A New Morphological Segmentation Method Applied for Character Extraction

Lixu Gu , Toyohisa Kaneko
Toyohashi University of Technology
1-1, Hibarigaoka, Tempaku-cho, Toyohashi 441,Japan
E-mail: ko(or kaneko) @mmip.tutics.tut.ac.jp

Naoki Tanaka
Kobe University of Mercantile Marine
1-1, 5-Chome, Fukae-minami-machi, Higashinada-ku, KOBE 658, Japan
E-mail: tanaka@athena.ti.kshosen.ac.jp

R.M.Haralick
Dept. Elec.Eng. University of Washington, Sealtle,WA98195,USA.
E-mail: haralick@george.ee.washington.edu

Abstract

Mathematical morphology provides us the theory and tools to capture geodetic information. Hence structure segmentation and shape representation have been popularly operated by morphological approaches. A primary morphology-based approach — top-hats transformation (TT) provides an excellent tool for extracting bright or dark objects from uneven background. But to many complicated segmentation problems, TT alone cannot provide satisfactory solutions. This paper describes a new morphological segmentation algorithm which derived from TT idea. When a series of structure elements with various sizes have been employed in our study, the gray-scale particle segmentation for some complex cases is realized successfully in an iterative manner. In addition, a complete system for extraction of characters from color scene images using our algorithm is presented. 486 characters out of 534 characters included in 12 source images are correctly extracted indicating a 91% extraction rate.

1 Introduction

Image segmentation via the methods of mathematical morphology may be divided into two categoris: texture segmentation and particle segmentation[1]. Texture segmentation partitions an image into several regions (micro-images) with particular features. These micro-image structures are usually viewed in terms of small texture primitives. Its application areas include segmentation of document images; medical image processing and aerial photograph analysis.

Particle segmentation is concerned with images of particles (objects). Interested objects should be extracted efficiently with defined conditions but extracted noise regions need to be removed consequently. A traditional approach called top-hats transformation (TT), which was originally proposed by Meyer [2], provides an excellent tool for extracting bright or dark objects from gray-scale images. Investigation of image skeleton and shape representation [3]–[6] provides many good ideas in morphological iterative operations. Recently considerable research efforts have been also addressed to the watershed analysis [7].

This paper presents a new approach for particle segmentation based on TT idea and morphological iterative operation. We apply this approach to deal with a character extraction problem for color scene images.

A scene image contains complicated information. The information carried by characters is considered to be the most essential. A number of researchers have noted the importance of this field and there have been several studies for character extraction[8]–[13], which are optimized for extraction of characters, but are limited to simpler background or other restrictive conditions.

Characters in scene images are extracted based upon the following characteristics:

1. They are uniformly brighter or darker against their complicated backgrounds.

2. They are regularly arranged together with various sizes, shapes and directions.

The whole procedure includes two distinct stages:

(a) Primary processing (particle segmentation)

(b) Extraction processing
- feature emphasis
- character extraction
- noise reduction

In the first stage, a new particle segmentation filter based on TT is implemented. This filter decomposes a gray-scale input image into a set of subimages according to the size of characters. In the second stage, we first employ a new morphological filter to emphasize characters' features in the subimages and remove most noises out of them, and then, the characters are extracted directly from the gray-scale subimages by a histogram method. Lastly, a morphological image refinement algorithm based on morphological conditional dilation is introduced to make the extracted character region distinct from noises. The resulting subimages are composed together to generate the final result in binary.

The organization of this paper is as follows: In section 2, after terminologies and notations are defined, a brief overview of typical morphological operations and definition of structure elements are presented. Section 3 describes the new algorithm based on TT. The details of our character extraction implementation for color scene images are presented in section 4. Experimental results are discussed in section 5, and the concluding remarks in section 6.

2 Conception and Notation

2.1 Morphological operation

Mathematical morphology is a set theoretical methodology for image analysis[14]–[15]. Images are considered as sets of points where set operations can be performed. By convention, we let $X-Q$, $X \oplus Q$, $X \ominus Q$, $X \circ Q$ and $X \bullet Q$ denote the set difference, dilation, erosion, opening and closing in binary operation, respectively, where X and Q denote image and structure element sets in Euclidean space R^2. In this paper, we denote E as a subset of Q.

Definition 1: The recursive dilation is defined as

$$X \overset{i}{+} E = \begin{cases} \{0\} & if\ i = 0 \\ \overbrace{((X + E) \cdots + E)}^{i} & if\ i = 1, 2, \cdots \end{cases} \tag{1}$$

Definition 2: The conditional dilation is defined as

$$(X \overset{i}{+} E)_c = (X \overset{i}{+} E) \bigcap C, \tag{2}$$

where C denotes a mask and the index i is the smallest one satisfying:

$$(X \overset{i}{+} E)_c = (X \overset{i-1}{+} E)_c \tag{3}$$

Definition 3: The morphological gray-scale operations can be extended from the binary operations(See details in [14]–[15]). In this paper we let $\oplus_g, \ominus_g, \circ_g, \bullet_g$ denote gray-scale dilation, erosion, opening and closing, respectively.

Definition 4: The top-hats transformation denoted by $T_i^{(i)}$ and defined as

$$T_i^{(i)} = \begin{cases} T_i & if\ WTT \\ T^i & if\ BTT \end{cases} \tag{4}$$

$$T_i = X_0 - X_0 \circ_g E \tag{5}$$

$$T^i = X_0 \bullet_g E - X_0, \tag{6}$$

The gray-scale original image X_0 opened by a structure element E can remove the bright areas which cannot contain the structure element, and subtracting the opened image from the original one yields an image where the bright objects clearly stand out. The transformation described in (5) is called "white top-hat" transformation (WTT). A closed original image in gray-scale subtracting original one allows us to extract dark objects from bright background, which is called "black top-hat" transformation (BTT) and defined in (6). An example of WTT is shown in Fig.1.

Definition 5: The threshold operation is denoted by $|D|_B$ and defined as

$$|D|_B = \begin{cases} 1 & if\ D(p) \geq B,\ p \in D \\ 0 & if\ D(p) < B,\ p \in D \end{cases} \tag{7}$$

2.2 Structure elements

Two standard shaped structure elements DISK and LINE are selected in our method. A DISK structure element is employed to detect objects by scale when a LINE structure element is utilized to detect edges in different directions. They are described in the following:

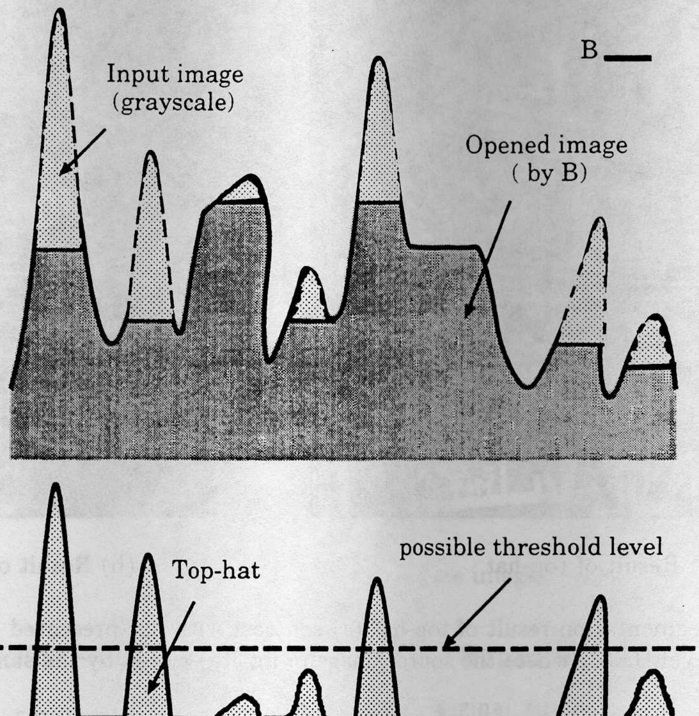


Figure 1: Peak regions extracted by the WTT and thresholded by possible gray lever

- A DISK structure element with its origin at the center and radius i is denoted by $r_i E_{disk}(r_i E_{disk} \in Q)$.
- A LINE structure element with sizes of $4i \times i$ (length $\times$ width) and a α angle direction is denoted by $L^{\alpha}_{4i\times i}(L^{\alpha}_{4i\times i} \in Q)$.

A set of particular structure elements which are derived from a 3×3 region of support in a recursive manner will be used in our proposed segmentation algorithm. For a given simplest structure element E, which may be a disk, a square, a triangle (in the Euclidean space R^2), a sphere, or a cube (in the Euclidean space R^3), etc.. A set of structure element $Q_i(Q_i \in Q)$ is defined by

$$Q_i = r_i E \tag{8}$$

When r_i is an integer, (8) is equivalent to the following relation, if E is bounded and convex:

$$Q_i = E \oplus E \oplus \cdots \oplus E \quad (i\ times), \tag{9}$$

which had been proven by I.Pitas and A.N. Venetsanpoulos[3].

With our apporach, the structure element $r_i E_{square}$ and $r_i E_{disk}$ will be obtained according to the equations (8) and (9). $r_i E_{square}$ can be directly produced in recurrence due to its simplity,

$$r_i E_{square} = r_1 E_{square} \oplus r_1 E_{square} \oplus \cdots \oplus r_1 E_{square} \quad (i\ times) \tag{10}$$

but $r_i E_{disk}$ has to be dealt by $r_1 E_{square}$ and $r_1 E_{rhombus}$ as

$$r_2 E_{disk} = r_1 E_{rhombus} \oplus r_1 E_{square}, \tag{11}$$

because it is not a simple structure element just derived from a 3×3 element. The i times enlarged disk $r_i E_{disk}$ is denoted as

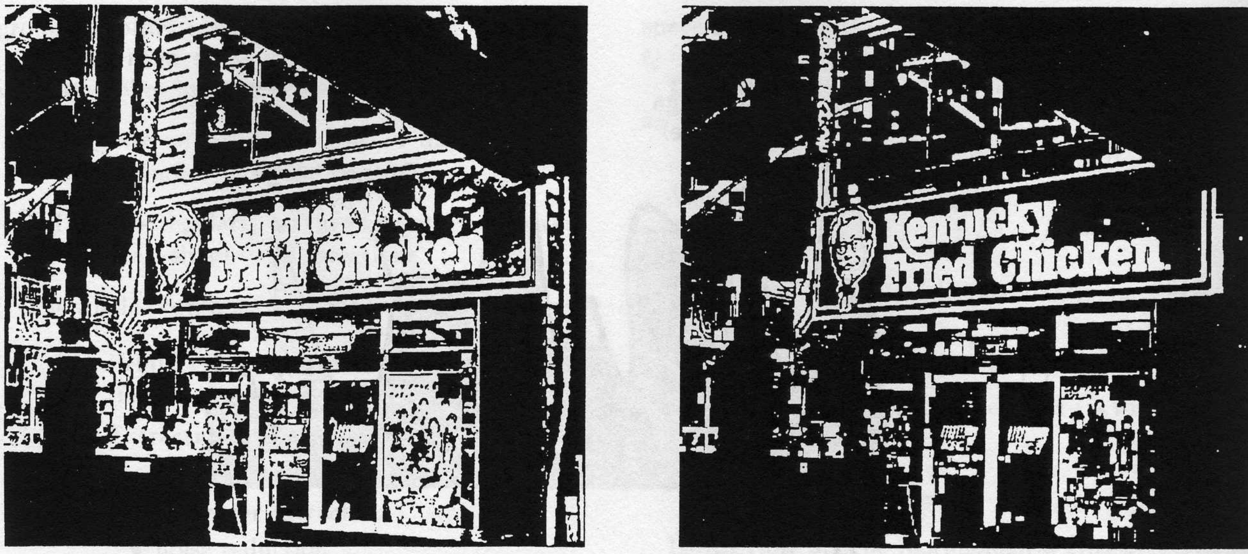
$$r_i E_{disk} = r_{i-1} E_{disk} \oplus M \quad (i\ times) \tag{12}$$

$$M = \begin{cases} r_1 E_{square} & if\ i = 2n \\ r_1 E_{rhombus} & if\ i = 2n-1 \end{cases} \quad (n \geq 1)$$

3 Segmentation Algorithm

3.1 Implementation of segmentation algorithm

For some complicated images, especially those in which the target objects are combined in a non-



(a) Result of top-hat (b) Result of presented algorithm

Figure 2: The segmentation result of top-hat (a) contrast with our presented algorithm (b) thresholded by the same gray level B=31, where the source image in fig.3(a) closed by the same structure element $r_6 E_{disk}$.

uniform background, it's difficult to segment interested particles satisfactorily. Clues for detecting features were discovered when we applied TT with different sizes of disk structure elements. The difference between $T_i^{(i)}$ and $T_{i-1}^{(i-1)}$ includes our interested objects, and that image can be easily thresholded to make features stand out.

$$\begin{aligned} X_i &= |T_i^{(i)} - T_{i-1}^{(i-1)}|_B - X'_{i-1} \\ X'_i &= \bigcup_{1 \le j \le i} X_j, \quad X'_1 = \emptyset \end{aligned} \tag{13}$$

where the original image is denoted by X_0, and segmented images which hold different sizes of objects are denoted by X'_i. The differences between the neighboring TT results up to i are united together in X'_i with a certain size of features.

An example of the performance of this segmentation algorithm contrasted with TT is shown in Fig.2.

3.2 Computational complexity

The computational cost of the segmentation algorithm is measured by counting the required number of set unions, set translations and set differences. We assume implementation on special purpose image processing hardware, where the morphological operation can be implemented with a series of shift and arithmetical operations, and an entire image can be processed in parallel. Also we assume that the dilation and erosion of a set by a structure element B requires $2(Card(E)-1)$ operations, where Card(.) denotes the cardinality of a set, and E is a 3×3 structure element. When we consider dilation of a set X by E as $X \oplus E = \cup_{E \in E}(X)_E$, the dilation requires $(Card(E)-1)$ set translations and $(Card(E)-1)$ set unions. The same conclusion can be found in erosion operation.

The opening and closing operation in (11) can be converted to another statement implemented only by dilation and erosion for speed up purpose:

$$\begin{aligned} X_0 \circ r_i E &= (X_0 \ominus r_i E) \oplus r_i E \\ &= (X_0 \ominus (\overbrace{E \oplus E \oplus E \cdots \oplus E}^{i-1})) \\ &\quad \oplus(\overbrace{E \oplus E \oplus E \cdots \oplus E}^{i-1}) \\ &= ((\ldots(\overbrace{X_0 \ominus E \ominus E \cdots \ominus E}^{i}) \\ &\quad \overbrace{\oplus E) \oplus E) \cdots \oplus E)}^{i} \\ &= (X_0 \overset{i}{\ominus} E) \overset{i}{\oplus} E \end{aligned} \tag{14}$$

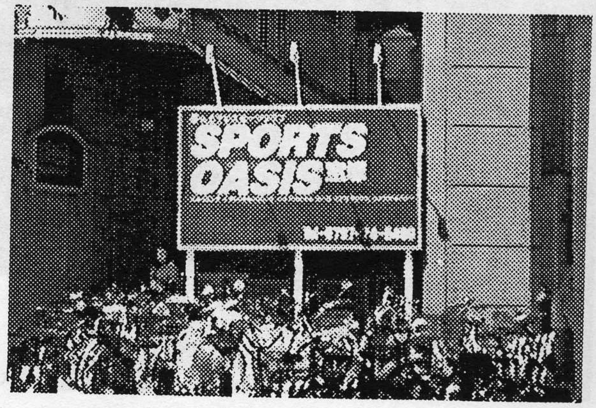
Therefore, the cost of $T_i^{(i)}$ is

$$O_{T_i^{(i)}} = 4Si(Card(E)-1)+1, \tag{15}$$

where S denotes the area of X_0. And

$$O_{X_i} = O_{T_i} + O_{T_{i-1}} + 3 \tag{16}$$

(a) P-1



(b) P-2

Figure 3: Examples of source images

$$= 4S(Card(E)-1)(2i-1)+5$$

As a result, the cost of the whole algorithm is

$$O_{X'_i} = \sum_{i=1}^{N} 4S(Card(E)-1)(2i-1)+5(N-1), \quad (17)$$

where N is the maximum of i, which is related to the maximal size of studying objects.

The worst case takes place when the 3×3 structure element E is a square. $Card(E) = 9$ and equation (16) is equivalent to the following relation:

$$O_{X'_i} \leq 32SN^2 + (32S+5)N - (32S+5) \quad (18)$$

The first term($32SN^2$) of this equation is the most essential part which shows that the computational complexity is proportional to the size of source images and the square of the maximal size of studying objects.

4 Extraction of Characters from Color Scene Image

The characters in scene images are intricate with a complicated background, and show various sizes, shapes and directions. This situation makes the extraction a difficult task. In this section, a complete system to extract characters from scene cover images using the presented segmentation algorithm is described.

4.1 Segmentation processing

The original color image is transformed to RGB images and the three gray-scale images are implemented independently. The source image is difficult to deal with in a general view. Thus we decompose it into simpler ones in this processing stage[16].

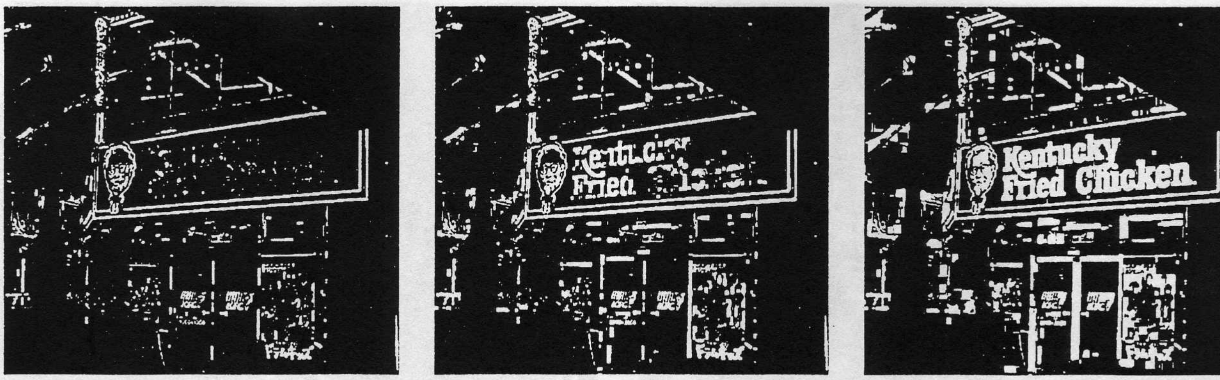
$$\begin{aligned} X_{ki} &= |T_{ki}^{(ki)} - T_{k(i-1)}^{(k(i-1))}|_B - X'_{k(i-1)} \\ X'_{ki} &= \bigcup_{1\leq j\leq i} X_{kj}, \quad X'_{k1} = \emptyset \\ T_{ki} &= X_{k0} - X_{k0} \circ_g r_i E_{disk} \qquad (19) \\ T^{ki} &= X_{k0} \bullet_g r_i E_{disk} - X_{k0} \\ X_0 &\Rightarrow X_{V0} = \{X_{R0}, X_{G0}, X_{B0}\}; \\ V &= \{R, G, B\}, \quad k \in V, \end{aligned}$$

where X_0, X_{k0} denote a source color image and RGB gray-scale images.

This processing stage starts with $r_1 E_{disk}$ and a series of subimages X'_{ki} are generated in a recursive manner. The WTT T_{ki} extracts bright areas from a dark background when the BTT T^{ki} does the opposite operation. For 256 gray level source images, the threshold operation is implemented by a value of $B = 31$ which was determined experimentally. The processing is stopped when $T_{kj}^{(kj)} = T_{k(j-1)}^{(k(j-1))}$ holds. The result of segmentation of "P-1" (Fig.3(a)) is shown in Fig.4.

4.2 Extraction processing

The decomposed subimages are processed by a morphological filter to emphasize the character region



(a) $i = 2$ (b) $i = 3$ (c) $i = 6$

Figure 4: Examples of segmented subimages

and suppress the small islands of noises by the operation:

$$F_{ki} = \begin{cases} ((X'_{ki} \circ r_{i-1}E_{disk}) \\ \oplus r_{i+1}E_{disk}) \\ \circ r_{2i}E_{disk} & (i \leq 10) \\ (X'_{ki} - X'_{k10}) & (i > 10) \end{cases} \tag{20}$$

Since the character regions are the main component in F_{ki}, they hold the peak values in the histogram. The peak values which are bigger than the average of all peak values are searched and $F_{ki} \times X_{m0}$ (see (21)) is thresholded by the selected peak values to extract characters standing out as:

$$H_{ki} = \bigcap_{m \in V \cap k^c} |F_{ki} \times X_{m0}|_B, \tag{21}$$

where, " $\times$ " denotes an arithmetic multiplication between two same size images for transforming binary image to gray-scale. The noise region attaching with the character candidate region may not hold the same gray level in all the three RGB images. This characteristic is used to distinguish the characters from the noises.

4.3 Refinement processing

The extracted characters in H_{ki} are broken and there remain some noises. A morphological filter derived from conditional dilation is implimented to refine the character candidate regions to get satisfactory results, where the gray level B is same with the extraction processing stage.

$$\begin{aligned} R_{ki0} &= ((H_{ki} \circ L^{0}_{i\times 4i}) \cap (H_{ki} \circ L^{\pi/2}_{i\times 4i})) \\ &\quad \cup ((H_{ki} \circ L^{\pi/4}_{i\times 4i}) \cap (H_{ki} \circ L^{3\pi/4}_{i\times 4i})) \\ R_{kin} &= (R_{ki0} \overset{n}{\oplus} r_5 E_{disk}) \cap |X_{k0}|_B \\ if\ R_{kin} &= R_{ki(n-1)}\ then\ stop \end{aligned} \tag{22}$$

Preliminarily, four direction lines are made use of to detect character edges, and then broken character regions are reconstructed by conditional dilation where the source image thresholded by the same gray level B with extraction processing stage is utilized as mask.

Finally, subimages R_{kin} are united to obtain the resulting image X_r, which is denoted by

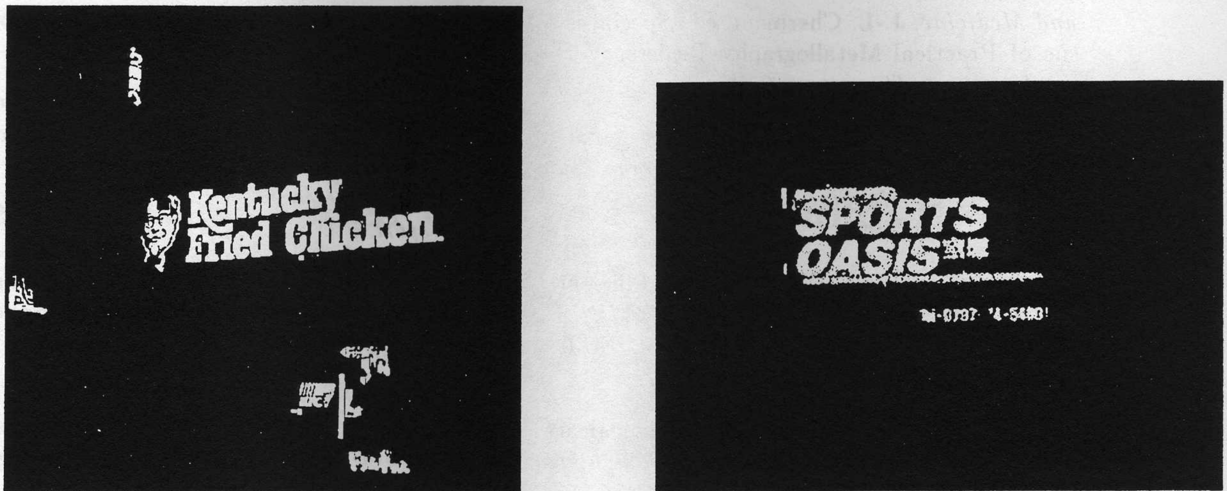
$$X_r = \bigcup_{\substack{1 \leq i \leq j \\ k \in \{R,G,B\}}} R_{kin} \tag{23}$$

5 Experimental Results

Color scene photos with various complicated background are used in our experiment to demonstrate the performance of our system. They are scanned in 75dpi with different sizes and computed by SUN SPARC STATION 20.

We concentrate on various scene photographs with a complicated background which are thought to be a typical representative of scene images. The two examples of 12 experimental images are shown in Fig.3.

As the result, 486 characters out of 534 characters are correctly extracted, indicating a 91% success rate. Two resultant examples shown in Fig.5 appear encouraging. In the scene image "P-1" (Fig.5(a)), 34 characters extracted from total 42 characters, showing a 83% extraction rate(ERA), when extracted characters(ECH) of "P-2" (Fig.5(b)) are 72 from total 74 characters with a 97% ERA. There is a same problem in the experiment. The extra small characters(ESM) can't be extracted so far. If the source images were scanned in a high resolusion, a 99% ERA could be obtained. The results of six examples are shown in Table 1.



(a) P-1

(b) P-2

Figure 5: Examples of result images

Table 1: The results of characters extraction

source	sort	totals	ESM	ECH	ERA	ERE	ERR
P-1	photo	42	5	34	83%	4	9%
P-2	photo	74	2	72	97%	1	1%
P-3	photo	32	4	28	88%	2	6%
P-4	photo	49	3	46	94%	3	6%
P-5	photo	22	3	19	89%	2	9%
P-6	photo	45	4	41	91%	2	5%

It's not easy to estimate the extraction error precisely because the error extracted regions are different in sizes and difficult to calculate exactly. In this paper, we introduce an "error extracted region"(ERE) to evaluate the performance of our approach in opposition. The "error extraction rate"(EER) appears the same mean in percent. The detail is decribed in Table 1. It is seen that 6% average EER was obtained with our method.

6 Conclusion

In this paper, a new approach for particle segmentation using mathematical morphology was presented. We described a new idea based on the TT and recursive morphological operations, analysed its computation complexity, and then related a complete system to extract characters from scene images. The new method can deal with more difficult segmentation problems than other known algorithms, especially with the objects existing in a complicated background with irregular sizes and directions. Furthermore, it also can be extended to three dimensional applications.

The proposed approach for detecting characters in scene images was found to be efficient in our experiment. We intend to proceed with our studing on character extraction from scene images to improve its accuracy, and to utilize it to other applications. We will suggest that a characters recognition system from scene images can be built with our approach.

References

[1] L.Vincent and E.R.Dougherty, "Morphological segmentation for textures and particles", *Digital Image Processing Methods*, pp. 43-102, Marcel Dekker, 1994.

[2] F.Meyer,"Contrast Feature Extraction", *Quantitative Analysis of Microstructures in Material Sciences, Biology*

and Medicine, J.-L. Chermant, ed., Special issue of Practical Metallography, Riederer Verlag, Stuttgart, Germany, 1978.

[3] I.Pitas and A.N.Venetsanpoulos, "Mathematical shape decomposition", *IEEE Trans. on Pattern Anal. Machine Intell.*, Vol. 12, No.1, pp.38–45, Jan., 1990.

[4] J.M.Reinhardt and W.E.Higgins, "Efficient Morphological Shape Representation", *IEEE Trans. on Image Processing*, Vol.5, No.1, pp.89–101, Jan., 1996.

[5] X.Zhang and R.M.Haralick, "Morphological structure element decomposition", *Computer Vision, Graphics*, Image Processing, Vol.35, pp.370–382, Sept. 1986.

[6] Hochong Park and Roland T.Chin, "Decomposition of arbitrarily shaped morphological structure elements", *IEEE Trans. on Pattern Anal. Machine Intell.*, Vol.17, No.1, pp.2–15, Jan. 1995.

[7] L.Vincent and P.Soille, "Watersheds in Digital Spaces:An Efficient Algorithm Based on Immersion Simulations", *IEEE Trans. on Pattern Anal. Machine Intell.*, Vol.13, No.6, pp.583-598, June, 1991.

[8] S.Liang and M.Ahmadi, "A Morphological Approach to Text String Extraction from Regular Periodic Overlapping Text/Background Images", Computer Vision, Graphics, Image Processing, Vol.56, No.5, pp.402-413, Sep., 1994.

[9] M.-Y.Chen, A.Kundu and S.N.Srihari, "Variable Duration Hidden Markov Model and Morphological Segmentation for Handwritten Word Recognition", *IEEE Trans. on Image Processing*, Vol. 4, No. 12, pp.1675-1688, Dec., 1995.

[10] M.Okamoto and H.Hayashi, "Character Extraction from Headlines with Background Patterns by Using Shrinking/Expanding Methods", *IEICE Technical Report*, PRU90-151, pp.47-54, 1991.

[11] J.Ohya, A.Shio and S.Akamatsu, "Recognizing character in scene images", *IEEE Trans. on Pattern Anal. Machine Intell.*, Vol.16, No.2, pp.214-220, 1994.

[12] A.Ueba, T.Takeda and Y.Okada, "The extraction of character images from color images based on processing if color contour line", *IEICE Technical Report*, PRU94-28, pp.9-16, 1994.

[13] S.Shuji, M.Michihiko, and I.Katsuo, "A method of extraction of character pattern from a color image based on the uniformity of the character color of a string", *IEICE Technical Report*, PRU94-29, pp.17-24, 1994.

[14] R.M.Haralick, S.R.Sternberg, and Zhang, "Image analysis using mathematical morphology", *IEEE Trans. on Pattern Anal. Machine Intell.*, Vol.9, No.4, pp.532–550, July, 1987.

[15] R.M.Haralick and L.G.Shapiro, *Computer and Robot Vision*, Addison-Wesly Publishing, 1992.

[16] L.Gu, N.Tanaka, T.Kaneko and R.M.Haralick, "The Extraction of Characters from Cover Images Using Mathematical Morphology", *The Trans. of IEICE D-II*, Vol.J80-D-II, No.10, pp.183-192, Oct., 1997.