# Supervised Range-Constrained Thresholding

Qingmao Hu, Zujun Hou, and Wieslaw L. Nowinski

***Abstract*—A novel thresholding approach to confine the intensity frequency range of the object based on supervision is introduced. It consists of three steps. First, the region of interest (ROI) is determined in the image. Then, from the histogram of the ROI, the frequency range in which the proportion of the background to the ROI varies is estimated through supervision. Finally, the threshold is determined by minimizing the classification error within the constrained variable background range. The performance of the approach has been validated against 54 brain MR images, including images with severe intensity inhomogeneity and/or noise, CT chest images, and the Cameraman image. Compared with nonsupervised thresholding methods, the proposed approach is substantially more robust and more reliable. It is also computationally efficient and can be applied to a wide class of computer vision problems, such as character recognition, fingerprint identification, and segmentation of a wide variety of medical images.**

***Index Terms*—Histogram, region of interest (ROI), robust thresholding, supervision, thresholding.**

## I. Introduction

THRESHOLDING is a popular technique for image segmentation and has been widely applied to a number of image processing and computer vision applications for preliminary segmentation. The implicit assumption is that the object and background are different in intensity values. The key issue is, thus, the threshold selection, which can be interactive or automatic, local, or global.

A simple method for automatic threshold selection is to detect the valley of the gray-level histogram [1]–[4]. More often, the threshold is determined by optimizing some criteria from the histogram. One criterion is to minimize the error between the segmented and the observed images. This error can be characterized in different ways, hence resulting in different thresholding methods. The Otsu thresholding method [5] makes use of the Euclidean distance. Li and Lee [6] employ the cross entropy. The minimum error thresholding method [7]–[11] minimizes the Bayes error. There are also efforts to assess the error using the index of fuzziness [12], [13]. Yan [14] has derived a unified formulation of the Otsu method [5], the cross entropy method [9], and the Huang and Wang method [12].

Another popular criterion is the maximum entropy principle, which was first applied by Pun [15] to threshold selection. After that, it has been improved in many ways. Wong and Sahoo [16] constrained the solution by maximizing the region homogeneity and the object boundary. Saha and Udupa [17] derived the optimum threshold by jointly maximizing the class uncertainty and the region homogeneity. Cheng *et al.* [18] applied fuzzy rules to derive the fuzzy-entropy thresholding methods.

The spatial information of image data is useful for object interpretation and has been utilized to guide the threshold selection. Ahuja and Rosenfeld [19] encoded the spatial information in the co-occurrence matrix. Other popular measures to characterize the spatial information include homogeneity, gradient and mean value. Wang *et al.* [20] selected the threshold based on the two-dimensional histogram spanned by the intensity and the mean value.

Generally, *a priori* knowledge can improve the performance of image processing applications. Here, knowledge refers to any kind of information that helps discriminate one class type from another [21]. Sezan *et al.* [22] presented an algorithm for digital chest radiography enhancement, where a lung/mediastinum threshold was determined based on rules derived from 50 training radiographs. This thresholding method facilitates anatomically selective modification of the image intensity and was demonstrated to be superior to the usual histogram-based thresholding methods. In the histogram of a document image, text is normally the darkest item. With this domain knowledge, Wu and Manmatha [4] proposed to select the threshold from the first valley of the histogram. The method was demonstrated to be even better than the adaptive thresholding method without domain knowledge [4].

Lee and Yang [23] proposed to handle multimodal distributed noise and small objects by trying to find the gray-level distribution of the background as well as the object. As pointed out in [23], no procedure can effectively separate the object from the background without some prior knowledge of their relationship and properties. A natural way to incorporate knowledge is through supervision which is more general than training.

In this paper, we propose a supervised thresholding approach which confines the analysis in the histogram of a region of interest (ROI), where the proportion of the background to the ROI varies in a range which can be estimated through supervision. Here, the supervision can be in the form of training of samples or approximation based on prior knowledge or visual assessment. The threshold is selected as the gray level minimizing the classification error within the frequency range of background. We have done quantitative and qualitative validation of the proposed approach against 54 brain MR images, as well as CT chest images and the Cameraman image. Comparison has been made with respect to the Otsu method [5], the minimum error thresholding method [9] and the fuzzy c-partition (FCP) method [18]. The results demonstrate that our approach is computationally efficient, consistent and reliable.

Manuscript received June 30, 2004; revised February 1, 2005. This work was supported by the Biomedical Research Council, Agency for Science, Technology, and Research, Singapore. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Eli Saber.

The authors are with the Biomedical Imaging Laboratory, Agency for Science, Technology, and Research, Singapore 138671 (e-mail: huqm@sbic.a-star.edu.sg; houzj@sbic.a-star.edu.sg; wieslaw@sbic.a-star.edu.sg).

Digital Object Identifier 10.1109/TIP.2005.860348

The paper is organized as follows. In Section II, we describe the algorithm for threshold selection from the gray-level histogram and formulate three range-constrained thresholding methods. Section III is devoted to approximating the parameters through supervision. Then, we present the validation as well as the comparison with other popular thresholding methods in Section IV. Discussion and conclusion are in Section V.

## II. Methods

This study addresses the problem of two-level thresholding, which can be regarded as a mapping from features to class labels. Without losing generality, we consider an image $f(x)$ having $L$ gray levels $r_0 < r_1 < \cdots < r_{L-1}$, where $f(x)$ is the gray level at position $x$. Let the class labels be denoted as $C_o$, $C_b$, which represent the classes of object and background, respectively. For convenience of description, it is assumed that *the object has higher intensity values than the background*. Then, the thresholding operation is defined as follows:

$$C_x = \begin{cases} C_o, & \text{if } \phi(x) > \theta \\ C_b, & \text{otherwise} \end{cases}$$

where $C_x$ is the label at position $x$, $\theta$ is the threshold, and $\phi(x)$ is the decision function. In this study, we aim to develop a simple yet robust thresholding method and the decision is purely based on the intensity information, that is, $\phi(x) = f(x)$. Then, the thresholding operation is equivalent to labeling the intensity level $r_i$. Thus, we arrive at the simplified thresholding operation

$$C_{r_i} = \begin{cases} C_o, & \text{if } r_i > \theta \\ C_b, & \text{otherwise} \end{cases}$$

where $i = 0, \cdots, L-1$.

To assess the performance of a thresholding method, we can identify two types of errors. Type I is the false negative (FN) error $\epsilon_1$, which classifies the object element to class $C_b$. Type II is the false positive (FP) error $\epsilon_2$, which labels the background element as class $C_o$. Let $P_o$, $P_b$ be the prior probabilities of object and background, respectively, the class probability density be $p(s|C_k)$, $k \in \{o, b\}$, and $\Omega_k$ be the decision region in which the element belongs to class $C_k$. Then type I error can be expressed as

$$\epsilon_1 = P_o \int_{\Omega_b} p(s|C_o) ds. \tag{1}$$

Similarly, type II error is

$$\epsilon_2 = P_b \int_{\Omega_o} p(s|C_b) ds. \tag{2}$$

An ideal threshold would minimize both types of classification errors. The minimum error threshold [8] minimizes the Bayes error $(\epsilon_1 + \epsilon_2)$. In real applications, the costs of the two classification errors are often different. For example, to apply a thresholding method to separating brain tissues from background, the FN error is usually not easy to be detected and corrected by subsequent processes, and always tends to result in negative bias upon tissue volume measurement. On the other hand, the impact of the FP error on tissue volume measurement could be decreased by subsequent processes if this error is in a reasonable range. Therefore, a more reasonable threshold might be that is derived from the Neyman–Pearson rule [24], that is, either to minimize $\epsilon_2$ with constraint on $\epsilon_1$

$$\theta_{N-P} = \min_{\theta} \{\epsilon_2 | \epsilon_1 = \alpha\}$$

where $\alpha$ is the tolerable upper bound of $\epsilon_1$, or to minimize $\epsilon_1$ with constraint on $\epsilon_2$, depending on which cost of classification error is higher. Alternatively, one can modify the implicit 0-1 cost in [8] to derive a generalized minimum error threshold.

In the following, we will present a supervision-based approach for threshold selection which implicitly takes into account the cost of classification error. Because the FN error $\epsilon_1$ is a monotonically increasing function of the threshold $\theta$, large FN error results from an "inappropriately" large threshold. In order to avoid the serious over segmentation of images (large FN error), we hope that there can be an upper bound $r_{\text{high}}$ for the threshold. Similarly, $\epsilon_2$ monotonically decreases with respect to $\theta$. We also hope that there can be a lower bound (LB) $r_{\text{low}}$ for $\theta$, so as not to select a too small threshold to yield a large FP error. In our approach, we will confine the threshold selection in the range of background, which naturally defines the lower and upper bounds of $\theta$. Details are given as follows.

First, the ROI is determined where the object is present. For instance, in the segmentation of the axial slice of the human head passing through the anterior commissure (AC) and the posterior commissure (PC) (denoted as AC-PC axial slice subsequently), the ROI is all the pixels enclosed by the skull. Because the measurement of gray matter (GM) and white matter (WM) will aid diagnosis of cerebral and cerebellar degenerations, toxic, metabolic and infective encephalopathies, the object within the ROI is the GM and WM, and the background is mainly the cerebrospinal fluid (CSF), the skull and scalp. The ROI can be seperated from the nonskull space through the following steps: 1) the usual histogram-based thresholding method is exploited to binarize the axial slice; 2) morphological closing operation is used to connect small gaps among the tissues; 3) the largest connected component is identified; and 4) holes within the component are filled. An original axial slice with serious intensity inhomogeneity is shown on the left of Fig. 1, where on the right is the ROI.

Next, the frequency range in which the proportion of the background to the ROI varies is estimated through supervision (see Section III). In the histogram of the ROI, let us denote the lower and the upper bounds of the background proportion as $H_l^b$ and $H_h^b$, respectively. The rationale behind the range estimation can be justified as follows. As our purpose is to differentiate the object from its background, different elements will play different roles in the process. Specifically, those object elements with high intensity and those background elements with low intensity are easy to identify, and the difficulty often comes from classifying the elements with medium gray levels. If the threshold is estimated based on the middle group, it will be less sensitive to outliers. In addition, more information from practice for a specific problem can be incorporated to tighten the proportion of these groups, which will help further reduce the chance for large type I/II error. Generally, the imaging

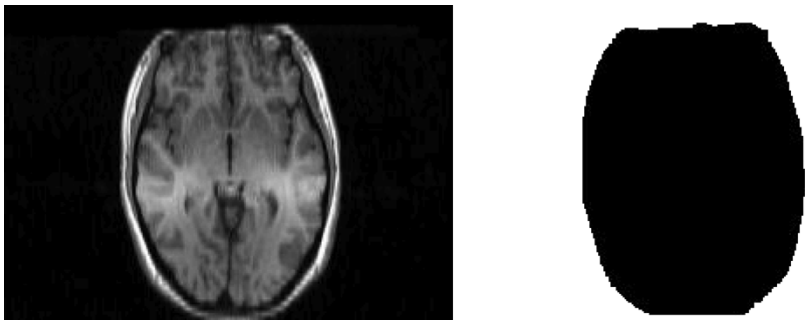


Fig. 1. (a) Axial slice. (b) The ROI is the space enclosed by the skull.

process and the image quality should be taken into account, because imperfect imaging, noise, or partial volume effect could dramatically affect the variability of ROI as well as the variability of background portion.

Finally, the threshold is determined by minimizing the classification error *within the frequency range of background* as estimated in the last step. For that purpose, we can accordingly modify the conventional histogram-based thresholding methods by incorporating into them the range constraint approach. Because the shape of histogram within the range is normally far from bimodal, the minimum error thresholding method will not be appropriate. For illustration, three methods are formulated and presented in the following. For all the three methods, the inputs are the ROI of the image to be segmented and the background lower and upper bounds in the ROI. Additional steps for each method are described below.

### A. Method 1: Range-Constrained Least Valley Detection Method (RCLVD)

If the frequency range of background is reasonably estimated, the estimated range will be located in the neighborhood of the valley that seperates the background and the object; thus, valley detection can be exploited to select the threshold.

Let $h(i)$ represent the frequency of gray level $r_i$, then the accumulative frequency $H(i)$ is $\sum_{i=r_0}^{r_i} h(i)$, and the frequency at interval $[r_m, r_n]$, $r_m < r_n$, is $\sum_{i=r_m}^{r_n} h(i)$. In real applications, due to pertubations such as noise and intensity inhomogeneity, a low-pass filtering is necessary to suppress the random fluctuation. The details are summarized as follows.

1) Specify the frequency interval (FI) $\delta h$.
2) Calculate $r_{\text{low}}$, which is the gray level corresponding to the background LB $H_l^b$
$$r_{\text{low}} = \min_i \left\{ i | H(i) \geq H_l^b \right\}.$$
3) Calculate $r_{\text{high}}$, which is the gray level corresponding to the background upper bound $H_h^b$
$$r_{\text{high}} = \min_i \left\{ i | H(i) \geq H_h^b \right\}.$$
4) Partition the region $[r_{\text{low}}, r_{\text{high}}]$ into intervals with equal frequency $\delta h$. For interval $j$, denote its left and right gray levels as $r_1^j$, $r_2^j$, respectively.
5) Calculate the average interval frequency, $\bar{h}^j$ for interval $j$, by $\bar{h}^j = (H[r_2^j] - H[r_1^j])/(r_2^j - r_1^j)$.
6) Let $J$ denote the interval with minimum average interval frequency. The threshold is selected as $\theta_{\text{RCLVD}} = (r_1^J + r_2^J)/2$.

### B. Method 2: Range-Constrained Otsu Method (RCOtsu)

Let $r_k \in (r_{\text{low}}, r_{\text{high}})$ and suppose that the pixels with intensity falling in the frequency range of background are dichotomized into classes, $C_1$ and $C_2$, where $C_1$ belongs to $C_b$ and consists of pixels with gray levels $[r_{\text{low}}, r_k]$, $C_2$ belongs to $C_o$ and is composed of pixels with gray levels $(r_k, r_{\text{high}}]$. The range-constrained Otsu method minimizes the pooled variance (within-class variance)

$$\theta_{\text{RCOtsu}} = \min_{r_k} \Pr(C_1) D(C_1) + \Pr(C_2) D(C_2) \tag{3}$$

where $\Pr(\cdot)$ denotes the class probability and $D(\cdot)$ denotes the class variance.

### C. Method 3: Range-Constrained Fuzzy c-Partition Thresholding Method (RCFCP)

Let $A_b/A_o$ be the fuzzy sets of fuzzy events, *background/object*, (which denotes a fuzzy partition of the set $\{r_{\text{low}}, \cdots, r_{\text{high}}\}$,) with membership function $\mu_{A_b}/\mu_{A_o}$, respectively. The probabilities of fuzzy events are defined by

$$P(A_i) = \sum_{j=r_{\text{low}}}^{r_{\text{high}}} \mu_{A_i}(j) h(j) \tag{4}$$

where $A_i \in \{A_b, A_o\}$, and the entropy with this fuzzy partition can be calculated as

$$S = P(A_b) \log P(A_b) + P(A_o) \log P(A_o). \tag{5}$$

Let $r_{\text{low}} \leq a < c \leq r_{\text{high}}$. The membership functions can be defined as follows:

$$\mu_{A_b}(x) = \begin{cases} 1, & r_{\text{low}} \leq x \leq a \\ \frac{x-c}{a-c}, & a < x < c \\ 0, & c \leq x \leq r_{\text{high}} \end{cases} \tag{6}$$

and

$$\mu_{A_o}(x) = \begin{cases} 0, & r_{\text{low}} \leq x \leq a \\ \frac{x-a}{c-a}, & a < x < c \\ 1, & c \leq x \leq r_{\text{high}}. \end{cases} \tag{7}$$

The optimum parameters $\hat{a}$ and $\hat{c}$ are chosen to maximize the entropy $S$, and the optimum threshold is

$$\theta_{\mathrm{RCFCP}} = \frac{(\hat{a}+\hat{c})}{2}. \tag{8}$$

The advantages of the proposed RCLVD, RCOtsu, and RCFCP over their conventional counterparts are threefold. First, the ROI is used instead of the whole image, so any irregularity outside the ROI will have no influence on estimating the threshold. Second, it provides a mechanism to handle cost differences of different types of classification error in response to practical requirements. Third, by appropriately specifying the lower and upper bounds of the background proportion within the ROI through supervision and calculating their threshold through the LVD, the Otsu or the FCP method within the constrained gray level range, the proposed three methods yield substantially more robust, and more reliable segmentation (see Section V for details).

## III. Supervision to Approximate Parameters

There are two parameters associated with the proposed range-constrained thresholding methods: $H_l^b$ and $H_h^b$, which specify the frequency range of background portion within the ROI. The parameters can be obtained from supervision, which can be in the form of training when a number of sample images with ground truth are available, or in the form of approximation based on any prior knowledge or visual judgement when sample images are not available.

When a number of sample images with ground truth are available, the background proportion in each of the ground truth of the sample images within the ROI can be calculated; thus, the minimum and maximum proportions of the background can be calculated and denoted as minP and maxP, respectively. Normally, the ROI is utilized instead of the whole image to calculate the two proportions in order to narrow the range of background proportions. The difference between minP and maxP may be due to the variation in physical contents of the background or due to the inaccuracy of the determination of the ROI. The estimated minimum background proportion should be smaller than minP, and the estimated maximum background proportion should be bigger than maxP, in order for the algorithm to handle images of the similar type with larger variations in the physical proportion of the background and to tolerate inaccuracy in determining the ROI from images to be segmented. Usually (minP-vP) and (maxP+vP) are taken as the estimated minimum and maximum background proportions, where vP is a positive constant and can be in the range of 1% and 5%, depending on how typical the sample images are.

When sample images with ground truth are not available, we can try to incorporate any prior knowledge about the relative relationship between the object and background in the image to be segmented. This relationship can be expressed as the minimum and maximum proportions (minP and maxP) of the background in the ROI. minP and maxP are taken as the estimated background bounds to segment the image.

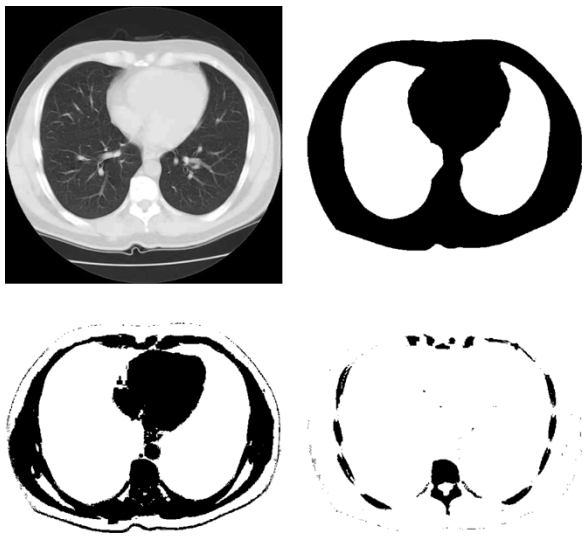


Fig. 2. On the top are (left) a CT chest image and (right) the ROI. On the bottom are (left) the thresholded results by Otsu and (right) RCOtsu.

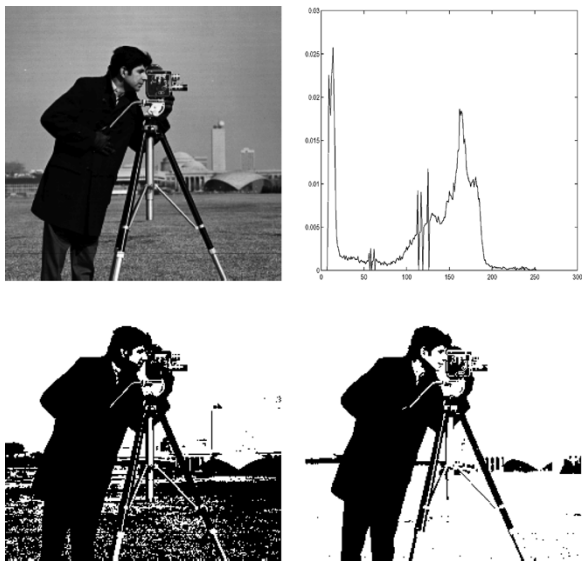


Fig. 3. On the top are (left) the Cameraman image and (right) the histogram. On the bottom are the thresholded results by (left) FCP (left) and (right) RCFCP.

When neither sample images with ground truth nor prior knowledge is available, visual judgement can be explored to estimate the background proportions within the ROI of the image to be segmented. Specifically, within the ROI, one can estimate the dark area (background) with respect to the bright area (object) by overlaying grids on the image. Generally, the more complex the shape of the background/object, the smaller grids needed to do the approximation. Suppose the approximated background proportion is backP, then the estimated minimum and maximum background proportions should be (backP-vP) and (backP+vP), respectively, where the positive constant vP is to tolerate the estimation error of backP and the gray level overlap between the background and object.

### A. Training From Sample Images With Ground Truth

For images with similar object contents, training can be used to obtain the two parameters. For a brain MR AC-PC axial

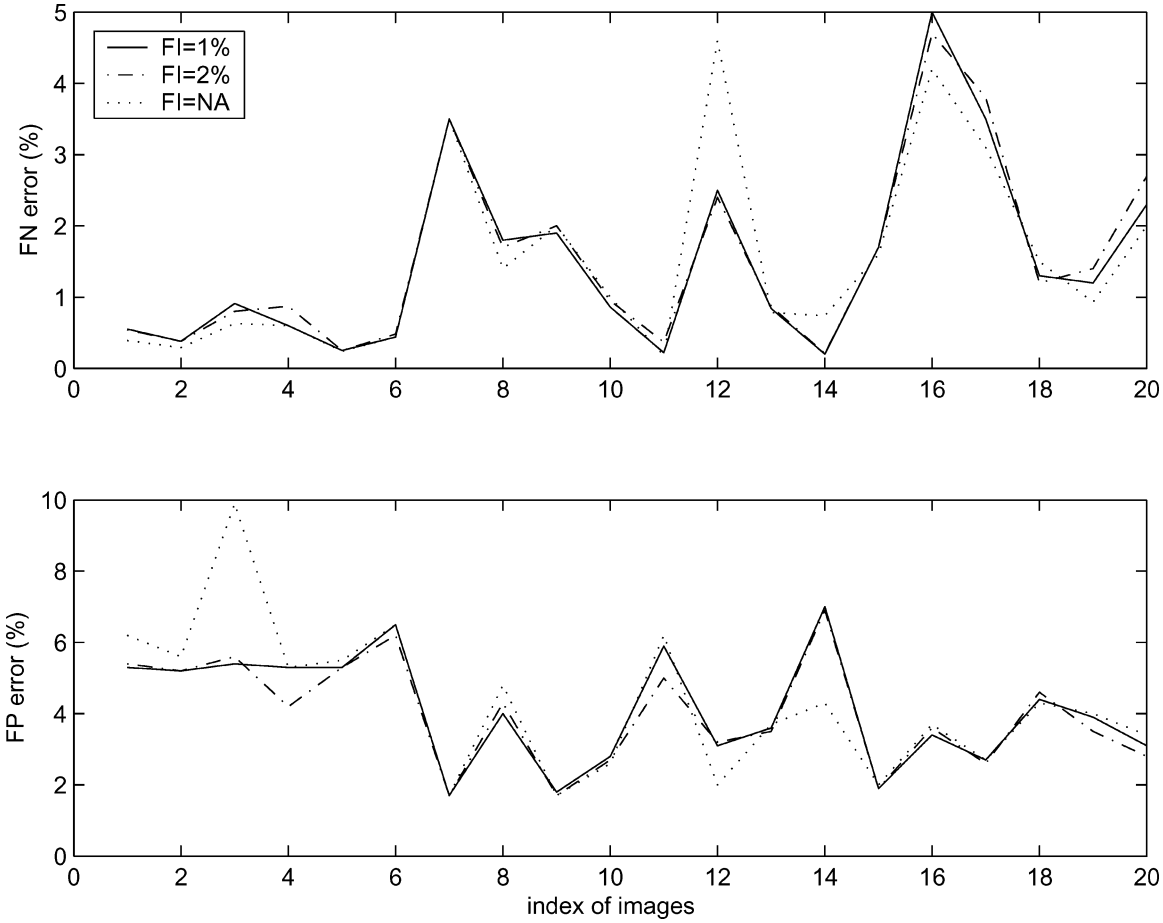



Fig. 4. Effect of the FI on the FN and FP error rates of the RCLVD method.

slice, anatomically, the proportion of GM/WM within the skull should vary within a narrow range. This anatomical range is dependant upon individuals being scanned. Noise and intensity inhomogeneity cause the failure of conventional methods in segmenting the GM/WM. Fortunately, the ground truth GM/WM for 20 normal data from the Internet Brain Segmentation Repository (IBSR, http://www.cma.mgh.harvar.edu/ibsr) and 18 phantom data [25] is available. It is found that the background portion largely falls in the range of $16 \sim 23\%$ ($\mathrm{minP} = 16\%$ and $\mathrm{maxP} = 23\%$). Considering the impact of pertubations of even more serious intensity inhomogeneity and noise, the background proportion was estimated as a more relaxed bound of $14 \sim 25$ ($vP = 2\%$), i.e., $H_l^b = 14\%$, $H_h^b = 25\%$, and the ROI is the pixels enclosed by the skull.

### B. Approximating Parameters Using Prior Knowledge

For an individual image to be segmented, any prior knowledge about the relative relationship between the object and background within the ROI can be used to approximate the two parameters. Take for example, a CT chest image (Fig. 2, top left) to segment ribs and vertebrae. Within the ROI (the dark region of Fig. 2, top right), there will be 1 to 3 pixels along the inner boundary (the dark pixels enclosing the two bright holes) of the ROI according to anatomy. Suppose the number of dark pixels in the ROI is $n_1$, and the number of inner boundary pixels is $n_2$. The lower background bound corresponds to the 3 rib/vertebra pixels along the inner boundary, i.e., $(1-3n_2/n_1)$. Similarly, the upper boundary bound corresponds to the 1 rib/vertebra pixel along the inner boundary, i.e., $(1 - n_2/n_1)$. For Fig. 2, the two bounds are calculated as $H_l^b = 80\%$ and $H_h^b = 94\%$.

### C. Approximating Parameters Using Visual Judgement

When no prior knowledge can be used to approximate the two parameters for segmenting an individual image, visual assessment can be applied. Fig. 3 shows the Cameraman image ($256 \times 256$, 8-bit resolution), where the original image is on the top left. The ROI is taken as the whole picture and the object is the cameraman. If we divide the image into four squares with equal size, we immediately find that the proportion of the object is around 25%. So, the background proportion is around 75%. In this case, the visual approximation is quite rough, and there are also gray level overlap between the object and the background; thus, we can allow for a 10% deviation ($\mathrm{vP} = 10\%$) to account for the roughness of estimation and gray level overlap, i.e., $H_l^b = 65\%$, $H_l^h = 85\%$. Segmentation is also performed with $\mathrm{vP} = 5\%$ and $\mathrm{vP} = 15\%$ to yield similar results.

## IV. Results

To evaluate the performance of the proposed supervised range-constrained thresholding methods, we have investigated in detail 54 brain MR images. In addition, we have explored their applications to CT chest images and a nonmedical image

(the Cameraman image). The brain MR images are AC-PC axial slices from a variety of brain MR data sets, including 18 phantom data, 20 IBSR, and 16 patient cases (collected in Singapore and Japan). The AC-PC axial slice is derived from the midsagittal plane of the human cerebrum [26] and contains four out of six modified Talairach landmarks [27] which are necessary for the Talairach transformation [28]. The latter is important in functional neurosurgery, neuroradiology, and human brain mapping [29].

In this study, we have investigated the three proposed range-constrained thresholding methods (RCOtsu, RCFCP, and RCLVD) and compared them with the Otsu method (Otsu) [5], the minimum error method (MinErr) [9], and the FCP method [18]. All methods are implemented on the histogram of ROI for two-level thresholding. The MinErr method exploits the algorithm in [6] and the bi-modal test in [8].

In this section, a detailed validation against the brain MR images is first presented, followed by the qualitative and quantitative comparisons. At last, applications to CT chest images and the Cameraman image are demonstrated.

For the RCLVD method, there is another parameter, the FI. The impact of varying the FI on the performance of this method is shown in Fig. 4, where the FI varies in the level of 1%, 2%, and NA (not applied, which means that we compare the frequency gray level by gray level). The computation of error rates will be detailed in the following subsection on quantitative comparison. From Fig. 4, it can be seen that three levels are overall very similar except for the NA level which results in large error for some images (#3 and #12).

Based on these observations, the parameters are set as $H_l^b = 14\%$, $H_h^b = 25\%$ and $\delta h = 1\%$ in the following comparisons and experiments of the brain MR images.

### A. BrainWeb Phantom Data Sets

Fig. 5 shows the results for the AC-PC axial slice of the phantom data without noise and inhomogeneity. Top left is the original image and top right is the histogram of the ROI. It can be seen that the histogram of the phantom data has sharp and well seperated peaks, which from left to right correspond to CSF, GM (including the small peak), and WM, respectively, from which an "optimum" threshold can be determined by an expert: a threshold for background removal would be appropriate with value around $60 \sim 70$. Fig. 5(a)–(e) displays the thresholding results for the method of (a) RCOtsu, (b) Otsu, (c) RCFCP, (d) FCP, and (e) RCLVD, and the thresholds are: (a) 61, (b) 84, (c) 58.5, (d) 117.5, and (e) 72, respectively. The MinErr method failed to calculate a threshold because the shape of the histogram is away from bimodal. The FCP method [Fig. 5(d)] obviously over-segmented the image and the result by the Otsu method [Fig. 5(b)] was slightly over-segmented. Comparatively, the thresholds selected by the three proposed range-constrained methods are more consistent with the intuitive threshold, and visual differences among them are small.

Fig. 6 is the results on the AC-PC axial slice of the phantom data with a 9% noise level (which is defined by the ratio of the standard deviation between the noise and the brightest tissue) and 40% inhomogeneity (which simulates a multiplicative field ranging from 0.8 to 1.2 over brain area). Although this example

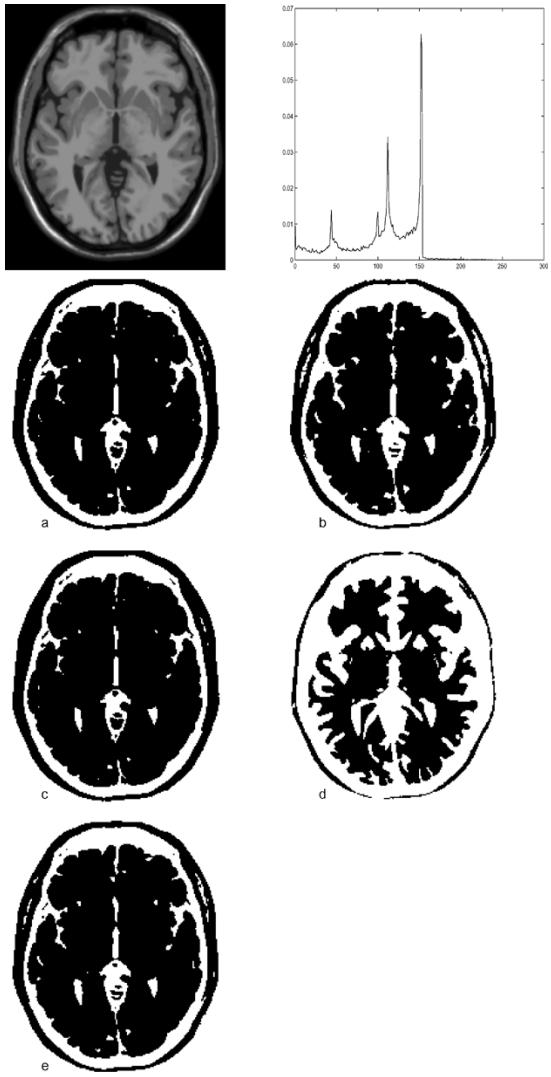



Fig. 5. Top left is the original phantom image without noise and inhomogeneity, and top right is the histogram of the ROI. (a)–(e) Thresholded results. (a) RCOtsu with threshold 61. (b) Otsu with threshold 84. (c) RCFCP with threshold 58.5. (d) FCP with threshold 117.5. (e) RCLVD with threshold 72.

simulates the case with imperfect imaging condition, it can be seen from the histogram of the ROI (top right) that the image quality is actually pretty good, as the peaks corresponding to GM/WM are distinctive from others. Intuitive examining reveals that a threshold of value around 60 is appropriate. Fig. 6(a)–(e) shows the thresholding results for methods of (a) RCOtsu, (b) Otsu, (c) RCFCP, (d) FCP, and (e) RCLVD, of which the thresholds are: (a) 54, (b) 79, (c) 55.5, (d) 107.5, and (e) 58, respectively. Again, the threshold by the FCP method is too large and the Otsu method misclassified some GM as background. The results produced by the three proposed range-constrained methods are pretty reasonable perceptually.

### B. IBSR Data Sets

For IBSR data sets, the MinErr failed to calculate the threshold for ten out of 20 data sets. Among the ten threshold-reported data, five thresholds determined by the MinErr method are very close to those found by the three range-constrained methods. One example is shown in Fig. 7. The histogram of the ROI

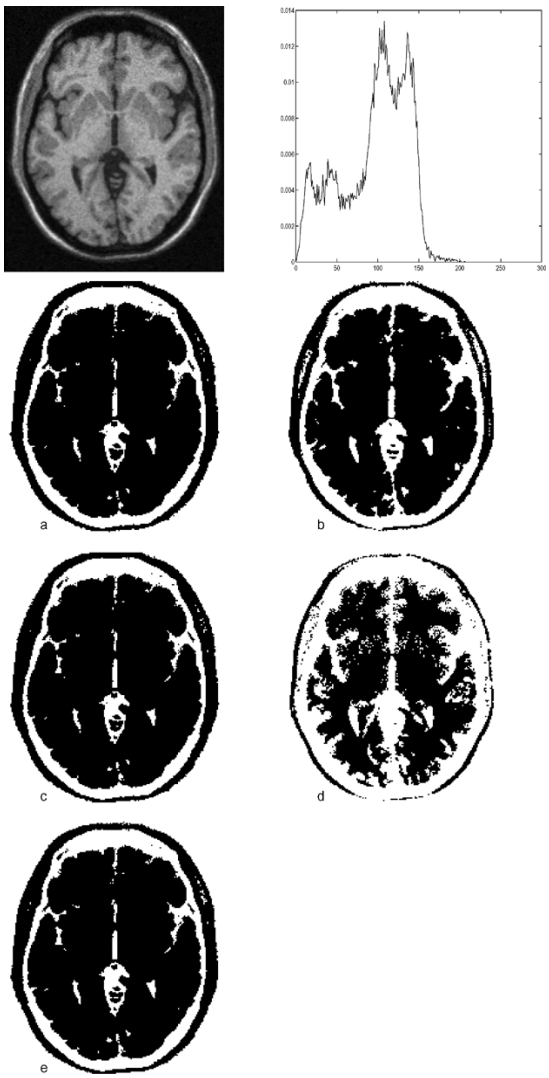



Fig. 6. Top left is the original phantom image with a 9% noise level and 40% inhomogeneity, and top right is the histogram of the ROI. (a)–(e) Thresholded results. (a) RCOtsu with threshold 54. (b) Otsu with threshold 79. (c) RCFCP with threshold 55.5. (d) FCP with threshold 107.5. (e) RCLVD with threshold 58.

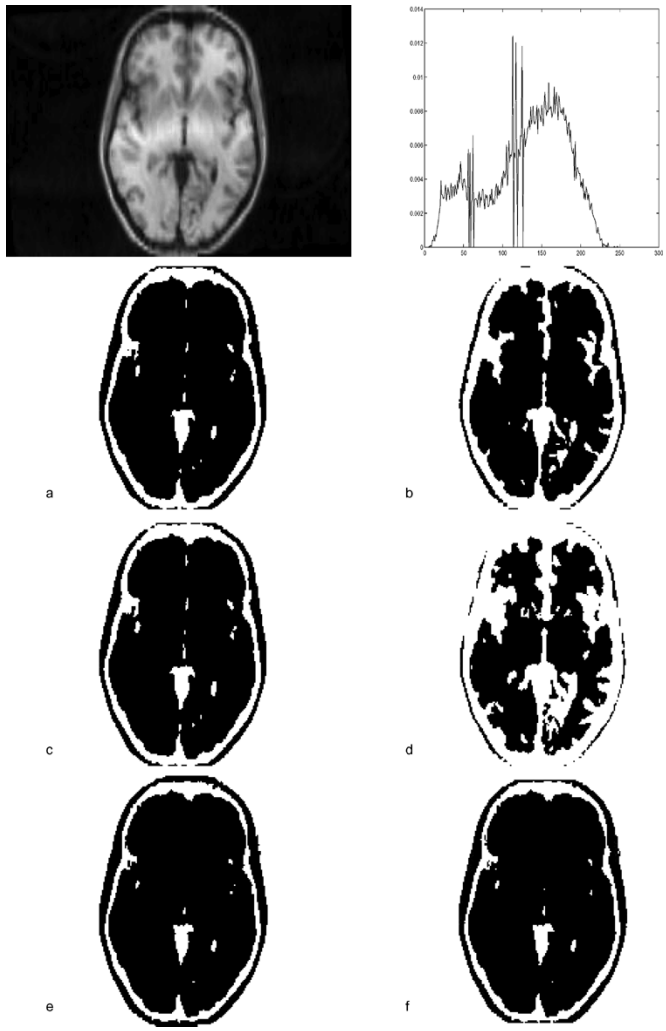



Fig. 7. Top left is the original AC-PC axial slice of an IBSR data, and top right is the histogram of the ROI. (a)–(f) Thresholded results. (a) RCOtsu with threshold 82.(b) Otsu with threshold 115. (c) RCFCP with threshold 84.5. (d) FCP with threshold 141.5. (e) RCLVD with threshold 68. (f) MinErr with threshold 72.

(top right) is jerky indicating that the image is noisy. There are also some sharp spikes which may be due to intensity inhomogeneity. However, these perturbations do not disturb an expert in selecting a right threshold around 75. The automatic thresholding results are presented in Fig. 7: (a) RCOtsu, (b) Otsu, (c) RCFCP, (d) FCP, (e) RCLVD, and (f) MinErr, and the thresholds are (a) 82, (b) 115, (c) 84.5, (d) 141.5, (e) 68, and (f) 72, respectively. The thresholds by the Otsu and the FCP methods are much larger than the expert's threshold and the resulting images are evidently over-segmented. The MinErr method agrees well with the intuitive segmentation and the difference from the RCLVD is small. The thresholds by the RCOtsu and the RCFCP methods are larger than the expert's threshold, however, the binarized images are quite reasonable.

Fig. 8 shows another example, where the image has severe intensity inhomogeneity. As a result, the intensity peak in the histogram of the ROI (top right) shifts to the left and in most regions the WM can not be differentiated from the GM. The MinErr method failed to calculate a threshold. The thresholded images by other methods are presented in Fig. 8: (a) RCOtsu, (b) Otsu, (c) RCFCP, (d) FCP, and (e) RCLVD, respectively. Obviously, both the Otsu and the FCP thresholded images are over segmented. On the contrary, the results by the proposed range-constrained methods are quite satisfactory.

### C. Patient Data Sets

For each of the 16 patient data sets, the MinErr method failed to identify a threshold. Both the Otsu and the FCP methods generally result in over-segmentation. Fig. 9 presents an example on the AC-PC axial slice of a volunteer data. From the histogram of the ROI (top right), it can be seen that the image quality is actually poorer than that in Fig. 6 of the phantom data. The portion corresponding to GM/WM is quite flat and the peaks of GM/WM are not clearly visible. Fig. 9(a)–(e) shows the thresholding results from methods of (a) RCOtsu, (b) Otsu, (c) RCFCP, (d) FCP, and (e) RCLVD, and the thresholds are (a) 89, (b) 125, (c) 94.5, (d) 152, and (e) 67, respectively. Evidently, the FCP and Otsu solutions [Fig. 9(b) and (d)] are over

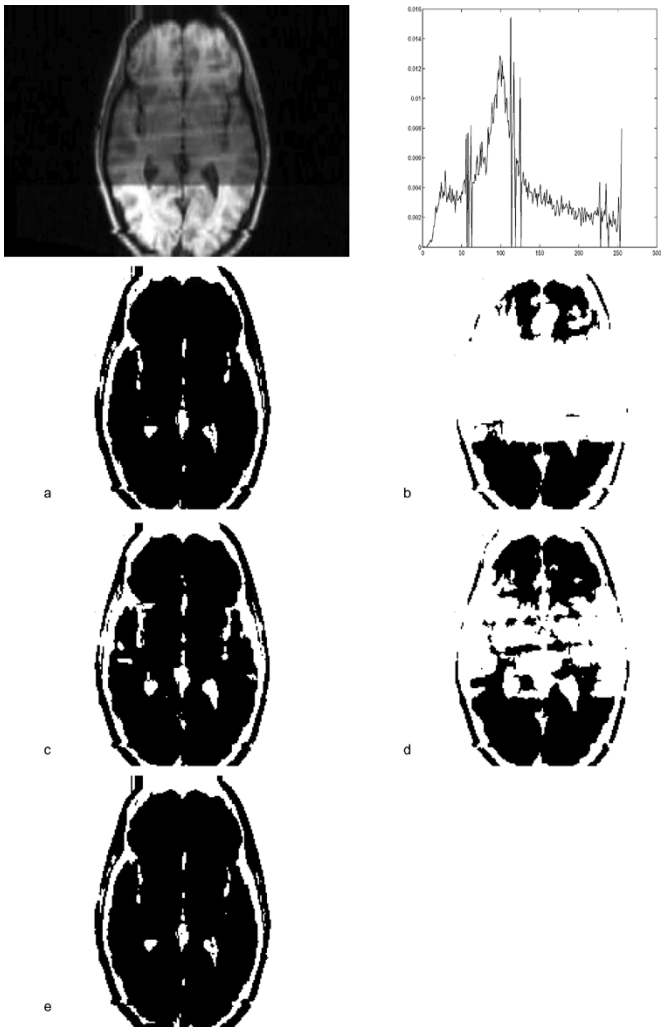



Fig. 8. Top left is the original AC-PC axial slice of an IBSR data with severe intensity inhomogeneity and top right is the histogram of the ROI. (a)–(e) Thresholded results. (a) RCOtsu with threshold 69. (b) Otsu with threshold 137. (c) RCFCP with threshold 72.5. (d) FCP with threshold 108. (e) RCLVD with threshold 59.

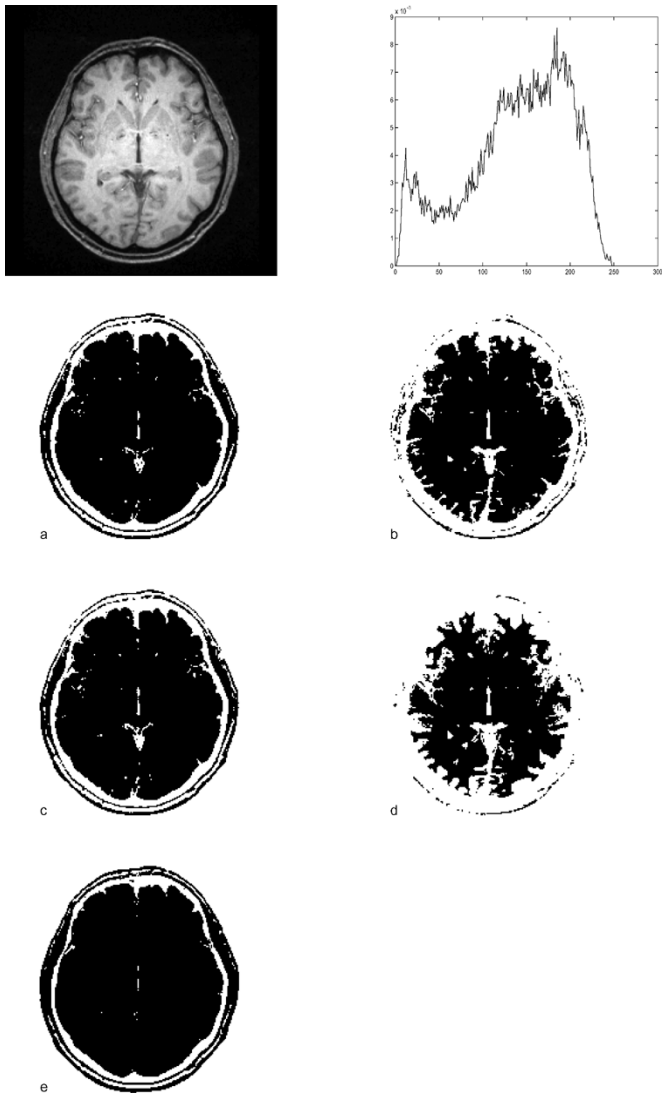



Fig. 9. Top left is the original AC-PC axial slice of a patient data, and top right is the histogram of the ROI. (a)–(e) Thresholded results. (a) RCOtsu with threshold 89. (b) Otsu with threshold 125. (c) RCFCP with threshold 94.5. (d) FCP with threshold 152. (e) RCLVD with threshold 67.

segmented. For the proposed three methods [Fig. 9(a), (c), and (e)], the results are quite reasonable perceptually. The thresholded images for the other 15 data sets can be found in Appendix for quick reference and comparison.

### *D. Quantitative Comparison*

The quantitative validation of segmentation algorithms can be done as the "ground truth" images of GM/WM are available for the BrainWeb phantom data sets and the IBSR data sets. For comparison, the FN error (1) and the FP error (2) were calculated for the thresholding methods used in this study. To do that, the scalp and the skull have to be removed. In this process, morphological opening operation is sometimes utilized to break the weak connections between the scalp and GM/WM. These connections may occasionally be strong due to severe under-segmentation and are manually broken accordingly. Fig. 10 shows an example of the segmented result after scalp/skull removal, where the ground truth of GM/WM is on the left.

*1) Phantom Data Sets:* As the FCP method significantly over-segmented the images and the MinErr failed to produce the threshold for all images, we only compare the error rates for the other four methods, as shown in Fig. 11, where the intensity inhomogeneity varies from 0% (row 1) to 20% (row 2) to 40% level (row 3). From Fig. 11, column 1, the Otsu method gives the largest FN error, mostly between $1 \sim 3\%$. The performances of the three proposed methods are very similar. The FN error is less than 0.2% at most times for these three methods. As for the FP error (column 2 in Fig. 11), the Otsu method attains the minimum error ($<2\%$). However, the FP error of the three proposed methods are also quite reasonable ($<8\%$).

*2) IBSR Data Sets:* Table I summarized the error rates (average and standard deviation), where the statistics of the MinErr method were calculated based on 10 threshold-reported images. The average FN error $(\epsilon_1)$ of the Otsu method is 17.19%, which is usually unacceptable for an initial segmentation. Its standard deviation is also large (21.08%), indicating that the performance of the Otsu method is highly variable for various images. With the range constraint, the average FN error reduces to 3.05%. Similarly, the FCP method is enhanced

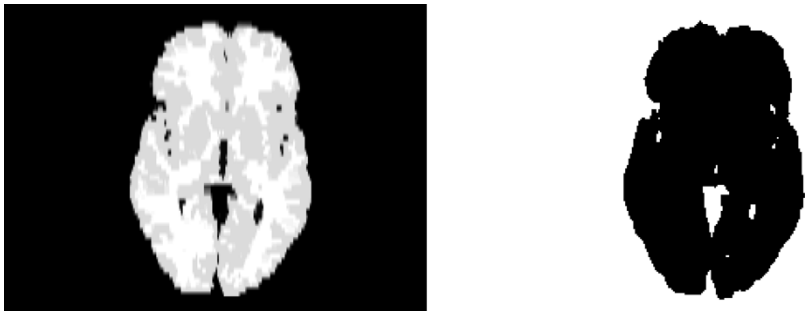


Fig. 10. Example of the segmented image after scalp/skull removal. On the left is the ground truth.

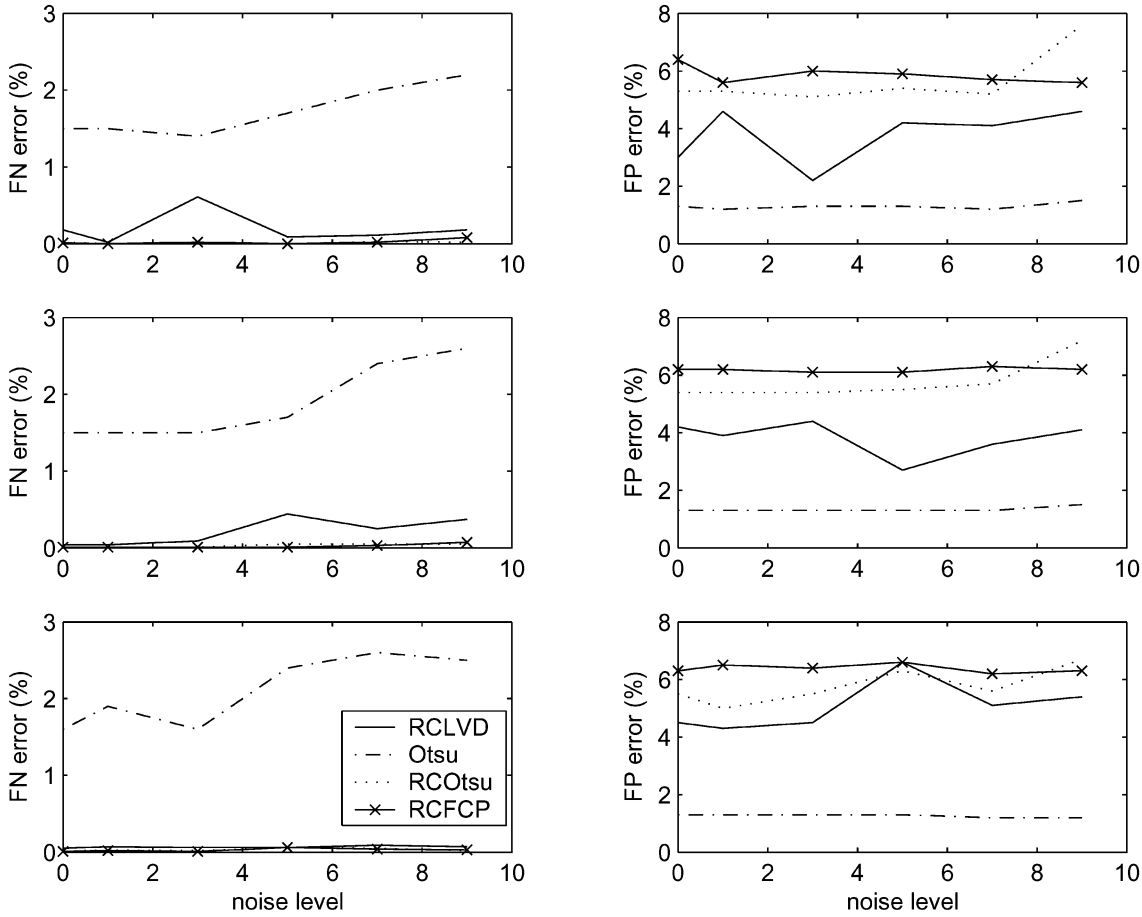



Fig. 11. FN and FP error rates of thresholding the phantom images, where the intensity inhomogeneity is 0 (row 1), 20% (row 2), and 40% (row 3).

with the range constraint to attain an average FN error of 3.7%. The RCLVD method achieves the minimum average FN error (1.5%) among the range-constrained methods, which is slightly larger than that of the MinErr method (1.37%). However, the MinErr method failed to produce a threshold for ten out of 20 images. On the contrary, the proposed range-constrained methods thresholded all images with a slightly larger average FN error ($1.5 \sim 3.7\%$). In addition, the average FP error of the proposed range-constrained methods ($2.3 \sim 4.11\%$) is also very reasonable.

### *E. Other Examples*

Fig. 3 shows the results on the Cameraman image. The thresholded image by the FCP method is shown on the bottom left of Fig. 3 with threshold of 140. From the histogram (top right), we can identify three peaks corresponding to the cameraman, the grassland, and the sky. The threshold by the FCP method falls in the region of grassland. Fig. 3 (bottom right) displays the result with threshold 79.5 by RCFCP, where the background proportion is estimated as $65 \sim 85\%$. It is clear that the cameraman is well seperated with both the grassland and the sky.

We have also applied the range-constrained thresholding method to identify ribs and vertebrae in CT chest images for computer-aided surgery (Fig. 2). The RCOtsu thresholding result is shown on the bottom right of Fig. 2, which is pretty reasonable. Comparatively, the Otus method misclassified quite a lot of background tissue as the bones (bottom left).

TABLE I
SUMMARY OF THE STATISTICS ON THE ERROR RATES OF THRESHOLDING THE IBSR IMAGES

| | RCLVD | Otsu | RCOtsu | MinErr[a] | FCP | RCFCP |
|---|---|---|---|---|---|---|
| $\epsilon_1(\%)$ | 1.50±1.31 | 17.19±21.08 | 3.05±2.33 | 1.37±1.14 | 35.45±5.21 | 3.70±2.92 |
| $\epsilon_2(\%)$ | 4.11±1.57 | 0.80±0.44 | 2.59±1.05 | 4.01±1.62 | 0±0 | 2.30±1.03 |

[a]calculated only for the cases when the method worked (as the method failed for 10 out of 20 cases studied).

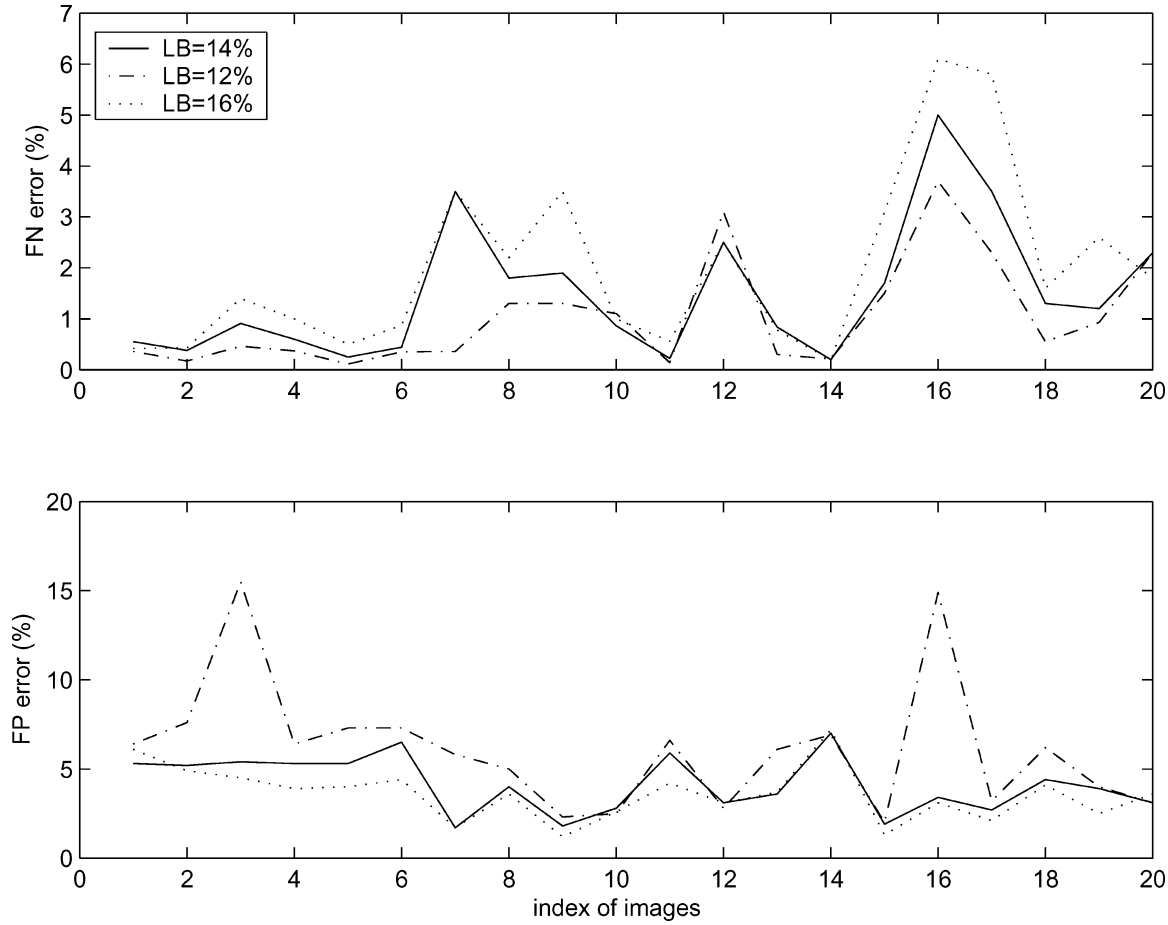



Fig. 12. Impact of background LB on the FN and FP error of the RCLVD method.

## V. DISCUSSION AND CONCLUSION

We have presented a supervision-based thresholding approach which confines the threshold selection within the frequency range of background. Supervision is adopted to estimate the range in which the proportion of the background to the ROI varies. By doing so, both the FN and FP errors could be kept from being large as the former is limited by the estimated maximum proportion of the background to the ROI, while the latter is bounded by the estimated minimum proportion. For illustration, three range-constrained thresholding methods are formulated, namely, RCOtsu, RCFCP, and RCLVD.

There are two parameters associated with the proposed algorithm, $H_l^b$ and $H_h^b$, to constrain the frequency range. Parameter $H_h^b$ sets an upper bound of threshold. When its value falls in the left side of the peak corresponding to the object, $H_h^b$ has a little effect on the position of the minimum average frequency and, consequently, a small impact on performance of the RCLVD method. However, it can be shown that the FN error of the RCOtsu and the RCFCP methods is an increasing function of $H_h^b$. Therefore, a tighter bound on $H_h^b$ will help reduce the FN error for the RCOtsu and the RCFCP methods.

As for parameter $H_l^b$, it sets a LB of threshold. It is obvious that the FN error of the RCOtsu and the RCFCP methods increases with the increase in $H_l^b$. Thus, a relaxed (smaller) bound of $H_l^b$ will result in a smaller FN error when the RCOtsu and the RCFCP methods are applied. To investigate its impact on the RCLVD method, Fig. 12 illustrates the effect of varying LB on the error rates for the IBSR images, where LB changes in three levels, 12%, 14%, and 16%. From Fig. 12, it can be seen that for most images the 16% LB level (dotted line) exhibits the largest FN error and the 12% LB level (dot-solid line) results in the minimum FN error. Although the 12% level is better in

terms of FN error, the corresponding FP error is obviously the largest in most images. The 14% level achieves a compromise of small FN error and reasonable FP error.

Most existing thresholding methods do not incorporate knowledge for threshold selection. The Otsu method seperates the image into classes of object and background, and determines the threshold by minimizing the within-class variance. The implicit assumption is that the intensity within each class is homogeneous. This could be true for simple images, but is often violated in real applications due to various reasons, such as the complexity of the scene, noise, and inhomogeneity (Fig. 8, for example). Consequently, the threshold selected by the Otsu method is often biased. As for the entropic thresholding, it is pointed out [18] that the method tends to classify the data into groups with equal probability, which is obviously contradictory to the intuitive segmentation. Despite coupling with fuzzy rules, we observe from our experiments that the selected threshold from the entropic method is systematically biased. For the MinErr method, it searches the threshold which minimizes the Bayes error and could achieve satisfactory performance for data with good quality. However, the assumed mixture-Gaussian model could be inapplicable for data modeling due to the complexity of real data. This kind of "model" bias is often encountered (in our experiment on brain MR images, 26 out of 36 real data) and makes the MinErr method fail to find a threshold for real data.

One may argue that more complex model can be used to approximate the real data such as a mixture of many Gaussian components, and even that an "intelligent" MinErr method could adjust the number of Gaussian components according to the complexity of data. Nevertheless, the task would not be easy, resulting in the increase of computation cost. Instead of designing a complex algorithm, we explicitly utilize prior knowledge, incorporate knowledge through supervision, and partition the histogram into three regions: background, mixture and object. The threshold is selected from the mixture region. The algorithm is simple yet efficient. It turned out that the selected threshold is consistent with that found by the MinErr method for data with good quality. For poor-quality data, the selected threshold is the most consistent with that found by experts.

Most existing thresholding methods do not differentiate the cost between the two types of errors. In the quantitative evaluation of segmentation algorithms, the Jaccard similarity

$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \tag{9}$$

and the $\kappa$ index

$$\kappa(S_1, S_2) = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|} \tag{10}$$

are frequently employed [30]–[32], where $S_1$ and $S_2$ are two binary images for comparison. However, information provided by these two metrics only characterizes the overall similarity between two segmentation methods. It is likely that two methods share a high similarity, but differ significantly from practical viewpoint. In medical applications, two costs are usually markedly different. In most cases, the cost to misclassify an ill person as a healthy one would be remarkably higher than the cost to misclassify the healthy as the ill. On the other hand, in security access applications, falsely granting access could result in an unpredictable disaster. For the present approach, we can conveniently handle these cost differences in the response to practical requirements. When the cost of the FN error is higher than that of the FP error, we can tighten (decrease) the upper bound or relax (decrease) the LB of the range in which the background proportion varies; and vice versa when the cost of the FP error is more expensive.

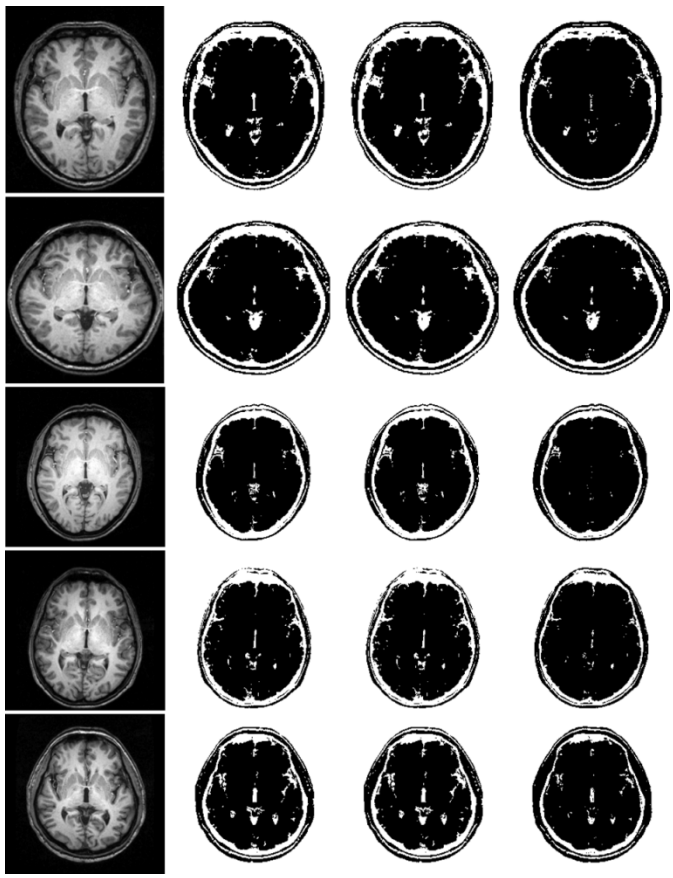


Fig. 13. Five clinical MR data (3T) and their thresholding: original (column 1), RCOtsu (column 2), RCFCP (column 3), and RCLVD (column 4).

It should be stressed that we cannot prove that the proposed supervised range-constrained thresholding will always yield a better threshold than a conventional thresholding method (like Otsu, FCP, and MinErr). As has been shown during the validation process, the MinErr method sometimes yields a better threshold in segmenting the AC-PC axial slices (smaller FN, Table I). What can be assured is that our proposed methods can provide a threshold with more consistent and robust binarization as compared with conventional thresholding methods. Further checking MinErr we can find that it fails to produce thresholds in 26 out of 36 images. Even for those good quality images, the best threshold yielded by MinErr just differs slightly from the thresholds by the proposed RCLVD, RCOtsu, and RCFCP. That is to say, when image quality is good, the proposed supervised range-constrained thresholding can provide a near-optimal or optimal solution; when image quality is not good, the proposed thresholding methods will yield a much better solution than conventional thresholding methods. This can be

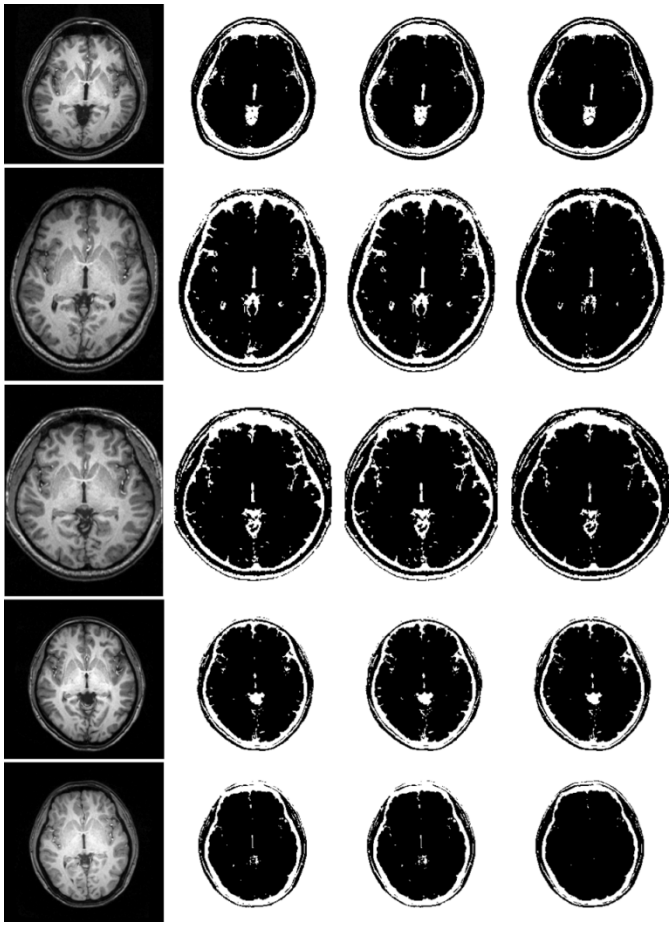


Fig. 14. Five clinical MR data (3T) and their thresholding: original (column 1), RCOtsu (column 2), RCFCP (column 3), and RCLVD (column 4).

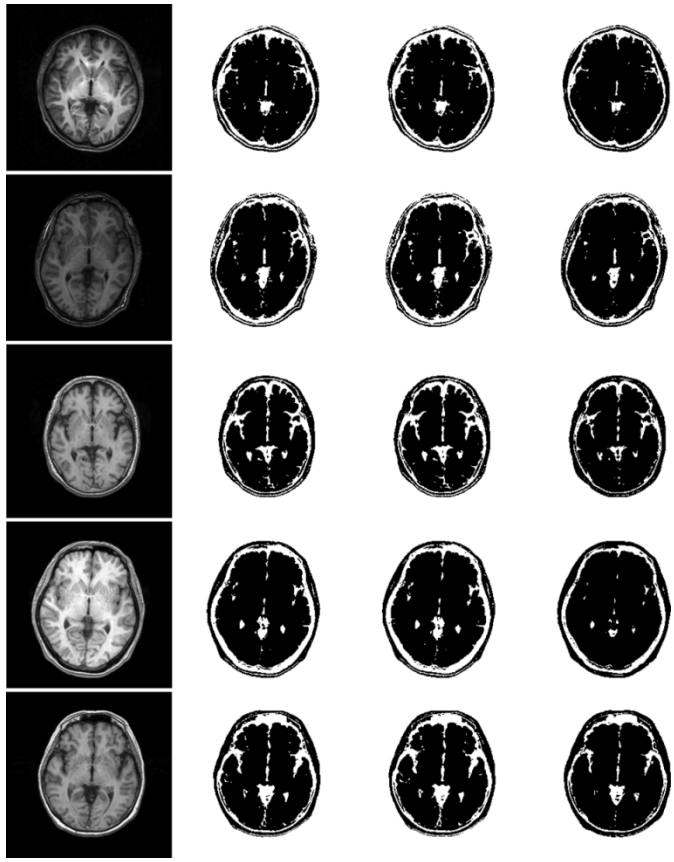


Fig. 15. Five clinical MR data (3T) and their thresholding: original (column 1), RCOtsu (column 2), RCFCP (column 3), and RCLVD (column 4).

reasoned as follows. We explicitly incorporate supervision to partition the intensity histogram within the ROI into three regions: background, mixture, and object. In our proposed framework, the threshold is estimated from the mixture region based on some minimum error criterion. For a good quality image, conventional thresholding methods may be able to find the optimal threshold. When the quality of the image is not good, conventional thresholding methods may find a threshold outside the mixture gray level range. No matter if the quality of images is good or not, our proposed methods ensure that the estimated threshold is in the mixture region. Minimizing classification error through the LVD, the Otsu or the FCP method within the mixture further chooses the optimum threshold in the range of mixture gray levels.

In summary, we have presented a novel thresholding approach which constrains the thresholding analysis within the variable proportion of the background to the ROI from the gray-level histogram of the ROI. It comprises three steps. The ROI is first determined from the image. Then, within the ROI, a range in the corresponding histogram is estimated from supervision, which represents the extent that the background portion varies. When a number of sample images with ground truth are available, the supervision can be in the form of training. When sample images are not available, the supervision can be in the form of approximation based on prior knowledge or visual assessment depending on whether a prior knowledge about the relationship between the object and the background within the ROI is available. Finally, the threshold is selected to minimize the classification error within this range. The method has been evaluated in detail against a medical image database which consists of 54 brain MR phantom or real images, with or without noise and/or inhomogeneity. Its applications to CT chest images and the Cameraman image have also been demonstrated. Experiments show that the proposed RCOtsu and RCFCP methods outperform their original counterparts (Otsu and FCP) and all the three range-constrained methods have comparable performance. It turned out that the proposed supervision-based thresholding approach is simple yet robust and reliable. The proposed approach is ideally applicable to images where background and object can be distinguished from their intensities (allowing for certain intensity overlapping). We believe that the proposed thresholding method can be employed to preprocess the images in a wide class of computer vision problems such as character recognition, fingerprint identification, and segmentation of a wide variety of medical images.

## ACKNOWLEDGMENT

The data for this study have kindly been provided by Massachusetts General Hospital, USA; Montreal Neurological Institute, Canada; Nagasaki University School of Medicine, Japan; and Singapore General Hospital, Singapore.

## REFERENCES


[1] J. M. S. Prewitt and M. L. Mendelsohn, "The analysis of cell images," *Ann. NY Acad. Sci.*, vol. 128, pp. 1035–1053, 1966.

[2] J. S. Weszka and A. Rosenfeld, "Histogram modification for threshold selection," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 38–52, Jan. 1979.

[3] A. Rosenfeld and P. D. L. Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, no. 2, pp. 231–235, Feb. 1983.

[4] V. Wu and R. Manmatha, "Document image clean-up and binarization," in *Proc. SPIE Document Recognition V*, 1998, pp. 263–273.

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

[6] C. H. Li and C. K. Lee, "Minimum cross entropy thresholding," *Pattern Recognit.*, vol. 26, pp. 617–625, 1993.

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

[8] Y. Nakagawa and A. Rosenfeld, "Some experiments on variable thresholding," *Pattern Recognit.*, vol. 11, pp. 191–204, 1979.

[9] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, pp. 41–47, 1986.

[10] S. Cho, R. Haralick, and S. Yi, "Improvement of Kittler and Illingworth's minimum error thresholding," *Pattern Recognit.*, vol. 22, pp. 609–617, 1989.

[11] Q. Ye and P. Danielsson, "On minimum error thresholding and its implementations," *Pattern Recognit. Lett.*, vol. 7, pp. 201–206, 1988.

[12] L. K. Huang and M. J. Wang, "Image thresholding by minimizing the measure of fuzziness," *Pattern Recognit.*, vol. 28, pp. 41–51, 1995.

[13] O. J. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1457–1465, Oct. 2002.

[14] H. Yan, "Unified formulation of a class of image thresholding techniques," *Patter Recognit.*, vol. 29, pp. 2025–2032, 1996.

[15] T. Pun, "Entropic thresholding: A new approach," *Comput. Vis. Graph. Image Process.*, vol. 16, pp. 210–239, 1981.

[16] A. K. C. Wong and P. K. Sahoo, "A gray-level threshold selection method based on maximum entropy principle," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 7, pp. 866–871, Jul. 1989.

[17] P. K. Saha and J. K. Udupa, "Optimum image thresholding via class uncertainty and region homogeneity," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, no. 5, pp. 689–706, May 2001.

[18] H. D. Cheng, J. Chen, and J. Li, "Threshold selection based on fuzzy c-partition entropy approach," *Pattern Recognit.*, vol. 31, pp. 857–870, 1998.

[19] N. Ahuja and A. Rosenfeld, "A note on the use of second-order gray-level statistics for thresholding selection," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 7, pp. 897–898, Aug. 1978.

[20] Q. Wang, Z. Chi, and R. Zhao, "Image thresholding by maximizing the index of nonfuzziness of the 2-D gray-scale histogram," *Comput. Vis. Image Understand.*, vol. 85, pp. 100–116, 2002.

[21] J. Giarratano and G. Riley, *Expert Systems: Principles and Programming*, 2nd ed. Boston, MA: PWS, 1994.

[22] M. I. Sezan, A. M. Tekalp, and R. Schaetzing, "Automatic anatomically selective image enhancement in digital chest radiography," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 154–162, Feb. 1989.

[23] J. S. Lee and M. C. K. Yang, "Threshold selection using estimates from truncated normal distributions," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 4, pp. 422–429, Apr. 1989.

[24] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypothesis," *Stat. Res. Memoirs*, vol. 1, pp. 1–37, 1936.

[25] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 463–468, Apr. 1998.

[26] Q. Hu and W. L. Nowinski, "A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human cerebrum from neuroimages based on local symmetry and outlier removal," *NeuroImage*, vol. 20, pp. 2154–2166, 2003.

[27] W. L. Nowinski, "Modified Talairach landmarks," *Acta Neurochirurgica*, vol. 143, pp. 1045–1057, 2001.

[28] J. Talairach and P. Tournoux, *Co-planar Stereotactic Atlas of the Human Brain*. New York: Georg Thieme Verlag/Thieme Medical, 1988.

[29] W. L. Nowinski, "Electronic brain atlases: features and applications," in *3D Image Processing: Techniques and Clinical Applications, Medical Radiology Series*, D. Caramella and C. Bartolozzi, Eds. New York: Springer-Verlag, 2002, pp. 79–93.

[30] Z. Y. Shan, G. H. Yue, and J. Z. Liu, "Automatic histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images," *NeuroImage*, vol. 17, pp. 1587–1598, 2002.

[31] V. K. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 897–908, Aug. 1999.

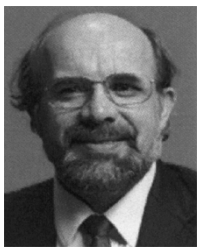
[32] D. W. Shattuck, R. S. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *NeuroImage*, vol. 13, pp. 856–876, 2001.


**Qingmao Hu** received the Ph.D. degree.

He is a Research Scientist with the Biomedical Imaging Laboratory, Agency for Science, Technology, and Research, Singapore. His research interests include neuroinformatics, image analysis, and pattern recognition.

**Zujun Hou** received the Ph.D. degree.

He is with the Biomedical Imaging Laboratory, Agency for Science, Technology, and Research, Singapore, as a Visiting Scientist. His current research interests include image and signal processing and pattern recognition.



**Wieslaw L. Nowinski** received the Ph.D. and D.Sc. degrees.

He is a Director and a Principal Scientist of the Biomedical Imaging Laboratory, Agency for Science, Technology, and Research, Singapore. His current research activities cover neuroinformatics, particularly brain atlases and atlas-based applications; virtual reality in medical intervention; and future directions in computer-aided surgery. He has published about 310 papers, developed 12 brain atlas global products, filed 29 patents, started two companies, and received international and national awards.