Decompose algorithm for thresholding degraded historical document images

Y. Chen and G. Leedham

Abstract: Numerous techniques have previously been proposed for single-stage thresholding of document images to separate the written or printed information from the background. Although these global or local thresholding techniques have proven effective on particular subclasses of documents, none is able to produce consistently good results on the wide range of document image qualities that exist in general or the image qualities encountered in degraded historical documents. A new thresholding structure called the decompose algorithm is proposed and compared against some existing single-stage algorithms. The decompose algorithm uses local feature vectors to analyse and find the best approach to threshold a local area. Instead of employing a single thresholding algorithm, automatic selection of an appropriate algorithm for specific types of subregions of the document is performed. The original image is recursively broken down into subregions using quad-tree decomposition until a suitable thresholding method can be applied to each subregion. The algorithm has been trained using 300 historical images obtained from the Library of Congress and evaluated on 300 'difficult' document images, also extracted from the Library of Congress, in which considerable background noise or variation in contrast and illumination exists. Quantitative analysis of the results by measuring text recall, and qualitative assessment of processed document image quality is reported. The decompose algorithm is demonstrated to be effective at resolving the problem in varying quality historical images.

1 Introduction

There are currently millions of historical documents stored in museums, libraries and government record offices all over the world. These documents contain important and interesting information that was written and recorded in handwritten letters and notes during the past few hundreds years, before the invention of typewriters and even printing. To allow the preservation of these delicate documents while also providing wider access to scholars and researchers the documents are frequently scanned and made available as high-resolution images. Given the current state of the art in computer recognition and processing of script, most of these historical documents are impossible to read automatically. To facilitate future automatic searching and analysis of the words and content in the documents, it is necessary to separate the useful pixels containing document content, such as handwriting, drawings, pictures and other information representing useful artefacts, from the background pixels of the paper.

This is a non-trivial task as the documents are frequently degraded due to poor storage and damage over time and are often difficult for a human to decipher. Historical handwritten document images frequently contain handwriting that was written by ink pen hundreds of years ago. The long storage time and adverse storage environments at some time in their past make the pen strokes fade or run and the quality of the document paper degrade, even producing some spots due to mould or bacterial growth.

Owing to the characteristic of pen strokes and spots on the paper, there are three main classes of subregion that can be defined: background region, faint region and heavy region. As shown in Fig. 1, a typical document image includes double-sided noise where writing on the reverse side of the paper soaks through and merges with writing on the front of the paper, ghosting noise where writing on the reverse side has soaked through giving the appearance of writing on the front of the paper, and varying background contrast. These are the general problems encountered in historical document images.

The quality of the thresholding result when separating foreground from background is decisive for subsequent analysis of the document content. It requires retention of the full information content on a clear white background. In some applications, such as forensic document analysis, or scholastic analysis of the writing style, we are interested in the detailed greyscale or colour variations of the pen strokes or printing. In others, such as studying the content of the documents, a binary image is sufficient.

In this paper we propose a decompose thresholding algorithm, which can be used effectively on degraded document images.

2 Background

Most of the existing thresholding methods can be allocated into four main classes.

The first class of threshold methods is *histogram-based techniques*. The structure of the peaks, valleys

IEE Proceedings online no. 20045054

doi: 10.1049/ip-vis:20045054

Paper first received 18th June 2004 and in revised form 22nd March 2005. Originally published online 3rd June 2005

The authors are with the School of Computer Engineering, Nanyang Technological University, N4-#2A-28 Nanyang Avenue, Singapore 639798

E-mail: p145806718@ntu.edu.sg

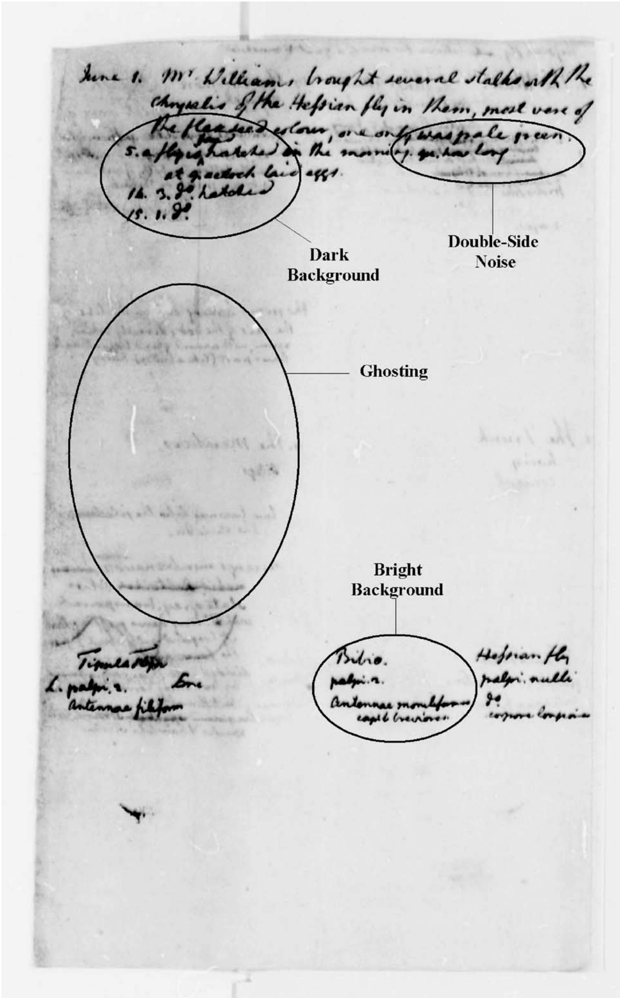

Fig. 1 *Example of typical historical document image*

and curvatures in the smoothed grey-scale histogram is analysed to determine the final global threshold value.

Otsu's method [1] is an early, but still popular histogram-based global threshold algorithm. It proposed a criterion for maximising the variance of the between-class pixel intensities to perform thresholding. However, Otsu's algorithm is very time-consuming for image binarisation because of its inefficient formulation of the between-class variance, and the performance varies with image types. Cheriet [2] applied a recursive version of Otsu's algorithm for bank cheque image segmentation. This technique is more flexible than Otsu's original method. At each recursion, the technique segments the object with the lowest intensity from the input image. This process continues until there is the darkest object left in the image. Don [3] proposed a histogram-based global technique that utilises noise attribute features from the image. It is based on a noise model to overcome the difficulty caused when some objects do not form prominent peaks in the histogram. Liu and Srihari [4] proposed a local thresholding method, which is based on texture features (stroke-width-based) to extract characters from the run-length featured texture (histogram-based) background. Solihin and Leedham [5] investigated two global techniques: native integral ratio (NIR) and quadratic integral ratio (QIR). These two histogram-shape-based methods were developed from a new class of global thresholding technique called integral ratio. The common disadvantage of histogram-based global methods is their ineffectiveness for images that do not have a bimodal histogram distribution, even when the images contain clear handwriting on a contrasting background.

The second class of thresholding methods is *entropy-based techniques*, where entropy is used to separate the global thresholding classes. For example, the optimal threshold value can be calculated by maximising the sum of the foreground and background entropies. i.e. maximally separating region intensities of the foreground and background. Shannon's entropy information theory has been used in image segmentation for many years. Pun [6] presented a maximum-entropy-based method. It used Shannon's concept to define the entropy of an image. Pun used this concept to derive an expression for an upper bound of the aposteriori entropy. The expression was finally used to threshold an image. Kapur *et al.* [7] reported an improvement of Pun's method. It is a histogram analysis and maximum-entropy-based global technique, which uses the maximum of the sum of the entropy of the grey-level distribution of the foreground and background. The technique has shown good performance for picture images. Pal and Pal [8] presented a higher order entropy method for object extraction and summarised several entropy methods used in image processing. Chang *et al.* [9] presented a spatial context thresholding algorithm that minimises relative entropy. It uses spatial dependence (co-occurrence probability) of the pixels. Wang *et al.* [10] applied a maximum-relative-entropy-based technique to picture images. Brink [11] proposed a minimum-spatial-entropy-based global thresholding algorithm.

The third class of thresholding methods is *local adaptive techniques*. Niblack [12] presented his method based on the calculation of the local mean and local standard deviation of grey scales. The method obtains an adaptive threshold by varying the slicing level according to the local variance of the image. Zhang and Tan [13] proposed another improved version of Niblack's method. In comparison with the Niblack algorithm, it performs well at shadow boundary detection. Bernsen [21] proposed a local thresholding technique based on neighbours of each pixel. It has proven to be a fast algorithm. The disadvantage is that it does not work well when the background regions have varying grey-level intensities and contain ghost images (see Fig. 1.) That is because a local manual variable called local contrast threshold (LCT) is used on the whole image. However, LCT is different for different local areas of an image, hence an LCT based on the whole image will only work on some image types. Eikvil *et al.* [22] proposed an adaptive thresholding technique in which the pixels inside a small window S are thresholded on the basis of clustering of the pixels inside a large window L. The principle of the technique is that a large window L with a small window S in the centre is moved across the image in a zig-zag fashion, in steps equal to the size of S. S is labelled as print or background on the basis of the clustering of the pixels inside L. Yanowitz and Bruckstein [20] proposed a local gradient-based thresholding algorithm. In this algorithm, the thresholding was determined by interpolating the image grey level at the points where the gradient is high, indicating probable object edges. The gradient map of the image was used to point at well-defined portions of object boundaries. It demonstrated that both the location and grey levels at these boundary points are a good choice for determining local thresholds. Wang and Pavlidis [17] assumed that the grey-scale image is a surface with top logical features corresponding to the shape features of the original image. Each pixel of the image was classified as: peak, pit, ridge, ravine, saddle, flat or hillside. Rules can be built based on the estimated first and second directional derivatives of the underlying image intensity surface. The characters can be extracted according to the rules and the features. Zhao and Yan [18] utilised geometrical features combined with grey-level information analysis to determine the local threshold values for blueprint images. Chen and Takagi [19] proposed a local thresholding technique, which was based on run length coding. Manjunath and Chellappa [20] presented a feature-extraction based local boundary detection technique. Djeziri *et al.* [21] proposed a local thresholding technique, which used *filiformity* as the close description of the local feature. Yasuda *et al.* [22] proposed a local-intensity-change-based local technique. It was determined by image contrast.

The fourth class of thresholding methods is *other global techniques*. In addition to histogram and entropy informatics theory, other global information can also be used in global thresholding. Gorman [23] presented a global method based on local features, which used a connectivity-preserving measure. The method determined threshold values at intensities between region levels where region break-up is least likely. Gu *et al.* [24] used the differential top-hat transform to extract characters from scene images. The main disadvantage is that the parameters used in the technique are fixed, so problems will arise when the input images have variable contrast. Weszka and Rosenfield [25] used the co-occurrence matrix to compute the sum of transitions between the object and background. Wang and Haralick [26] and Kohler [27] proposed an edge-analysis based global multi-threshold technique.

All previously reported thresholding methods have been demonstrated to be effective for certain classes of document images. None, however, has proven effective for all examples of 'difficult' document images as are frequently found in historical document. A summary of previously thresholding techniques is shown in Table 1.

In this paper, a decompose algorithm, which recursively decomposes a document image into subregions until appropriate weighted values can be used to select an

Table 1*a*: Histogram-shaped-based thresholding techniques

	Histogram-based Technique author	Year	Main features	Major field
1	Otsu	1979	Class separability method	Scene image
2	Don	1995	Noise attribute	Printed and mail image
3	Liu and Srihari	1997	Stroke-based	Printed image
4	Cheriet	1998	Recursive Otsu's algorithm	Cheque image
5	Solihin and Leedham	1999	IR-class based	Handwritten Image

Table 1*b*: Entropy-based thresholding techniques

	Entropy-based techniques Technique author	Year	Main features	Major field
1	Pun	1981	Maximum Shannon's entropy method	Scene image
2	Kapur *et el.*	1985	Entropy-based	Scene image
3	Pal and Pal	1988	Higher-order entropy based	Scene image
4	Chang *et al.*	1994	Minimise relative entropy based	Scene image
5	Brink	1995	Spatial information analysis; Minimum spatial entropy based	Handwritten image with pattern
6	Wang *et al.*	2002	Maximum relative entropy based	Scene image

Table 1*c*: Local adaptive thresholding techniques

	Local adaptive techniques Technique author	Year	Main features	Major field
1	Yasuda *et al.*	1980	Based on local intensity change (max/min and contrast)	Scene image
2	Bernsen	1986	Locally based on neighbours	Scene image
3	Niblack	1986	Local mean and local standard deviation	Scene image
4	Yanowitz and Bruckstein	1986	Using the threshold surface	Scene image
5	Eikvil *et al.*	1991	The pixels inside a small window S are threshold on the basis of clustering of the pixels inside a large window L	Printed text image
6	Manjunath and Chellappa	1991	Feature extraction based	Images (boundary detection)
7	Chen and Takagi	1993	Run-length-coding based	Scene image
8	Wang and Pavlidis	1993	Adaptive thresholding	Handwritten image

Table 1*c*: (*continued*)

	Local adaptive techniques Technique author	Year	Main features	Major field
9	Djeziri *et al.*	1998	Use *filiformity* as the closest description of the local feature	Check image
10	Zhao and Yan	1999	Utilise geometrical features combined with grey-level analysis	Blueprint image
11	Zhang and Tan	2001	Improvement of Niblack's algorithm	Handwritten image

Table 1*d*: Other global thresholding techniques

	Entropy based techniques Technique author	Year	Main features	Major field
1	Kohler	1981	Edge analysis	Scene image
2	Wang and Haralick	1984	Edge analysis	Scene image
3	Weszka and Rosenfield	1985	Co-occurrence matrix is used to compute the sum of transition between the object and the background	Scene image
4	Gorman	1994	Connectivity-preserving measure	Technical journal cover image
5	Gu *et al.*	1998	Differential-top-Hat-transform based	Scene image

appropriate single-stage thresholding algorithm for each region, is proposed. A new mean-gradient-based method to select the threshold for each subregion is also proposed. The effectiveness of the algorithm is assessed on 300 'difficult' document images extracted from the Library of Congress online database of historical documents. The six thresholding algorithms used for comparative evaluation are four local thresholding methods: improved Niblack's method proposed by Zhang and Tan [13], Yanowitz and Bruckstein's method [16], Bernsen's method [14], ETM [15] and two global methods: Otsu's method [1] and QIR [5]. The relative performance of the six methods have been evaluated and compared previously [28, 29].

3 Proposed decompose algorithm

The proposed decompose algorithm is shown in Fig. 2. The initial step of the algorithm tests whether the image has a bimodal histogram. If it does, then histogram-based global thresholding methods, which have been proven to have outstanding results for bimodal histogram images, can be applied. Global methods are particularly effective at saving processing time (complexity is equal to $O(n)$). The decompose algorithm focuses on 8-bit grey-scale images of historical handwritten documents, which do not have a non-bimodal histogram. It is a local adaptive analysis method, which uses local feature vectors to find the best approach for thresholding a local area. Appropriate weighted values are selected automatically for the specific types of document image regions under investigation. The original image is recursively broken down into subregions using quad-tree decomposition until an appropriate weighted thresholding method can be applied to each subregion.

The outline of the decompose algorithm is:

1. bimodal testing
2. decompose image into four equal size local regions
3. extract feature vectors from each local region
4. local region classification
5. repeat steps 2, 3 and 4 until all regions are classified
6. smoothing of the edges of each region
7. threshold method is applied to each region.

If the image is not bimodal it is recursively decomposed using quad-tree decomposition into smaller regions and a local threshold method with appropriate weighted values is applied to each different class region. This continues until the whole image has been decomposed and a threshold technique assigned to each region. In this paper, a weighted gradient-based thresholding method is applied. The grey-scale values of the foreground pixels can, if required, be restored by setting each black pixel to its associated grey-level value in the original image.

The new algorithm analyses the feature information of the local regions with different sizes, and applies a new mean-gradient-based threshold method with appropriate weighted values to obtain the best result.

The complexity of this algorithm can be sufficiently described by the function $g(n) = n^2$, where n is number of pixels in the input grey-scale historical image. Formally, this algorithm is 'of order $O(n^2)$'. This function expresses the worst-case scenario.

3.1 Image bimodality testing

The peaks, valleys and curvatures in the smoothed histogram of the historical handwritten document grey-scale values are analysed using shape-based techniques to determine the final global threshold value. Numerous methods have been described previously that use histogram-based global thresholding methods. They work well for grey-scale images, which have a bimodal histogram [5, 29, 30]. Figure 3 shows two examples of historical images illustrating the difference between 'unimodal' and 'bimodal' histograms. Figure 3*a* is the unimodal histogram of Fig. 3*b*, which is referred to as a unimodal

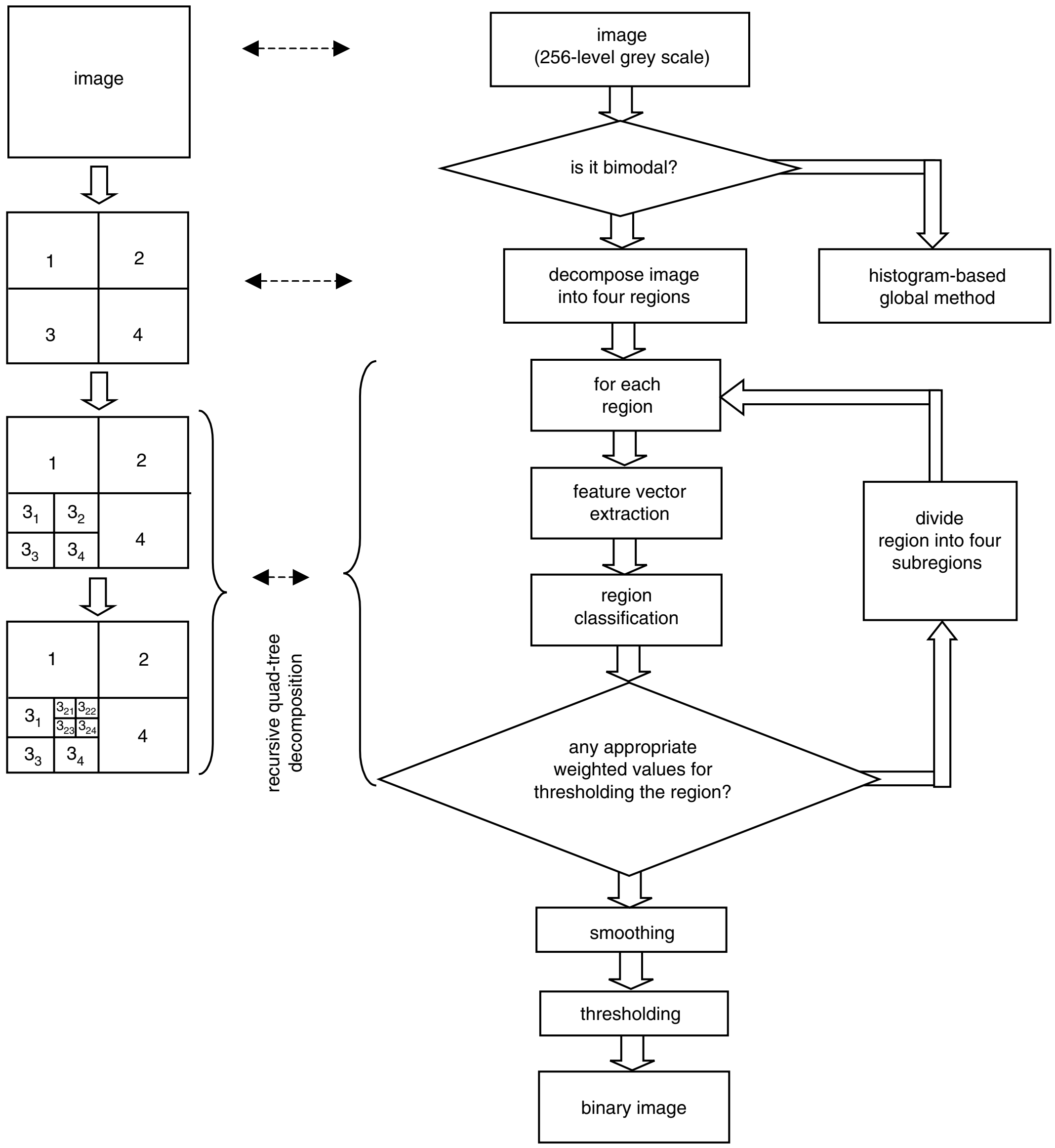

Fig. 2 *Flow chart of decompose thresholding algorithm*

historical document image. Figure 3*c* is the bimodal histogram of Fig. 3*d*, which is referred to as a bimodal historical document image. The unimodal historical image has various characteristics in different areas, which are difficult to binarise accurately using a global threshold value. The bimodal historical image, on the other hand, contains balanced characteristics all over the whole image and is easier to threshold using an appropriate global threshold.

A test of bimodality test is performed on the grey-scale histogram to determine whether a global thresholding method can be effectively applied. For an image that has an obvious bimodal histogram, for example dark black ink on clean white paper, an existing histogram-based global thresholding method can usually be applied to the whole image. There are two branches following the bimodal testing. One is to threshold the input image using the QIR [5] or other global method. The other branch is to apply the decompose approach and analyse local feature vectors to find the best approach for thresholding each subdivided local area.

A simple histogram classification method is introduced: if the histogram of the image contains two obvious peaks, and the valley between the two peaks is located around the middle of the histogram range (0–255), then the histogram is bimodal; otherwise the histogram is unimodal. The bimodal testing is carried out in three steps:

1. histogram smoothing
2. peak number calculation
3. testing whether the valley between two peaks is in the middle of the range 0–255.

If the peak number is equal to two and the valley between the two peaks is in the middle range of 0–255, the image has a bimodal histogram.

3.2 Image decomposition

Contrast variation is a major problem in historical document images, and causes many of the previously reported thresholding methods to fail. Contrast variation is therefore a major characteristic, which can be used to determine whether a thresholding algorithm can be applied or whether the image needs to be further decomposed. The input image is first decomposed into four equal size subimages if the contrast of the decomposed input image $C \geq$ threshold, T (T is currently empirically set at a grey-scale value of 180 based on observations of test images.), and then the subimage is further decomposed. It can be achieved in five steps:

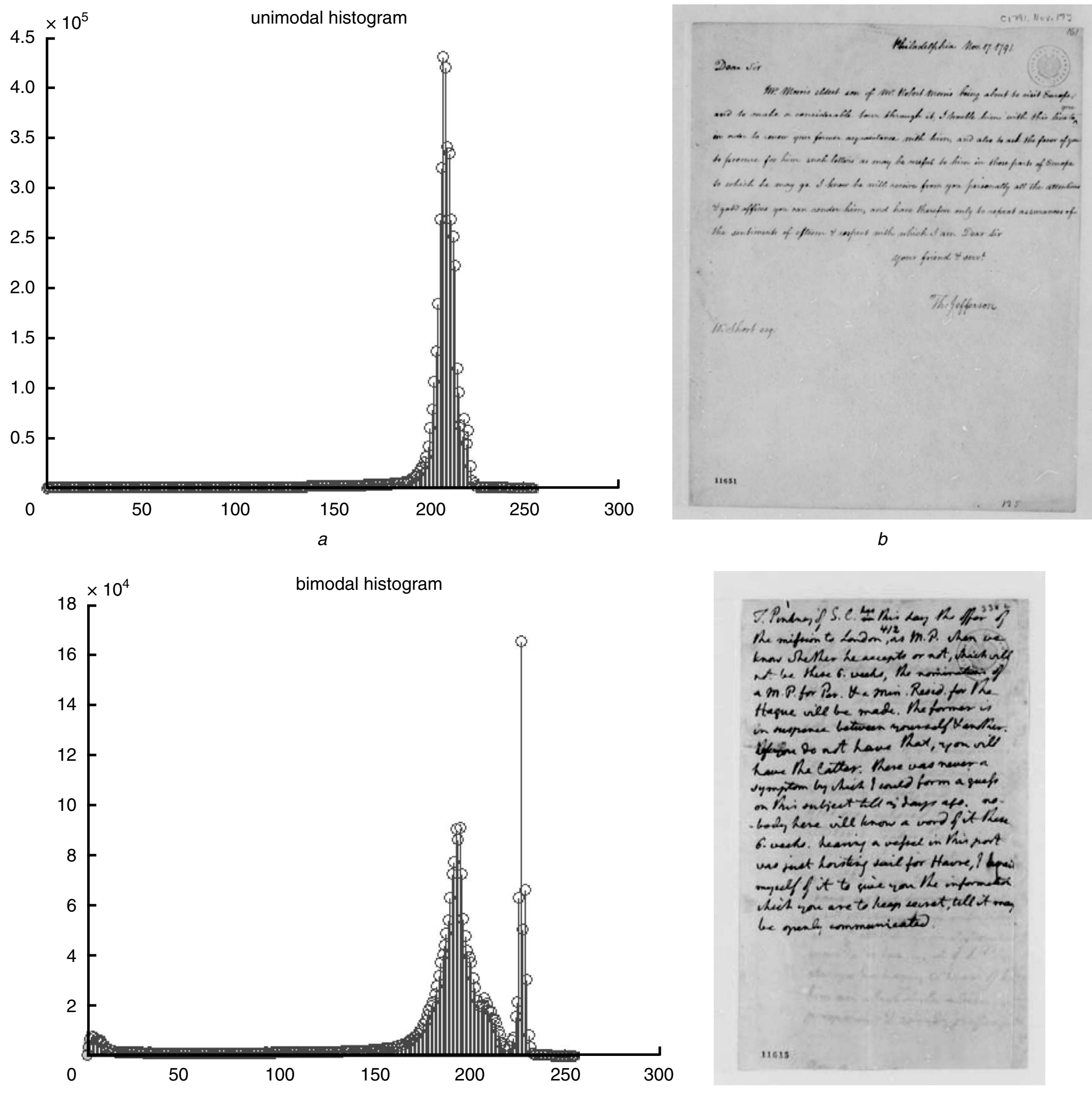

Fig. 3 *Example of unimodal and bimodal histogram*

a Unimodal histogram of image *b*
b Historical Image
c Bimodal histogram of image *d*
d Historical image

1. Set the input image as sub-image.
2. Calculate the mean value of the 24 neighbour of the pixels that lie at co-ordinates $(2M+1,\ 2N+1)$ of the subimage, where

$$\begin{cases} M = 1, 2, \ldots, \frac{1}{2}(\text{row number of subimage}) \\ N = 1, 2, \ldots, \frac{1}{2}(\text{column number of subimage}) \end{cases}$$

3. Find the maximum mean value *maximean* and minimum mean value *minimean* of the subimage.
4. Contrast = *maximean* − *minimean*. If contrast $C \geq T$, the subimage is decomposed into four smaller equal-sized subimages.
5. Repeat steps 2 to 4 until $C < T$, or when the subimage reaches 64×64. A subimage, which is smaller than 64×64, has insufficient information for further feature extraction.

3.3 Feature extraction

Feature vectors are extracted to measure useful information from the decomposed subimages. Many feature vectors have been used in document image binarisation. Most of them were applied to printed documents with clean (white) backgrounds but did not work well for degraded images. Three feature vectors are proposed in this paper, which focus on handwritten document images with messy background and faded writing.

By observing handwritten document images, it can be seen that edge and variance measures change with stroke or word direction based on the characteristics of the handwriting. These two feature vectors are presented as WDES and WDVAR, and are described in Section 3.3.1. Since degraded historical image strokes are frequently affected by noise, the mean-gradient feature vector is good at describing the effect of stroke noise. These three important features are considered in region classification.

3.3.1 Word-direction-based grey-level co-occurrence matrix (WD-GLCM) feature vectors: A grey-level co-occurrence matrix (GLCM) contains information about the positions of pixels having similar grey-level values. A co-occurrence matrix is a two-dimensional array, $\boldsymbol{G}$, in which both the rows and the columns represent a set of possible image values.

A GLCM $\boldsymbol{G_d}[i,j]$ is defined by first specifying a displacement vector $\boldsymbol{d} = (dx, dy)$ and counting all pairs of pixels separated by $\boldsymbol{d}$ having grey levels i and j ($i = 1$–n; $j = 1$–n, where n is the number of grey levels in the image). The GLCM is defined by:

$$\boldsymbol{G_d}[\boldsymbol{i},\boldsymbol{j}] = \boldsymbol{n_{ij}} \tag{1}$$

where $\boldsymbol{n_{ij}}$ is the number of occurrences of the pixel values (i, j) lying at distance $\boldsymbol{d}$ in the image.

The co-occurrence matrix $\boldsymbol{P_d}$ has dimension $n \times n$, where n is the number of grey levels in the image.

Lam [31] described the grey-level gradient co-occurrence matrix (GLGCM). In GLGCM, $\boldsymbol{n_{ij}}$ was the number of occurrences of the pixel values (i, j) lying at distance $\boldsymbol{d}$ in the four directions: 0, 45, 90 and 135°.

For the three fragments of handwritten document images shown in Fig. 4, it can be seen that there are three main directions in the slant of the handwritten words: left top to right bottom, top to bottom, and right top to left bottom. In counter-clockwise direction, they are at 45° (or 225°), 0° (or 180°,) and 135° (or 315°.)

Figure 5 shows the eight directions of words in counter-clockwise direction. Referring to Fig. 5, the matrices of word directions can be defined as in Fig. 6. The mean direction of the input handwritten document image is determined by using $G0$–$G7$. The word-direction based GLCM can be determined after the stroke direction has been calculated:

$$\text{WDGLCM} = \text{occurrences of pixel}(i, j) \text{ lying at } [\,(i, j-d) \quad \ldots \quad (i, j) \quad \ldots \quad (i, j+d)\,], \text{when } G0$$

$$\text{WDGLCM} = \text{occurrences of pixel}(i, j) \text{ lying at } \begin{bmatrix} \ldots & \ldots & \ldots & \ldots & (i-d, j+d) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & (i, j) & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ (i+d, j-d) & \ldots & \ldots & \ldots & \ldots \end{bmatrix}, \text{when } G1$$

$$\text{WDGLCM} = \text{occurrences of pixel}(i, j) \text{ lying at } \begin{bmatrix} (i-d, j) \\ \ldots \\ (i, j) \\ \ldots \\ (i+d, j) \end{bmatrix}, \text{when } G2$$

$$\text{WDGLCM} = \text{occurrences of pixel}(i, j) \text{ lying at } \begin{bmatrix} (i-d, j-d) & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & (i, j) & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & (i+d, j+d) \end{bmatrix}, \text{when } G3$$

where d is half of the typical stroke width of the word in the input document image, i, j are the co-ordinates of the pixels.

The co-occurrence matrix WDGLCM has dimension $n \times n$, where n is the number of grey levels in the image.

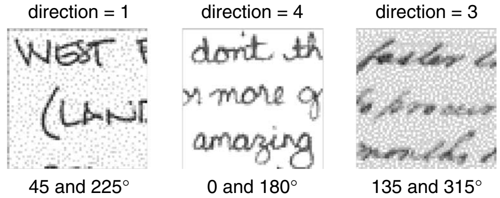

Fig. 4 *Three main word directions*

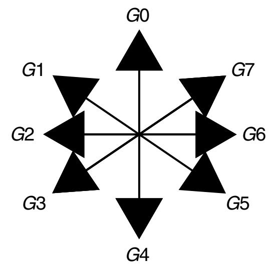

Fig. 5 *Eight directions of word*

$$Go\ (0° \text{ direction}) = \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

$$G1\ (45° \text{ direction}) = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$$

$$G2\ (90° \text{ direction}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$G3\ (135° \text{ direction}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

$$G4\ (180° \text{ direction}) = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix}$$

$$G5\ (225° \text{ direction}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$G6\ (270° \text{ direction}) = \begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$G7\ (315° \text{ direction}) = \begin{bmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Fig. 6 *Direction matrices*

- Word-direction-based-edge-strength (WDES)

Edge Strength can be defined based on the WD-GLCM vector as

$$\text{WDES} = \frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} (i-j)^2 \times \text{WDGLCM}(i, j) = \text{mean}[(i-j)^2 \times \text{WDGLCM}(i, j)] \tag{2}$$

where i and j are the co-ordinates of WDGLCM, and K is the number of grey levels in the input image.

Direction-based edge strength measures the grey-level gradient differences in a certain direction determined by the conditions of the input image. It can provide more useful information for further analysis and works more effectively than simple edge strength based on GLCM.

- Word-direction based variance (WDVAR)

Word-direction-based variance (WDVAR) measures the variability of grey value differences and hence coarseness of texture. A large value of variance indicates large local variation. Word-direction-based variance is defined by WDGLCM as:

$$\mathrm{WDVAR}^2 = \frac{1}{K-1}\sum_{i=1}^{K}\sum_{j=1}^{K}[\mathrm{WDGLCM}(i,j)-\mu]^2 \quad (3)$$

where $\mu = (1/K^2)\sum_{i=1}^{K}\sum_{j=1}^{K}\mathrm{WDGLCM}(i,j)$

- Mean-gradient

Gradient is the change of image texture along some direction in the image. From the definition, the gradient of the intensity image $I(x, y)$ at location (x, y) is:

$$G(x,y) = \sum_{x=0}^{i-1}\sum_{y=0}^{j-1}\frac{[\partial I(x,y)/\partial x, \partial I(x,y)/\partial y]}{xy} \quad (4)$$

where $\boldsymbol{x}$, $\boldsymbol{y}$ are the co-ordinates of the pixel of input image. G_N is the mean-gradient value of the sub-block in direction N given by

$$G_N(x,y) = \sum_{x=0}^{i-1}\sum_{y=0}^{j-1}\frac{[\partial I(x,y)/\partial x_N, \partial I(x,y)/\partial y_N]}{xy} \quad (5)$$

where x, y is the size of the subblock image region, N is the mean direction of words in the local region (0–7) as in the Fig. 5. Mean gradient is sensitive to small variance between strokes; it can be used to detect the faint strokes between heavy strokes.

3.4 Sub-block classification and thresholding

The three feature vectors described in Section 3.3.1 were used to test the local regions and classify them into three types: background, faint strokes or heavy strokes. Typical examples of these three types of region are shown in Fig. 7. The background of a document does not contain any useful content information. A background area typically has lower values of edge strength and variance. A noise-free background also has a small mean-gradient value. Faint stroke areas contain faint strokes, which are very difficult to detect from the background. This kind of area typically has a medium value of edge strength and mean gradient but less variance. Heavy stroke areas have strong edge strength, more variance and larger mean-gradient value. The proposed weighted gradient thresholding method is applied to the different classes of subblock:

3.4.1 Background area:

This area is simply set to white (grey-scale value 255).

3.4.2 Faint handwritten image:

(*a*) *Enhancement:* Enhancement of faint strokes is necessary for further processing. To avoid the enhancement of noise, a Wiener filter was first applied. The enhancement can be divided into two steps.

1. Use a 3×3 window to enhance the image by finding the maximum and minimum grey value in the window using (6) and (7):

$$\mathrm{mini} = \min(\text{elements in the window}) \quad (6)$$

$$\mathrm{maxi} = \max(\text{elements in the window}) \quad (7)$$

2. Compare 'pixel – mini' and 'maxi – pixel', where 'pixel' is the pixel-value. If the former is larger, the 'pixel' is closer to the highest grey value than the lowest value in this window; hence the value of 'pixel' is set to the highest grey value ('pixel' = 'maxi'). If the former is smaller, then the value of 'pixel' is set to the lowest grey value ('pixel' = 'mini').

(*b*) *Thresholding:* A new weighted method based on mean-gradient direction is proposed for thresholding faint strokes. Handwritten English or Western-style scripts normally contain strokes written in several directions. In this method a different matrix is used to detect different mean gradients in eight different directions ($G0$–$G7$), as described in Section 3.3 and shown in Fig. 5.

The matrices $G0$–$G7$ are convolved with the subblock to discover the maximum mean-gradient value. The subblock direction is the one in $G0$–$G7$ that produces the largest mean-gradient value. For example, consider the subblock shown in Fig. 6. The largest mean-gradient value exists at the convolution of $G3$ with the subblock, indicating that the direction of strokes in Fig. 7 is 135°.

The mean directions of the area's strokes are calculated as N, and then the mean-gradient method is applied using N to obtain the binary output.

$$T(x,y) = wM(x,y) + kG_N(x,y) \quad (8)$$

where T is the threshold value, M is the mean value of subblock, w and k are weighted value, G_N is the mean-gradient value of the sub-block at direction N given by

$$G_N(x,y) = \sum_{x=0}^{i-1}\sum_{y=0}^{j-1}\frac{[\partial I(x,y)/\partial x_N, \partial I(x,y)/\partial y_N]}{xy} \quad (9)$$

3.4.3 Heavy handwritten image:

There are two subclasses in the heavy stroke class. One contains heavy strokes only; the other contains some faint connected strokes alongside heavy strokes. The proposed weighted method is

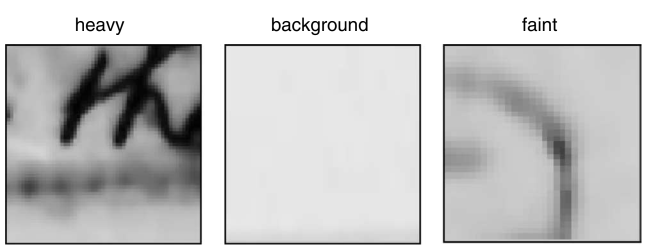

Fig. 7 *Examples of subregions containing heavy strokes, background (no strokes) and faint strokes*

applied to these two subclasses with different weighted values w and k. Practically, the formula for thresholding with a different weighted value for specific cases is:

$$T(x,y) = wM(x,y) + kG_N(x,y)$$
$$\text{where} \begin{cases} \text{if 'background'}, & w=0, k=0 \\ \text{if 'faint'}, & w=0.5, k=-1.1 \\ \text{if 'heavyonly'}, & w=0.7, k=-0.8 \\ \text{if 'heavywithfaint'}, & w=0.7, k=-1.1 \end{cases} \tag{10}$$

To avoid blocking effects at the boundary of the subregions, a smoothing matrix

$$A = \frac{1}{25}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

was convoluted with the area, which is a 5×5 window centred on the edges of each region before the mean-gradient-based thresholding method is applied.

4 Experimental results

The algorithm was trained using 300 historical images obtained from the Library of Congress, which contained considerable background noise or variation in contrast and illumination. There are two main steps in the proposed decompose algorithm:

1 Determination of stroke direction
Direction of stroke slant in the words in the handwritten historical document images is an important factor to WDES and WDVAR features. The mean direction calculated from each region affect the value of WDES and WDVAR directly.
2 Classification of local region
Handwritten historical document images always have different content information in different regions. To apply the best appropriate threshold method on different regions, it is necessary to classify the regions. They can be classified into three classes: A Region with no strokes – background region; B Region with faint strokes – faint region; or C Region with heavy strokes – heavy region.

4.1 Stroke direction of image

The stroke direction of an image is measured by convolving the eight direction matrixes (shown in Fig. 6: $G0$–$G7$) with the local area of the image one by one to obtain matrices $M0$–$M7$, respectively. Matrices $M0$–$M7$ are the same size matrixes and contain grey-scale value with direction information on each pixel (point). The tracing of the direction at each pixel (point) in pseudocode is shown in Fig. 8.

for $i = 1:1:x$,
  for $j = 1:1:y$,

  $\text{TempltMatrix}(i,j)$
   $= \text{Maximum}\{\, a(i,j) \quad b(i,j) \quad c(i,j) \quad d(i,j)$

  $e(i,j)f(i,j)g(i,j)h(i,j)\}$;

  if $\text{TempltMatrix}(i,j) = a(i,j)$, then $N(i,j) = 0$;
  if $\text{TempltMatrix}(i,j) = b(i,j)$, then $N(i,j) = 1$;
  if $\text{TempltMatrix}(i,j) = c(i,j)$, then $N(i,j) = 2$;

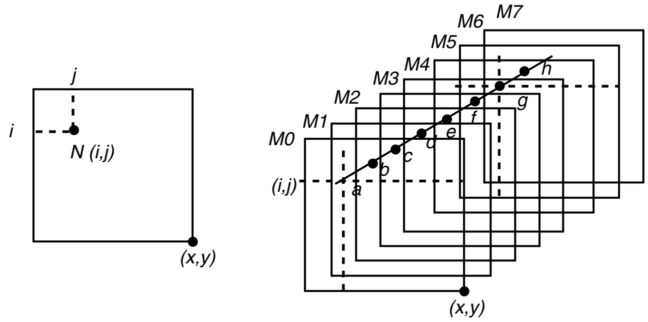

Fig. 8 *Tracing of point direction*

  if $\text{TempltMatrix}(i,j) = d(i,j)$, then $N(i,j) = 3$;
  if $\text{TempltMatrix}(i,j) = e(i,j)$, then $N(i,j) = 4$;
  if $\text{TempltMatrix}(i,j) = f(i,j)$, then $N(i,j) = 5$;
  if $\text{TempltMatrix}(i,j) = g(i,j)$, then $N(i,j) = 6$;
  if $\text{TempltMatrix}(i,j) = h(i,j)$, then $N(i,j) = 7$;
 end
end

N is pint direction matrix.

The three subimages in Fig. 4 are 64×64 pixel local areas. They are in 256-grey-level TIFF image format. The comments in Fig. 4, 'Direction = 1' means the direction of word stroke is $G1$ (refer to Fig. 5), which is equal to 45°; 'Direction = 4' means the direction of word stroke is $G4$, which is equal to 180°; 'Direction = 3' means the direction of word stroke is $G3$, which is equal to 135°.

Practically, in the experiment, 64×64 is the best size to measure the stroke direction. The result is not accurate if the size of the local area is too small, and too many calculations will be required to get a very small accurate result if the local area size is too big.

4.2 Local region classification

Classification of the local decomposed subregions is a major step of the decompose algorithm. Historical handwritten document images normally contain varying contrast, noise and shade of strokes through the image. To apply the best appropriate threshold method to each region, classification of each local region is needed. Some experimental results from the 300 training images are shown in Table 2 and Fig. 9.

The left part of Fig. 9 shows two background class images, which contain only noise and no stroke information. It was observed that background areas exhibit low edge strength and low mean-gradient value, but may have high variance value, which is usually produced by noise or meaningless information.

The middle part of Fig. 9 shows two faint stroke areas, which include noise and faint handwritten strokes. This area contains stronger edge strength, variance and higher mean-gradient value than a background area.

The right part of Fig. 9 shows two areas that contain heavy strokes. These regions have strong edge strength, strong variance and high mean-gradient value. The upper image in the heavy class contains only heavy strokes. Compared to the upper image, the lower image contains not only heavy strokes but also a light stroke connecting 'e' and 'n'. This faint stroke results in a lower mean-gradient value.

4.3 Decompose algorithm

From an aesthetic and subjective point of view, the decompose structure performs better than other single-stage local methods. It detects feature vectors of different areas

Table 2: Experiment result of area classification from training image group

	Class name		Heavy strokes	
Feature name	Background	Faint strokes	With some faint strokes	Only heavy strokes
Edge strength	$1 \le ES \le 13$	$14 \le ES \le 40$	$ES \ge 41$	
Variance	$1 \le V \le 30$	$10 \le V \le 44$	$V \ge 45$	
Mean gradient	$1 \le G \le 2$	$3 \le G \le 10$	$3 \le G \le 10$	$G \ge 10$

Note: *ES*=edge strength, V=variance; G=mean gradient

and then applies appropriate methods to avoid losing important useful information.

The decompose structure is highly effective for the images that contain different conditions at different locations. The proposed mean-direction gradient method can retain faint strokes by extracting the direction information from the image. The new method has superior performance on all test images compared to the other six methods evaluated. Some results of the decompose algorithm and other six methods when applied to the original image of Fig. 1 are shown in Fig. 10. Other detailed images can be viewed at http://www.ntu.edu.sg/home5/p145806718/Decompose.htm

Of these six methods, Bernsen's method works well on clear background and high contrast historical images, but it is a contrast-based method and so is sensitive to the noise in the images. ETM's method uses a fixed value to determine the difference between two windows. It works well for both faint and heavy handwriting, but failed when there was a noisy background. Yanowitz's method can retain very detailed strokes but still includes much useless noise. The improved Niblack technique can retain detail stroke but is sensitive to noise. QIR and Otsu's techniques only work well on bimodal histogram images and so do not perform well on these degraded document images.

300 historical images were selected from the Library of Congress online database to train the algorithms. The images were chosen to have varying resolutions, sizes and contrast to ensure correct comparison of performance between the algorithms. In these selected images, some historical images are still having acceptable quality even through they were created many years ago. The images were characterised by high resolution of the scanned images with varying contrast of the handwriting.

The decompose algorithm was evaluated and compared with these six thresholding algorithms on a further 300 historical images selected from the Library of Congress.

The standard measure, recall [30], was used to quantitatively compare the relative performance of the proposed methods at retaining the word information in the documents. Recall is defined as:

$$\text{recall} = \frac{\text{correctly detected words}}{\text{total words}}$$

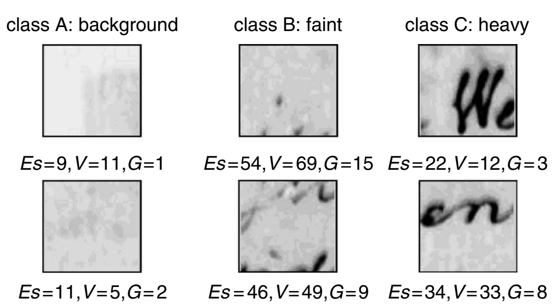

Fig. 9 *Local region classification and feature extraction*

Table 3 shows the recall value of the seven threshold methods. In the Table, the 300 example historical images are classified into six groups, with 50 images per group. The grouping was random. Average Recall values are presented as mean values using 50 images per group for each of the seven thresholding methods. The Table also shows the average recall value of the 300 example images. From Table 3 and Fig. 11, it is apparent that the decompose algorithm produced significantly better recall results than the other six methods.

5 Discussion and conclusion

A number of techniques have previously been proposed for thresholding document images. However, none of them can provide ideal results for degraded historical handwritten document images. Because of varying contrast and noise conditions across the image, applying the same processing to the whole image is not flexible and will not produce good results.

In this paper a new thresholding structure called the decompose-threshold approach is proposed and compared against some existing global and local algorithms. The proposed approach is a local adaptive analysis method, which uses local feature vectors to find the best approach for thresholding a local area. Appropriate weighted values are selected automatically for thresholding specific types of document image under investigation. The original image is recursively broken down into subregions using quad-trees until an appropriate weighted thresholding method can be applied to each of the subregion.

The proposed decompose algorithm analyses the features extracted from local regions to determine the appropriate weighted value for a threshold method based on mean gradient. The three feature vectors can effectively describe the texture characteristic of the local subregions.

A set six thresholding algorithms [1, 5, 13-16], which have shown a superior level of performance on 'difficult' images were chosen for incorporation in this work for performance comparison purposes.

The improved Niblack method [13] works well on heavy handwriting image blocks, because it is sensitive to edge information. Bernsen's method [14] also works well on heavy handwriting image blocks because a heavy image has high contrast. Compared to the other five methods, it also works well on faint handwriting images. Yanowitz and Bruckstein's [16] method works well on various kinds of document image, but is not good when there is a noisy background, especially in images where there is double-sided noise (where the ink has seeped through the paper from the other side). ETM [15] works well for both faint and heavy handwriting images. Otsu's method [1] can achieve good performance with simple documents where the background and foreground are clearly distinct in the histogram. However, Otsu's algorithm is very time-consuming for image binarisation because of its inefficient formulation of the between-class variance,

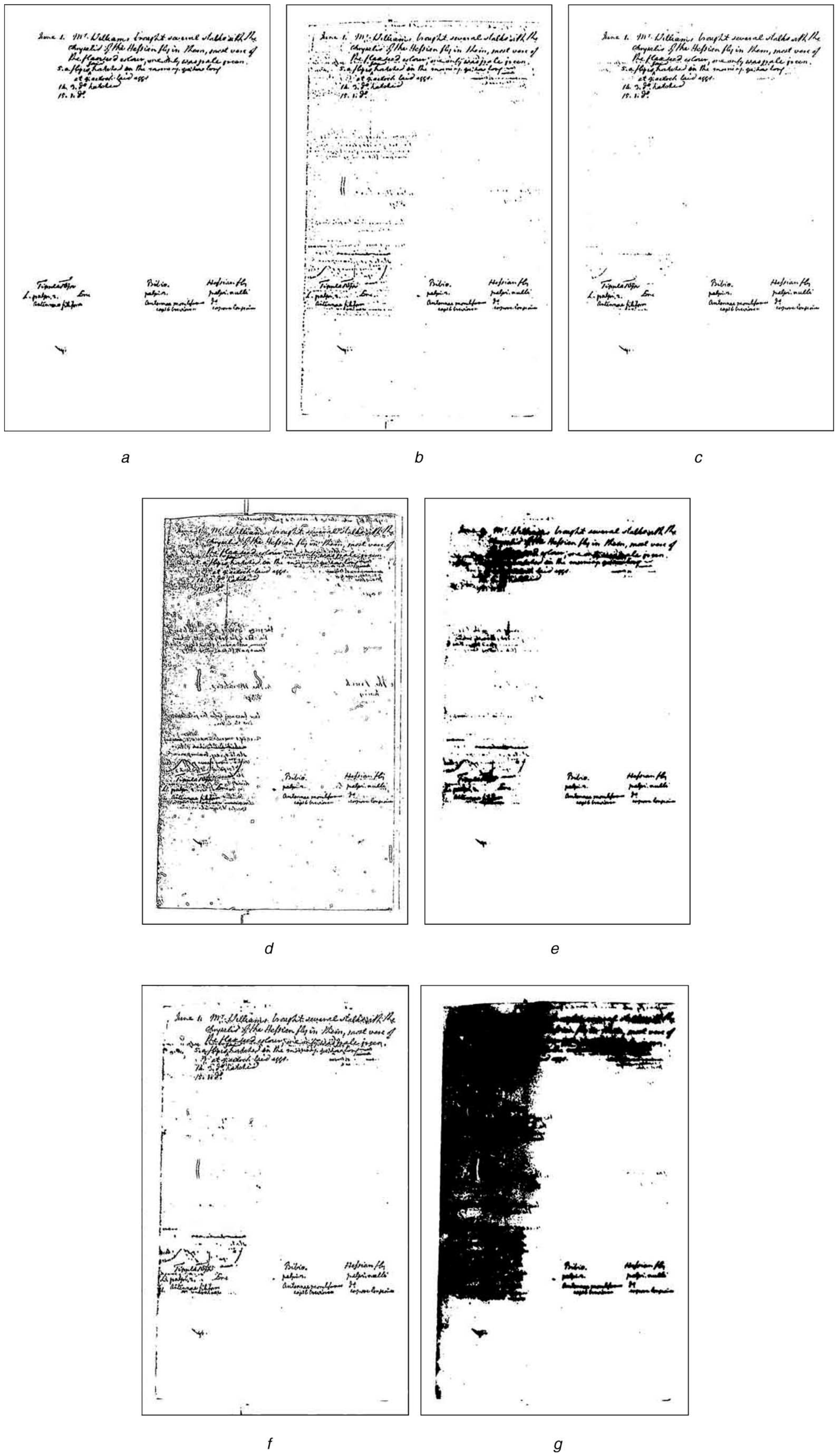

Fig. 10

a Decompose algorithm
b ETM algorithm
c Improved Niblack's technique
d Bernsen's method
e Otsu's technique
f Yanowitz's algorithm
g QIR technique

Table 3: Comparison of Recall for the 7 Algorithms

Image number	Seven algorithms Bersen	ETM	Im-Niblack	Proposed decompose	Otsu	QIR	Yanowitz
0–49	84%	89%	94%	97%	89%	91%	94%
50–99	62%	77%	87%	91%	70%	72%	87%
100–149	66%	78%	84%	89%	68%	74%	87%
150–199	65%	77%	83%	93%	67%	74%	87%
200–249	60%	75%	80%	92%	63%	70%	86%
250–299	56%	71%	81%	93%	59%	68%	89%
Average	65%	78%	85%	92%	69%	75%	88%

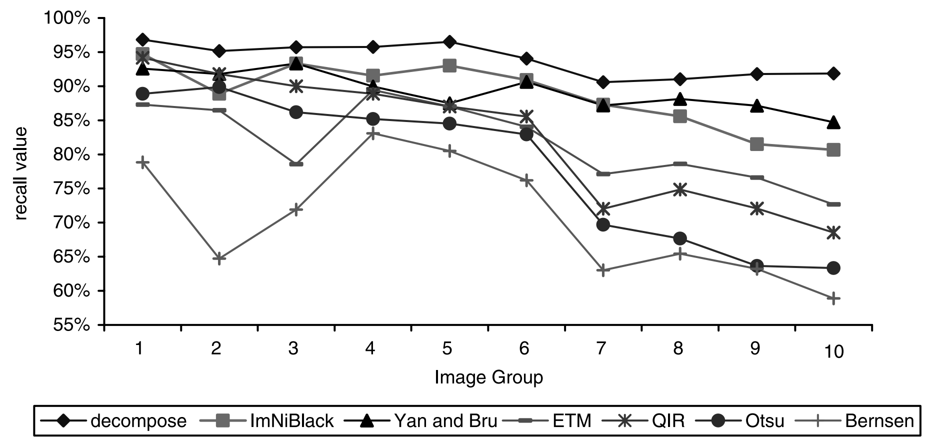

Fig. 11 *Evaluation of seven algorithms in ten groups (30 images/group)*

and the performance varies with data sets. QIR works well for bimodal histogram images.

Bersen's, Yanowitz's, Niblack's, Otsu's, QIR and the ETM thresholding methods have been evaluated and compared in [28, 29].

Degraded historical images often exhibit varying image qualities. From oberservations during the experiments reported in this paper, degraded historical handwritten images normally contain the following characteristic: varying contrast; varying stroke quality; many marks or blotches which do not contain any information. Satisfactory thresholding results can rarely be obtained if only a single global or local method is applied to the whole image. The decompose algorithm is demonstrated as effective at improving the result. It uses local feature vectors to analyse and find the best approach to threshold a local area. Instead of employing a single thresholding algorithm, automatic selection of an appropriate algorithm for specific types of subregions of the document is performed. The original image is recursively broken down into subregions using quad-tree decomposition until a suitable thresholding method can be applied to each subregion. The three feature vectors proposed in this paper can accurately classify the different regions so that the appropriate weighted value can be applied to the mean-gradient-based threshold method. The new weighted mean-gradient-based threshold method is based on the local mean-gradient value for the word direction. The new threshold method is an improved version of that by Leedham *et al.* [28].

Future work will concentrate on feature vectors to describe accurately the local texture properties. The algorithm works well on handwritten document containing text but does not work well on document images with big patterns or pictures. More complete evaluation of the decompose method could investigate its applicability to a wider range of difficult document image, such as bank cheques and newspaper images.

6 References

1 Otsu, N.: 'A threshold selection method from grey level histogram', *IEEE Trans. Syst. Man Cybern.*, 1979, **9**, (1), pp. 62–66
2 Cheriet, M.: 'Extraction of handwritten data from noisy grey-level images using a multi-scale approach', *Int. J. Pattern Recognit. Artif. Intell.*, 1999, **13**, (5), pp. 665–685
3 Don, H.S.: 'A noise attribute thresholding method for document image binarization'. Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), 1995, pp. 231–234
4 Liu, Y., and Srihari, S.N.: 'Document image binarization based on texture features', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (5), pp. 540–544
5 Solihin, Y., and Leedham, C.G.: 'Integral ratio: a new class of global thresholding techniques for handwriting images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999, **21**, (8), pp. 761–768
6 Pun, T.: 'Entopic thresholding: a new approach', *Comput. Vis., Graph. Image Process.*, 1981, **16**, pp. 210–239
7 Kapur, J.N., Sahoo, P.K., and Wong, A.K.C.: 'A new method for gray-level picture thresholding using the entropy of the histogram', *Comput. Vis., Graph. Image Process.*, 1985, **29**, pp. 273–285
8 Pal, N.R., and Pal, S.K.: 'Object extraction from image using higher order entropy'. Proc. Int. Conf. on Pattern Recognition (ICPR), 1988, Vol. 1, pp. 348–350
9 Chang, C.I., Chen, K., Wang, J., and Althouse, M.L.G.: 'A relative entropy-based approach to image thresholding', *Pattern Recognit.*, 1994, **27**, (9), pp. 1275–1289
10 Wang, J.W., Yingzi, E., Du Chein, I.C., and Thouin, P.D.: 'Relative entropy-based methods for image thresholding'. IEEE Int. Symp. on Circuits and Systems (ISCAS), 2002, Vol. 2, pp. 265–268
11 Brink, A.D.: 'Minimum spatial entropy threshold selection', *IEE Proc. Vis. Image Signal Process.*, 1995, **142**, (3), pp. 128–132
12 Niblack, W.: 'An introduction to digital image processing' (Prentice Hall, Englewood Cliffs, NJ, USA, 1986), pp. 115–116
13 Zhang, Z., and Tan, C.L.: 'Restoration of images scanned from thick bound documents'. Proc. Int. Conf. on Image Processing (ICIP), 2001, Vol. 1, pp. 1074–1077
14 Bernsen, J.: 'Dynamic thresholding of grey level images'. Proc. Int. Conf. on Pattern Recognition (ICPR), 1986, pp. 1251–1255
15 Eikvil, L., Taxt, T., and Moen, K.: 'A fast adaptive method for binarization of document images'. Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), 1991, pp. 435–443

16 Yanowitz, S.D., and Bruckstein, A.M.: 'A new method for image segmentation', *Comput. Vis. Graph. Image Process.*, 1989, **46**, (1), pp. 82–95
17 Wang, L., and Pavlidis, T.: 'Direct grey-scale extraction of features for character recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1993, **15**, pp. 1053–1067
18 Zhao, M., and Yan, H.: 'Adaptive thresholding method for binarization blueprint images'. Proc. 5th Int. Symp. on Signal Processing and Its Applications (ISSPA), 1999, Vol. 2, pp. 931–934
19 Chen, T., and Takagi, M.: 'Run length coding based new approach to automatic image thresholding'. IEEE Int. Symp. on Circuits and Systems (ISCAS), 1993, pp. 555–558
20 Manjunath, B.S., and Chellappa, R.: 'A computational approach to boundary detection'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), 1991, pp. 358–363
21 Djeziri, S., Nouboud, F., and Plamondon, R.: 'Extraction of signatures from check background based on a filiformity criterion', *IEEE Trans. Image Process.*, 1998, **7**, pp. 1425–1438
22 Yasuda, Y., Dubois, M., and Huang, T.S.: 'Data compression for check processing machine', *Proc. IEEE*, 1980, **68**, (7), pp. 874–885
23 Gorman, L.O.: 'Binarization and multi-thresholding of document images using connectivity', *Graph. Models Image Process*, 1994, **56**, (6), pp. 494–506
24 Gu, L., Kaneko, T., Tanaka, N., and Haralick, R.M.: 'Robust extraction of characters from colour scene image using mathematical morphology'. Proc. Int. Conf. on Pattern Recognition (ICPR), 1998, pp. 1002–1004
25 Weszka, J.S., and Rosenfield, A.: 'Threshold evaluation techniques', *IEEE Trans. Syst. Man Cybern.*, 1985, **29**, pp. 377–393
26 Wang, S., and Haralick, R.M.: 'Automatic multi-threshold selection', *Comput. Vis. Graph. Image Process*, 1984, **25**, pp. 46–67
27 Kohler, R.: 'A segmentation system based on thresholding', *Comput. Graph. Image Process.*, 1981, **15**, pp. 319–338
28 Leedham, C.G., Chen, Y., Takru, K., Tan, J., and Li, M.: 'Comparison of some thresholding algorithms for text/background segmentation in difficult document images'. Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), 2003, Vol. 2, pp. 859–865
29 Trier, O.D., and Jain, A.K.: 'Goal-directed evaluation of binarization methods', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1995, **17**, (12), pp. 1191–1201
30 Junker, M., and Hoch, R.: 'On the evaluation of document analysis components by recall, precision, and accuracy'. Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), 1999, pp. 713–716
31 Lam, S.W.: 'Texture feature extraction using grey level gradient based co-occurrence matrices'. Proc. IEEE Int. Conf. on Systems, Man and Cybernetics, 1996, Vol. 1 (14-17), pp. 267–271