

Multi-modal gray-level histogram modeling and decomposition

Jeng-Horng Chang^a, Kuo-Chin Fan^{b,*}, Yang-Lang Chang^b

^aHuafan University, 1 Huafan Road, Shihtin Hsiang, Taipei County, 223, Taiwan, ROC

^bInstitute of Computer Science and Information Engineering, National Central University, Chung-Li 32054, Taiwan, ROC

Received 11 May 2000; received in revised form 15 November 2001; accepted 22 November 2001

Abstract

In this paper, we present a novel multi-modal histogram thresholding method in which no a priori knowledge about the number of clusters to be extracted is needed. The proposed method combines regularization and statistical approaches. By converting the approaching histogram thresholding problem to the mixture Gaussian density modeling problem, threshold values can be estimated precisely according to the parameters belonging to each contiguous cluster. Computational complexity has been greatly reduced since our method does not employ conventional iterative parameter refinement. Instead, an optimal parameter estimation interval was defined before the estimation procedure. This predefined optimal estimation interval reduces time consumption while other histogram decomposition based methods search all feature space to locate an estimation interval for each candidate cluster. Experimental results with both simulated data and real images demonstrate the robustness of our method. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Image thresholding; Multi-modal histogram analysis; Gaussian mixture density; Histogram decomposition; Parameter estimation

1. Introduction

Essentially, image thresholding [1–3] is a pixel classification problem. Its basic objective is to classify the pixels of a given image into two classes: those pertaining to objects and those pertaining to background. While one class includes pixels with gray values that are below or equal to a certain threshold value, the other class includes those with gray values above the threshold value. Thresholding is a popular tool for image segmentation. It is widely used in halftone reproduction, automatic target recognition, design of visual navigation system for autonomous land vehicle, industrial application of computer vision, biomedical image analysis, and content-based image retrieval system.

In general, threshold selection techniques can be broadly divided into two categories: global and local thresholding. A global thresholding technique [4] is the technique that thresholds the entire image with a single threshold value, whereas a local thresholding technique [5,6] is the one that partitions a given image into sub-images and determines a threshold value for each of these sub-images. Global thresholding techniques may be point-dependent or region-dependent. A thresholding method is point-dependent if the thresholding value is determined solely from the pixels'

gray tone as represented by gray-level histogram and is independent of the gray tone of the neighborhood of a pixel. On the other hand, a thresholding method is called region-dependent if the threshold value is determined from the local property within the neighborhood of a pixel.

The threshold selection techniques can be divided into two groups, which are bilevel and multi-level. Bilevel thresholding is used if an image has only one object which is distinct from the background. The gray-level histogram of these images will always be a bimodal distribution located apparently far from each other. Hence, one threshold value is sufficient to segment images into object and background. On the other hand, the gray-level histogram of an image is composed of several distributions if the image has a number of objects distinct from the background. Multi-level thresholding techniques [7–11] have to be used to segment multi-modal images into sub-images with those representing unaffiliated objects in these images. The major difficulty for multi-thresholding techniques is that the number of objects is always unknown. Nevertheless, threshold values selection will be a crucial problem if objects in images are intermixed with each other.

In this paper, we propose a novel method to thresholding multi-modal gray-level images. This proposed automatic multi-thresholding method is a global and point-dependent approach. In our method, gray-level histogram is smoothed first in order to obtain existing local-minimum points as our initial thresholding points. The smoothed gray-level

* Corresponding author. Tel.: +886-3-4227151x4453; fax: +886-3-4222681.

E-mail address: kcfan@csie.ncu.edu.tw (K.-C. Fan).

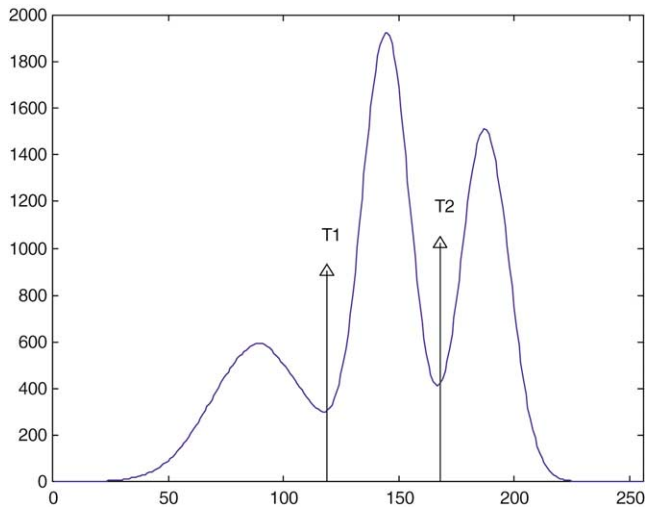


Fig. 1. Three distinctly distributed histograms. They must have two threshold values, T1 and T2, which separate this histogram into three non-overlapping distributions.

histogram is identical to *mixture Gaussian distribution* that has been formulated and proved by Zhuang et al. [12]. The only assumption in Zhuang's method is that the gray-level histogram of a certain image is constructed by several Gaussian distributions. That is, each object in an image will be a Gaussian-like distribution in this gray-level histogram with different mean, variance and probability values. Therefore, the histogram decomposition problems can be considered as fundamental issues relating to computer vision [17]. We use statistical approach and some heuristic parameters to decompose this histogram into non-overlapping distributions without a priori knowledge about the number of objects. We must emphasize that our method does not employ conventional parameter estimation procedure by iterative parameter refinement. Instead, by estimating the initial non-exact mean and variance values as cues for determining the initial threshold value, the *skewness* of certain interval in this candidate distribution is calculated to quickly locate the deterministic optimal estimation interval. After optimally estimating the mean and variance values of each distribution in the histogram, a maximum-likelihood based decision criterion is applied to determine the optimal threshold values among distributions. Finally, images are segmented into several non-overlapping sub-images with respect to the specified threshold values.

The rest of this paper is organized as follows: in Section 2, we state the basic representation of multi-modal gray-level histogram. Some multi-thresholding techniques are also described in this section. We will introduce a Gaussian mixture densities decomposition method in this section, which is the basis of our histogram decomposition algorithm. A concrete algorithm of automatic multi-modal gray-level histogram thresholding is presented in Section 3. In Section 4, experiments conducted on both simulated data and real images are

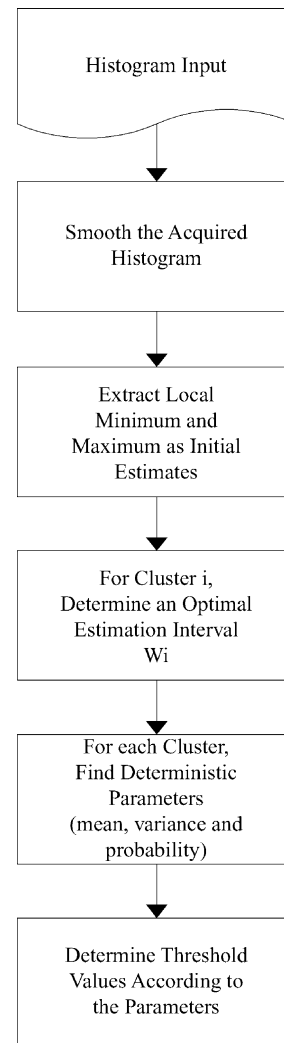


Fig. 2. An overview of automatic multiple threshold selection algorithm.

demonstrated. Comparisons are also made with some approaches in literatures. Finally, conclusions and discussions are summarized in Section 5.

2. Multi-modal gray-level histogram analysis

Let I be a $p \times q$ gray-scale fingerprint image with G gray levels and $\text{gray}(x,y)$ be the gray value of pixel $I(x,y)$ with $x = 1, \dots, p$ and $y = 1, \dots, q$. Then, the gray-level histogram H of image I is of the form $H = \{H(j) | j \in [1, G]\}$. $H(1)$ to $H(G)$ represent the histogram probabilities of the observed gray values from 1 to G , and $H(g) = \#\{\text{gray}[I(x,y)] = g, g = 1, \dots, G, x = 1, \dots, p \text{ and } y = 1, \dots, q\}$. Generally, there will exist a number of 'mountains' in the histogram if it is a multi-modal distribution. Each distribution in the histogram will map to an object in the image. Fig. 1 shows an example of three distinctive distributions in a gray-level histogram. For any gray-level histogram with n distributions, the multi-thresholding techniques are to automatically determine

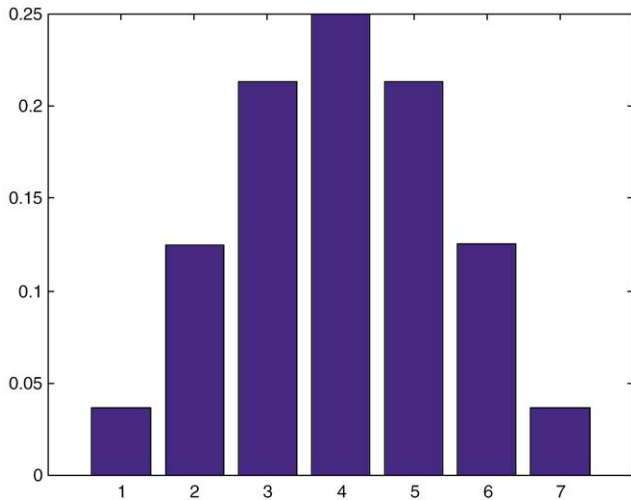
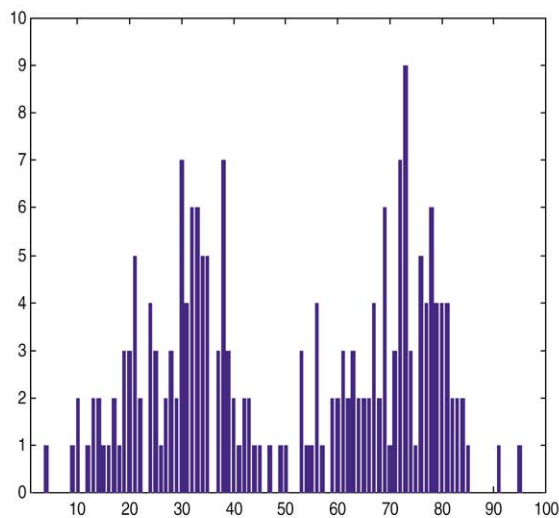


Fig. 3. An example which shows a Gaussian masking window with seven bins. Each bin in this window is calculated by $b_k = 0.5(1 - \cos(\pi k/p))$.

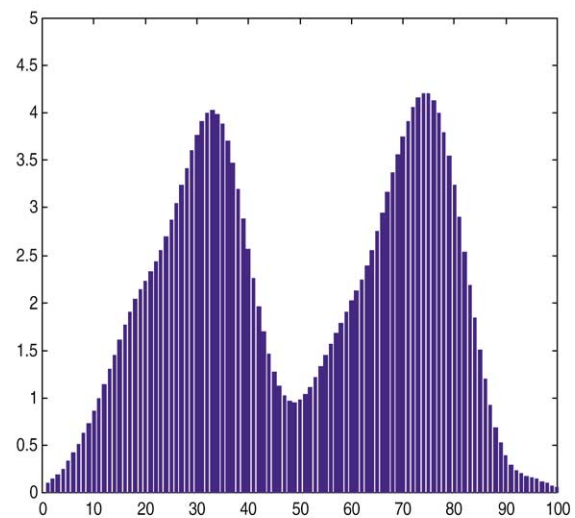
$n - 1$ threshold values that are used to separate this multi-modal histogram into n non-overlapping distributions.

Many automatic multi-thresholding techniques have been proposed for various applications with different approaches. Pal and Bhattacharyya [13] proposed a regularization approach with both continuous and discrete formulations for multi-modal histogram analysis. In their approach, gray-level histogram is smoothed first. Then, the peaks and valleys of the histogram can be extracted by a predefined score function. Carlotto [14] also solved the multi-thresholding problems by regularization approach. The method obtains a multi-scale description of the histogram by convoluting it with the second derivation of the Gaussian filter with gradually increasing width.

These two regularization methods obtain a set of local maximums and local minimums, always peaks and valleys, as their cues to select the best possible thresholds by some predefined post-processing criteria. Statistical approaches are the most widely used methods for determining multiple threshold values. Based on the understanding of the information extracted from the histogram distributions, a set of threshold values can be determined according to the parameters of each distinctively distributed cluster in the histogram. Many statistical approaches concerned with the problem of determining the parameters of mixture distributions. In other words, multi-modal histogram can be decomposed into several non-overlapping distributions with respect to the estimated parameters, which come from the histogram itself. Render and Walker [15] have provided an overview of various methods developed for the parameter estimation of mixture densities. The generalized minimum volume ellipsoid (GMVE) clustering algorithm [16], which identifies the least volume region containing in the data space, is also introduced in the field of computer vision. The MF-Estimator [12] is another mixture density decomposition algorithm, which is similar to the GMVE method. Both GMVE and MF-Estimator estimate the parameters of distributions by modeling the histogram with normal distributions. The Kolmogorov–Smirnov test of fit is employed for the comparison between the shape of extracted distribution and that of an ideal cluster generated by a Gaussian density. The best-fit cluster, which passed Kolmogorov–Smirnov test, is then removed from the feature space and a new iteration of cluster extraction starts. The main drawbacks of these mixture decomposition algorithms are their high computational complexity due to the iterative parameter refinement until



(a)



(b)

Fig. 4. (a) Original histogram distribution; and (b) smoother histogram after convoluting with a Gaussian masking window.

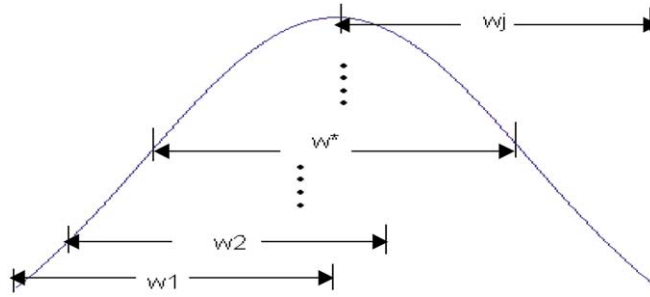


Fig. 5. Optimal interval for parameter estimation. For each cluster in the histogram distribution, the searching window w with fixed length r slides from the leftmost point toward the right end of the cluster. The interval with minimum absolute skewness value will become the optimal interval w^* .

convergence. Moreover, the randomly selected initial estimates probably will converge to incorrect values if the initial estimates are not close enough to the desired solution for the iterative systems.

From the literatures described earlier, we know that there exist some major difficulties, which make multi-thresholding an uneasy problem.

- (A) The number of distinctive distributions, which represent objects in an image, is always unknown and is given by supervision in some approaches.
- (B) High computational complexity in determining threshold values due to iterative parameter refinement for each candidate cluster.
- (C) The correctness of thresholding results still cannot be measured by any criterion because the histogram is randomly distributed.

In order to overcome these problems, we propose an exceptional statistical approach. In our method, no iterative parameter refinement is necessary because our method finds the optimal estimation interval of each candidate cluster first instead of random searching of estimation region. Thus, our method can reduce computational complexity comparing to the other histogram decomposition based methods.

For natural scene with large samples, we assume that the observation comes from a mixture of $n + 1$ Gaussian distri-

butions, name f , having respective means and variances $(m_1, \sigma_1^2), \dots, (m_{n+1}, \sigma_{n+1}^2)$ with respective proportions P_1, \dots, P_{n+1} . Therefore, the mixture distributions reflected in the histogram will be in the form of

$$f(k) = \sum_{i=1}^{n+1} \frac{P_i}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{k - m_i}{\sigma_i}\right)^2\right\}$$

Our objective is to find the parameters, i.e. means, variances, and proportions, to satisfy the minimization

$$\min(|f - H|)$$

Then, by using these parameters, the optimal multi-threshold values can be determined to make suitable segmentation.

An overview of our automatic multiple threshold selection algorithm is presented in Fig. 2. First, smooth the histogram by convoluting it with a Gaussian filter. This procedure prevents the detection of a lot of fake local minimums and maximums. Then, we extract genuine local minimums as our initial threshold values for further parameter estimation. For each cluster, an optimal interval for deterministic parameter estimation is determined beforehand by calculating a measurement factor called skewness. For any cluster in the histogram, there exists only one estimation interval nearing the cluster center with the smallest absolute skewness value. By using this optimal estimation interval,

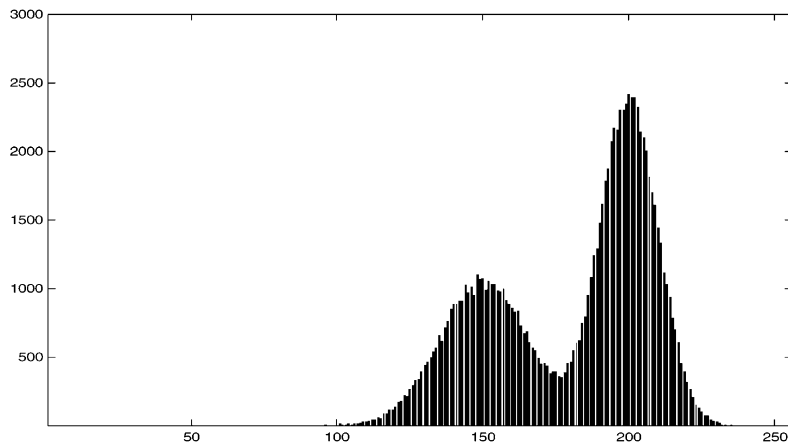


Fig. 6. Two distinctly distributed histogram.

Table 1

Parameters	Generated data	B-Spline	MF-Estimator	Our method
Mean 1	150	148.28	148.80	149.26
Variance 1	225	180.58	190.43	199.69
Probability 1	0.4	0.3793	0.3870	0.3935
Mean 2	200	199.33	199.64	200.05
Variance 2	100	108.37	101.83	99.19
Probability 2	0.6	0.6207	0.6130	0.6065
Threshold	–	174	176	178

Table 2

	Generated data	B-Spline	MF-Estimator	Our method
Mean 1	90	87.02	87.88	89.59
Variance 1	400	304.65	324.12	368.00
Probability 1	0.3	0.2784	0.2858	0.2999
Mean 2	145	143.74	144.70	144.90
Variance 2	100	108.99	104.71	76.83
Probability 2	0.3	0.3208	0.3190	0.2982
Mean 3	188	187.77	188.39	188.08
Variance 3	100	97.44	92.04	102.48
Probability 3	0.4	0.4063	0.3952	0.4019
Threshold	–	119, 165	122, 168	127, 166

we can predict the parameters representing clusters, i.e. means, variances and probabilities, without iterative parameter refinement. After determining these parameters, a maximum-likelihood based decision criterion is employed to optimally determine a set of threshold values.

3. Multi-modal histogram decomposition algorithm

In order to decompose a gray-level histogram into several non-overlapping distributions, we have to find the local minimums first and then perform further parameters estimation tasks. However, the histogram distribution, which was acquired from real world scene, is always anomalously distributed. Hence, a histogram smoothing process is necessary before performing the decomposition process. In our method, we use a Gaussian kernel to smooth the histogram.

Let W_g be a Gaussian masking window with $2p + 1$ bins and $b_k, k = 1, 2, \dots, 2p + 1 (p \geq 0, b_k \geq 0, \sum b_k = 1)$ be the elements of W_g .

The new gray-level histogram is calculated as the convolution of H and W_g .

$$\tilde{H} = H \otimes W_g$$

where ‘ \otimes ’ denotes the convolution operation. Thus, $\tilde{H} = \{\tilde{H}(j) | j \in [1, G]\}$ forms the smoothed histogram,

where

$$\tilde{H}(i) = \frac{1}{2p + 1} \sum_{u=-p}^p b_{p+1+u} H(i + u)$$

for $i = p + 1$ to $G - p$. Fig. 3 is an example which shows a Gaussian masking window with seven bins. For $2p + 1$ bins masking window, each bin can be calculated by

$$b_k = 0.5(1 - \cos(\pi k/p))$$

Fig. 4 shows the histograms before and after convoluting with a Gaussian masking window of 21 bins ($p = 10$).

After the smoothed histogram \tilde{H} has been obtained, the peaks and valleys in the histogram can be determined by the following rule: for any gray value $i, i \in [1, G]$, $\tilde{H}(i)$ is a peak if $\tilde{H}(i) > \tilde{H}(i - 1)$ and $\tilde{H}(i) \leq \tilde{H}(i + 1)$. On the other hand, $\tilde{H}(i)$ is a valley if $\tilde{H}(i) < \tilde{H}(i - 1)$ and $\tilde{H}(i) \geq \tilde{H}(i + 1)$.

Suppose there exists n distinct Gaussian clusters $C_i, i = 1, \dots, n$, \tilde{H} must have n peaks, denoted by $R(1), \dots, R(n)$, and $n - 1$ valleys, denoted as $V(1), \dots, V(n - 1)$. Then, the interval of C_i in the smoothed histogram \tilde{H} will be $[V(i - 1), V(i) - 1]$, with $V(0) = 1$ and $V(n) = G + 1$. We will define an optimal estimation interval within each cluster

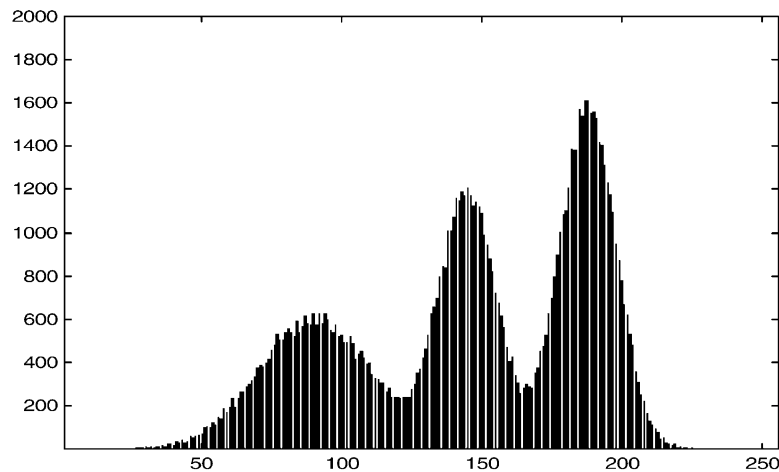


Fig. 7. Three distinctively distributed histogram.

Table 3

	Generated data	B-Spline	MF-Estimator	Our method
Mean 1	75	68.2811	71.8638	73.5629
Variance 1	400	235.8946	288.1821	315.7168
Probability 1	0.2	0.1632	0.1884	0.2005
Mean 2	128	130.2739	127.8450	127.8056
Variance 2	225	301.5959	181.5720	156.0945
Probability 2	0.5	0.6381	0.5340	0.5068
Mean 3	170	180.0706	173.7722	172.6219
Variance 3	324	143.2174	205.1889	218.9188
Probability 3	0.3	0.1987	0.2776	0.2927
Threshold	–	92, 164	99, 153	102, 151

to estimate the parameters that can represent the distribution of the clusters.

For a certain Gaussian cluster C_i , $i = 1, \dots, n$, the range has been defined previously by $[V(i - 1), V(i) - 1]$. The conventional confidential interval of cluster mean \bar{m} can be predicted by sample mean \bar{m}_i and sample standard deviation $\bar{\sigma}_i$ with student- t distribution and confidence coefficient α for s samples.

$$\bar{m}_i - t_{(1-\alpha/2)} \frac{\bar{\sigma}_i}{\sqrt{s}} \leq \bar{m} \leq \bar{m}_i + t_{(1-\alpha/2)} \frac{\bar{\sigma}_i}{\sqrt{s}}$$

This confidential interval \bar{m} can be a sample interval to estimate the initial mean and variance of cluster C_i .

Unfortunately, the conditions of mixtures are always unknown. For the distributions which are too close to be separated, the aliasing problem will occur while utilizing this confidential interval \bar{m} as the initial guess. Nevertheless, the sample mean \bar{m}_i and the sample variance $\bar{\sigma}_i$ are obtained from the range of C_i without any limitations. It will probably lead to a biased initial guess and will lose some genuine cluster members. Thus, the interval of estimation should be selected carefully for precise estimation.

For any univariate distribution $f(x)$ as introduced in the statistical theory, the *central moments* are defined as $\mu_n = E\{(x - m)^n\} = \int_{-\infty}^{\infty} (x - m)^n f(x) dx$.

If the random variable x is a discrete type with unknown mean value, μ_n can be rewritten by its sample mean, \bar{m} , as $\mu_n = \sum_i p_i (x_i - \bar{m})^n$, with $\bar{m} = \sum_i p_i x_i$ where p_i is the occurrence probability of x_i .

The skewness, β_1 , involves the second and third order central moments and can be defined as $\beta_1 = \mu_3 / \sqrt{\mu_2^3}$. Skewness is a symmetric measurement of distributions. $\beta_1 > 0$ means that the distributions are left-biased, and $\beta_1 < 0$ means that the distributions are right-biased. For univariate normal distributions $N(m, \sigma^2)$, since $f(-x) = f(x)$, the odd-order central moments will all be zeros. That is, $\mu_3 = E\{(x - \bar{m})^3\} = 0$. Thus, β_1 will be zero if this distribution is a normal distribution.

For mixture Gaussian distributions with clusters C_i , $i = 1, \dots, n$, the overall skewness for the range of each cluster is meaningless because the neighborhood clusters at the margins of both sides contaminate each Gaussian cluster. However, the skewness is also close to zero at a certain interval which is near the center of each cluster. Therefore, this special interval should be determined first to obtain optimal initial estimation of the cluster centers and then perform other further estimations and decisions.

For the purpose of choosing the optimal interval, the length should be determined first. It can be easily proved that the length of this interval has no effect to the mean value estimation if this length is sufficiently large. For random variable $y = g(x)$, the mean value is given by $E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f(x) dx$. Suppose that $f(x)$ is negligible outside an interval $(m - \varepsilon, m + \varepsilon)$ and $g(x) \cong g(m)$ for this specified interval if x is concentrated near its mean m . In this case, the mean of $g(x)$ becomes

$$E\{g(x)\} \cong g(m) \int_{m-\varepsilon}^{m+\varepsilon} f(x) dx \cong g(m)$$

By *Tchebycheff inequality*, for any $\varepsilon > 0$ we have $P\{|x - m| \geq \varepsilon\} \leq (\sigma^2 / \varepsilon^2)$.

This means the probability that x is outside an arbitrary

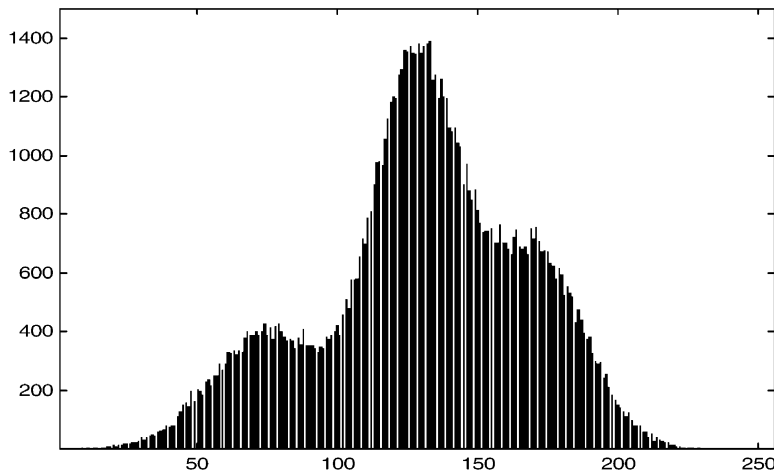


Fig. 8. Three non-distinctly distributed histogram.

interval $(m - \varepsilon, m + \varepsilon)$ is negligible if the ratio σ/ε is sufficiently small.

For each cluster C_i in Gaussian mixture, we select the interval for initial cluster's mean value estimation by length $r_i = (1/2)(V(i) - V(i - 1) - 1)$. This length, r_i , possesses the properties of $r_i \gg \sigma_i$ and $r_i \propto \sigma_i$. The length of interval satisfies Tchebycheff inequality and can be used as the length of optimal parameter estimation interval.

Now, let us define a searching window w with length r_i to search the location of optimal estimation interval of cluster C_i . The searching window w which starts by placing the leftmost point at $V(i - 1)$ slides toward the end of cluster $V(i) - 1$ by moving one bin at a time. The searching process stops if the rightmost point reaches the end of cluster $V(i) - 1$. There will be $V(i) - V(i - 1) - 1 - r_i$ searching windows. Meanwhile, the skewness β_1 is calculated for each searching window w_j , $j = 1, 2, \dots, V(i) - V(i - 1) - 1 - r_i$, denoted by $\beta_1(w_j)$. For our problem, the skewness of each searching window can be calculated as

$$\tilde{m}_{w_j} = \frac{\sum_{i \in w_j} iH(i)}{\sum_{i \in w_j} H(i)},$$

$$\mu_n(w_j) = \frac{\sum_{i \in w_j} (i - \tilde{m}_{w_j})^n H(i)}{\sum_{i \in w_j} H(i)} \quad \text{and} \quad \beta_1(w_j) = \frac{\mu_3(w_j)}{\sqrt{\mu_2^3(w_j)}}$$

Therefore, the optimal interval w^* for estimating the mean and variance of each cluster is determined by the interval which has minimum absolute skewness value. That is

$$w^* = \min_j |\beta_1(w_j)|$$

The optimal estimation interval w_i^* of cluster C_i will be located at $[a_i, b_i)$. Fig. 5 is an example which shows the position of searching windows w and the location of optimal estimation interval w^* for one distribution.

Then the initial mean, variance and proportion of this cluster can be optimally determined by

$$\hat{m}_i = \frac{\sum_{k=a_i}^{b_i} kH(k)}{\sum_{k=a_i}^{b_i} H(k)},$$

$$\hat{\sigma}_i^2 = \frac{\sum_{k=a_i}^{b_i} (k - \hat{m}_i)^2 H(k)}{\sum_{k=a_i}^{b_i} H(k)} \quad \text{and} \quad \hat{P}_i = \frac{\sum_{k=a_i}^{b_i} H(k)}{\sum_{u=1}^g H(u)}$$

For the i th observation $H(i)$, it is more likely generated by

cluster C_k if

$$\frac{\hat{P}_k}{\sqrt{2\pi\hat{\sigma}_k}} \exp\left\{-\frac{1}{2}\left(\frac{i - \hat{m}_k}{\hat{\sigma}_k}\right)^2\right\} > \frac{\hat{P}_j}{\sqrt{2\pi\hat{\sigma}_j}} \exp\left\{-\frac{1}{2}\left(\frac{i - \hat{m}_j}{\hat{\sigma}_j}\right)^2\right\}$$

for $1 \leq j \leq n$, $1 \leq k \leq n$, and $j \neq k$.

If there are n clusters, we will obtain $n - 1$ threshold values T_i , $i = 1, 2, \dots, n - 1$.

Therefore, the i th threshold T_i can be determined as follows:

$$T_i = \max\{k : H(k) \text{ is generated by the } i\text{th Gaussian cluster}\}$$

Finally, for each cluster C_i , $i = 1, 2, \dots, n$, the range becomes $[T(i - 1), T(i) - 1]$ with $T(0) = 1$ and $T(n) = g$. The mean, variance and proportion of the cluster can then be determined by the following equations:

$$m_i = \frac{\sum_{i \in C_i} iH(i)}{\sum_{i \in C_i} H(i)},$$

$$\sigma_i^2 = \frac{\sum_{i \in C_i} (i - m_i)^2 H(i)}{\sum_{i \in C_i} H(i)} \quad \text{and} \quad P_i = \frac{\sum_{i \in C_i} H(i)}{\sum_{i=1}^g H(i)}$$

4. Experimental results

To verify the validity of the proposed histogram decomposition algorithm, several experiments were conducted on both artificial randomly generated data and real image data. It must be emphasized that no a priori knowledge about the number of clusters is assumed in our algorithm. As will be seen, the algorithm can automatically find a cluster in the feature space if there indeed exists one.

4.1. Experiments with simulated data

In this sub-section, we demonstrate the performance of the proposed histogram decomposition algorithm by using randomly generated data sets. We will present some experimental results with pure Gaussian mixture distributions as our simulated gray-level histogram. As will be seen, the proposed histogram decomposition algorithm can find those parameters, i.e. $m_1, \sigma_1, P_1, \dots, m_i, \sigma_i, P_i$, no matter whether these Gaussian distributions are distinct or not. Then, use these parameters to determine the threshold values without a priori knowledge about the cluster number.

In our experiments, the point numbers of each cluster is its probability times 100,000 points. For example, a cluster with probability 0.5 has 50,000 points. The parameter estimation results are compared with two well-known methods,

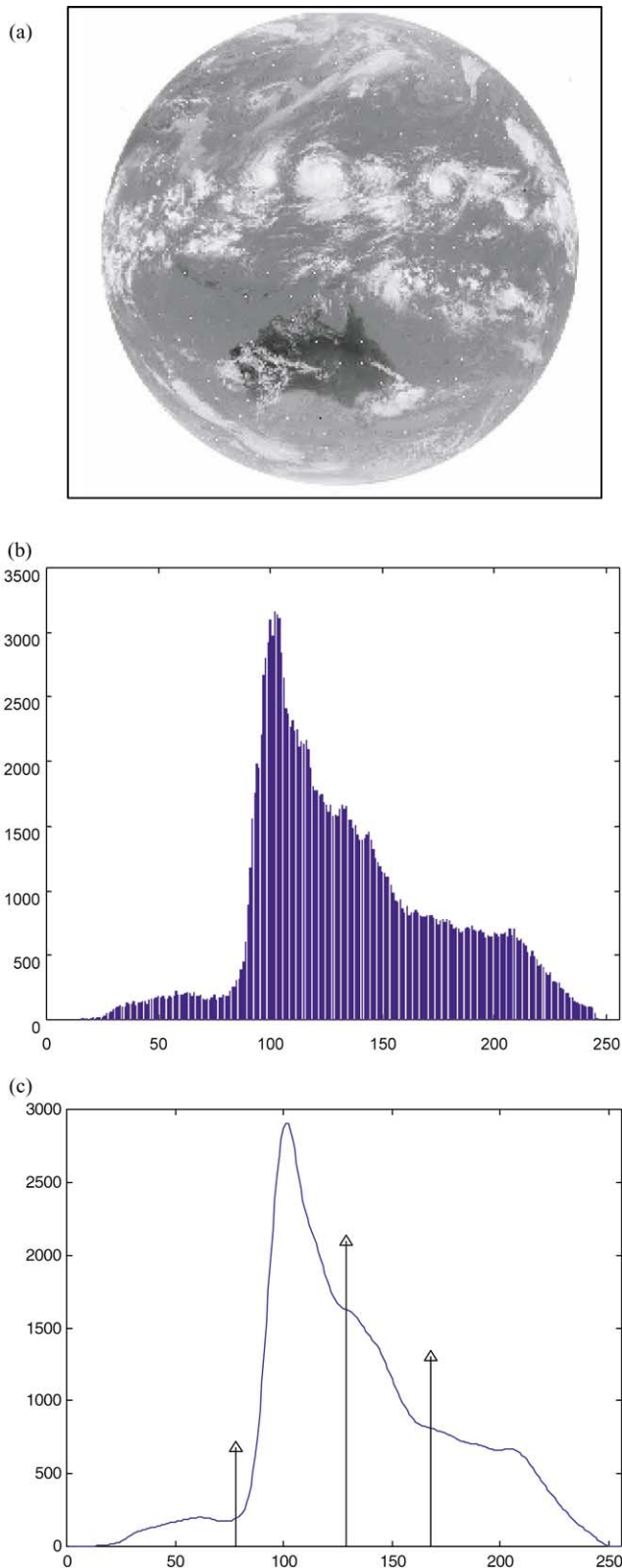


Fig. 9. (a) Original image of the earth; (b) gray-level histogram of the image; and (c) smoothed histogram. The threshold values are 78, 129 and 168. (d) The land; (e) the sea; (f) clouds; and (g) typhoon.

the MF-Estimator and B-Spline methods. Since most of the clusters are considered to be overlapped, the estimated probability of each cluster will be an important indication about the correctness for the threshold value. It means that the thresholding processes have found the optimal threshold values if the estimated probability of each cluster is close to the designed probability.

As shown in Table 1, two clusters were artificially generated with unequal probabilities in the feature space by the designed means and variances. The distribution of these points is shown in Fig. 6. This is a distinctively distributed case. The parameter estimation results after applying the histogram decomposition algorithm are tabulated in Table 1. The estimated means and variances generated by our method differ slightly comparing to the designed data due to the randomness of the random data generation. The estimated probability of each cluster is almost the same as the designed value. This means that the histogram decomposition algorithm has found the best threshold values for this case. On the other hand, the parameters that are estimated by either MF-Estimator or B-Spline methods are not as precise as our method.

As seen in Fig. 7 and Table 2 that the randomly generated simulated data are three distinctively distributed clusters and only two of these three clusters have equal probability. The thresholding result and the estimated parameter values are tabulated in Table 2.

Table 3 shows the designed parameters of three non-distinctively distributed clusters with unequal probabilities. Shown in Fig. 8 is the histogram of the distributions. The estimated parameters are almost perfectly correct with this obscure distribution and the estimation and segmentation results are superior to the other two methods. In this case, the small estimated parameter difference will result in large cluster probability inaccuracy because clusters 2 and 3 are highly overlapped. As summarized in Table 3, our method can always estimate those parameters precisely and find the best threshold values no matter these distributions are distinctively distributed or not.

In the experiments of simulated data, B-Spline method finds local maximum and local minimum points within the histogram distribution. However, local minimums, i.e. the threshold values generated by B-Spline method, can only represent the turning over points between two contiguous distributions. From the result of estimated parameters, local minimums cannot be used as the threshold values because the probability of each candidate distribution is far from the generated data sets especially for non-distinctively distributed case.

For another maximum likelihood estimation based method—the MF-Estimator, the parameter estimation results are also not as accurate as our method. The major disadvantage of MF-Estimator is mainly due to its randomly selected initial estimation point. With this unpredictable initial guess, the interval for further estimation tasks is uncontrollable, even this estimation interval can pass the

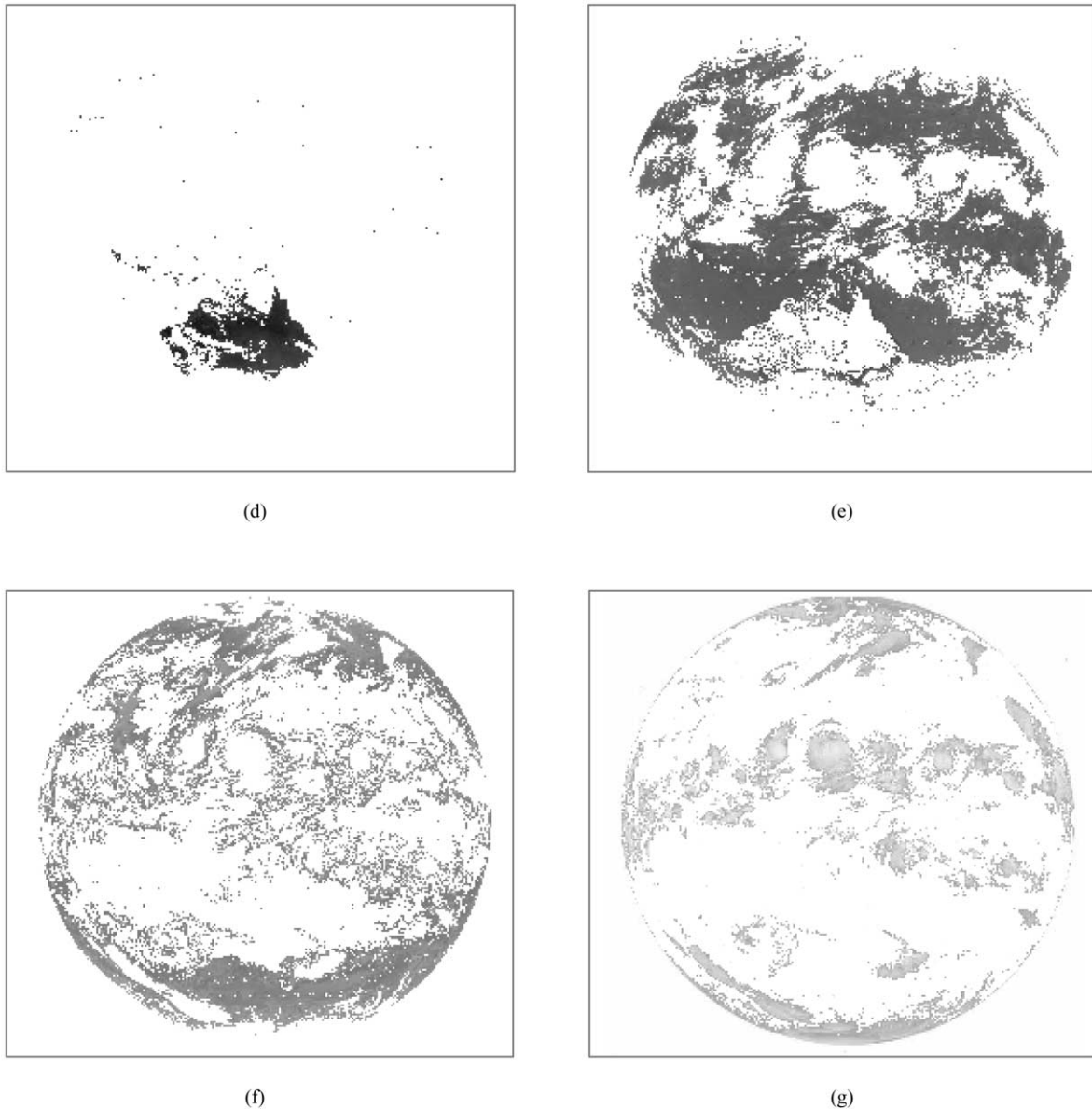


Fig. 9. (continued)

normality test, i.e. the Kolmogorov–Smirnov test [17]. This property makes the estimation results of MF-Estimator can only be close to the genuine data but not exact.

4.2. Experiments with real image data

In this sub-section, we present some experimental results with real image data. Each image was taken and quantized into 256 gray levels. Gray level 1 represents the darkest and gray level 256 represents the brightest pixels. It must be emphasized that the employed images are not especially well suited for multi-thresholding. The gray-level histograms of these image, which are acquired from the real world scene, are used to conduct the experiments to verify that the gray-level histogram comply with the mixture Gaussian distribution, i.e. our assumption, and can be thresholded

by using the proposed method. There are three images that we used in our experiments. They are the earth, blood cells and cameraman as shown in Figs. 9–11, respectively. In each figure, (a) is the original image and (b) is the gray-level histogram of the image. Each histogram of the image is smoothed by convoluting it with a 21 bins Gaussian masking window ($p = 10$) as shown in (c). For multi-thresholding, the smoothed gray-level histogram has been decomposed into a few non-overlapping clusters characterized by their cluster centers m_i , variances σ_i^2 and probability P_i . Then, the optimal threshold values are determined after applying the statistical decision rules as described in Section 4.1. The rest of the images in each figure shows the unaffiliated images generated after thresholding. The result is just like a global segmentation of the image. Let the original image be I and the histogram decomposition algorithm

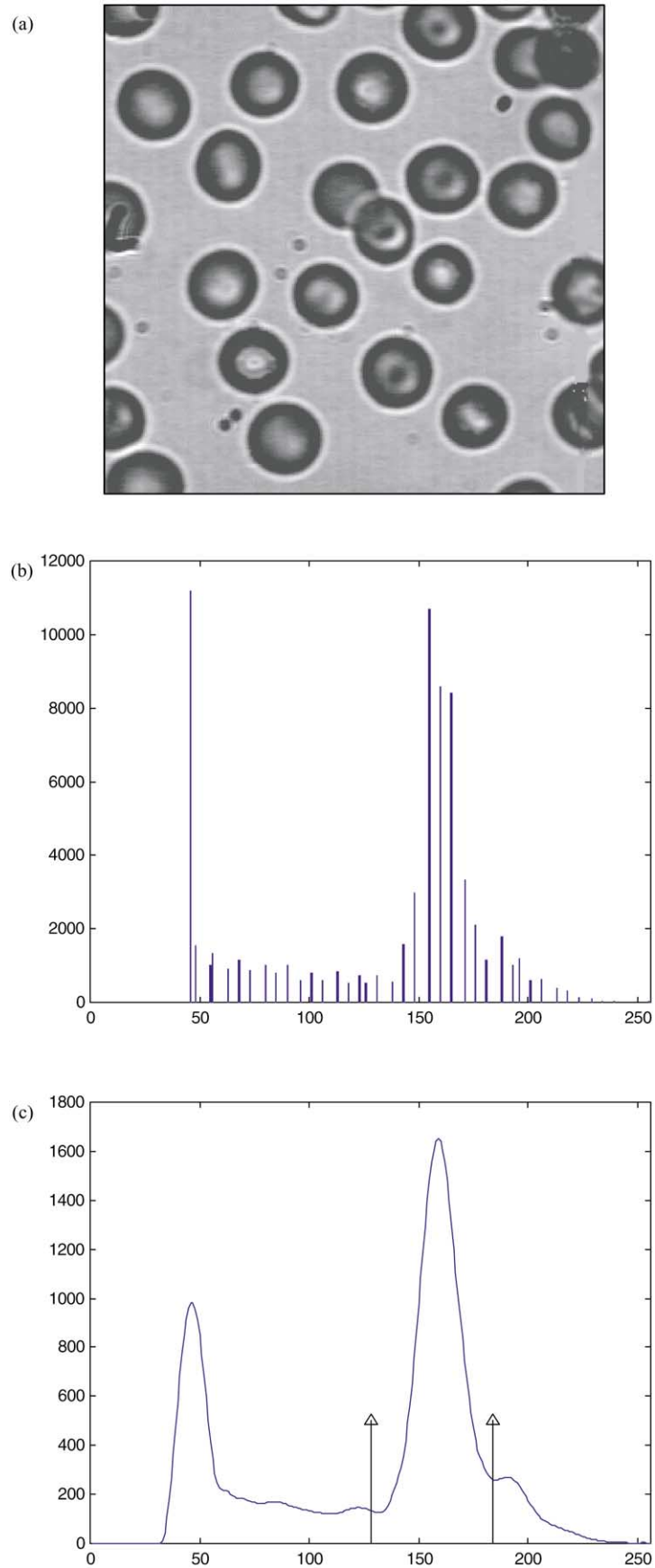
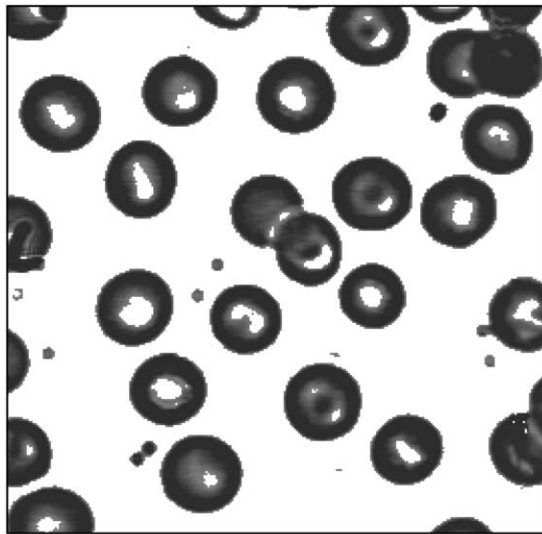
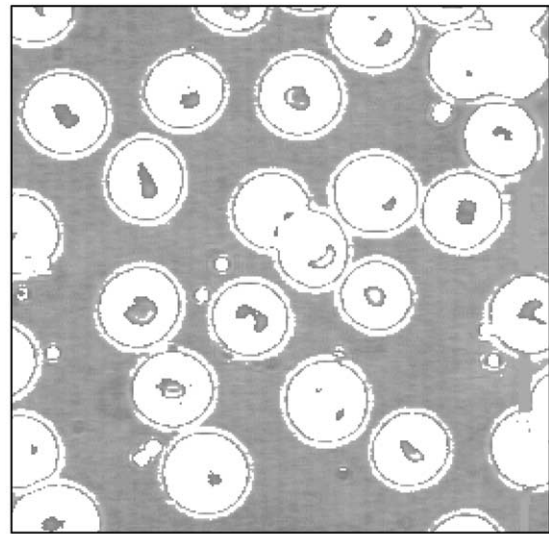


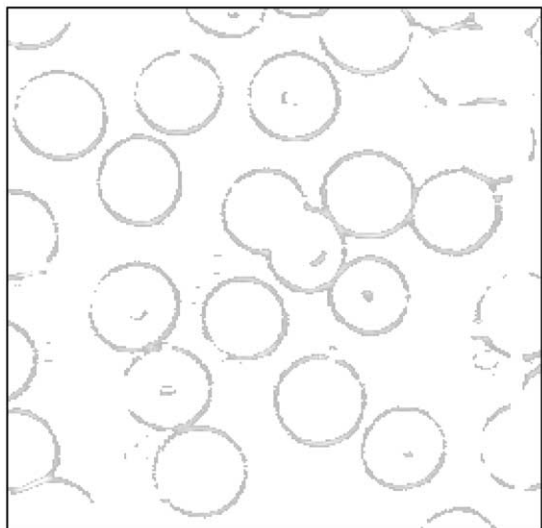
Fig. 10. (a) Original image of blood; (b) gray-level histogram of the image; and (c) smoothed histogram with threshold values. The threshold values are 128 and 184. (d) The blood cells; (e) blood-plasma; and (f) the cell membrane.



(d)



(e)



(f)

Fig. 10. (continued)

has found n clusters. They will have n non-overlapping sub-images, namely I_i $i = 1, \dots, n$, which are separated by $n - 1$ threshold values. We can reconstruct the original image by interlapping all sub-images. That is, $I = \sum_{i=1}^n I_i$. Note that the number of thresholds used, i.e. the number of clusters, is not known a priori since different histograms require different decompositions. It must be emphasized that the employed images are not especially well suited for multi-thresholding.

5. Conclusions

In this paper, we present a multi-modal histogram thresholding method based on the combination of regularization and statistical approaches. The multi-modal histogram is

decomposed into several non-overlapping distributions by modeling it with a mixture Gaussian density. Our method employs a predefined optimal estimation interval to predict the parameters of each cluster in order to reduce the high computational complexity of conventional iterative parameter refinement procedure. Although the histogram is highly contaminated by the contiguous distributions, experiments with simulated data demonstrate that our method can always find the best threshold values and the parameters of predefined distributions, i.e. means, variances and probabilities, of each cluster in the histogram comparing to other methods. Experiments with real images reveal the robustness of the proposed method because the histograms of real world scene generally do not come from genuine Gaussian mixture densities as our assumption. We believe that our method can also be applied to other applications while the

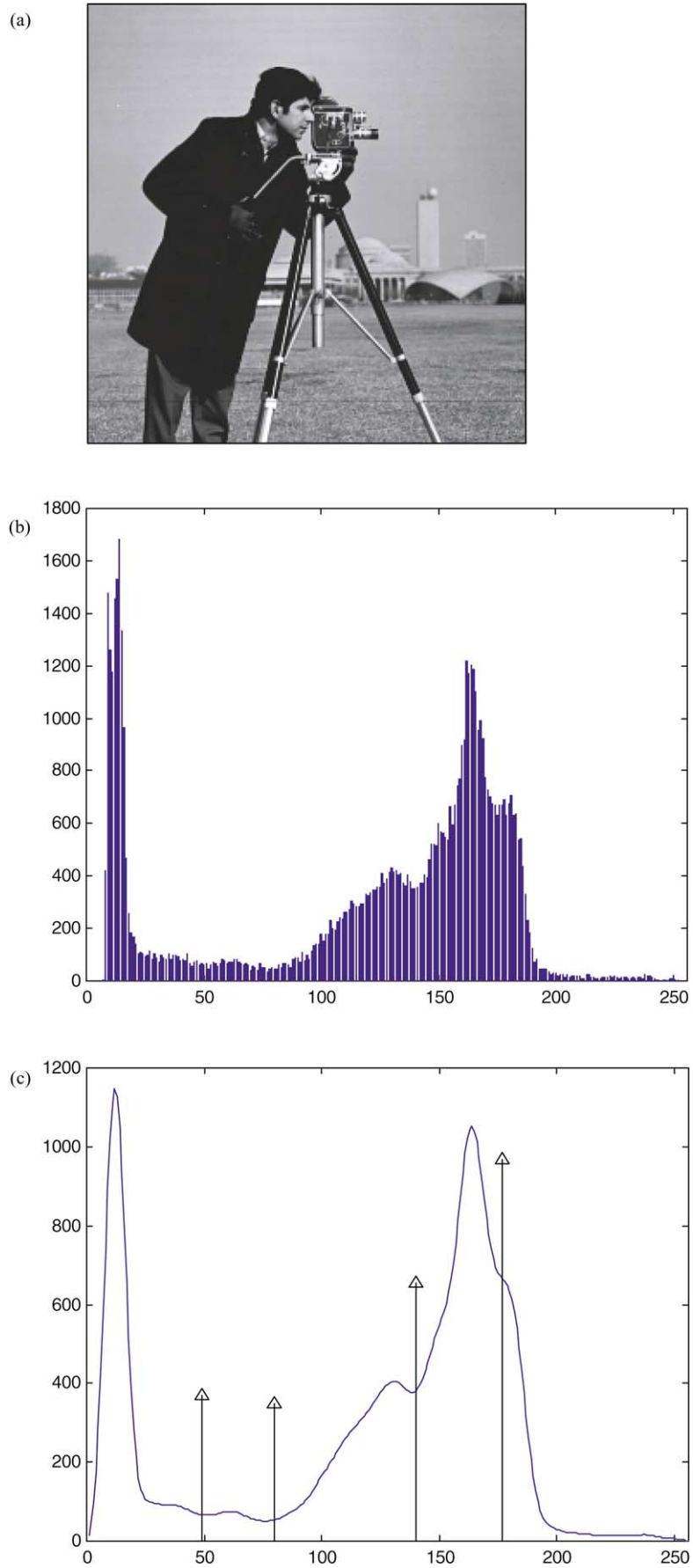
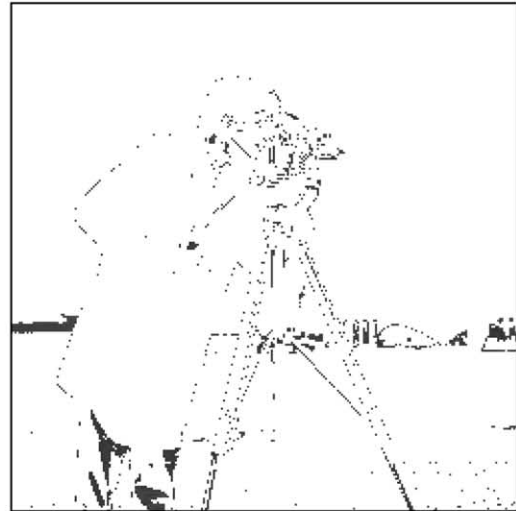


Fig. 11. (a) Original image of the cameraman; (b) gray-level histogram of the image; and (c) the smoothed histogram. Threshold values are 49, 80, 140 and 177. ((d)–(h)) Sub-images of these five clusters.



(d)



(e)



(f)



(g)



(h)

Fig. 11. (continued)

estimation of parameters is necessary. There exists one issue that needs to be discussed in this paper. That is, the factor for smoothing a histogram is heuristically determined which will probably result in different results if the factor changes. Actually, our method chooses the particular factor that can conform to almost all gray-level images and then can generate the best results. Our further research will focus on theoretical analysis and automatically determine different smoothing factors for distinct images.

References

- [1] J.S. Weszka, A survey of threshold selection techniques, *Comput. Vis. Graph. Image Process.* 7 (1978) 259–265.
- [2] P.K. Sahoo, S. Soltani, A.K.C. Wang, Y.C. Chen, A survey of the thresholding techniques, *Comput. Vis. Graph. Image Process.* 41 (1988) 233–260.
- [3] K.S. Fu, J.K. Mui, A survey on image segmentation, *Pattern Recog.* 13 (1981) 3–16.
- [4] S.U. Lee, S.Y. Chung, A comparative performance study of several global thresholding techniques for segmentation, *Comput. Vis. Graph. Image Process.* 52 (1990) 171–190.
- [5] A.Y. Wu, T. Hong, A. Rosenfeld, Threshold selection using quad-tree, *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-4 (1982) 90–93.
- [6] R.L. Kirby, A. Rosenfeld, A note on the use of (gray level, local average gray level) space as an aid in threshold selection, *IEEE Trans. Systems Man Cybernet.* SMC-9 (1979) 860–864.
- [7] L. Hertz, R.W. Schafer, Multilevel thresholding using edge matching, *Comput. Vis. Graph. Image Process.* 44 (1988) 279–295.
- [8] T. Kurita, N. Otsu, N. Abdelmalek, Maximum likelihood thresholding based on population mixture models, *Pattern Recog.* 25 (1992) 1231–1240.
- [9] N. Papamarkos, B. Gatos, A new approach for multilevel threshold selection, *CVGIP: Graph. Models Image Process.* 56 (1994) 357–370.
- [10] J.G. Postaire, M. Ameziane, A pattern classification approach to multilevel thresholding for image segmentation, in: L. Sapiro, A. Rosenfeld (Eds.), *Computer Vision and Image Processing*, 1992, pp. 307–328.
- [11] S.C. Sahasrabudhe, K.S. Das, A valley-seeking threshold selection technique, in: L. Sapiro, A. Rosenfeld (Eds.), *Computer Vision and Image Processing*, 1992, pp. 55–65.
- [12] X. Zhuang, T. Wang, P. Zhang, A highly robust estimator through partially likelihood function modeling and its application in computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 14 (1) (1992) 19–35.
- [13] S. Pal, P. Bhattacharyya, Multipeak histogram analysis in region splitting: a regularisation problem, *IEE Proc.-E* 138 (4) (1991) 285–288.
- [14] M.J. Carlotto, Histogram analysis using a scale-space approach, *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-9 (1) (1987) 121–129.
- [15] R. Render, H. Walker, Mixture densities, maximum likelihood, and the Em algorithm, *SIAM Rev.* 26 (2) (1984).
- [16] J. Jolion, P. Meer, S. Bataouche, Robust clustering with application in computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (8) (1991) 791–802.
- [17] X. Zhuang, Y. Huang, K. Palaniappan, Y. Zhao, Gaussain mixture density modeling, decomposition, and applications, *IEEE Trans. Image Process.* 5 (9) (1996) 1293–1302.

Further reading

- A. Rattarangsi, R.T. Chin, Scale-based detection of corners of planar curves, *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (1992) 430–449.
- L. Gupta, T. Sortrakul, A Gaussian-mixture-based image segmentation algorithm, *Pattern Recog.* 31 (3) (1998) 315–325.