images whose characters were large enough to be read by observers. In (4), $T_t = 7$ was used. An example of this extraction using the same original image as that for Fig. 5(b) is shown in Fig. 6, where all 21 characters in the original image were extracted as character pattern candidates, which are indicated by a black triangle at the upper-left corner of each CPC cell. Similarly, all of the 1045 characters in the 50 images were included in the set of character pattern candidates.

High similarity patterns were selected from 3164 character pattern candidates extracted from 15 images. These images included 213 characters whose shapes fit relatively well the fonts which the character recognition process could read. The dictionary contained 3417 categories, and the $T(C_i)$'s in (6) are set as shown in Table II.

An example of the selection for Fig. 5(b) is shown in Fig. 6, where notations H and L indicate selected and rejected candidates, respectively. Among the 21 character pattern candidates corresponding to characters, 20 candidates were correctly extracted as high similarity patterns, and one was incorrectly rejected. Among the 29 candidates corresponding to noncharacters, 23 were correctly rejected, and six were incorrectly extracted. The results for the 15 images are shown in Table III, where the correct judgment rate, that is, the rate of extracting real characters and rejecting noncharacters, was 71.1% and 93.4% of the 213 characters were correctly extracted.

Among the 20 high similarity patterns corresponding to real characters in Fig. 6, 11 patterns were correctly recognized. For the 15 images, the character recognition rate (the correct category has the highest similarity among the candidates) was 34.3%, and the accumulated recognition rate at the 10th order [17], which is defined as the percentage ratio where the correct category is contained in the candidate categories above the 10th order, was 66.5%. These values are lower than those of OCR's currently in use, but since the dictionary used in the experiments is not turned to all of the fonts of the characters in the images like in current-use OCR's, the results are considered appropriate.

To confirm the effectiveness of the relaxational approach for determining character patterns, the correct categories were added for

CPC*1		—	U	C	0	3	31	33	3	43	1	14	4
Sel.*2	L	L	H	H	H	H	L	L	H	L	H	L	H
Reg*3			u	c	B	3			3		1		4
Det*4			C	C	C	C			C		C		C
Eval*5	☆	☆	○	○	△	◎	☆	☆	◎	☆	○	☆	◎
CPC	3	13	1	3	S	1		3	43	31	0	10	4
Sel.	H	H	H	H	H	L	L	H	L	L	H	H	H
Reg	3	n	1	3	3			3			o	m	4
Det	C	R	C	C	C			C			C	R	C
Eval	◎	☆	○	◎	○	×	☆	◎	☆	☆	○	☆	◎
CPC	1	S	US	U				(1)			(2) i	c	(3)
Sel.	H	H	L	H	L	L	L	H	H	L	H	H	H
Reg	1	s		u				n	m		i	c	1
Det	C	C		C				R	R		C	C	R
Eval	○	○	☆	○	☆	☆	☆	☆	☆	☆	◎	◎	☆
CPC	ic	ic	8	83	3								
Sel.	L	L	H	L	H	L	L	L	L	L	L		
Reg			8		3								
Det			C		C								
Eval	☆	☆	◎	☆	◎	☆	☆	☆	☆	☆	☆		

Fig. 6. Example of extracting and recognizing characters [original image is that of Fig. 7(b)]. *1: Image of character pattern candidate (CPC). Black triangle at the upper left corner indicates CPC corresponding to a character. *2: Results of selecting high similarity patterns. H—Selected as high similarity pattern. L—Rejected as low similarity pattern. *3: Recognition results (the first order category). *4: Determining character pattern with relaxation. C—Character pattern. R—Rejected. *5: Evaluation of the results. For CPC corresponding to a character: ◎—correct reading (1st order), ○—correct selection, × —incorrect reject. For CPC corresponding to a noncharacter: ⋆—correct reject, △—incorrect selection.

TABLE II
SIMILARITY THRESHOLD FOR EACH CATEGORY

Category	Threshold
Il	0.85
'o'.° /\''o	0.95
Others	0.80

the 33.5% of the patterns that did not include the correct categories in their candidate categories. Five iterations were performed for (7) and (8). For example, the high similarity patterns (1)–(3) in Fig. 6, which belonged to an indirect relating pattern group, had similarity values of 0.880, 0.802, and 0.875, respectively. As a result of the relaxational operation, they were updated to 0.776, 0.912, and 0.901, and (2) was correctly extracted. In Fig. 6, notations C and R indicate extracted and rejected patterns by the relaxational approach, respectively. The evaluation of the results is indicated by ◎, ○ / ⋆ for correct selection/rejection (◎: correct recognition) and × / △ for incorrect rejection/selection. Here, only two patterns were misjudged. For the 15 images, the correct judgement rate was increased to 85.7%, but the correct extraction of characters decreased to 85.4%, as shown in Table III.

The incorrect extraction of noncharacters is mainly caused by the indirect relating pattern groups all of whose patterns correspond to noncharacters. Most of the mistaken rejections of characters occurred with Chinese characters whose direct relating patterns have simple structures and whose resolution is not high enough. It is easier to increase the similarity of simple structure categories than to increase that of degraded complex Chinese characters. Moreover, one incorrect extraction of a noncharacter pattern often caused several incorrect rejections of real character patterns in the same group. These errors

TABLE III
RESULTS OF CHARACTER PATTERN EXTRACTION

Character Pattern Candidate (CPC)[a]		The Number of CPC	Correct Judgment	
			Without Relaxation	With Relaxation
What the CPC Correspond To	Character	213	199 (93.4%)	182 (85.4%)
	Noncharacter	2951	2051 (69.5%)	2531 (85.8%)
Total		3164	2250 (71.1%)	2713 (85.7%)

[a] CPC: Character pattern candidate.

will be reduced if more precise information, such as the positions of character segments and holes in each category, is used for interactions (A) and (B) in Section IV.

VI. CONCLUSION

A method for recognizing characters in scene images has been presented. The main experimental results using 100 scene images are as follows.

1) By an image segmentation method based on local thresholding and an evaluation of the gray-level differences between regions and their back ground, all character segments in 95 images were correctly detected as character candidate regions.

2) The character recognition process made the correct selection/rejection judgement for 71.1% of the 3164 character pattern candidates, and 93.4% of the 213 characters were correctly extracted. The accumulated character recognition rate at the 10th order was 66.5%.

3) The relaxational operation using the interactions between candidate categories of related patterns increased the correct judgment rate to 85.7%.

The method developed in this correspondence can be employed to read variously formatted characters with unknown size, font, and gray-level under uncontrolled lighting. It appears promising for application in various robotic systems dealing with scene images. Remaining problems include the development of an algorithm that can deal with characters with 3-D distortions such as slant, tilt, and shape of objects on which the characters are printed.

ACKNOWLEDGMENT

The authors would like to thank Dr. R. Matsuda of SONY Corp, S. Yamazaki of Nitsuko Corp., K. Komori of NTT-IT Corp., Dr. H. Yasuda, Dr. Y. Kobayashi, T. Sakai, and Dr. K. Ishii of NTT Human Interface Laboratories, and T. Kishimoto of Engineering Strategy Planning Headquarters of NTT for their support in this work.

REFERENCES

[1] T. Kawatani, T. Tsutsumida, S. Akamatsu, and N. Nakajima, "A new precise recognition method for handprinted Kanji," in *Proc. CVPR'85*, 1985, pp. 538–543.

[2] I. Masuda, N. Hagita, T. Akiyama, T. Takahashi, and S. Naito, "Approach to smart document reader system," in *Proc. CVPR'85*, June 1985, pp. 550–557.

[3] Y. Nakamura *et al.*, "An optical character recognition system for industrial applications: TOSEYE-1000," in *Proc. Int. Workshop Industrial Applications of Machine Vision and Machine Intelligence*, 1987, pp. 364–368.

[4] M. Takatoo *et al.*, "Gray scale image processing technology applied to vehicle license number recognition system," in *Proc. Int. Workshop Industrial Applications of Machine Vision and Machine Intelligence*, 1987, pp. 76–79.

[5] L. C. Howington, "Automated license plate reading," *Advanced Imaging*, vol. 4, no. 8, pp. 46–50, Sept. 1989.

[6] H. Hoshino *et al.*, "A method for extracting specific characters from complex background," in *Nat. Convention Rec.*, IECE Japan, vol. 1152, 1983 (in Japanese).

[7] K. Funakoshi *et al.*, "Study of character extraction from scene images," in *Nat. Convention Rec.*, IEICE Japan, vol. 1503, 1987 (in Japanese).

[8] A. Perez and R. Gonzalez, "An iterative thresholding algorithm for image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 6, pp. 742–751, Nov. 1987.

[9] K. V. Mardia and T. J. Hainsworth, "A spatial thresholding method for image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 6, pp. 919–927, Nov. 1988.

[10] T. Taxt, P. J. Flynn, and A. K. Jain, "Segmentation of document images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 12, pp. 1322–1329, Dec. 1989.

[11] J. Ohya and A. Shio, "A study of recognizing characters in scene images," *Trans. IEICE Japan*, vol. J71-D, no. 6, pp. 1037–1047, June 1988 (in Japanese).

[12] J. Ohya, A. Shio, and S. Akamatsu, "A relaxational extracting method for character recognition in scene images," in *Proc. CVPR'88*, June 1988, pp. 424–429.

[13] C. K. Chow and T. Kaneko, "Automatic boundary detection of the left ventricle from cineangiograms," *Comput. Biomed. Res.*, vol. 5, pp. 388–410, 1972.

[14] S. W. Zucker, "Region growing: Childhood and adolescence," *Comput. Graphics Image Processing*, vol. 5, no. 3, pp. 382–399, 1976.

[15] N. Otsu, "A threshold selection method from gray-scale histograms," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-9, no. 1, pp. 62–66, 1979.

[16] J. Ohya and A. Shio, "Detection of character candidate regions from scene images," in *Nat. Convention Rec.*, IECE Japan, vol. S13-1, 1986 (in Japanese).

[17] M. Umeda, "Recognition of multi-font printed Chinese characters," in *Proc. 6th Int. Conf. Pattern Recognition*, 1982, pp. 793–796.

[18] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, no. 6, pp. 420–433, 1976.

A Tight Upper Bound on the Bayesian Probability of Error

W. A. Hashlamoun, P. K. Varshney, and V. N. S. Samarasooriya

***Abstract*—In this paper, we present a new upper bound on the minimum probability of error of Bayesian decision systems. This new bound is continuous everywhere and is shown to be tighter than several existing bounds such as the Bhattacharyya and the Bayesian bounds. Numerical results are also presented.**

***Index Terms*—Ali–Silvey distance measures, Bayesian decision systems, divergence, minimum probability of error, probability of error bounds, statistical pattern recognition.**

I. INTRODUCTION

In statistical pattern recognition, system performance is expressed in terms of the probability of error (the probability of misrecognition or the probability of misclassification). However, it is often difficult

Manuscript received August 29, 1991; revised October 25, 1992. This work was supported by the Rome Laboratory under Contract F30602-89-C-0082. Recommended for acceptance by Editor-in-Chief A. K. Jain.

W. A. Hashlamoun is with the Department of Electrical Engineering, Birzeit University, Birzeit, P.O. Box 14, West Bank—Via Israel.

P. K. Varshney and V. N. S. Samarasooriya are with the Department of Electrical and Computer Engineering, Syracuse University, Syracuse, NY 13244.

IEEE Log Number 9212244.