Integral Ratio: A New Class of Global Thresholding Techniques for Handwriting Images

Yan Solihin and
C.G. Leedham, *Senior Member, IEEE*

Abstract—In this paper, we propose a new class of histogram based global thresholding techniques called Integral Ratio. They are designed to threshold gray-scale handwriting images and separate the handwriting from the background. The following tight requirements must be met: 1) all the details of the handwriting are to be retained, 2) the writing paper used may contain strong colored and/or patterned background which must be removed, and 3) the handwriting may be written using a wide variety of pens such as a fountain pen, ballpoint pen, or pencil. A specific application area which requires these tight requirements is forensic document examination, where a handwritten document is often considered as legal evidence and the handwriting must not be tampered with or modified in any way. The proposed class of techniques is based on a two stage thresholding approach requiring each pixel of a handwritten image to be placed into one of three classes: foreground, background, and a fuzzy area between them where it is hard to determine whether a pixel belongs to the foreground or the background. Two techniques, Native Integral Ratio (NIR) and Quadratic Integral Ratio (QIR), were created based on this class and tested against two well-known thresholding techniques: Otsu's technique and the Entropy thresholding technique. We found that QIR has superior performance compared to all the other techniques tested.

Index Terms—Document image thresholding, background removal, preprocessing.

1 Introduction

MANY computer vision applications, such as Optical Character Recognition, Signature Verification, Automatic Document Processing, and Forensic Document Examination, process optically scanned gray scale images acquired from a flat-bed scanner or camera arrangement. The first preprocessing step is usually to threshold the image to separate the foreground (handwriting or text) from the background.

In this study, we were particularly concerned with the imaging of questioned documents for computer-assisted forensic examination. In Forensic Document Examination, sections of handwritten text extracted from documents are frequently used as evidence in a court of law. Any imaged version of the handwriting must be a true and accurate representation of the original writing or it may be dismissed as tampered evidence. Therefore, all preprocessing facilities of the scanner must be turned off and all subsequent preprocessing carried out under careful control. In this application, the role of thresholding is to remove the background as much as possible without modifying the handwriting. In particular, the thresholding algorithm must:

1. Retain all detail of the handwriting, including faint skate-on and skate-off penstrokes at the beginning and end of strokes,
2. Remove the background paper image which may contain dark colored and/or patterned background, and
3. Retain handwriting produced by a wide variety of pens such as a fountain pen, ballpoint pen, fiber-tip pen, and pencil.

The algorithm is to provide an automatic first try at background removal with later manual refinement if necessary. In order to represent all the details of handwriting, a handwritten document needs to be scanned at appropriate spatial and gray-scale resolutions. Assuming that the color information of handwriting is not important to forensic document examiners (usually that information is used, if necessary, as a separate forensic exercise together with the type and age of the paper and the ink), a 256 gray scale image is appropriate [4]. A spatial resolution of 600 × 600 dpi is also sufficient to represent the significant details of the handwriting [18].

The wide variety of writing paper frequently causes scanned gray-scale images of handwriting to contain unwanted background detail from the paper. In addition, the scanning process may also introduce noise due to the defects of the scanner or inappropriate parameter tuning. If the inking is bright (e.g., a red pen or a bright highlighter) or if it has a wide variation of brightness (e.g., a graphite pencil or ballpoint pen which produces uneven inking), then the task of separating the handwriting from the background is even more problematic.

Separation of foreground from background in a document image is commonly performed by a global thresholding technique. Global techniques are usually effective because the majority of documents have relatively constant contrast. Local adaptive thresholding may be appropriate in some instances. However, local thresholding tends to emphasize the patterns in the background when there is no handwriting present in the local region. We restrict our consideration here to global thresholding techniques. Numerous global techniques have been proposed [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. Otsu's thresholding technique [9] has been cited as an effective technique [2], [3], [7], [10], [13] and has been used in numerous document processing applications [3], [5]. In Trier and Jain's study [10], four global thresholding techniques were compared using a performance criterion based on the ability of an OCR module to recognize numbers from hydrographic images. In their study, Otsu's technique [9] performed the best, followed, in order, by Kapur et al.'s Entropy technique [6], Abutaleb's entropy technique [1], and Kittler and Illingworth's minimum error technique [7].

Although many thresholding techniques have been proposed, none were designed specifically for handwriting images. There has not been any rigorous study in comparing the performance of these techniques in thresholding handwriting images. Most comparisons consider the problem in a more general image perspective [15]. This paper presents a new class of thresholding technique called Integral Ratio and two of its techniques, Native Integral Ratio (NIR) and Quadratic Integral Ratio (QIR). The results of rigorous experimentation using several thresholding techniques on handwriting images are presented. In the experimentation, the NIR and QIR algorithms were compared against Otsu's algorithm and Kapur's entropy thresholding technique.

2 Two Stage Thresholding

A *global n-stage thresholding technique* is one that performs thresholding in n stages. The traditional thresholding approach is basically a one-stage thresholding approach where an image is separated into two classes of pixels: the foreground pixels and background pixels. Global thresholding techniques attempt to find a single threshold value T that best separates the two classes of pixels in an image. Our class of technique is a global two stage thresholding approach. In the first stage, the image is divided into three classes of pixels (instead of two): foreground, background,

• *The authors are with the School of Applied Science, Nanyang Technological University, N4-#2C-77 Nanyang Avenue, Singapore 639798. E-mail: solihin@cs.uiuc.edu, asgleedham@ntu.edu.sg.*

Manuscript received 20 Mar. 1997; revised 13 Feb. 1999.
Recommended for acceptance by M. Mohiuddin.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107548.

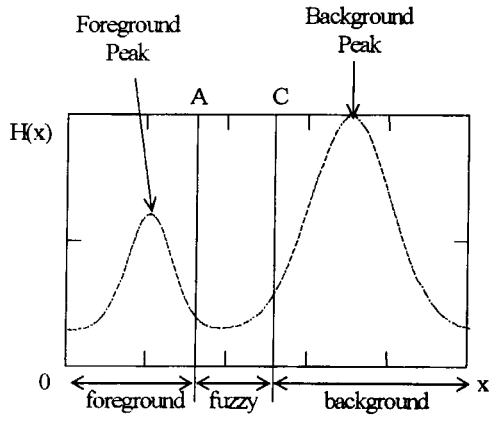

Fig. 1. Three pixel classes in the Integral Ratio thresholding approach.

and a fuzzy class where it is hard to determine whether a pixel actually belongs to the foreground or the background. Fig. 1 shows the three pixel classes where x denotes the pixel gray-scale intensity value and H(x) denotes the number of pixels in the image which have intensity x. Two important parameters that separate these classes are A, which separates the foreground and the fuzzy classes, and C, which separates the fuzzy and the background classes.

If, however, a pixel has an intensity greater than A but less than C, it belongs to the fuzzy class and more information from the image is needed to decide whether it belongs to the foreground or the background.

The first stage of the technique produces a range of threshold values delimited by the parameters A and C ($T_1 = [A, C]$). During the second stage, a final threshold value is chosen between A and C. If the parameter C is chosen as the threshold value T, there is a risk of not removing the background completely. This is a *pessimistic thresholding* scheme in that we will retain the foreground but will also retain some background. If the parameter A is chosen as the threshold value T, all the background will be removed, but there is a risk of removing some part of the foreground. This is an *optimistic thresholding* scheme. Pessimistic thresholding is usually preferred in thresholding for forensic document examination purposes because it is better to include some small parts of the background rather than losing some significant part of the handwriting. It is a fundamental weakness if some part of the handwriting is lost because it cannot be recovered. But, not clearing all the background is not a fundamental weakness since the remaining background can be removed in subsequent manual processing steps. However, the threshold value should be large enough to remove as much background as possible while retaining the foreground.

Ideally, during the second stage, the final threshold value ($T \in [A, C]$) should make use of the knowledge of the handwriting media used because the type of media determines the distribution of the intensity. For example, in general, a pencil has more intensity variation than a ballpoint pen and a ballpoint pen has more intensity variation than a fiber tipped pen. It is not uncommon to have foreground pixels with intensities very near to the background's peak for pencils and ballpoint pens during points of poor inking or low pencil pressure. The type and color of the writing implement determines the location of the peak of the foreground and the variation around the peak. However, information about the type of writing implement is not readily available and is not used explicitly in this study.

3 Integral Ratio Class of Technique

An *Integral Ratio function* is defined as a function $f : D \rightarrow R$, where $D \subset R$ and

$$f(u) = \frac{\int_{g_3(u)}^{g_2(u)} h(x)dx}{\int_{g_3(u)}^{g_4(u)} h(x)dx}, \tag{1}$$

where u is its variable and u_0, g_1, g_2, g_3, g_4, and h are its parameters. g_1, g_2, g_3, g_4 are all real functions with variable u and parameter u_0. h is a real function.

$$g_i : D_i \rightarrow R, \quad D_i \subset R, \text{and } i = 1, 2, 3, 4$$
$$h : D_h \rightarrow R, \quad D_h \subset R.$$

An *equally spaced adjacent interval Integral Ratio function* is defined as

$$f(u) = \frac{\int_{u_0}^{u_0+au} h(x)dx}{\int_{u_0+au}^{u_0+2au} h(x)dx}. \tag{2}$$

u_0 is called a pivot of g_i, $i = 1, 2, 3, 4$.

The Integral Ratio class of thresholding techniques are a collection of all two stage thresholding techniques that divide an image into three pixel classes: foreground, background, and a fuzzy class separated by parameters A and C; and estimate A and C using Integral Ratio functions.

3.1 Native Integral Ratio (NIR) Technique

The NIR technique finds the value of parameters A and C using NIR estimators. NIR estimators make use of the intensity histogram of the image to be thresholded.

Let us assume that p_f denotes the intensity value of the local maximum of foreground pixels. If the background has only one peak, let p_b represent the peak of the background. If the background has more than one peak, let p_b represent the background peak with the darkest intensity. All the images are assumed to be 256 gray scale images. The NIR technique can be used to estimate the values of parameter A and C using the relationship:

$$A = p_f + \underset{u=1..(p_b-p_f)/2}{\arg\max} f(u) - 1, \tag{3}$$

where f(u) is an NIR estimator of the form:

$$f(u) = \frac{\sum_{x_i=p_f}^{x_i=p_f+u-1} h(x_i)}{\sum_{x_i=p_f+u}^{x_i=p_f+2u-1} h(x_i)} \tag{4}$$

and

$$C = p_b - \underset{u=1..(p_b-p_f)/2}{\arg\max} f(u) + 1, \tag{5}$$

where f(u) is an NIR estimator of the form:

$$f(u) = \frac{\sum_{x_i=p_b-u+1}^{x_i=p_b} h(x_i)}{\sum_{x_i=p_b-2u+1}^{x_i=p_b-u} h(x_i)}. \tag{6}$$

Note that x_i denotes gray scale intensity value ranges from 0 to 255, with 0 as black and 255 as white. These estimators were used in our previous study [16], [17]. They proved to be robust for light

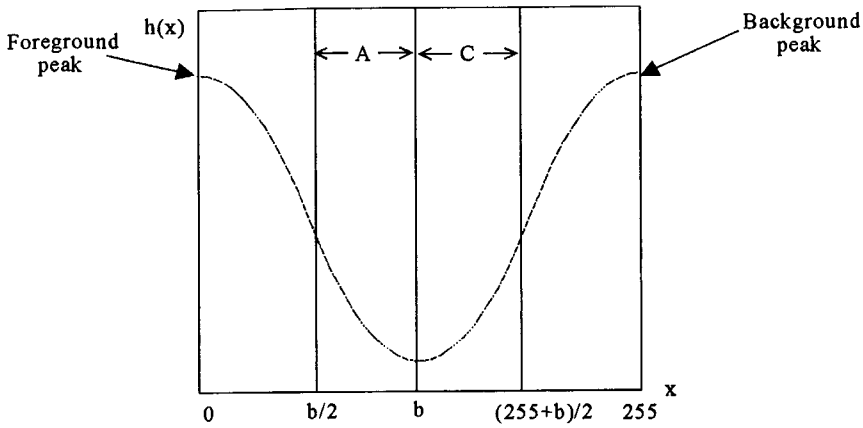

Fig. 2. Desired location of parameters A and C.

and dark backgrounds. In hypothetical distributions (Gaussian), they are robust to local fluctuation. However, when applied to real intensity histograms of handwriting images, they are sensitive to local fluctuations, which sometimes leads to unexpected results. This is especially the case when the handwriting is faint.

3.2 Quadratic Integral Ratio (QIR) Technique

To overcome the NIR algorithms sensitivity to local fluctuations, a new technique called QIR was created. Before describing the QIR technique, we will describe the desired properties of Integral Ratio estimators. Suppose that we are to find the values for A and C manually by looking at a bimodal distribution of intensity histogram shown in Fig. 2. The left peak is the foreground peak and the right peak is the background peak. We want to separate this into three classes, the foreground, the background, and the fuzzy area, by finding the value of parameters A and C. Ideally, we want the fuzzy class to be as small as possible, consisting of only the common intensity values that some foreground pixels and some background pixels have.

There are two desired properties of any Integral Ratio estimators:

1. The resulting parameters should range between the midpoint and the peak ($1/2b$ to b for A and b to $(b+255)$ for C in Fig. 2).
2. If the degree of overlap is high, the fuzzy class should be large (i.e., A and C should be far apart). If the degree of overlap is small, the fuzzy class should be small (i.e., A and C are close). A should increase as the degree of overlapping decreases, while C should decrease as the degree of overlapping decreases.

Instead of working on the real intensity histogram, as in the NIR approach, the QIR approach deals with a quadratic approximation of the real intensity histogram. That is, if h(x) is a quadratic function $h(x) = a(x-b)^2 + c$, where $a > 0, b > 0$, and $c > -1/3ab^2$, then the right and left simple equally spaced adjacent interval integral ratio functions (which are the estimators of A and C) are called *Quadratic Integral Ratio estimators (functions)*. Note here that a quadratic function cannot approximate the whole histogram, hence, the approximation is only made near the maxima. By working with a quadratic function, we can control the performance of the estimators. By using quadratic approximation, an explicit formula for the estimator can be extracted. This has enabled observations on the estimators. The theorems below will explain why QIR estimators are robust.

For the sake of simplicity without losing generality, we use a QIR estimator with $u_0 = 0$.

Theorem 1. *The explicit formula for* $U = \arg\max_{u\in(0,\infty)} f(u)$, *where* $h(x) = a(x-b)^2 + c$ *and* $u_0 = 0$ *is*

$$U = \frac{3(ab^2 + c) - \sqrt{3}\sqrt{a^2b^4 + 4ab^2c + 3c^2}}{2ab}. \tag{7}$$

Theorem 1 expresses the explicit formula of U. See [18] for the derivation of this formula; it is useful since other mathematical properties can be derived from it.

Theorem 2. $1/2b < U \leq b$.

Theorem 2 gives an upperbound and a lowerbound of U. This theorem says that the value of U will be between b (the lowest point separating two peaks), and $1/2b$ (midpoint of the peak and the lowest point). As Theorem 5 states below, they are in fact the tight upperbound and lowerbound of U. If $u_0 = 0$, U is actually the estimator of parameter A. Theorem 2 is important because it says that QIR estimators satisfy the first desired property of Integral Ratio estimators, i.e., it lies in the desired range.

Theorem 3. *If* $ab^2 \to \infty$ *but* b *is finite, or if* $c \to 0$, *or, if* $ab^2 >> c$, *then* $U \to 1/2(3-\sqrt{3})b$. *Also, if* $c = 0$, *then* $U = 1/2(3-\sqrt{3})b$.

Theorem 4. *If* $c >> ab^2$ *or if* $c \to \infty$, *then* $U \to 1/2b$.

Theorem 3 shows that the value of U converges to $1/2(3-\sqrt{3})b$ (approximately $5/8b$) when c is relatively small or close to 0. Theorem 4 shows that the value of U converges to $4/8b$ if c is relatively large. This value is the tight lowerbound of U.

Theorem 5. *If* c *is considered as the variable of* U *(a and b as parameters of U), then U is a monotonically decreasing function, intersecting at* $(c, U) = (-1/3ab^2, b)$, $(0, 1/2(3-\sqrt{3})b)$, *and* $(c \to \infty, U \to 1/2b)$.

Theorem 5 implies that b is the tight upperbound of U and $1/2b$ is the tight lowerbound of U. If the value of c is limited to positive or zero, then $1/2(3-\sqrt{3})b$ becomes the upperbound of c. Theorem 5 is important because it implies that QIR estimators satisfy the second desired property of Integral Ratio estimators. Large c implies a high degree of overlap. In this case, U should be small, close to its lowerbound. This is satisfied by the intersection at $(c \to \infty, U \to 1/2b)$. On the other hand, if c is small, the degree of overlap is small. Then, U should be large, close to its upperbound. If c is a small positive number, then U is close to $1/2(3-\sqrt{3})$. If c is

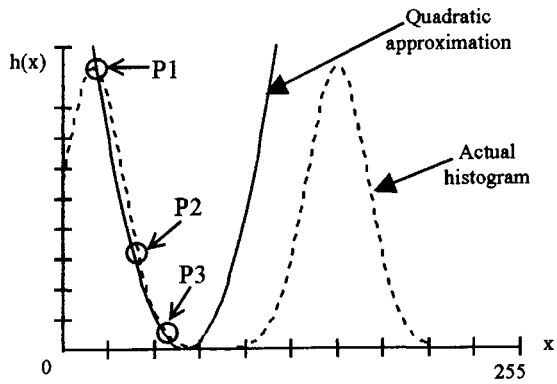

Fig. 3. Quadratic approximation to the foreground's curve in the intensity histogram.

a negative number, U will converge to b. This is satisfied by the intersection of c and U at $(-1/3ab^2, b)$.

Theorems 1 through 5 prove that QIR estimators satisfy all the desired properties of Integral Ratio estimators. Therefore, theoretically, QIR estimators provide a good and robust means of thresholding an image. The experiments described below show this to be the case.

To describe how QIR estimators can be used for thresholding, suppose we have a bimodal histogram, as shown by the dashed line in Fig. 3. Three points are selected from the histogram, say $P1 = (x_1, y_1)$, $P2 = (x_2, y_2)$, and $P3 = (x_3, y_3)$. Then, from these three points, a quadratic curve can be generated by the following set of equations:

$$a' = \frac{\frac{y_3 - y_1}{x_3 - x_1} - \frac{y_2 - y_1}{x_2 - x_1}}{x_3 - x_2} \tag{8}$$

$$b' = \frac{y_2 - y_1}{x_2 - x_1} - a'(x_2 + x_1) \tag{9}$$

$$c' = y_1 - bx_1 - ax_1^2 \tag{10}$$

$$a = a' \quad b = -\frac{b'}{2a'} \quad c = \frac{4a' - b'^2}{4a'} \tag{11}$$

and the quadratic function is $h(x) = a(x - b)^2 + c$. The value of U can then be calculated. The final value of $A = U + x_1$. A similar procedure can be used to obtain threshold C. Smoothing the histogram can be performed beforehand to improve the result of the approximation.

In the current implementation, the second stage of thresholding selects the empirically obtained final threshold value as

$$T = T_2 = \begin{cases} C - 1/2(C - A), & \text{if the writing implement is a felt tipped pen} \\ C - 1/10(C - A), & \text{if the writing implement is a ballpoint pen} \\ C, & \text{if the writing implement is a pencil} \\ C - 1/10(C - A), & \text{if the writing implement is not specified.} \end{cases} \tag{12}$$

4 Experimental Results

4.1 Design of Experiments

To compare the thresholding techniques, a total of 150 images were collected from 10 writers and examined. Each writer wrote the sentence "Handwriting sample for testing FUZZY RATIO" on 15 different paper sheets. The paper sheets differed in material, pattern, and colors. Some of the sheets had background images. The pens used included fiber tip pens, ballpoint pens, fine line markers, and pencils. Most of the papers used contain noisy to very noisy background.

Each page was scanned using a Hewlett Packard HP ScanJet 4c flatbed scanner. The pages were scanned with a resolution of 256 gray scales and 600 × 600 dpi spatial resolution. All preprocessing which can be performed by the scanner was turned off, yielding true raw scanned images with every detail of the handwriting represented.

After the 150 images were collected, they were fed into four thresholding modules (NIR, Otsu [9], Entropy [6], and QIR). To measure and compare the performance of the thresholding techniques, several measures were collected from each thresholded image and each thresholding technique:

1. *Threshold value:* This is the value T produced by each thresholding technique.
2. *Number of over-thresholded images:* In an over-thresholded image, the background is usually removed completely and the handwriting is damaged. If the over-thresholding is not severe, the handwriting may appear thin and contain

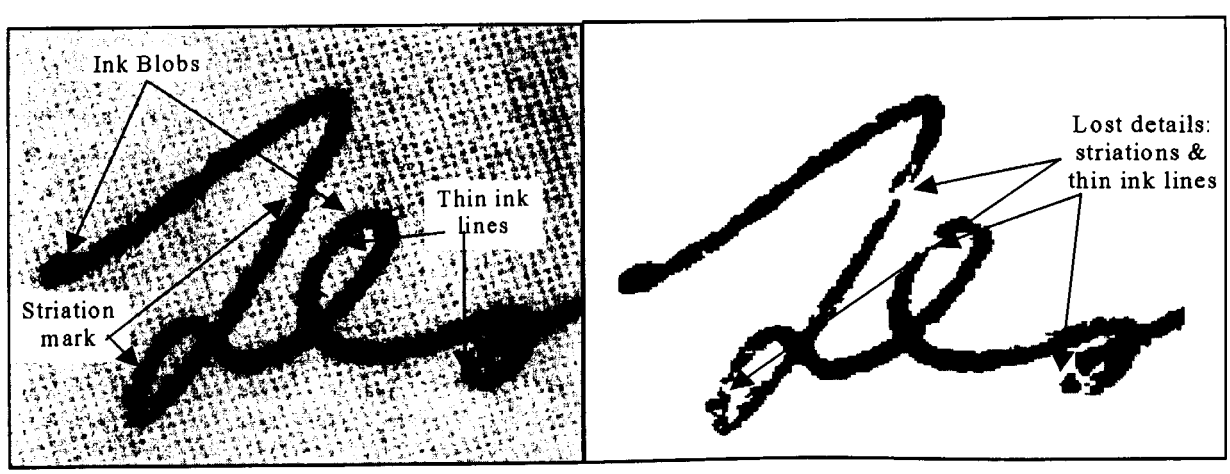

Fig. 4. Original image (left) and its over-thresholded image, showing broken strokes and loss of details (right).

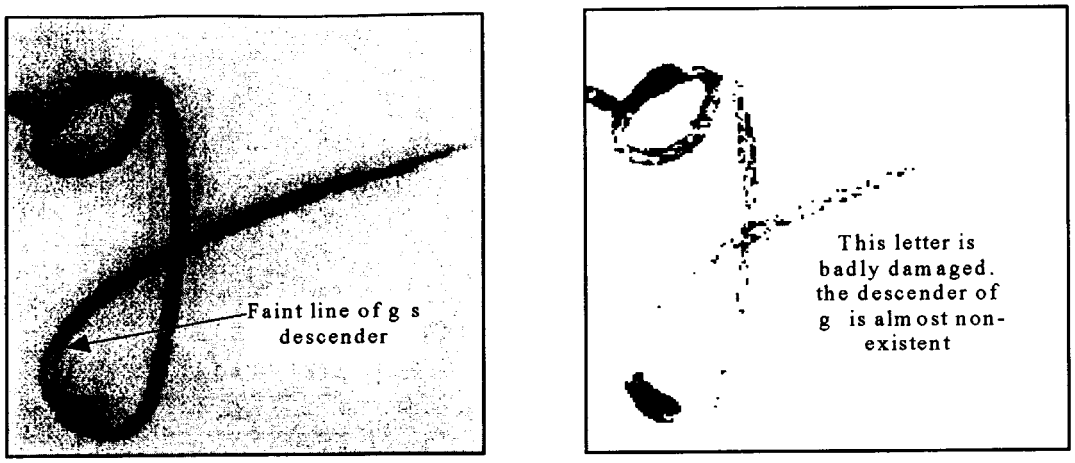

Fig. 5. Original image (left) and its over-thresholded image, showing a badly deformed 'g.'

many broken strokes (which were connected in the original handwriting). An image is categorized as over-thresholded if the handwriting has been significantly thinned or more than half of the letters contain broken strokes.

3. *Number of under-thresholded images:* In an under-thresholded image, the background is usually not completely removed. If the under-thresholding is not severe, most of the background is removed but some strokes which were separated in the original writing became connected. If the under-threshold is very severe, the loops (as in letters o, a, g, etc.) are covered by the background and some letters may be completely surrounded by the background. The letters appear as if they were smeared by the background color. An image is categorized as under-thresholded is a significant amount of the background remains.
4. *Number of correctly thresholded images:* If the image is not over-thresholded and not under-thresholded, it is categorized as correctly thresholded.
5. *Number of letters with broken strokes:* This measurement, together with the number of deformed letter-forms, measures the severity of over-thresholding. The number of characters that contain broken strokes, and not the number of broken strokes, were counted. Fig. 4 shows an illustration of letters that contain broken strokes.
6. *Number of deformed letter forms:* This measures the degree of severe over-thresholding. Severe over-thresholding can damage letter forms so that it is hard to recognize the letters they represent. Lost parts or all major strokes may cause them to become indistinguishable from other letters or cause them to become unreadable, as, for example, an 'O' may look like a 'C,' 'j' or 'g' may lose their descenders. Fig. 5 (right) shows a 'g' that has lost most of its descender.
7. *Number of letter with connected strokes:* This measurement, together with the number of smeared letters/loops, measures the severity of under-thresholding. If two letters that were originally separated appear connected, they were considered to be in this category. The number of letters that became connected with other letters were counted. Therefore, this number depends on the formation of the writing. If it is a purely cursive script, many letters were originally connected, so the number of letters with connected strokes is low. However, if the letters are written as discrete separated characters, this number can be high.
8. *Number of smeared letters/loops:* If the under-thresholding is severe, the letters will not only be connected, but will also be surrounded completely by the background. Loops where the holdes are filled by the background and letters that are surrounded totally by the background are categorized in this group.

TABLE 1
Distribution of Threshold Cases

	NIR	Otsu	Entropy	QIR
Under-thresholded	9	33	98	38
Over-thresholded	29	47	7	1
Correctly thresholded	112	70	45	111

4.2 Results

Table 1 summarizes the distribution of under/over-thresholded and correctly thresholded iamges. Two points should be noted about the results:

1. The higher the proportion of correctly thresholded images, the better the performance of the thresholding algorithm. However, this measure is not the only measure of quality of a thresholding algorithm. The severity of under-thresholding or over-thresholding in incorrectly thresholded images is also important. Other factors, such as the average number of defects and intraclass standard deviation, should also be taken into account.
2. A high proportion of over-thresholded and under-thresholded images indicates a poor thresholding algorithm. It should be noted that if a thresholding algorithm tends to over-threshold images, it may not be improved by simply increasing the threshold values by a constant. Though increasing the threshold values will reduce the number of over-thresholded images, some of the correctly thresholded images may become under-thresholded, thus reducing the number of correctly thresholded images. Therefore, in general, it is not easy to improve the performance of a thresholding algorithm. The same restriction applies to a thresholding algorithm that tends to under-threshold an image.

TABLE 2
Average Number (and Percentage) of Incorrectly Thresholded Characters per Image

Type of defects	NIR	Otsu	Entropy	QIR
Letters with broken strokes	6.86	9.43	1.32	0.61
	(18.1%)	(24.8%)	(3.5%)	(1.6%)
Deformed letters	3.42	4.35	0.64	0.01
	(9.0%)	(11.4%)	(1.7%)	(0.026%)
Connected letters	0.01	0.03	3.35	0.75
	(0.026%)	(0.08%)	(8.8%)	(2.0%)
Smeared letters/loops	0.05	0.22	2.54	0.8
	(0.13%)	(0.58%)	(6.7%)	(2.1%)

Total number of characters per image is 38.

TABLE 3
Average Threshold and Intraclass Variance

	NIR	Otsu	Entropy	QIR
Average threshold value	122.5	125.4	169	155.7
Average intraclass variance	20.1	22.3	8.1	5.3

From Table 1, it can be seen that the NIR and Otsu algorithms had the largest proportions of over-thresholded images. The NIR algorithm tended to over-threshold an image and the proportion of under-thresholded images was small. Otsu's method resulted in a high proportion of both over-thresholded and under-thresholded images, suggesting that its performance varies considerably and is not predictable. The Entropy method did not perform well and under-thresholded most of the images. This may be due to the sensitivity of the technique to the unequal number of pixels in the handwriting and the background. The QIR method tended to under-threshold images but it correctly thresholded most of the images. Overall, the QIR and NIR algorithms achieved the best performance, correctly thresholding some 74 percent of the images. Of the remaining images, the QIR algorithms tended to under-threshold and the NIR algorithm tended to over-threshold.

Table 2 shows the average number and percentage of incorrectly thresholded characters per image for each thresholding technique. Points to note about the results are:

1. The number of *letters with broken strokes* and number of *deformed letters* are measurements associated with over-thresholded cases. The number of *connected letters* and number of *smeared letter/loops* are measurements for under-thresholded cases.
2. While the number of *letters with broken strokes* can indicate over-thresholding that is both subtle and severe, the number of *deformed letters* can only measure severe over-thresholding.
3. The number of *deformed letters* with deformed form is always higher than the number of letters with broken strokes since a deformed letter also has broken strokes. However, the number of *smeared letters/loops* may not be higher than the number of *connected letters* because a letter can be smeared without causing it to be connected to another letter.

Table 2 shows that both the NIR and Otsu algorithms have high number of *letters with broken strokes* and *deformed letters* with deformed form but low number of *connected* and *smeared letters/loops*, which means that they tend to over-threshold an image. Both NIR and Otsu's methods often over-thresholded images severely (on average, Otsu's technique badly damaged 11.4 percent of the characters per document). The Entropy technique tends to under-threshold an image, although it occasionally over-thresholded them. The Entropy technique also has a high proportion of *smeared letters/loops* (6.7 percent), which means that it often severely under-thresholded an image. The severity of under-thresholding of the Entropy method is approximately equal to the severity of over-thresholding of the NIR method (3.35 (8.8 percent) versus 3.42 (9.0 percent) characters per image). The QIR method, on the other hand, almost never over-thresholded an image severely (shown by the very low number of *deformed letters*).It seldom under-thresholded an image but, when it did, it under-thresholded severely. However, the performance of the QIR method is still excellent, with leass than 1.0 (2.6 percent) incorrectly thresholded characters for each measurement. Based on these results, it is concluded that the WIR method performs best as it makes the most correct thresholds and tends toward under-thresholding rather than over-thresholding, and the NIR method is better than Otsu's technique. The Entropy method is comparable in performance to the NIR method.

It should be noted that, although Otsu's technique performed relatively poorly in this study, its performance may still be acceptable in some applications. This is because many of the images in this study are of difficult cases (dark patterned background, faint handwriting) and are more difficult to threshold than cases encountered in many document image applications. However, for forensic document examination where tight requirements are to be satisfied, neither Otsu's, Entropy, nor, NIR methods gave acceptable performance. It should be noted that some of the images could not be adequately segmented using a

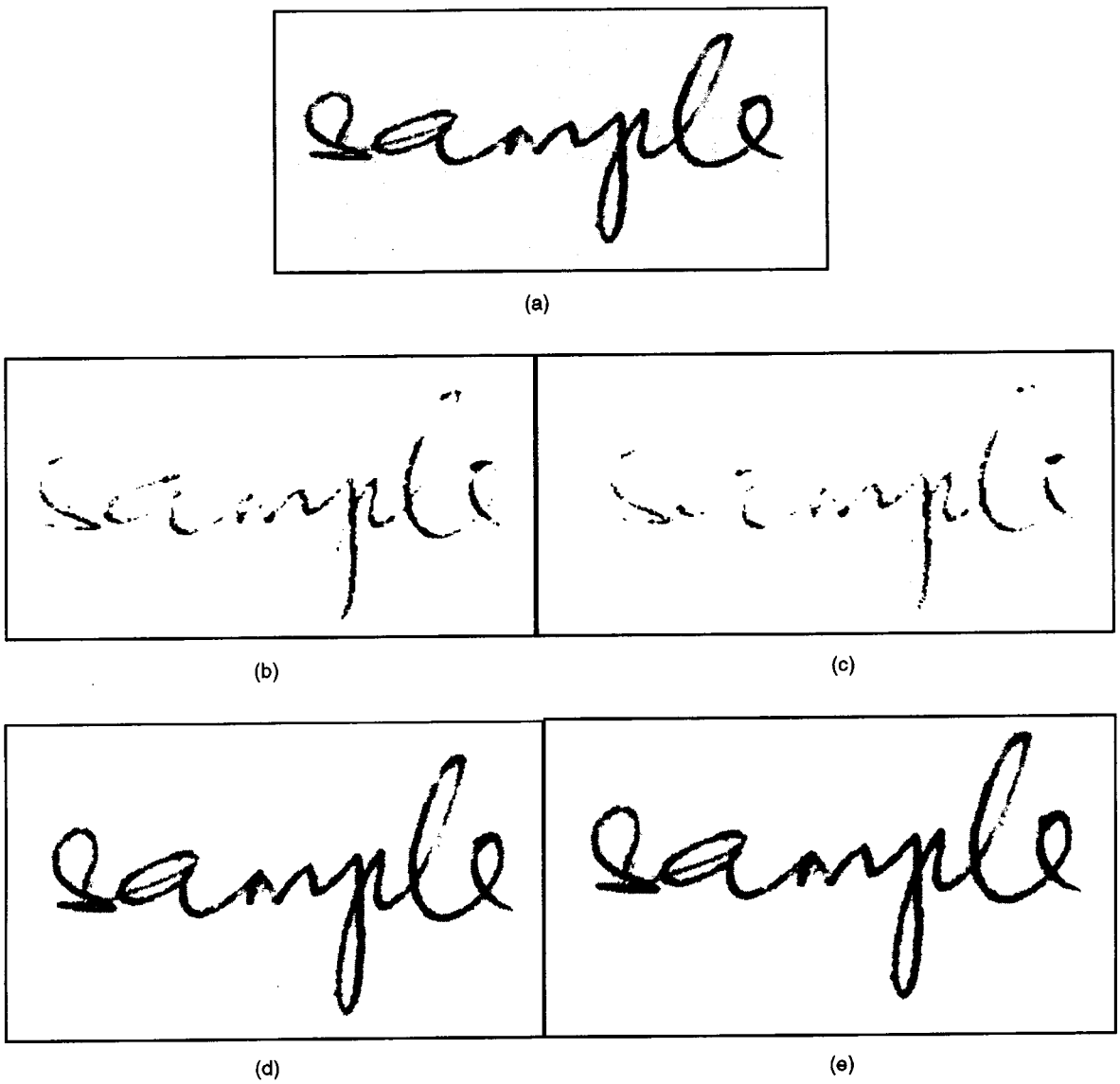

Fig. 6. (a) Original image showing faint handwriting. (b) Over-thresholded by NIR. (c) Over-thresholded by Otsu's technique. (d) Correctly thresholded by Entropy. (e) Correctly thresholded by QIR.

single threshold value and would need later manual image editing to be adequate. In these cases, it is always better to under-threshold the images.

Table 3 shows the average threshold value of each thresholding technique. Points to note about the results are:

1. The average threshold value can be used together with results in Table 2 to conclude whether a technique tends to over-threshold or under-threshold an image.
2. The average intraclass standard deviation measures the variation of threshold values that have different images that have the same background (paper). If the background of several images is the same, since the goal of the techniques is to remove the background, the threshold values should be centered on a constant value (except for a few patterned papers where the patterns are not equally distributed). Therefore, a good thresholding technique has a small average intraclass standard deviation, depending on the paper.

Fig. 6a shows an example of an image where the handwriting is faint on a relatively dark background. In this case, both NIR and Otsu's techniques often over-threshold an image, sometimes severely. Fig. 6b shows the results of thresholding by NIR technique. Fig. 6c shows the result of over-thresholding by Otsu's technique. Entropy and QIR are both good in thresholding handwriting images that have faint handwriting. Although the ascender of the 'l' is borken, overall both Fig. 6d and Fig. 6e show correctly thresholded images.

5 Conclusions

Certain applications, such as Forensic Document Examination, have tight requirements on the performance of a thresholding technique. They require that the background of handwriting images is removed as much as possible without losing any details of the handwriting, while the handwriting can be written using a wide variety of writing instruments on a wide variety of paper.

Two Integral Ratio thresholding techniques, Native Integral Ratio (NIR) and Quadratic Integral Ratio (QIR), were presented.

The QIR technique used to separate these classes was shown analytically to satisfy all the desired properties of Integral Ratio estimators. Experimentally, it was compared against NIR, Otsu's, and the Entropy thresholding techniques. Detailed examination of the resulting thresholded images showed that the QIR algorithm exhibited the best performance and is a sound technique for global thresholding of handwriting images, especially under tight requirements.

REFERENCES

[1] A.S. Abutaleb, "Automatic Thresholding of Gray-Level Pictures Using Two Dimensional Entropy," *Computer Vision, Graphics, and Image Processing*, vol. 47, pp. 22-32, 1989.

[2] A.D. Brink and N.E. Pendock, "Minimum Cross-Entropy Threshold Selection," *Pattern Recognition*, vol. 29, no. 1, pp. 179-188, 1996.

[3] B. Yu, A.K. Jain, and M. Mohiuddin, "Address Block Location on Complex Mail Pieces," *Proc. Int'l Conf. Document Analysis and Recognition*, pp. 897-901, Ulm, Germany, 18-20 Aug. 1997.

[4] D.S. Doermann and A. Rosenfeld, "Recovery of Temporal Information from Static Images of Handwriting," *Proc. IEEE CS Conf. Computer Vision and Pattern Recogition*, pp. 162-168, 1992.

[5] H. Bunke, M. Roth, and E.G. Schukat-Talamazzini, "Off-Line Cursive Handwriting Recognition Using Hidden Markov Model," *Pattern Recognition*, vol. 28, no. 9, pp. 1,399-1,413, 1995.

[6] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, "A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, pp. 273-285, 1985.

[7] J. Kittler and J. Illingworth, "Minimum Error Thresholding," *Pattern Recognition*, vol. 19, no. 1, pp. 41-47, 1986.

[8] L.K. Huang and M.J.J. Wang, "Image Thresholding by Minimizing the Measures of Fuzziness," *Pattern Recognition*, vol. 28, no. 1, pp. 41-51, 1995.

[9] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 9, pp. 62-66, 1979.

[10] O.D. Trier and A.K. Jain, "Goal-Directed Evaluation of Binarization Methods," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1,191-1,201, Dec. 1995.

[11] N.R. Pal and S.K. Pal, "Entropic Thresholding," *Signal Processing*, vol. 16, pp. 97-108, 1989.

[12] S.K. Pal and A. Rosenfeld, "Image Enhancement and Thresholding by Optimization of Fuzzy Compactness," *Pattern Recognition Letters*, vol. 7, pp. 77-86, 1988.

[13] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum Likelihood Thresholding Based on Population Mixture Models," *Pattern Recognition*, vol. 25, no. 10, pp. 1,231-1,240, 1992.

[14] T. Ridler and S. Calvard, "Picture Thresholding Using an Iterative Selection Method," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 8, pp. 630-632, 1978.

[15] F. Albregtsen, "Non-Parametric Histogram Thresholding Methods—Error Versus Relative Object Area," *Proc. Eighth Scandinavian Conf. Image Analysis*, pp. 273-280, Tromsø, Norway, 1993.

[16] Y. Solihin, C.G. Leedham, and V.K. Sagar, "A Fuzzy Based Handwriting Extraction Technique for Handwriting Document Preprocessing," *Proc. IEEE Region 10 Conf.: Digital Signal Processing Applications*, pp. 927-932, Perth, Australia, 27-29 Nov. 1996.

[17] Y. Solihin and C.G. Leedham, "Mathematical Properties of FUZZY_RATIO Handwriting Extraction Technique," *Proc. Int'l Conf. Document Analysis and Recognition*, pp. 1,102-1,106, Ulm, Germany, 18-20 Aug. 1997.

[18] Y. Solihin, "A Toolset of Image Processing Algorithms for Forensic Document Examination," Master's thesis, School of Applied Science, Nanyang Technological Univ., Singapore, 1997.

Robust Rotation Angle Estimator

Whoi-Yul Kim, *Member, IEEE*, and Young-Sung Kim

Abstract—The conventional method of estimating the rotation angle of a pattern using the principal axes is not suitable for circular symmetric patterns since their eigenvalues are similar in both directions. In this paper, a new and robust method of estimating a rotation angle using the phase information of Zernike moments is presented. The experimental results show that the proposed method estimates the rotation angle of the circular symmetric patterns more accurately than the principal axes method, even in the presence of noise.

Index Terms—Rotation angle, zernike moment, phase, principal axes.

1 INTRODUCTION

IN many practical applications of machine vision systems, the recognition of a symbol or a pattern must be done invariant to its orientations. In addition, the rotation angle of the pattern needs to be computed reliably. For example, when a robot arm manipulates an object on a conveyor belt, the robot must first recognize the object, regardless of its orientation angle, and then estimate its rotation angle for manipulation.

The conventional method of finding the rotation angle of a pattern using the principal axes method makes use of the major and minor axes which are the eigenvectors of the covariance matrix of a pattern. The largest eigenvector of the covariance matrix is in the direction where the maximal spread of the object occurs and the degree of spread is measured by the corresponding eigenvalue [1]. This method is suitable for patterns whose degrees of spread are distinct in one direction. As will be demonstrated in the experiment, however, this method fails with circular patterns since their eigenvalues are similar in both directions [2].

Rotation invariant pattern recognition using the magnitude of Zernike moments has been extensively studied [3], [4], [5], [6], [7]. When the magnitudes of Zernike moments of the rotated pattern and that of the reference are the same, the pattern and the reference can be considered the same, regardless of their orientations. The phase differences of the Zernike moments between the recognized pattern and that of its reference account for the rotation angle of the pattern. In this paper, a new and robust method of estimating the rotation angle of a pattern using the phase information of Zernike moments is presented.

2 ZERNIKE MOMENTS

Zernike moments are the mapping of a pattern onto a set of complex basis functions which have two distinctive properties, the orthogonality and the rotation invariance of their magnitude [1]. The first property ensures that the contribution of each moment to the pattern is unique and independent. The second property allows the magnitude of Zernike moments extracted from a pattern to be the same at any orientation.

Zernike moments are defined inside the unit circle. In (ρ, θ) polar coordinate, the radial polynomials $\{R_{nm}(\rho)\}$ are defined as

• *The authors are with the Electronic Engineering Department, Hanyang University, Heangdang-Dong, Sungdong-Ku, Seoul 133-791 Korea. E-mail: wykim@email.hanyang.ac.kr, yskim@vision.hanyang.ac.kr.*

Manuscript received 17 Oct. 1998; revised 22 Dec. 1998.
Recommended for acceptance by H.R. Keshavan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 108081.

0162-8828/99/$10.00 © 1999 IEEE