Received 8 November 2020
Revised 27 November 2020
Accepted 27 November 2020

Precise and parallel segmentation model (PPSM) via MCET using hybrid distributions

Soha Rawas and Ali El-Zaart
Department of Maths and Computer Science, Beirut Arab University, Beirut, Lebanon

Abstract

Purpose – Image segmentation is one of the most essential tasks in image processing applications. It is a valuable tool in many oriented applications such as health-care systems, pattern recognition, traffic control, surveillance systems, etc. However, an accurate segmentation is a critical task since finding a correct model that fits a different type of image processing application is a persistent problem. This paper develops a novel segmentation model that aims to be a unified model using any kind of image processing application. The proposed precise and parallel segmentation model (PPSM) combines the three benchmark distribution thresholding techniques to estimate an optimum threshold value that leads to optimum extraction of the segmented region: Gaussian, lognormal and gamma distributions. Moreover, a parallel boosting algorithm is proposed to improve the performance of the developed segmentation algorithm and minimize its computational cost. To evaluate the effectiveness of the proposed PPSM, different benchmark data sets for image segmentation are used such as Planet Hunters 2 (PH2), the International Skin Imaging Collaboration (ISIC), Microsoft Research in Cambridge (MSRC), the Berkley Segmentation Benchmark Data set (BSDS) and Common Objects in COntext (COCO). The obtained results indicate the efficacy of the proposed model in achieving high accuracy with significant processing time reduction compared to other segmentation models and using different types and fields of benchmarking data sets.

Design/methodology/approach – The proposed PPSM combines the three benchmark distribution thresholding techniques to estimate an optimum threshold value that leads to optimum extraction of the segmented region: Gaussian, lognormal and gamma distributions.

Findings – On the basis of the achieved results, it can be observed that the proposed PPSM–minimum cross-entropy thresholding (PPSM–MCET)-based segmentation model is a robust, accurate and highly consistent method with high-performance ability.

Originality/value – A novel hybrid segmentation model is constructed exploiting a combination of Gaussian, gamma and lognormal distributions using MCET. Moreover, and to provide an accurate and high-performance thresholding with minimum computational cost, the proposed PPSM uses a parallel processing method to minimize the computational effort in MCET computing. The proposed model might be used as a valuable tool in many oriented applications such as health-care systems, pattern recognition, traffic control, surveillance systems, etc.

Keywords Minimum cross-entropy thresholding, Hybrid distributions, Precise segmentation, Parallel computing

Paper type Research paper

1. Introduction

Image processing is one of the most challenging issues in an image analysis, used in different fields including medical diagnostics, pattern recognition, etc. Image segmentation techniques have been developed, considering different intensity distributions. A total of four broad

Applied Computing and Informatics
Vol. 20 No. 3/4, 2024
pp. 262-278
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1108/ACI-11-2020-0123

segmentation methods include region- and boundary-based, thresholding and hybrid techniques [1].

Thresholding is one of the most commonly used segmentation techniques due to its simplicity [2]. Different algorithms, i.e. artificial bee colony (ABC), locust swarms (LS), cuckoo search, particle swarm optimization and metaheuristics, have been successfully applied in image thresholding [1, 3]. Though thresholding techniques provide adequate solutions, they are not precise [3]. Metaheuristic methods involve a stochastic process, executing random operations which lead to slow execution. Entropic thresholding, proposed by Pun [4], is, however, widely used due to its ease of implementation. The entropic method, originally proposed by Pullback [5], is used to minimize the cross entropy (minimum cross-entropy thresholding [MCET]) between the original and the segmented images through selecting an optimum threshold between two probabilistic distributions. The perfection of MCET results is affected by predicted distribution type and the number of threshold points. One of the drawbacks of the MCET method is that its complexity increases with the increase in the number of thresholds points [6]. Distribution type predicted using the image histogram, i.e. Gaussian, gamma, lognormal, etc. plays an important role in determining the optimum threshold point. This leads to an accurate thresholding since each image histogram is made using different distributions [7]. However, this technique is time-consuming.

In this paper, a novel hybrid segmentation model is constructed exploiting a combination of Gaussian, gamma and lognormal distributions using MCET. To provide an accurate and high-performance thresholding with minimum computational cost, an improved precise and parallel segmentation model (PPSM) is proposed. The proposed model uses a parallel processing method to minimize the computational effort in MCET computing in locating the optimum threshold in an image histogram.

This paper deals with

(1) Developing a PPSM model using a hybrid of thresholding techniques (Gaussian, gamma and lognormal methods) based on MCET.

(2) Modeling PPSM segmentation for optimization.

(3) Developing and implementing a parallel boosting segmentation algorithm to improve the performance of the PPSM model.

(4) Extensive simulation using different benchmark segmentation data sets to test the effectiveness of the proposed model, using different fields such as the International Skin Imaging Collaboration (ISIC), Planet Hunters 2 (PH2), Common Objects in COntext COCO, Microsoft Research in Cambridge (MSRC) and the Berkley Segmentation Benchmark Data set (BSDS).

2. Related work

The MCET-based thresholding method invented by Li *et al.* [8] has been an easy method used widely to find the optimum thresholding. It has been known for its robustness in finding enhanced solutions for variety and different types of image classes. Chakraborty *et al.* [9] proposed an improved particle swarm optimization (IPSO)-based thresholding method. IPSO was an MCET-based thresholding technique to obtain optimal threshold with a high convergence rate. Mittal *et al.* [10] proposed an exponential *k*-best gravitation search algorithm. El-Zaart *et al.* [11] developed a novel segmentation model for brain cancer detection using Gaussian distribution. The authors used Gaussian distribution if the magnetic resonance imaging (MRI) brain cancer image histogram has symmetric and gamma distributions for nonsymmetric histograms. El-Hajj-Chehade *et al.* [12] proposed fingerprint image segmentation using gamma distribution. The hybrid cross-entropy thresholding using

Gaussian and gamma distributions (HCET-G^2), proposed by Rawas *et al.* [7], combined both the Gaussian and gamma distributions for optimum thresholding of skin cancer images.

Parallel image processing has been implemented on a single compute node using several auto-parallelization tools. Satpute *et al.* [13] proposed a segmentation model based on cross modality using a graphics processing unit (GPU) acceleration platform that provided high performance. In Ref. [14], the authors studied the power of GPU in accelerating microscopy image segmentation. This study points out that the segmentation algorithm needs to exhibit coarse-grained parallelism in order to benefit from GPU parallelization. Liu *et al.* [15] proposed a fuzzy *c*-means (FCM) parallel algorithm based on the Spark-distributed computing platform to solve the image processing and analysis problem for big data of an agricultural image.

Most of the proposed MCET-based segmentation methods follow a single distribution model and ignore the computational complexity that increases significantly with thresholding levels [6]. Consequently, the contribution of this paper over the aforementioned ones is folded in two areas: (1) constructing a robust MCET-based segmentation technique suitable for different application and image classes using the combination of three benchmark segmentation methods (Gaussian, gamma and lognormal) and (2) optimize the developed segmentation model by applying parallel computing technology to boost its performance. To the best of our knowledge, this paper is the first research study that introduces the combination of the three benchmark distributions to extract the optimum thresholds.

3. Image segmentation and thresholding

Image segmentation is used to separate the main image object from its background [6]. Thresholding techniques elaborate in outlining a set of thresholds based on the defined image characteristics such as intensity, texture, position, etc. [16]. It can be categorized into two classes: bimodal and multimodal thresholding. Bimodal segments the image into two separate regions, whereas multimodal type divides an image into a number of separate regions [17], which is out of the scope of this paper.

3.1 Bimodal thresholding

The bimodal technique separates the image into two classes using the image histogram local minimum as shown in Figure 1 [16]. Let $h(x)$ be the histogram of an image f and t^* is the optimum threshold that separates the two different regions $c1$ and $c2$, where $c1$ and $c2$ are class 1 (mode 1) and class 2 (mode 2), respectively. Bimodal thresholding works as follows:

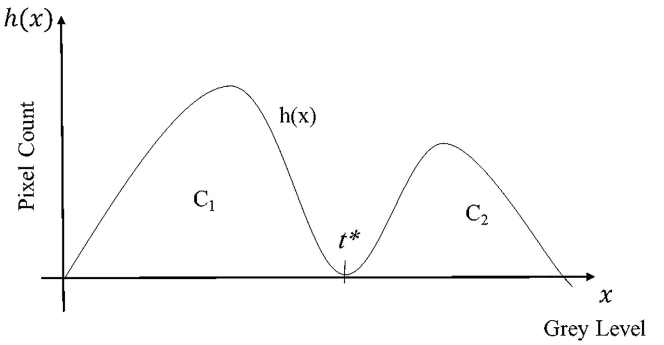

Figure 1.
Bimodal image histogram

$$g(x, y) = \begin{cases} 0, & f(x, y) < t^* \\ 255, & f(x, y) \geq t^* \end{cases} \quad (1)$$

where $g(x, y)$ represents the image intensity (pixel value).

Through applying the bimodal algorithm on $h(x)$, we can reach t^* using the following equation:

$$t^* = \text{bimodal Threshold } (0, 255, h) \quad (2)$$

where t^* is between the image histogram gray level [0, 255].

4. Minimum cross-entropy thresholding and probabilistic distributions

MCET that is based on Gaussian distribution is one of the widely used entropy methods [4]. However, when considering the nonsymmetric images, i.e. those that provide a nonsymmetric histogram, Gaussian distribution will not be an effective method [18]. Accordingly, nonsymmetric distributions to provide more proficiency in diagnosing different gray-level images are needed, such as K, gamma and beta distributions [7].

4.1 Cross-entropy thresholding

In 1968, Kullback [18] proposed the cross-entropy method to evaluate the similarity between two probabilistic distributions. Let f and g be two probabilistic distributions on the same set such that $f = \{f_1, f_2, \ldots, f_n\}$ and $g = \{g_1, g_2, \ldots, g_n\}$ and such that f_i and g_i come from the same location in the image space. Consequently, the information–theoretical distance between the two distributions (f and g) is denoted as a cross entropy and calculated using the following function:

$$D(f, g) = \sum_{i=1}^{n} f_i * \log\left(\frac{f_i}{g_i}\right) \quad (4)$$

Let $I(x, y)$ be an original image and $I_t(x, y)$ be a thresholded image such as

$$I_t(x, y) = \begin{cases} \mu_1(1, t) = \mu_a(t), & I(x, y) < t \\ \mu_2(t, L+1) = \mu_b(t), & I(x, y) \geq t \end{cases} \quad (5)$$

where t denotes the obtained threshold that divides the image into two different classes (A and B as Figure 2 reveals), $\mu_a(t)$ and $\mu_b(t)$ are the mean value of classes A and B, respectively (i.e. dark and bright regions), as shown in Figure 2. Therefore, the cross entropy between $I(x, y)$ and $I_t(x, y)$ is defined by the following equation:

$$D(I, I_t) = \sum_{i=0}^{t-1} i * h(i) * \log\left(\frac{i}{\mu_a(t)}\right) + \sum_{i=t}^{L} i * h(i) * \log\left(\frac{i}{\mu_b(t)}\right) \quad (6)$$

where $h(i)$ is the image histogram for $i = 1, 2, \ldots, L$, such that $L = 255$ being the number of gray levels.

4.2 Minimum cross-entropy thresholding

The MCET technique developed by Li *et al.* [8] determines the optimal threshold t^* through minimizing the cross entropy as follows:

$$t^* = \arg\min_t(D(I, I_t)) = \arg\min_t(D(t)) \quad (7)$$

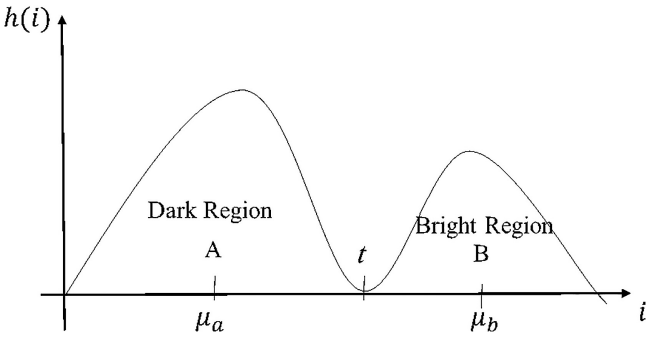

Figure 2.
Image histogram

Noting the distance between the original and the thresholded images $D(I, I_t)$ that can be written as $D(t)$ is determined as follows:

$$D(t) = \sum_{i=1}^{L} i * h(i) * \log(i) - \sum_{i=1}^{t-1} i * h(i) * \log(\mu_a(t)) - \sum_{i=t}^{L} i * h(i) * \log(\mu_b(t)) \quad (8)$$

Since $\sum_{i=1}^{L} i * h(i) * \log(i)$ is constant for a given image, the objective function can be redefined using the following equation:

$$n(t) = -\sum_{i=1}^{t-1} i * h(i) * \log(\mu_a(t)) - \sum_{i=t}^{L} i * h(i) * \log(\mu_b(t)) \quad (9)$$

Therefore, the objective function can be rewritten as follows:

$$n(t) = A(t) * \log(\mu_a(t)) + B(t) * \log(\mu_b(t)) \quad (10)$$

where $A(t) = -\sum_{i=1}^{t-1} i * h(i)$ and $B(t) = -\sum_{i=t}^{L} i * h(i)$.

4.3 Probabilistic distributions

In image segmentation, the optimum thresholding depends on predicting the best distribution type that forms the pixels of the image segment [7].

Gaussian distribution: The normal or Gaussian probability distribution is often labeled as a bell-shaped curve [19]. In image segmentation, if the data in the image are assumed to be modeled by Gaussian distribution, therefore, $\mu_a(t)$ and $\mu_b(t)$ can be estimated from two Gaussian distributions as follows [18]:

$$\mu_a(t) = \frac{\sum_{i=0}^{t-1} i * h(i)}{\sum_{i=0}^{t-1} h(i)} \quad (11)$$

$$\mu_b(t) = \frac{\sum_{i=t}^{L} i * h(i)}{\sum_{i=t}^{L} h(i)} \quad (12)$$

where μ is the distribution mean, t is the obtained threshold that divides the image into two different regions, L is the image gray levels and h is the image histogram.

Gamma distribution is known as a general type of statistical distribution [7]. Based on El-Zaart *et al.*'s derivation [20], if the data in the image are assumed to be modeled by gamma distribution, $\mu_a(t)$ and $\mu_b(t)$ can be calculated as follows:

$$\mu_a(t) = \sqrt{\frac{\sum_{i=0}^{t-1} h(i) * i^2 * q^2}{\sum_{i=0}^{t-1} h(i)}} \tag{13}$$

$$\mu_b(t) = \sqrt{\frac{\sum_{i=t}^{L} h(i) * i^2 * q^2}{\sum_{i=t}^{L} h(i)}} \tag{14}$$

where

$$q = \frac{\mu(N + 0.5)}{\sqrt{N} * \mu(N)}$$

such that N is the shape of the distribution.

Lognormal distribution: In probability theory, lognormal distribution is a statistical distribution of the logarithmic value from an associated normal distribution. Thus, if the data in the image are assumed to be modeled by lognormal distribution, therefore, $\mu_a(t)$ and $\mu_b(t)$ can be calculated from two lognormal distributions as follows [21]:

$$\mu_a(t) = \frac{\sum_{i=0}^{t-1} \log(i) * h(i)}{\sum_{i=0}^{t-1} h(i)} \tag{15}$$

$$\mu_b(t) = \frac{\sum_{i=t}^{L} \log(i) * h(i)}{\sum_{i=t}^{L} h(i)} \tag{16}$$

4.4 Performance measure

The performance measurement metric is an important issue to find the best segmentation method [22]. According to the literature, there is no standard and benchmark evaluation metrics used by the researcher [23]. Therefore, to evaluate the proposed PPSM objectively and since we are developing a new entropy-based thresholding method, we followed the following metrics:

Image uniformity (IU): IU is a well-known distance metric [23]. It has been investigated as an evaluation metric for entropy-based segmentation methods. The IU method measures quantitatively inter and intra-region uniformity differences between the original and segmented images. For a segmented image, the IU value ranges between [0, 1], such that 1 and 0 indicate perfect and bad segmentation outputs, respectively. For a given threshold t, the image uniformity IU(t) is defined as follows:

$$\mathrm{IU}(t) = 1 - \frac{\sigma_1^2(t) + \sigma_2^2(t)}{C} \tag{17}$$

where $\sigma_1^2(t)$ and $\sigma_2^2(t)$ represent the class variance of the object and its background, respectively, such that

$$\sigma_1^2(t) = \frac{\sum_{i=0}^{t-1} (i - \mu_1(t))^2 * h(i)}{\sum_{i=0}^{t-1} h(i)} \tag{18}$$

$$\sigma_2^2(t) = \frac{\sum_{i=t}^{L} (i - \mu_2(t))^2 * h(i)}{\sum_{i=t}^{L} h(i)} \tag{19}$$

where h represents the image histogram, $\mu_1(t)$ and $\mu_2(t)$ are is the mean values of dark and bright regions, respectively (as shown in Figure 3), calculated using Gaussian distribution as indicated in Eqns (11–12) and L indicates the different image gray level L [0, 255]. Where C represents half of the squared difference between the maximum and minimum original gray-level value and is calculated as follows:

$$C = \frac{(g_{\max} - g_{\min})^2}{2} \tag{20}$$

such that $g_{\max}$ and $g_{\min} \in [0 \ldots L]$.

Region contrast (RC): The RC metric defines the inter-region disparity that can be calculated using the absolute difference of the object mean values of dark and bright regions, respectively, divided by the sum of their average mean [23], as shown in Eqn (21):

$$\mathrm{RC}(t) = \frac{|\mu_1(t) - \mu_2(t)|}{\mu_1(t) + \mu_2(t)} \tag{21}$$

Noting that, and similar to the IU metric, for a given value of threshold t RC(t), this metric value ranges between [0, 1], such that 0 and 1 indicate bad and perfect segmentation performance, respectively.

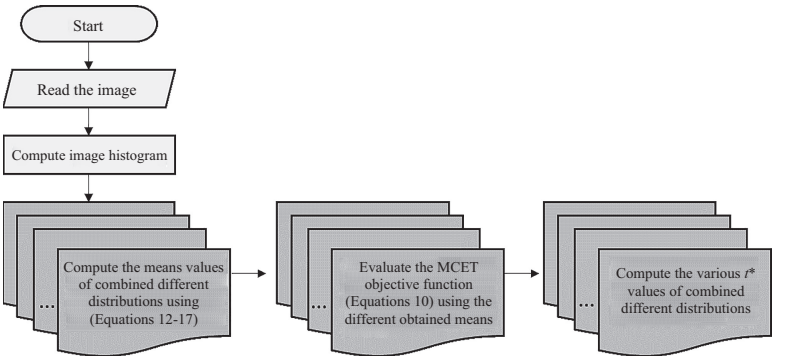

Figure 3.
Workflow architecture of the parallelized PPSM algorithm

5. PPSM formulation

This section provides generalized modeling of the PPSM to derive an accurate segmentation model based on finding ideal distributions forming the segmented image histogram.

5.1 Modeling the image segmentation problem

Different distributions have been used in minimizing the MCET such as Gaussian, gamma and lognormal distributions. However, finding the best type that leads to optimum thresholding t^* drag the image segmentation problem to be a nondeterministic polynomial (NP)-hard optimization problem.

Let $I(x, y)$ be the original image to be segmented. Since the MCET technique to find an optimum threshold depends on Eqn (7), the PPSM, which is an entropy-based model, has the following as an objective function:

$$\text{Minimize}(t^*) \tag{22}$$

subject to the subsequent quality constraints (QCs) that aim to maximize the accuracy of the thresholded output through maximizing the IU and its RC using Eqns (17) and (21) as follows:

$$\text{Maximize}\begin{pmatrix} \text{IU}(t^*) \\ \text{RC}(t^*) \end{pmatrix} \tag{23}$$

$$\text{IU}(t^*),\ \text{RC}(t^*) \in [0,\ 1]$$

Noting that t^* could be any threshold between [0, 255].

5.2 PPSM formulation

The PPSM depends on optimizing and developing an accurate thresholding t^* through determining the best distribution type that forms the original image histogram. Therefore, the PPSM proposes that the image $I(x,y)$ is composed of hybrid distributions using Eqns (12)–(17). Thus, the image histogram could be modeled using the following function:

$$h(x) = P_a * \text{dist}_a(x,\ \mu_a(t)) + P_b * \text{dist}_b(x,\ \mu_b(t))$$

such that (1) $\text{dist}_a(x,\ \mu_a(t))$ is the first distribution and can be Gaussian/gamma/lognormal and (2) $\text{dist}_b(x,\ \mu_b(t))$ is the second distribution type and can be gamma/Gaussian/lognormal. Noting that P_a, P_b are two *prior* probabilities such that $P_a + P_b = 1$.

6. Solving the optimization problem and PPSM algorithm

In this section, a PPSM algorithm is proposed to find an optimum threshold t* that leads to an accurate extraction of the image from its background.

6.1 PPSM algorithm

Algorithm 1 shows a high-level pseudocode of the proposed PPSM algorithm based on MCET using a hybrid distributions combination of Gaussian, gamma and lognormal. The algorithm starts with reading the original image (line 1) to compute its histogram (line 2). Then, the PPSM algorithm works iteratively on each image histogram point (i.e. [0, 255]) to attain the optimum MCET t^* (line 3–10). As shown in algorithm 1, different mean values are computed using different distributions (lines 4–6). The obtained means used in their various combinations (line 7) to compute nine hybrid thresholds that reduce the MCET (line 8). The PPSM algorithm selects the optimum threshold t^* based on the average of the

performance measure metrics (lines 11–12). Finally, the algorithm returns the optimum threshold t^* that satisfies QC maximization for best segmentation performance.

Proposed Algorithm 1 PPSM Algorithm

Input: $I(x, y)$// the original image
Output: t*
Processing:
1: Read image $I(x, y)$
2: Compute the image I histogram h(i), i=0, …., 255
3: for j = 1 : 255 do
4: Compute $\mu_{a1}(j)$ and $\mu_{b1}(j)$ using Equations (12-13)
5: Compute $\mu_{a2}(j)$ and $\mu_{b2}(j)$ using Equations (14-15)
6: Compute $\mu_{a3}(j)$ and $\mu_{b3}(j)$ using Equations (16-17)
7: Using Equation 10, Compute $n_1(i)$ based on $\mu_{a1}(j)$ and $\mu_{b1}(j)$
Compute $n_{11}(i)$ based on $\mu_{a1}(j)$ and $\mu_{b2}(j)$
Compute $n_{111}(i)$ based on $\mu_{a1}(j)$ and $\mu_{b3}(j)$
Compute $n_2(i)$ based on $\mu_{a2}(j)$ and $\mu_{b2}(j)$
Compute $n_{22}(i)$ based on $\mu_{a2}(j)$ and $\mu_{b1}(j)$
Compute$n_{222}(i)$ based on $\mu_{a2}(j)$ and $\mu_{b3}(j)$
Compute $n_3(i)$ based on $\mu_{a3}(j)$ and $\mu_{b3}(j)$
Compute $n_{33}(i)$ based on $\mu_{a3}(j)$ and $\mu_{b1}(j)$
Compute $n_{333}(i)$ based on $\mu_{a3}(j)$ and $\mu_{b2}(j)$
8: Compute t1*, t2*, …, t9* using the different distribution combinations (line 8)
i.e. If $n\_x(i)$ < min-value_x
min-value_x = $n\_x(i)$
t_x* = j
End if
9: j = j + 1
10: End for
11: Compute Q(t1*), Q(t2*), …, Q(t9*) using the average sum of the performance measure of Equation 23
12: best(t) = Maximum(Q(t1*), Q(t2*), …, Q(t9*)) // select the best output
13: t*=t
14: End if
15: Return t*

The computational complexity to obtain a threshold t^* is $O(L^2)$ using a homogenous distribution environment [17]. However, obtaining the optimum threshold could be exhaustive and time-consuming under hybrid distribution scenarios. Thus, for n-thresholding problem, the PPSM algorithm computational complexity will be $O(L^{n+1})$. For that reason, this paper proposes to use the parallel processing technology, so that most of the used computational resources could be consumed correctly and our key aim in achieving minimum processing time could be attained.

7. Parallel processing

MCET computation is recognized as a complex task due to the exhaustive search to obtain the best thresholds. Although recursive programming is used widely in MCET computation, this will not be effective using simultaneous computation [24]. Therefore, dividing the larger problem into smaller ones and applying parallel processing technology can save the available computing resources and boost PPSM processing time.

Parallel processing technology has been used to speed up application performance through manipulating their essential parallelism [24]. Our aim is to benefit from this technology and improve PPSM algorithm performance using the embedded parallel central processing units (CPUs) and GPUs. Accordingly, the proposed PPSM algorithm has been parallelized among the available computing resources to overcome the drawback of its processing time. However, to take the advantage of the multithreading technique using a multiprocessor computing architecture, the application dependency constraint shall be applicable.

Independency: One of the critical requirements to design a parallel algorithm is that the parallelized problem type shall contain independent iterations [25]. Let Seg_i and Seg_j be two

program segments, therefore, to execute these two segments concurrently, the following should be satisfied.

Consider that Input_i and Output_i are the input and output variables, respectively, for Seg_i and Input_j and Output_j are the input and output variables, respectively, for Seg_j. Therefore, Seg_i and Seg_j segments are considered to be parallel if they satisfied the following conditions:

$$\text{Input}_i \cap \text{Output}_j = \emptyset$$

$$\text{Input}_j \cap \text{Output}_i = \emptyset$$

$$\text{Output}_i \cap \text{Ouput}_j = \emptyset$$

7.1 The methodology to parallelize the PPSM algorithm

The parallelization process that has been applied to the PPSM algorithm described is generic and can be applied to any application that satisfies the independency conditions. In particular, the parallelized system is divided into the following three stages:

(1) Means distribution: calculate the different means distribution using the combination of different Eqns (12)–(17). In this stage, the means could be divided into k parts (k = different types of distributions, where $k = 3$), such as each part is treated as a separate task.

(2) MCET evaluation: on each available processor, the MCET objective function Eqn (10) was evaluated using the combination of the resultant means.

(3) Threshold computation: evaluate the obtained different thresholds using the combination of different distributions to obtain the optimal t^* that suits to form the final segmentation.

Using the above methodology, we can speed up the proposed PPSM algorithm up to p times using p parallel processors in one computing machine.

Figure 3 illustrates the workflow architecture of the used methodology to parallelize the PPSM algorithm, such that the dark gray shapes are the one that represent the tasks running in parallel.

Algorithm 2 shows a high-level pseudocode for the parallel implementation of the PPSM algorithm using lines 3–10.

Algorithm 2 PPSM parallel implementation

```
for j = 1 : 255 do
    Parallel compute of different mean values using Equations 12-17
    Parallel evaluate of MCET objective function using obtained means
    Parallel compute of various thresholds (t*)
End for
```

8. Performance evaluation

This section evaluates the efficacy of the proposed PPSM using MATLAB R2019b with 3.4 GHz Intel(R) Core(TM) i7-4770 machine of 8 GB RAM. Multithreaded implementation has been developed using a MATLAB parallel computing toolbox, which offers the parallelization technology on a multicore computing platform. Using a parallel processor computing node, the *parpool* function provided by the MATLAB parallel toolbox assigns the threads of the proposed PPSM algorithm to the available cores [26]. Noting that, the used Core i7 desktop processors feature four cores with eight concurrent threads.

8.1 Data sets

To test the validity of the proposed PPSM–MCET-based segmentation model, five different benchmark data sets have been used to objectively evaluate the model's accuracy and performance.

ISIC: The ISIC is a dermoscopic data set and an open-source public archive of skin images [27]. It has been used as a benchmark for developing and testing the segmentation methods used in the automated diagnostics system of melanoma diseases. The ISIC data set contained more than 2,000 skin lesion images collected from different clinical centers all over the world. The ISIC data set is the largest publicly available benchmark dermoscopic data set for skin cancer detection and supported by the International Society for Digital Imaging of the Skin (ISDIS).

PH2: The PH2 data set is one more dermoscopic benchmark skin lesion data set developed by the Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal [28]. It consists of currently 200 melanocytic lesions. These images comprise 80 common and 80 atypical nevi and 40 melanomas.

MSRC: MSRC is a static scene parsing data set provided from Microsoft Research in Cambridge [22]. It is one of the benchmark segmentation data sets that are made up of 591 images divided into 23 object classes.

BSDS: The BSDS is a large data set of natural images developed by the Berkeley University of California [29]. The BSDS comprises 500 natural images. It is one of the most difficult benchmark segmentation data sets as it includes various object classes such as posture and background variations.

COCO: COCO is the largest benchmark segmentation data set provided by Microsoft [30]. The COCO data set provides about 328k images for holistic scenes divided into 91 object classes (Table 1).

8.2 Experimental results

In this section, numerous experiments are conducted to validate the effectiveness of the PPSM and to validate the efficacy of the multithreading system in achieving fast and accurate segmentation. In order to prove the importance of the proposed algorithm, three benchmarks and classic MCET-based segmentation algorithms are used for comparison: Gaussian, gamma and lognormal approaches. To evaluate the PPSM performance, two performance metrics have been recorded: IU and RC metrics as discussed in Section 4.4. For parallelism evaluation, the total processing time required to execute the multithreaded PPSM algorithm is compared to the sequential processing.

8.2.1 Accuracy. The merits of the proposed PPSM are validated using the output comparison in Tables 2 and 3. Tables 2 and 3 contain a detailed comparison of various combination distribution means used in our proposed PPSM using dermoscopic and miscellaneous data sets, respectively. As shown in Tables 2 and 3, 99% of the tested benchmark data sets

Table 1.
Summary of the used benchmark data sets' specifications

Data set name	Number of images	Image category
ISIC	2,000	Dermoscopic
PH2	200	Dermoscopic
MSRC	591	Natural scene
BSDS	500	Natural scene
COCO	328K	Miscellaneous

confirmed that the image histogram is a combination of different distribution types. Consequently, these results validate that the proposed PPSM is a robust model and overtakes the other baseline segmentation methods with 99% accuracy. Noting that for the huge data sets that contain more than 500 images, i.e. PH2 and COCO data sets, 500 images are selected randomly for testing. One of the interesting outcomes, as shown in Tables 2 and 3, is that the lognormal–Gaussian combination attained the best results with an average 90% compared to other combinations and competing segmentations algorithms. Accordingly, this could provide evidence and draw a new conclusion that image histograms mostly followed lognormal and gamma distributions at their start and endpoints, respectively. Moreover, classical methods that used single distributions such as Gaussian, gamma or lognormal face great difficulties to find the best threshold values as shown in the obtained results.

The PPSM algorithm is tested over a various number of thresholds to check the correctness and stability of the proposed segmentation model. Figure 4 shows (1) a selected

PPSM joint distributions	ISIC # of best segmented images	ISIC Best performance (%)	PH2 # of best segmented images	PH2 Best performance (%)
Gaussian	0	0.0	0	0.0
Gamma	1	0.2	0	0.0
Lognormal	2	0.4	0	0.0
Gaussian–gamma	14	2.8	15	7.5
Gaussian–lognormal	5	1.0	6	3.0
Gamma–Gaussian	22	4.4	1	0.5
Gamma–lognormal	0	0.0	0	0.0
Lognormal–Gaussian	456	91.2	178	89.0
Lognormal–gamma	0	0.0	0	0.0

Table 2. Hybrid distribution combinations results of PPSM model using dermoscopic ISIC and PH2 datasets

PPSM joint distributions	MSRC # of best segmented images	MSRC Best performance (%)	BSDS # of best segmented images	BSDS Best performance (%)	COCO # of best segmented images	COCO Best performance (%)
Gaussian	1	0.2	0	0	1	0.2
Gamma	0	0.0	0	0	0	0.0
Lognormal	0	0.0	2	0	0	0.0
Gaussian–gamma	1	0.2	42	8	10	2.0
Gaussian–lognormal	2	0.4	19	4	1	0.2
Gamma–Gaussian	2	0.4	38	8	3	0.6
Gamma–lognormal	0	0.0	0	0	0	0.0
Lognormal–Gaussian	450	98.7	399	80	485	97.0
Lognormal–gamma	0	0.0	0	0	0	0.0

Table 3. Hybrid distribution combinations results of PPSM model using miscellaneous images of MSRC, BSDS, and COCO datasets

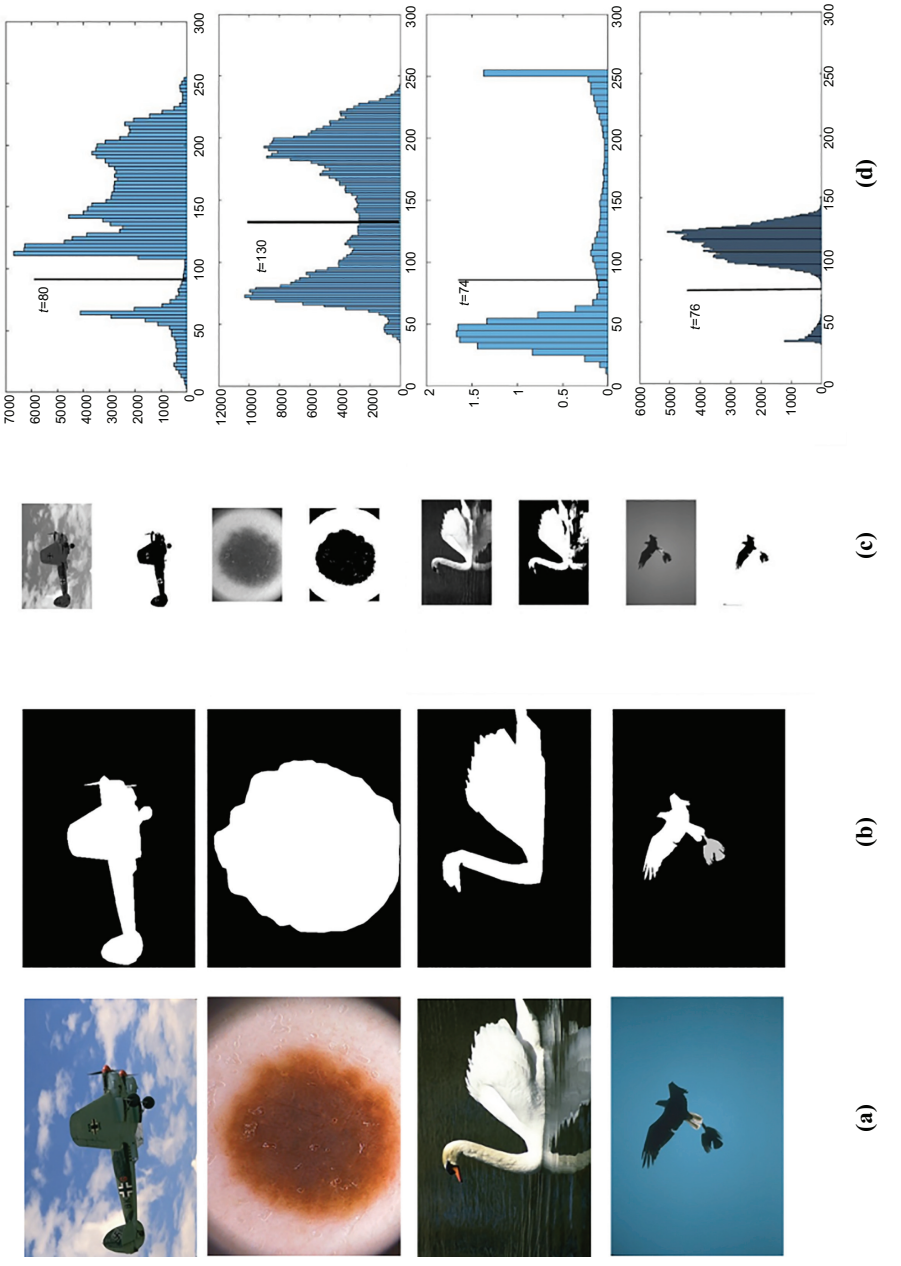

Figure 4.
Selected lesion segmentation using PPSM (a) Original image, (b) Image ground truth, (c) PPSM segmented output and (d) Thresholded image histogram

number of original images from the used five different data sets, (2) the image ground truth, (3) the PPSM result and (4) the thresholded image histogram. The segmentation output compared with the ground truth shows promising results in finding the optimal thresholds. This clearly demonstrates the obtained results in Tables 2 and 3 and reflects the fact that the proposed PPSM outperforms the benchmark segmentation models by 99%.

8.2.2 Parallel processing efficiency. The primary aim of the proposed PPSM is to provide optimum efficiency. This could be represented by the total processing time taken by the CPU to complete the segmentation process. Using Core i7 desktop processors features four cores with eight concurrent threads; Table 4 presents an effective comparison in the computation time (in secs) between sequential and parallel processing methods using the PPSM algorithm. As shown in Table 4, the computation time performance under parallel processing is superior to the sequential method with an average of 58%. An important notice is that the proposed PPSM is designed for program execution on parallel cores due to the PPSM data-dependent properties. It is also worth to be noted that our proposed approach successfully increases the efficiency using a single core in terms of performance and multithreading. This reflects the conventional wisdom in academics and industry through workload distribution rather than scaling up systems.

The experimental results recorded in Table 4 validate the importance and efficiency of the proposed parallel processing model. However, for more accuracy and to shed light on the statistical importance of the applied results, a statistical test has been conducted using the Wilcoxon test [31]. Accordingly, the difference in performance gain using the PPSM parallel model and sequential one shows statistically significant results using the Wilcoxon test that gives a p-value of 0.0079, confirming the significance of this improvement.

The accuracy of the segmentation model is verified using the average of the performance measure of IU and RC as discussed in Section 4. According to these metrics, higher values indicate an accurate segmentation model. Consequently, the accuracy of the PPSM is obviously shown using Figure 5 that includes a randomly selected average of various images

Data set name	Sequential (secs)	Parallel computing (secs)	Speed-up saving (%)
ISIC	346.35	187.62	45.8
PH2	263.88	136.96	48.1
MSRC	283.38	98.79	65.1
BSDS	375.82	145.07	61.3
COCO	487.92	164.16	66.3

Table 4.
Performance gain using PPSM parallel computing model implementation

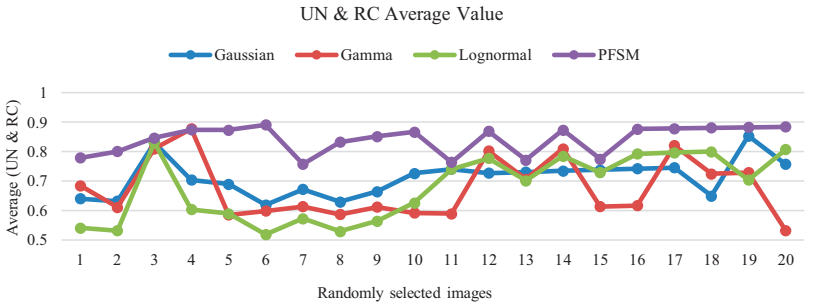

Figure 5.
Performance metric measurement comparing PPSM accuracy with benchmark Gaussian, gamma and lognormal segmentation methods

using different data sets. This depicts that the proposed PPSM leads to optimum thresholding with high processing performance.

On the basis of the above results, it can be observed that the proposed PPSM–MCET-based segmentation model is a robust, accurate and highly consistent method with high-performance ability. Moreover, predicting the multidistributions of the one-image histogram is more appropriate for optimum thresholding and the image histogram analysis.

9. Conclusion and future work

This paper proposed a novel PPSM for optimum thresholding of different application types and image classes. The proposed MCET-based segmentation model was built to be a combination of different benchmark distribution models (i.e. Gaussian, gamma and lognormal). The PPSM was formulated and implemented to fit bimodal thresholding techniques. The accuracy of the proposed model, which has been formulated as an optimization function, is tested using five benchmark data sets and compared with the other three related approaches. It has been shown that it outperforms the three widely known MCET-based segmentation techniques that used Gaussian, gamma and lognormal for image histogram distribution. The PPSM applied a novel and parallel boosting algorithm to minimize the model computation time and improve its performance. The experimental results have indicated that the PPSM algorithm preserves an accurate segmentation output and is able to find an optimum threshold even with different types of image classes.

The future aim is to extend the PPSM to cover the multimodal thresholding techniques. Moreover, extending the performance metric measure to improve the quality and accuracy of the segmented images is one of the important goals that should be studied carefully.

References

1. Liang H, Jia H, Xing Z, Ma J, Peng X. Modified grasshopper algorithm-based multilevel thresholding for color image segmentation. IEEE Access. 2019; 7: 11258-95.
2. Rundo L, Tangherloni A, Cazzaniga P, Nobile MS, Russo G, Gilardi MC, Vitabile S, Mauri G, Besozzi D, Militello C. A novel framework for MR image segmentation and quantification by using MedGA. Comput Methods Progr Biomed. 2019; 176: 159-72.
3. Oliva D, Abd Elaziz M, Hinojosa S. Multilevel thresholding for image segmentation based on metaheuristic algorithms. Metaheuristic Algorithms for image Segmentation: Theory and Applications; Cham: Springer; 2019: 59-69.
4. Pun T. Entropic thresholding, a new approach. Comput Graph Image Process. 1981; 16(3): 210-39.
5. Kullback S. Information theory and statistics. New York: Dover: 1968.
6. Oliva D, Hinojosa S, Osuna-Enciso V, Cuevas E, Pérez-Cisneros M, Sanchez-Ante G. Image segmentation by minimum cross entropy using evolutionary methods. Soft Comput. 2019; 23(2): 431-50.
7. Rawas S, El-Zaart A. HCET-G 2: dermoscopic skin lesion segmentation via hybrid cross entropy thresholding using Gaussian and gamma distributions. 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS): IEEE; 2019.
8. Li, CH, Lee CK. Minimum cross entropy thresholding. Pattern Recogn. 1993; 26(4): 617-25.
9. Chakraborty R, Rama Sushil R, Garg ML. An improved PSO-based multilevel image segmentation technique using minimum cross-entropy thresholding. Arabian J Sci Eng. 2019; 44(4): 3005-20.

10. Mittal H, Saraswat M. An optimum multi-level image thresholding segmentation using non-local means 2D histogram and exponential Kbest gravitational search algorithm. Eng Appl Artif Intell. 2018; 71: 226-35.
11. Chehade WEH, Abdel Kader R, Ali El-Zaart A. Segmentation of MRI images for brain cancer detection. 2018 International Conference on Information and Communications Technology (ICOIACT). IEEE; 2018.
12. El-Hajj-Chehade W, Kader RA, Kassem R, El-Zaart A. Image segmentation for fingerprint recognition. 2018 IEEE Applied Signal Processing Conference (ASPCON). IEEE; 2018: 314-19.
13. Satpute N, Naseem R, Pelanis E, Gómez-Luna J, Cheikh FA, Elle OJ, Olivares J. GPU acceleration of liver enhancement for tumor segmentation. Comput Meth Progr Biomed. 2020; 184: 105285.
14. Roels J, De Vylder J, Saeys Y, Goossens B, Philips W. Decreasing time consumption of microscopy image segmentation through parallel processing on the GPU. International Conference on Advanced Concepts for Intelligent Vision Systems. Springer: Cham; 2016: 147-59.
15. Liu B, He S, He D, Zhang Y, Guizani M. A spark-based parallel fuzzy $ c $-Means segmentation algorithm for agricultural image big data. IEEE Access. 2019; 7: 42169-80.
16. Jia H, Ma J, Song W. Multilevel thresholding segmentation for color image using modified moth-flame optimization. IEEE Access. 2019; 7: 44097-134.
17. Kau T, Singh Saini B, Gupta S. Optimization techniques for the multilevel thresholding of the medical images. Medical Data Security for Bioengineers. IGI Global. 2019: 166-84.
18. Al-Ajlan A, El-Zaart A. Image segmentation using minimum cross-entropy thresholding. 2009 IEEE International Conference on Systems, Man and Cybernetics. 2009: 1776-78.
19. Song W, Zheng N, Zheng R, Zhao XB, Wang A. Digital image semantic segmentation algorithms: a survey. JIHMSP. 2019; 10(1): 196-211.
20. Al-Osaimi G, Ali El-Zaart A. Minimum cross entropy thresholding for SAR images. 3rd International Conference on Information and Communication Technologies: From Theory to Applications; April 2008; 2008: 1-6.
21. AlSaeed DH, Bouridane A, ElZaart A, Sammouda R. Two modified Otsu image segmentation methods based on Lognormal and Gamma distribution models. 2012 International Conference on Information Technology and e-Services. IEEE; 2012: 1-5.
22. Zhu H, Meng F, Cai J, Lu S. Beyond pixels: a comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. J Vis Commun Image Represent. 2016; 34: 12-27.
23. Beneš M, Zitova B. Performance evaluation of image segmentation algorithms on microscopic image data. J. Microsc. 2015; 257(1): 65-85.
24. Azmat S, Wills L, Wills S. Parallelizing multimodal background modeling on a low-power integrated GPU. J Signal Process Syst. 2017; 88(1): 43-53.
25. Hwang K, Jotwani N. Advanced computer architecture, 3e. McGraw-Hill Education; 2016.
26. Weiss A, Elserbeni A, Demir V, Hadi M. Accelerating the FDTD algorithm on CPUs with MATLAB's parallel computing toolbox. 2019 International Applied Computational Electromagnetics Society Symposium (ACES): IEEE; 2019: 1-2.
27. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, Helba B, Kalloo A, Liopyris K, Marchetti M, Kittler H. Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv: 1902.03368: 2019.
28. Mendonca TF, Celebi ME, Mendonca T, Marques JS. PH2: a public database for the analysis of dermoscopic images. Dermoscopy image analysis. CRC Press, 2015.
29. Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms. Department of Electrical Engineering and Computer Sciences University of California, Berkeley.

30. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft coco: common objects in context. European Conference on Computer Vision. Cham: Springer; 2014: 740-55.

31. García S, Molina D, Lozano M, Herrera F. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. J Heurist. 2009; 15(6): 617.

Corresponding author
Ali El-Zaart can be contacted at: elzaart@bau.edu.lb