Determining Optimal Filters for Binarization of Degraded Grayscale Characters Using Genetic Algorithms

Yusuke Ojima, Satoshi Kirigaya, and Toru Wakahara
Faculty of Computer and Information Sciences
Hosei University
3-7-2 Kajino-cho, Koganei-shi, Tokyo, 184-8584 Japan
E-mail: wakahara@k.hosei.ac.jp

Abstract

Optimal binarization of degraded grayscale characters is a crucial step to subsequent character recognition. This paper proposes a new, promising binalization technique of grayscale characters using genetic algorithms (GA) to search for an optimal sequence of filters from among a set of rather simple, representative image processing filters. First, we classify degraded samples of grayscale characters into several categories. Then, in the learning stage, by selecting a training sample from each degradation category we apply GA to the combinatorial optimization problem of determining a sequence of filters that maximizes the fitness value between the filtered training sample and its target image ideally binarized by humans. Finally, in the testing stage, we apply the optimal sequence of filters thus obtained to remaining test samples for each degradation category. Experiments using the public ICDAR 2003 robust OCR dataset demonstrate promising results of binarization of grayscale characters against a wide variety of degradation causes.

1. Introduction

Recognition of characters captured both in indoor and outdoor scenes is very challenging and is most expected to play an essential role in intelligent robot realization. Most of recognition difficulties are due to the presence of complex background, variations of lighting conditions, and a variety of degradations in captured images. In particular, separation of character and background is the key to the success of subsequent recognition. This is a typical segmentation or *binarization* problem.

There have been many intensive researches aiming at optimal binarization of degraded printed documents [1], [2]. Baird's work on image defect models [3] was pioneering in analyzing degraded documents and might be extended to analysis of degraded characters in natural scenes. Also, we have sophisticated filtering techniques for image enhancement or restoration such as Fast Fourier Transform (FFT), Gabor filters, and wavelet transforms. However, due to uncontrolled environment in natural scenes, it might be very difficult for a single filtering technique to deal with a variety of image degradations even if controlling parameter values are optimally chosen. On the other hand, if we search for a combination of several filtering techniques, we must be faced with a difficult problem of combinatorial explosion.

To resolve this problem, our paper proposes a new binarization technique of degraded grayscale characters in natural scenes. Nagao [4], [5] used genetic algorithms (GA) [6] to construct an optimal sequence of image processing filters for problems of extracting characters from maps, cell walls from micrographs, skin regions from face images, etc. We apply an extended version of Nagao's works to the very binarization problem.

We use the public ICDAR 2003 robust OCR dataset [7] containing images of single characters in natural scenes. We classify, not automatically at present but manually, a variety of samples subject to degradation into several categories. Then, we determine an optimal sequence of filters by GA that maximizes the fitness value between a representative training sample for each degradation category and its target image ideally binarized by humans. Finally, the thus determined optimal sequence of filters is applied to binarization of test samples belonging to the same degradation category.

Section 2 explains ICDAR 2003 character dataset. Section 3 describes categorization of character images in natural scenes for binarization. In Section 4, we describe how to determine an optimal sequence of filters for binarization using GA. Section 5 shows promising experimental results.

2. ICDAR 2003 character dataset

Several datasets used in ICDAR 2003 robust reading competitions are available for download from the website [7]. We use the robust OCR dataset containing JPEG images of single characters in natural scenes.

In particular, to tackle the problem of binarization of degraded grayscale characters with nonuniform lighting and/or complex backgrounds we selected a total of 970 images from the robust OCR dataset. All color images in

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

the JPEG format were converted to grayscale images in the PGM format with 256 gray levels.

3. Categorization of character images in natural scenes for binarization

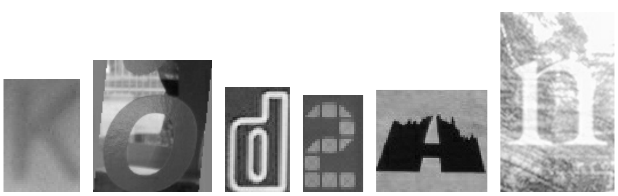
A total of 970 selected images are subject to an intensive degree of image degradation and background complexity. By carefully examining those images we classified them into six categories. The criterion of the categorization at present is rather subjective just to show a wide variety of binarization problems. Of course, for a practical application categories should be selected automatically, and also an automatic method is necessary to decide which category the input image belongs to.

Table 1 shows the result of the manual sorting of 970 images into six categories for binarization.

Figure 1 shows examples in each of six categories.

Table 1. Categorization of character images for binarization.

Category	Number of images
extreme blurring	298
background with pattern	179
character with rims	143
character with pattern	137
partial defects	113
nonuniform lighting	100
total	970



(a) (b) (c) (d) (e) (f)

Figure 1. Examples of character images. (a) Extreme blurring. (b) Background with pattern. (c) Character with rims. (d) Character with pattern. (e) Partial defects. (f) Nonuniform lightning.

4. Determination of an optimal sequence of filters for binarization via GA

We try to determine the optimal filter sequences for each category of images based on the two key ideas.

The first one is use of a filter bank composed of a number of rather simple, representative image processing filters. It is clear that our heuristics on how to combine these filters for optimal binarization has its limitations. Also, exhaustive trial and error in search of an optimal sequence of filters is impractical. Hence, the second key idea is application of genetic algorithms (GA) [6] to this combinatorial optimization problem. Here, in GA application, a chromosome encodes a sequence of filters selected from a filter bank where any filter can be called multiple times. As a result, each chromosome has a variable length.

Figure 2 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.

In Fig. 2, we use GA in search of an optimal sequence of filters, equivalent to the image transformation L^*, that approximates the ideal image transformation, denoted by L. The degree of approximation is evaluated in terms of the fitness value calculated between the target image and the filtered image.

First, we prepare a typical training sample of degraded character image and its target image ideally binarized by humans. Then, GA determines an optimal sequence of filters that transforms the original, training image so as to maximize the value of fitness between the target image and the resulting filtered image. Finally, the optimal sequence of filters thus obtained is applied to binarization of test character images belonging to the same degradation category as that of the training sample.

The details are explained in the following subsections.

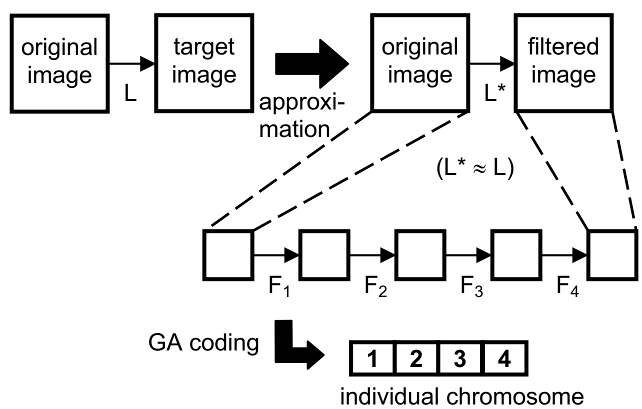


Figure 2. Total flow of image transformation by a sequence of filters.

4.1. Filter bank

A filter bank is designed to cover a wide range of rather simple, representative image processing filters. We excluded sophisticated filters deliberately.

Table 2 shows a list of filters in our filter bank.

We expanded the list used in Nagao's work [4]. Most of filters are not computationally demanding. It is easy to add more filters to the present filter bank. A wide variety of filters in the filter bank actually enhance the ability of

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

image transformation while the learning time in GA application increases steadily.

Table 2. List of filters in a filter bank.

ID no.	Filter name	Function
1	Thresholding (1)	threshold T = 50
2	Thresholding (2)	threshold T = 100
3	Thresholding (3)	threshold T = 150
4	Thresholding (4)	threshold T = 200
5	Mean	local mean in a 3×3 window
6	Min	local min in a 3×3 window
7	Max	local max in a 3×3 window
8	Sobel (1)	horizontal differential
9	Sobel (2)	vertical differential
10	Sobel (3)	the norm of differential
11	LightEdge	Laplacian
12	DarkEdge	Laplacian + 255
13	Erosion	morphological erosion
14	Dilation	morphological dilation
15	Inversion	255 – g; g = pixel value
16	Subtract	subtract the original image from the filtered one
17	LightPixel	if g < total mean, g = 0
18	DarkPixel	If g ≥ total mean, g = 255
19	Median	median filter
20	Sharpen	sharpening with Laplacian
21	Otsu_th	binarization by Otsu's method
22	HighPass	Gaussian highpass filtering
23	LowPass	Gaussian lowpass filtering
24	Hilditch	thinning by Hilditch's algorithm
25	Lin_hst	histogram linear stretching
26	Eql_hst	histogram equalization

4.2. Gene encoding and specifications of GA

We use genetic algorithms (GA) [6] 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.

First, we explain gene encoding that represents a sequence of filters selected from the filter bank described in **4.1**. We specify individual filters by their respective ID numbers as shown in Table 2 where each ID number can be represented by 8-bit integer. Hence, a chromosome is encoded as a string of 8-bit integers that represents a corresponding filter sequence. Also, the total number of bits in a chromosome is variable in proportion to the number of constituent filters. Here, we set the maximum number of constituent filters at 50.

Next, specifications of GA are as follows. The initial population of 800 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 90%. Mutation also exchanges every constituent filter ID number within a chromosome for a different filter ID number with the rate of 0.01%.

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. The obtained sequence of filters generates an optimally filtered image that approximates the target image as closely as possible. This completes a process of determining optimal filters for binarization by means of GA.

4.3. Image representation and fitness value

We calculate a fitness value between a target image and all filtered images in the given generation.

First, we represent gray levels of target and filtered images by $T(x, y)$ and $F(x, y)$ at (x, y), respectively. Also, we assume two images, $\boldsymbol{T} = \{T(x, y)\}$ and $\boldsymbol{F} = \{F(x, y)\}$, have the same size of $W_x \times W_y$ in width and height.

Here, a target image, $\boldsymbol{T}$, represents a result of ideal binarization of an original image, and is carefully designed by humans as follows.

$$T(x,y)=\begin{cases} 0 & \text{if } (x,y)\in \text{character} \\ 255 & \text{if } (x,y)\in \text{background} \end{cases} \qquad (1)$$
$$1\le x\le W_x, \quad 1\le y\le W_y .$$

Then, we define a fitness value between a target image, $\boldsymbol{T}$, and a filtered image, $\boldsymbol{F}$, as an average difference of gray levels given by

$$f(\boldsymbol{T},\boldsymbol{F})=1-\frac{\sum_{x=1}^{W_x}\sum_{y=1}^{W_y} w(x,y)\left|T(x,y)-F(x,y)\right|}{\sum_{x=1}^{W_x}\sum_{y=1}^{W_y} w(x,y)\times 255}, \qquad (2)$$
$$w(x,y)=\begin{cases} 1.0 & \text{if } T(x,y)=0 \text{ (black)}, \\ 0.0 & \text{otherwise}, \end{cases}$$

where $w(x, y)$ is a weighting function that controls the matching area with variable weights. Here, we adopt the form of $w(x, y)$ that is just equal to a support of $T(x, y)$.

A chromosome with the maximum value of $f(\boldsymbol{T}, \boldsymbol{F})$ is selected as an elite one in the given generation.

This definition of a fitness value is a very simple choice. It is to be noted that a resulting sequence of filters heavily depends on how to define a fitness values, $f(\boldsymbol{T}, \boldsymbol{F})$, and a weighting function, $w(x, y)$, in GA application.

5. Experimental results

In this section, we show promising experimental results of construction of an optimal filter sequence for binarization using a filter bank and GA made on a set of 970 degraded character images described in Section 3.

5.1. Learning of optimal sequence of filters

We selected one representative training sample from a group of samples belonging to each of six degradation categories shown in Table 1.

Next, we carefully generated respective target images by handwork using Adobe Photoshop.

Figure 3 shows one example of a training pair of original ant its target images.

(a) (b)

Figure 3. Example of a pair of original and its target images belonging to "extreme blurring" category in Table 1. (a) Original image. (b) Target image ideally binarized by humans.

Then, following the procedure described in Section 4, we searched for an optimal sequence of filters for binarization as applied to the image of Fig. 3 (a) to yield the closest approximation to Fig. 3 (b).

Figure 4 shows one example of learning result of an optimal sequence of filters in the case of Fig. 3.

It is clear that the obtained optimal sequence of filters for binarization shown in Fig. 4 (a) is beyond the range of humans' ability of combining filters by experience and heuristics.

{6, 5, 13, 13, 13, 13, 14, 13, 14, 6, 6, 14, 14, 1, 7, 7, 6, 3, 13, 7, 7, 13, 5, 5, 14, 13}

(a)

(b)

Figure 4. Example of learning result in the case of Fig. 3. (a) Obtained optimal sequence of filters. (b) Binarized image of Fig. 3 (a).

5.2. Test of binarization by optimal filters

In the learning stage, we obtained all of six optimal sequences of filters for binarization corresponding to six respective image degradation categories in Table 1.

Then, in the testing stage, we apply each optimal sequence of filters to binarization of test samples belonging to the same image degradation category.

Figure 5 shows examples of binarization results of training samples vs. test samples in all of six image degradation categories.

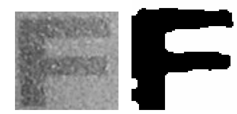
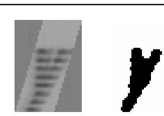
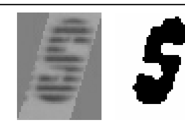
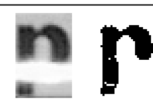
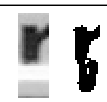
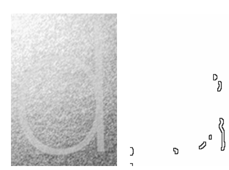
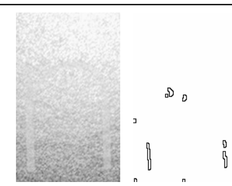
Category	Binarization of training samples	Binarization of test samples
(a)		
(b)		
(c)		
(d)		
(e)		
(f)		

Figure 5. Examples of binarization results of training samples vs. test samples. (a) Extreme blurring. (b) Background with pattern. (c) Character with rims. (d) Character with pattern. (e) Partial defects. (f) Nonuniform lighting.

From Fig. 5, it is first found that binarization of test samples is remarkably successful even if embedded characters in training and test samples are totally different in shape.

It is also found that it is very difficult to construct optimal filters for binarization when there is little contrast between character and background due to nonuniform lighting as shown in Fig. 5 (f).

In order to evaluate the ability of binarization in a more quantitative manner, we calculated a matching measure between binarizeded images by optimal sequences of filters and their respective target images ideally binarized by humans.

Here, we adopted a correlation value between binary images [8] as a matching measure. A correlation value,

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

$C(\boldsymbol{T},\boldsymbol{F})$, in binary matching between a target image, $\boldsymbol{T}$, and a filtered image, $\boldsymbol{F}$, is given by

$$C(\boldsymbol{T},\boldsymbol{F}) = \frac{n_{11}n_{00} - n_{10}n_{01}}{\left(n_{1.}n_{0.}n_{.1}n_{.0}\right)^{1/2}}, \tag{3}$$

$$\text{where } n_{11} = \#\{(x,y)\,;T(x,y)=\text{black} \;\&\; F(x,y)=\text{black}\},$$
$$n_{10} = \#\{(x,y)\,;T(x,y)=\text{black} \;\&\; F(x,y)=\text{white}\},$$
$$n_{01} = \#\{(x,y)\,;T(x,y)=\text{white} \;\&\; F(x,y)=\text{black}\},$$
$$n_{00} = \#\{(x,y)\,;T(x,y)=\text{white} \;\&\; F(x,y)=\text{white}\},$$
$$n_{i.} = n_{i1} + n_{i0}, \quad n_{.j} = n_{1j} + n_{0j}.$$

As mentioned before, we used only one training sample for each image degradation category. On the other hand, in evaluation of correlation values, we selected test samples subjected to almost the same kind of image degradation as that of training samples. The total number of test samples thus evaluated was fifty three.

Figure 6 shows relation between correlation values and image degradation categories obtained from both training and test samples against their target images. Correlation values for test samples are averaged ones.

From Fig. 6, we can confirm that optimal sequences of filters determined by GA have the marked ability to perform successful binarization of test samples subjected to image degradation categories (a), (b), and (d).

Figure 6. Relation between correlation values vs. image degradation categories. (a) Extreme blurring. (b) Background with pattern. (c) Character with rims. (d) Character with pattern. (e) Partial defects. (f) Nonuniform lighting.

6. Conclusion

Recognition of degraded character images in natural scenes is one of the most challenging topics in the arena of pattern recognition research.

We proposed a new, promising technique of determining optimal filters for binarization of degraded grayscale characters using genetic algorithms. The key idea is to search for an optimal sequence of filters selected from among a number of rather simple, representative image processing filters by means of GA. In GA application, the fitness value between a filtered image of original degraded training sample and its target image ideally binarized by humans is maximized. Then, the optimal sequence of filters thus obtained is applied to binarization of test samples.

In experiments, we used grayscale images of single characters in natural scenes converted from JPEG images in the public ICDAR 2003 robust OCR dataset. We classified degraded samples of grayscale characters into six categories manually at present.

Experimental results showed that the proposed technique successfully constructed optimal sequences of filters for binariation in respective image degradation categories. It is surprising that just a sequence of rather simple filters can approximate ideal binarization against a wide variety of image degradations.

Future work is to automatically select degradation categories and decide which category the input image belongs to. Also, it is necessary to see how the OCR performs on images binarized by the proposed method. Moreover, it is very interesting to reinforce this technique so as to deal with direct binarization of color images of degraded characters in natural scenes.

References

[1] D. Trier and A. K. Jain, "Goal-directed evaluation of binarization methods," *IEEE Tran. Pattern Anal. Machine Intel.*, vol. PAMI-17, pp. 1191-1201, 1995.

[2] L. O'Gorman, "Binarization and multithresholding of document images using connectivity," *Graphical Models and Image Processing*, vol. 56, pp. 494-506, 1994.

[3] H. S. Baird, "Document image defect models," in H. S. Baird, H. Bunke, and K. Yamamoto (Eds.), *Structured Document Image Analysis*, Springer-Verlag, New York, 1992, pp. 546-556.

[4] S. Masunaga and T. Nagao, "Automatic construction of image transformation processes using genetic algorithm," *Proceedings of International Conference on Image Processing '96 (ICIP-96)*, pp. 731-736, 1996.

[5] S. Aoki and T. Nagao, "Automatic construction of tree-structural image transformation," *Proceeding of the 1999 International Conference on Image Processing (ICIP-99)*, pp. 276-279, 1999.

[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, 1989.

[7] ICDAR 2003 Competitions, "Robust reading datasets," http://algoval.essex.ac.uk/icdar/Datasets.html.

[8] J. D. Tubbs, "A note on binary template matching," *Pattern Recognition*, vol. 22, no. 4, pp. 359-365, 1989.

1520-5263/05 $20.00 © 2005 **IEEE**