Skin Cancer Segmentation with Entropy PAL MCET using Gaussian Distribution

Nancy A. Zreika
Math and Computer Science Department
Beirut Arab University
Beirut, Lebanon
nancy_zreika@hotmail.com

Ali El Zaart
Math and Computer Science Department
Beirut Arab University
Beirut, Lebanon
elzaart@bau.edu.lb

Abdallah El Chakik
Math and Computer Science Department
Beirut Arab University
Tripoli, Lebanon
a.alshakik@bau.edu.lb

Toufic El Arwadi
Math and Computer Science Department
Beirut Arab University
Beirut, Lebanon
t.elarwadi@bau.edu.lb

***Abstract*—Skin cancer is the most common cancer diagnosis and it is the most preventable cancer. Diagnosis of skin cancer would be improved if an accurate skin image segmentation is available. The process of image segmentation is a fundamental step in many applications of image processing, yet current methods and techniques for image segmentation necessitate particular domain knowledge to define well the region of the cancer. To estimate an optimal threshold for skin cancer images, thresholding is used as the principal approach of segmentation. We propose a new method for skin cancer segmentation using a Minimum Cross Entropy Thresholding (MCET) method. We applied this method on bimodal skin cancer images and obtained promising experimental results. The resulting segmented skin cancer images yielded better estimation of the optimal threshold than does the same MCET method with Poisson distribution.**

Keywords— Image Segmentation, Threshold, Minimum Cross Entropy, Skin cancer, Melanoma, Gaussian distribution, Computer-Aided Diagnosis, Bi modal technique.

I. INTRODUCTION

Skin cancer may have many causes and ranging degrees of malignancy. Skin cancer (Melanoma) is one of the deadliest cancers, but when diagnosed early, it can be cured. There are two major types of skin cancer, name malignant melanoma and non-melanoma. Melanoma is more dangerous and can be fatal if not treated. A non-melanoma tumor is a benign tumor and it is unable to spread throughout the body.

Skin cancer recognition process is initiated by image segmentation. This is refined by subtracting image background from the skin cancer (foreground) [1]. An algorithm is used to process image segmentation, which is based on one of two main properties of intensity values, discontinuity and similarity. In one of the analysis, partitioning of an image is based on sudden changes in intensity. Alternatively, the image is divided into regions of similarity based on some criteria [2]. Despite the above processes, there is not a single image segmentation algorithm that reliably gives results for every image [3]. Hence, best image segmentation results are achieved by considering the type of image application.

The existence of several general-purpose methods and techniques for image segmentation, poses a problem necessitating combining these techniques with a particular domain knowledge to overcome this issue [4].

We propose a novel method, an enhancement to MCET, that uses the Minimum Cross Entropy Thresholding, based on Gaussian distribution for estimating an optimal threshold for skin cancer images. By comparing performance of several skin cancer images with the original MCET we obtained better estimation of the optimal threshold than does the method of MCET Poisson [5].

This article introduces the segmentation process in general and the thresholding technique in particular, including a brief review of cross entropy thresholding and Gaussian distribution. The section that follows it describes our proposed technique followed by experiments and results. Finally, we state our conclusion.

II. RELATED WORK

A variety of methods have been used for image processing, in each method a different technique was applied to obtain a good image segmentation and an optimal threshold [17,18,19,20,21]. Among these methods, some are unsupervised methods (fully automated) while others are supervised methods (semi-automatic) [6]. The fully automated methods use grouping of pixels based on some criteria for image processing. On the other hand, supervised methods require input from the user to guide the segmentation process. To estimate an optimal threshold for image segmentation, entropy based thresholding is one broad type of thresholding techniques used to separate objects from the background [7, 8]. To distinguish a dark object image f (x, y) from the light background, a threshold T would lie in the valley of a bimodal histogram in which the foreground object can be extracted by comparing pixels with the threshold T. If f (x, y) > T, then the pixel (x, y) is belonging to the object class, otherwise, it belongs to the background class [2].

A. What is Cross-Entropy Thresholding?

Kullback[14] proposed to measure the distance between two distributions by using Cross Entropy. He assumed that the minimization of the cross entropy forces the total intensity in the threshoded image to be identical to that of the original

image in both the object and the background regions. The more similar the distribution of two variables, the smaller cross entropy is, and vice versa. This cross entropy is defined as [9]:

$$D(p,q)= \sum_{i=1}^{L} p(i).log\frac{p(i)}{q(i)} \quad (1)$$

where, p and q are two different sources of information, p(i) and q(i) are the two statistical distributions of the two sources while L refers to the number of information values.

Brink and Pendock in [10] added a new definition of the information theoretic distance in this way:

$$D(p,q)= \sum_{i=1}^{L} p(i).log\frac{p(i)}{q(i)} + \sum_{i=1}^{L} q(i).log\frac{q(i)}{p(i)} \quad (2)$$

To find an optimal threshold value, Li and Lee in [9] proposed a sequential method using MCET based on Gaussian distribution. In [15], Lee and Tam developed an iterative algorithm using MCET based on Gaussian distribution. In [16], Brink and Pendock improve Li and Lee's work by developing a sequential method using Gaussian distribution.

B. Cross-Entropy Method of Pal [5]

Pal wrote the formula of total cross entropy of the segmented image as it follows:

$$D(s) = D_O(s) + D_B(s)$$

$$= \sum_{i=1}^{s} p_i^o \log\left(\frac{p_i^o}{q_i^o}\right) + \sum_{i=1}^{s} q_i^o \log\left(\frac{q_i^o}{p_i^o}\right) + \sum_{i=s+1}^{L} p_i^B \log\left(\frac{p_i^B}{q_i^B}\right) + \sum_{i=s+1}^{L} q_i^B \log\left(\frac{q_i^B}{p_i^B}\right) \quad (3)$$

Where $p_i^o = \frac{h_i}{p_s}$, *i=1, 2, ….., s;* (4)

$$p_s = \sum_{i=1}^{s} h_i; \quad (5)$$

$$p_i^B = \frac{h_i}{MN - p_s}, \quad i=s+1, s+2, \ldots, L; \quad (6)$$

p_i^o is the pixel frequency of object in the original image.
p_i^B is the pixel frequency of background in the original image.
q_i^o is the pixel frequency of object in the segmented image.
q_i^B is the pixel frequency of background in the segmented image.

Pal used Poisson distribution model to estimate Q_O and Q_B. He assumed that the gray value in the object region follows the Poisson distribution.

III. THE PROPOSED MCET ALGORITHM USING GAUSSIAN DISTRIBUTION

Pal assumed that that image histograms are modeled by a mixture of Poisson distributions. He used Poisson distribution in order to estimate the threshold values; however, the data in image does not always follow Poisson distribution. In this study, our goal is to propose an improvement of Pal method using Gaussian distribution instead of Poisson distribution. In the proposed method, the histogram (h) of a skin cancer image is assumed to be a mixture of several Gaussian distributions, since the modes of the histogram are symmetric. This property of Gaussian distribution model will lead to more accurate estimations.

A. Gaussian Distribution

The number of pixels in a skin cancer image at each intensity value form the histogram graph. Two or more distinct modes derive from the grey level histogram. A bimodal histogram consists of two dominant modes whereas if three or more dominant modes define the image histogram, this is called multi-modal histogram.
In probability theory, Gaussian distribution is a continuous probability distribution and has a bell-shaped probability density function, known as the Gaussian function [11]:

$$f(x,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (7)$$

Where x is the pixel's intensity level; σ is the standard deviation and μ is the mean.
A Gaussian distribution models the data in an image $I_t(x,y)$. Therefore, the means $\mu_1(t)$ and $\mu_2(t)$ can be estimated as:

$$\mu 1(t) = \frac{\sum_{i=1}^{t} i.h(i)}{\sum_{i=1}^{t} h(i)} \quad (8)$$

$$\mu 2(t) = \frac{\sum_{i=t}^{L} i.h(i)}{\sum_{i=t}^{L} h(i)} \quad (9)$$

Where $\mu_1(t)$ represents the mean value of the background region in the segmented image and $\mu_2(t)$ represents the mean value of the object region.

B. Bimodal Thresholding:

To segment an image into background and object, a bimodal thresholding is used. The bimodal distribution always represents the grey level histogram of the image where one threshold value is enough to segment this image. To determine the optimal threshold t^*, we have used Minimum Cross Entropy Thresholding technique (MCET). MCET minimizes the cross entropy between original image I and resulting thresholded image I_t.

$$I_t(x,y) = \begin{cases} \mu 1(t), & I(x,y) < t \\ \mu 2(t, L+1), & I(x,y) \geq t \end{cases} \quad (10)$$

The bimodal algorithm is described as following:

Input: image *I (x, y)*
Compute histogram *h(i)* of the input image, where

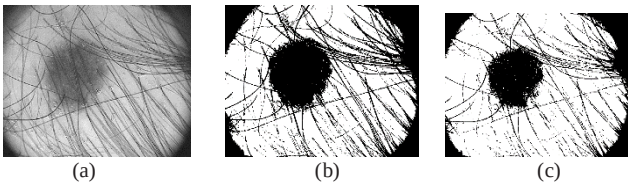
i=0…255.
Compute μ_O using Gaussian distribution.
Compute μ_B using Gaussian distribution.
Compute p_i^o, *i=1, 2, …..., s* using equation(4).
Compute p_i^B, *i=s+1, s+2, …..., L* using equation(6).
Compute q_i^o , q_i^B using equation (7).
Compute D(S) using equation (3).
if (min > D(S))
Begin
min = D(S)
threshold = S
end
end
end
Output: optimal threshold: S

IV. EXPERIMENTAL RESULTS

300 real skin images with different qualities and gray levels were subjected to the proposed Gaussian distribution method. The range of the search space was set from 0 to 255 [start, end]. Further, the skin cancer images were subjected to a bimodal thresholding method, and our Gaussian distribution method was compared against the original MCET method using Entropy-Pal with Poisson distribution.

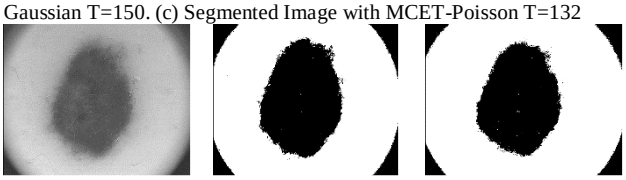
Samples of skin cancer images are shown in figures-(1...9). Image uniformity and inter-region contrast measures are employed to evaluate the performance of our proposed method [12, 13]. Evaluation of 30 test images, the threshold value for all versions of the method and the value of the evaluation measure, are shown in Tables-(1,2). The two methods are compared and ranked for each image based on their performance. As shown, our proposed method estimates better the optimal threshold value, in that 87% of tested images have better segmentation results. Further, with the proposed method 87% of the cases inter-region contrast is higher with MCET-Gaussian thresholding while for uniformity measure, MCET-Poisson was ranked 1st in 83% of the images (See Table-3).

.



(a) (b) (c)

Fig. 1. (a)Original Image (SC002.bmp). (b) Segmented Image with MCET-Gaussian T=150. (c) Segmented Image with MCET-Poisson T=132



(a) (b) (c)

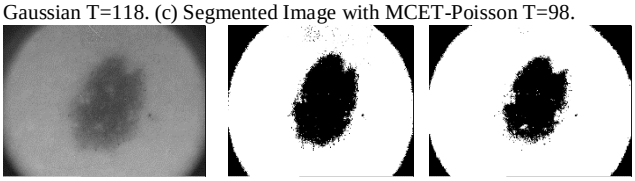
Fig. 2. (a)Original Image (SC007.bmp). (b) Segmented Image with MCET-Gaussian T=144. (c) Segmented Image with MCET-Poisson T=127.

(a) (b) (c)

Fig. 3. (a)Original Image (SC008.bmp). (b) Segmented Image with MCET-Gaussian T=128. (c) Segmented Image with MCET-Poisson T=123.

(a) (b) (c)

Fig. 4. (a)Original Image (SC0013.bmp). (b) Segmented Image with MCET-Gaussian T=118. (c) Segmented Image with MCET-Poisson T=98.



(a) (b) (c)

Fig. 5. (a)Original Image (SC0015.bmp). (b) Segmented Image with MCET-Gaussian T=116. (c) Segmented Image with MCET-Poisson T=104

(a) (b) (c)

Fig. 6. (a)Original Image (SC0017.bmp). (b) Segmented Image with MCET-Gaussian T=130. (c) Segmented Image with MCET-Poisson T=142.

(a) (b) (c)

Fig. 7. (a)Original Image (SC0022.bmp). (b) Segmented Image with MCET-Gaussian T=151. (c) Segmented Image with MCET-Poisson T=132.

(a) (b) (c)

Fig. 8. (a)Original Image (SC0023.bmp). (b) Segmented Image with MCET-Gaussian T=104. (c) Segmented Image with MCET-Poisson T=106.

TABLE 1. Comparison of Performance Measures between MCET-Gaussian and MCET-Poisson Methods (Images 1-15)

Image	Method	Threshold	Uniformity	Rank	Inter-Region Contrast (RC)	Rank	Avg: $\frac{NU+RC}{2}$	Rank
SC001	MCET-Gaussian	132	0.999903	2	0.679273	1	0.839588	1
	MCET-Poisson	114	0.99991	1	0.65535	2	0.82763	2
SC002	MCET-Gaussian	150	0.99978	2	0.718909	1	0.859344	1
	MCET-Poisson	132	0.999798	1	0.679859	2	0.839828	2
SC003	MCET-Gaussian	160	0.999891	2	0.639789	1	0.81984	1
	MCET-Poisson	128	0.999903	1	0.592551	2	0.796227	2
SC004	MCET-Gaussian	121	0.999982	2	0.780515	1	0.890249	1
	MCET-Poisson	104	0.999984	1	0.737691	2	0.868837	2
SC005	MCET-Gaussian	135	0.999729	2	0.654338	1	0.827033	1
	MCET-Poisson	126	0.999735	1	0.647243	2	0.823489	2
SC006	MCET-Gaussian	135	0.999729	2	0.843202	1	0.921597	1
	MCET-Poisson	128	0.999992	1	0.833107	2	0.91655	2
SC007	MCET-Gaussian	144	0.999875	2	0.675524	1	0.837699	1
	MCET-Poisson	127	0.999884	1	0.656695	2	0.828289	2
SC008	MCET-Gaussian	128	0.999741	2	0.643491	1	0.821616	1
	MCET-Poisson	123	0.999745	1	0.639296	2	0.819521	2
SC009	MCET-Gaussian	157	0.999897	2	0.713093	1	0.856495	1
	MCET-Poisson	138	0.999905	1	0.682198	2	0.841052	2
SC010	MCET-Gaussian	166	0.999888	2	0.682248	1	0.841068	1
	MCET-Poisson	138	0.999902	1	0.634014	2	0.816958	2
SC011	MCET-Gaussian	166	0.999942	2	0.651115	1	0.825529	1
	MCET-Poisson	130	0.999951	1	0.593516	2	0.796734	2
SC012	MCET-Gaussian	154	0.99996	2	0.68883	1	0.844395	1
	MCET-Poisson	127	0.999965	1	0.641344	2	0.820655	2
SC013	MCET-Gaussian	118	0.999987	2	0.659826	1	0.829906	1
	MCET-Poisson	98	0.999988	1	0.6258	2	0.812894	2
SC014	MCET-Gaussian	157	0.99994	2	0.706893	1	0.853417	1
	MCET-Poisson	126	0.999951	1	0.625195	2	0.812573	2
SC015	MCET-Gaussian	116	0.99996	2	0.711518	1	0.855739	1
	MCET-Poisson	104	0.999963	1	0.679225	2	0.839594	2

TABLE 2: Comparison of Performance Measures between MCET-Gaussian and MCET-Poisson Methods (Images 16-30)

Image	Method	Threshold	Uniformity	Rank	Inter-Region contrast (RC)	Rank	Avg: $\frac{NU+RC}{2}$	Rank
SC016	MCET-Gaussian	167	0.999859	2	0.751409	1	0.875634	1
	MCET-Poisson	150	0.99987	1	0.727213	2	0.863541	2
SC017	MCET-Gaussian	130	0.999912	1	0.663484	1	0.831698	1
	MCET-Poisson	142	0.999909	2	0.663412	2	0.831661	2
SC018	MCET-Gaussian	118	0.999911	2	0.626965	1	0.813438	1
	MCET-Poisson	107	0.999915	1	0.607494	2	0.803705	2
SC019	MCET-Gaussian	166	0.999922	2	0.666553	1	0.833237	1
	MCET-Poisson	126	0.999936	1	0.585614	2	0.792775	2
SC020	MCET-Gaussian	144	0.999956	2	0.654285	1	0.827121	1
	MCET-Poisson	107	0.999964	1	0.586185	2	0.793074	2
SC021	MCET-Gaussian	114	0.999968	1	0.762454	2	0.881211	2
	MCET-Poisson	119	0.999967	2	0.763862	1	0.881915	1
SC022	MCET-Gaussian	151	0.999818	2	0.656874	1	0.828346	1
	MCET-Poisson	132	0.99983	1	0.633693	2	0.816762	2
SC023	MCET-Gaussian	104	0.999878	1	0.606068	2	0.802973	2
	MCET-Poisson	106	0.999877	2	0.608436	1	0.804157	1
SC024	MCET-Gaussian	90	0.999907	1	0.53396	2	0.766933	2
	MCET-Poisson	99	0.999905	2	0.537698	1	0.768801	1
SC025	MCET-Gaussian	119	0.999933	2	0.704319	1	0.852126	1
	MCET-Poisson	105	0.999938	1	0.678782	2	0.83936	2
SC026	MCET-Gaussian	160	0.999953	2	0.771452	1	0.885702	1
	MCET-Poisson	132	0.999959	1	0.704486	2	0.852222	2
SC027	MCET-Gaussian	129	0.999935	2	0.659081	1	0.829508	1
	MCET-Poisson	112	0.99994	1	0.634401	2	0.81717	2
SC028	MCET-Gaussian	101	0.999909	2	0.619878	1	0.809893	1
	MCET-Poisson	94	0.999911	1	0.607427	2	0.803669	2
SC029	MCET-Gaussian	100	0.999909	1	0.579566	2	0.789738	2
	MCET-Poisson	109	0.999908	2	0.585029	1	0.792469	1
SC030	MCET-Gaussian	125	0.999946	2	0.648778	1	0.824362	1
	MCET-Poisson	106	0.999951	1	0.603372	2	0.801661	2

TABLE 3. Evaluation of Performance Measures for the 2 Methods (All Images). R/M = Rank/Method. MGau = MCET-Gaussian, MPoi = MCET-Poisson.

Performance Metric	R/M	MGau	MPoi
UN: Uniformity	1st	17%	83%
	2nd	83%	17%
RC:Inter-region Contrast	1st	87%	13%
	2nd	13%	87%
Average: $\frac{NU+RC}{2}$	1st	87%	13%
	2nd	13%	87%

V. CONCLUSION

Many image processing applications rely on image segmentation as an essential tool. A critical issue is to obtain accurate segmentation of skin cancer images for a better and more efficient segmentation that ultimately results in better diagnosis. In this work, we presented a new bimodal gray level thresholding algorithm based on minimum cross entropy thresholding with Pal-Gaussian and Poisson distribution. We conclude based on the evaluation of the resulting images, that the proposed method yields better images.

VI. REFERENCES

[1] R.B. Dubeya, M. Hanmandlub, and S.K. Guptac, "A comparison of two methods for the segmentation of masses in the digital Mammograms," Computerized Medical Imaging and Graphics, vol34, pp. 185–191, 2010.

[2] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing.*, Prentice Hall, 2nd edition, 2002.

[3] D. L. Pham, C. Xu, and J. L. Prince, "A survey of current methods inmedical image segmentation," *In Annual Review of Biomedical Engineering*, pp. 318–338, 2000.

[4] Hui-Lan Luo, Wei-Wang, and Jing-LI, "A new method for image segmentation based on integration technique," in *MVHI*, Kaifeng, China, 2010, pp. 342–345.

[5] Pal, N., "On Minimum cross entropy thresholding," *Pattern Recognition Letter*, vol. 29 no. 4, pp. 575–580, 1996.

[6] Ali El Zaart, Djemel Ziou, Shengrui Wang, and Qingshan Jiang, "Segmentation of sar images.," *Pattern Recognition*, pp. 713–724, 2002.

[7] Juyong Zhang, Jianmin Zheng, and Jianfei Cai, "A diffusion approach to seeded image segmentation," in *CVPR*, San Francisco, CA, 2010, pp. 2125 – 2132

[8] Liang Chen, Lei Guo, and Ning Yang Yaqin Du, "Multi-level image thresholding. based on histogram voting.," in *CISP*, Tianjin, 2009.

[9] C. Li and C. Lee, "Minimum cross entropy thresholding," *Pattern Recognition*, vol. 26, no. 4, pp. 617–625, 1993.

[10] Mehmet Sezgin and Bulent Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging.*, pp. 146–168, 2004.

[11] G.McPherson,"Statistics in scientific investigation: its basis, application and interpretation", Springer-Verlag,1990.

[12] Liane C. Ramaca and Pramod K. Varshney, "Image thresholding based on ali-silvey distance measures," *Pattern Recognetion*, pp. 1161–1174, 1997.

[13] C.-I. Chang et al., "Survey and comparative analysis of entropy and relative entropy thresholding techniques," in *IEEE Proceedings - Vision, Image and Signal Processing*, Dec 2006.

[14] S. Kullback, Information Theory and Statistics. Wiley, New York (1959).

[15] C. Li, and P. Tam, "An Iterative Algorithm for Minimum Cross Entropy Thresholding," Pattern Recognition Letter - Elsevier, vol.19, no. 8, 1998, pp. 771-776.

[16] A. Brink, and N. Pendock, "Minimum Cross-entropy Threshold Selection," Pattern Recognition Letter - Elsevier, vol. 29, no. 1, 1996, pp. 179-188.

[17] D. AlSaeed, A. Bouridane, A. El-Zaart, A. "A novel fast Otsu digital image segmentation method", International Arab Journal of Information Technology, Volume 13, Issue 4, pp. 427-344, 2016

[18] A. El-Zaart, A. Ghosn, "SAR images thresholding for oil spill detection", 2013 Saudi International Electronics, Communications and Photonics Conference, Fira, Greece, 27-30 April 2013.

[19] A. El-Zaart, "Synthetic Aperture Radar Images Segmentation Using Minimum Cross entropy with Gamma Distribution", International Journal of Signal and Image Processing. Volume 6, Issue 4, pp.19-31, 2015.

[20] M. Al-Bayati, A. El-Zaart, "Mammogram images thresholding for breast cancer detection using different thresholding methods", Advances in Breast Cancer Research, Volume 2, Issue 3, pp.72-77, 2013.

[21] A. el sayed, A. el chakik, H. Alabboud, A.Yassin, "3D face detection based on salient features extraction and skin colour detection using data mining", Imaging science journal, Volume 65, Issue 7, pp.393-408, 2017.