Maximum Likelihood Thresholding Algorithm Based on Four-Parameter Gamma Distributions

Peter De-Ford, Geovanni Martinez
Image Processing and Computer Vision Research Laboratory (IPCV-LAB)
Escuela de Ingeniería Eléctrica, Universidad de Costa Rica
Apartado Postal 11501-2060 UCR, San José, Costa Rica
gmartin@eie.ucr.ac.cr

***Abstract*—In this contribution, we present a segmentation algorithm based on thresholding to subdivide an intensity image in the regions of object and background. The optimal threshold is found by maximizing a likelihood function derived from a novel intensity probability density function model, which consists of the sum of two weighted four-parameter gamma distributions, as a more flexible alternative to currently used models consisting of the sum of two weighted two-parameter Gaussian distributions. According to our experiments with 132 images, the proposed algorithm is in average slightly better than the best found in the scientific literature, performing particularly good in low contrast images. The additional parameters and complexity of its likelihood function resulted in an increase of the processing time by a factor of 3, from 0.003 sec/image to 0.009 sec/image.**

I. Introduction

In the field of image processing, algorithms are commonly needed to subdivide an intensity image (grayscale image) that shows an object in front of a background in the regions of object and background. These algorithms are known in the literature as image segmentation algorithms. An image segmentation algorithm classifies each pixel of the intensity image in one of two possible classes: the class of the pixels that conform the background region (class 0) or the class of the pixels that conform the object region (class 1), or vice versa. Among many segmentation algorithms found in the literature [11], [20], [22], [29]–[32], some of the most popular segmentation algorithms are those based on thresholding, which make the classification by comparing the intensity values of the pixels with a reference intensity value called threshold. If the intensity value of the pixel is less or equal to the threshold, the pixel is classified as belonging to the background region (class 0), otherwise it is classified as belonging to the object region (class 1), or vice versa. The algorithm output is a binary image, usually known as segmented image, with the same dimensions as the original intensity image. This image has the value of 0 (black) in the pixels classified as belonging to the background region and the value of 255 (white) in the pixels classified as belonging to the object region, or vice versa.

We are particularly interested in those segmentation algorithms based on thresholding, which automatically obtain the optimal threshold by maximizing a likelihood function [20], [22]–[29]. In these algorithms, the optimal threshold th_{op} is the one that maximizes a likelihood function which consists on the natural logarithm of the conditional probability $p\left(I|th\right)$ of image I given a threshold th, this is:

$$\ln\left(p\left(I|th_{op}\right)\right) = max_{th}\left[\ln\left(p\left(I|th\right)\right)\right], \forall\ th \in 0, 1, ..., 255 \quad (1)$$

where $p\left(I|th\right)$ is obtained by assuming that a memory-less information source generates each of the N intensity values i_n, $n = 0, 1, ..., n, ..., N-1$, of the image I by choosing a statistically independent number from a discrete alphabet of 256 possible intensity values $\{0, 1, .., i, .., 255\}$ according to the following discrete probability density function described by the parameter set P:

$$\begin{aligned} PDF^S(i; P : \{th, w_0, w_1, S_0, S_1\}) = PDF^S(i; P) = \\ ([u(0) - u(th)] * w_0 * PDF_0^S(i; S_0)) + \\ ([u(th+1) - u(255)] * w_1 * PDF_1^S(i; S_1)) \end{aligned} \quad (2)$$

Note that, the first part of (2) is the probability density function $PDF_0^S(i; S_0)$ of the intensity values of the pixels of class 0, which is described by parameter set S_0, weighted by w_0 and rectangular windowed in the domain $[0, th]$ by the unit step functions $u(0)$ and $u(th)$. The second part is a probability density function $PDF_1^S(i; S_1)$ of the intensity values of the pixels of class 1, which is described by parameter set S_1, weighted by w_1 and rectangular windowed in the domain $[th+1, 255]$ by the unit step functions $u(th+1)$ and $u(255)$, where $w_0 + w_1 = 1$. Then, using the PDF^S of (2), the probability of the image given the threshold $p\left(I|th\right)$ can be calculated as the product of the probabilities of occurrence of each of the N image intensity values assuming that these are statistically independent as follows:

$$p\left(I|th\right) = \prod_{n=1}^{N} PDF^S(i_n; P) \quad (3)$$

The set of parameters $P : \{th, w_0, w_1, S_0, S_1\}$ is usually unknown. However, for a given threshold th, the other parameters $\{w_0, w_1, S_0, S_1\}$ can be estimated using the PDF^I (intensity probability density function of the image) derived from the histogram $h(i)$ of the intensity image I as $PDF^I(i) = h(i)/N$, where $h(i)$ gives the relative frequency (number of occurrences) of any of the 256 possible intensity values i in the image. It is very important to mention that the PDF^S associated with the optimal threshold th_{op}, which is called $PDF^{S,op}$, represents the best approximation of the PDF^I derived from the histogram of the image.

The most popular thresholding algorithms based on maximum likelihood found in the scientific literature are those

978-1-4799-6230-3/14/$31.00

proposed by Kittler *et al.* in [20] (see also [16], [17]), Kurita *et al.* in [22] and Otsu *et al.* in [29]. The three of them assume that the probability density functions $PDF_0^S(i;S_0)$ and $PDF_1^S(i;S_1)$ of (2) are Gaussian distributions of two parameters each (mean and variance). Kittler assumes also that the weights, means and variances are different. However, Kurita constraints the variances to be equal and Otsu constraints not only the variances to be equal, but the weights as well.

In the last exhaustive survey on thresholding algorithms [32], 40 renamed thresholding algorithms were applied to text document images and Non Destructing Testing (NDT) images (ultrasonic images, eddy current images, thermal images, x-ray computed topography, endoscopic images, laser scanning confocal microscopy, etc.). Among the 40 algorithms, Kittler's algorithm was undisputed number one in both application areas. Despite Kittler's algorithm supremacy over the rest, it had medium and low performance in some tested images. This happened because the two-parameter Gaussian distributions were not capable to describe some complex shapes of probability density functions found in those images. Thus, a more flexible PDF^S is required in order to increase the performance of the maximum likelihood thresholding algorithms. To accomplish this, we propose here a new maximum likelihood thresholding algorithm based on a more flexible PDF^S to model the probability density function of the intensity values PDF^I of image I. Instead of assuming that PDF_0^S and PDF_1^S are two Gaussian distributions of two parameters each (mean and variance), we assume that they are two gamma distributions of four parameters each (shape, scale, location and reflection). The asymmetry and quantity of parameters of these gamma distributions make the PDF^S of (2) more flexible for modeling complex shapes of probability density functions. This increase in flexibility will make the proposed algorithm perform better than the other maximum likelihood thresholding algorithms. Related approaches can be found in the scientific literature [1]–[10], but using the classical two-parameter gamma distribution, which as the two-parameter Gaussian distribution lacks of sufficient flexibility to model complex shapes of probability density functions.

This paper is organized as follows. In Section II, the proposed algorithm is described. In Section III, the experimental results are given. Finally, in Section IV, the summary and conclusions are found.

II. Proposed algorithm

In this work, we propose to use in (2) the following two four-parameter gamma probability density functions:

$$PDF_0^S(i;S_0) = \frac{1}{\beta_0 * \Gamma(\alpha_0)} \left(\frac{\lambda_0(i-\varphi_0)}{\beta_0}\right)^{\alpha_0-1} e^{\frac{-\lambda_0(i-\varphi_0)}{\beta_0}} \quad (4)$$

$$PDF_1^S(i;S_1) = \frac{1}{\beta_1 * \Gamma(\alpha_1)} \left(\frac{\lambda_1(i-\varphi_1)}{\beta_1}\right)^{\alpha_1-1} e^{\frac{-\lambda_1(i-\varphi_1)}{\beta_1}} \quad (5)$$

where the parameter α_j (real positive), $j = \{0,1\}$, describes the shape, β_j (real positive) describes the scale, φ_j (real positive or negative) describes the location, and λ_j (either 1 or -1) describes whether the PDF_j^S is non reflected or reflected over the y axis, respectively. The set of parameters that describe this model is composed by 11 parameters: $P : \{\, th, w_0, w_1, S_0 : \{\, \lambda_0, \varphi_0, \beta_0, \alpha_0 \,\}, S_1 : \{\, \lambda_1, \varphi_1, \beta_1, \alpha_1 \,\} \,\}$.

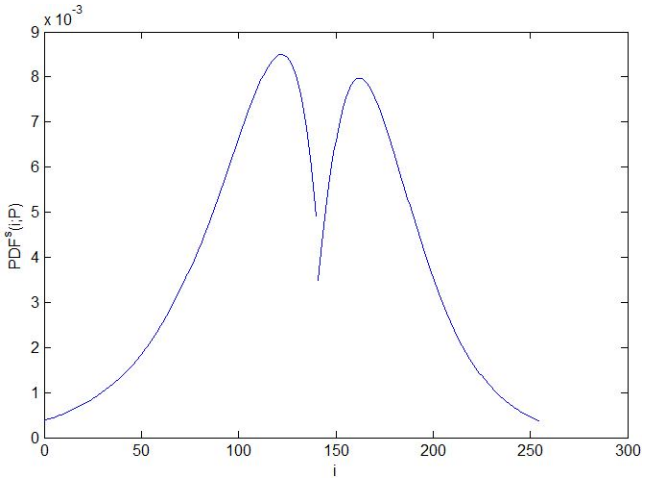

Fig. 1. Example of the proposed $PDF^S(i;P)$ with parameters $P : \{\, th = 140, w_0 = 0.55, w_1 = 0.45, S_0 : \{\, \lambda_0 = -1, \varphi_0 = 147, \beta_0 = 25, \alpha_0 = 2 \,\}, S_1 : \{\, \lambda_1 = 1, \varphi_1 = 130, \beta_1 = 16, \alpha_1 = 3 \,\} \,\}$.

Fig. 1 depicts an example of the proposed $PDF^S(i;P)$.

Inserting (4) and (5) into (2) the natural logarithm of (3) becomes the following likelihood function:

$$\begin{aligned} \ln(p(I|th)) = {} & [c_{0,th} N \ln(w_0) + c_{1,th} N \ln(w_1)] + \\ & [\sum_{i=0}^{th} (h(i)[(\alpha_0 - 1)\ln(\lambda_0(i-\varphi_0)) - \frac{\lambda_0(i-\varphi_0)}{\beta_0}]) \\ & - c_{0,th} N \ln(\beta_0^{\alpha_0} * \Gamma(\alpha_0))] + \\ & [\sum_{i=th+1}^{255} h(i)[(\alpha_1 - 1)\ln(\lambda_1(i-\varphi_1)) - \frac{\lambda_1(i-\varphi_1)}{\beta_1}] \\ & - c_{1,th} N \ln(\beta_1^{\alpha_1} * \Gamma(\alpha_1))] \end{aligned} \quad (6)$$

where $c_{0,th} = \sum_{i=0}^{th} h(i)/N$, $c_{1,th} = \sum_{i=th+1}^{255} h(i)/N$, $c_{0,th} + c_{1,th} = 1$ and $h(i)$ is the histogram of the intensity image I. For the maximization of (6), we apply a downhill simplex based method [14], [15].

It is important to mention that (6) represents a general likelihood function from which Otsu's [29], Kurita's [22] and Kittler's [20] likelihood functions (as described in [22]) can be derived as special cases. For instance, based on [12], [13], if the parameters sets $S_j : \{\lambda_j, \varphi_j, \beta_j, \alpha_j\}$, $j = \{0,1\}$, are written as follows

$$S_j : \{\, \lambda_j = 1, \varphi_j \to -\infty, \beta_j^2 = \frac{\sigma_j^2}{\mu_j - \varphi_j}, \alpha_j = \frac{\sigma_j^2}{\beta_j^2} \,\} \quad (7)$$

and the Stirling's Formula [18] for approximating a gamma function for large shape values α_j is used, then (6) becomes Kittler's likelihood function, where μ_j and σ_j^2, $j = \{0,1\}$, are the means and variances of the PDF_j^S based on Gaussian distributions, respectively. Furthermore, assuming that the variances of the Gaussian distributions have the same values, then (6) becomes Kurita's likelihood function. Finally, if also both w_0 and w_1 are set to 0.5, then (6) becomes Otsu's likelihood function.

III. Experimental results

The proposed algorithm, along with Kittler's [20], Kurita's [22] and Otsu's [29] algorithms, were implemented in the

programming language C under Visual C++ 2008. To assess the performance of the proposed algorithm against Kittler's, Kurita's and Otsu's algorithms, three different experiments on 132 images were carried out. The 132 images, along with their manual segmentations, are available on a data base called *IPCV-Lab Image Data Base*. These images are subdivided into 7 different categories: 12 images whose histograms were synthetically generated using random number generators [14], [21]; 8 images generated by computer graphics; 13 images captured using imaging techniques (thermography, radiography, x-rays and remote sensing); 33 images containing letters; 42 photographs containing objects, people, animals and nature; 4 images of small objects; and 20 low contrast images. The average size of the images was about *600 pixels x 400 pixels*. In the following, the three experiments will be described.

In the first experiment, for each of the 132 images, the F1-scores [19], [32] between each of the segmented images provided by the four algorithms and the corresponding manually segmented image were computed. Table I depicts the average F1-scores for all of the 7 categories and the average F1-scores for the low contrast image category. Since the proposed

TABLE I. F1-SCORE EVALUATION

Image Category	Proposed	Kittler	Kurita	Otsu
Low contrast	0.79	0.63	0.61	0.69
All categories	0.92	0.87	0.89	0.85

algorithm average F1-score for all of the 7 categories was slightly higher than the averages obtained by Kittler's, Kurita's, and Otsu's algorithms, we can conclude that the proposed algorithm is in average slightly more precise than the other three. In addition, since in the low contrast image category the average F1-score of the proposed algorithm is significantly higher than the average F1-scores of the other three algorithms, the proposed algorithm can be considered the best option for segmenting low contrast images. This can be seen in part A of Fig. 2, which shows the segmentation results of a low contrast intensity image depicting a coin. Note that the segmented image obtained by the proposed algorithm (see sub-Fig. 2-A.5) is much better than those obtained by the other algorithms (see sub-Figs. 2-A.6, 2-A.7 and 2-A.8).

In the second experiment, for each of the 132 images, the mean square errors between the optimal models $PDF^{S,op}$ obtained by each of the four algorithms and the intensity probability density function PDF^{I} derived from the image histogram were calculated [24]. Table II depicts the average mean square errors for all of the 7 categories and the average mean squared errors for the low contrast image category (scaled by a factor of 10^7). As it can be seen, the proposed

TABLE II. MEAN SQUARE ERROR EVALUATION (SCALED BY A FACTOR OF 10^7)

Image Category	Proposed	Kittler	Kurita	Otsu
Low contrast	43	51	66	190
All categories	84	113	178	377

algorithm achieves the lowest average mean square error for all of the 7 categories. This is because the proposed gamma distributions used in the PDF^{S} allow better approximations of the PDF^{I} than the ones achieved using Gaussian distributions. An example of this can be appreciated in part B of Fig. 2, which shows the segmentation results of an image depicting a buoy. Note that the optimal model $PDF^{S,op}$ obtained by the proposed algorithm (see sub-Fig. 2-B.9) approximates better the PDF^{I} than those obtained by the other algorithms (see sub-Figs. 2-B.10, 2-B.11 and 2-B.12).

In the third experiment, several people of different ages, genders, careers, etc., subjectively evaluated the quality of the segmentation results obtained with all of the 132 images. Every person was asked to evaluate each segmentation result with a number between 0 and 100, where 0 represented a very bad segmentation result and 100 an excellent segmentation result. As a reference, the manual segmentation of each image (ground truth) was also delivered to the evaluators. Table III depicts the results of this subjective evaluation. From these

TABLE III. SUBJECTIVE EVALUATION

Image Category	Proposed	Kittler	Kurita	Otsu
Low contrast	65	49	42	53
All categories	81	78	78	74

results, we can infer that the proposed algorithm is slightly more precise than the other three, which was also inferred in the F1-score evaluation of Table I.

The segmentation of the 132 images was made in a Windows 7 laptop with an intel core i3 at 4 GHz speed and 4 GB RAM. Table IV shows the average time that each of the four algorithms spent for segmenting one image (in seconds). It can be seen how the proposed algorithm requires in average

TABLE IV. AVERAGE PROCESSING TIME FOR THE 132 IMAGES (IN SECONDS)

Algorithm	Proposed	Kittler	Kurita	Otsu
Time	0.009	0.003	0.003	0.003

3 times the processing time required by Kittler's, Kurita's or Otsu's algorithm for segmenting one image. The increase in processing time is a result of the additional parameters and complexity of the proposed likelihood function. However, since the processing times are very low (in the millisecond scale), this difference is not very significant even for real-time applications.

IV. SUMMARY AND CONCLUSIONS

In this contribution, we presented a maximum likelihood thresholding algorithm, whose likelihood function was derived from a new intensity probability density function model. This new model consists of the sum of two weighted gamma distributions of four parameters each (shape, scale, location and reflection), and represents a more precise alternative to currently used models found in the scientific literature, consisting on the sum of two weighted Gaussian distributions of two parameters each (mean and variance). It is important to mention that the likelihood function obtained from this new model was demonstrated to be a general likelihood function from which the Gaussian based likelihood functions [20], [22], [29] (as described in [22]) can be derived as special cases.

The proposed algorithm, along with Kittler's [20], Kurita's [22] and Otsu's [29] algorithms, were applied to 132 images. The total average F1-score achieved by the proposed algorithm was 5%, 3% and 7% higher when compared to those obtained by the Kittler's, Kurita's and Otsu's algorithm, respectively.

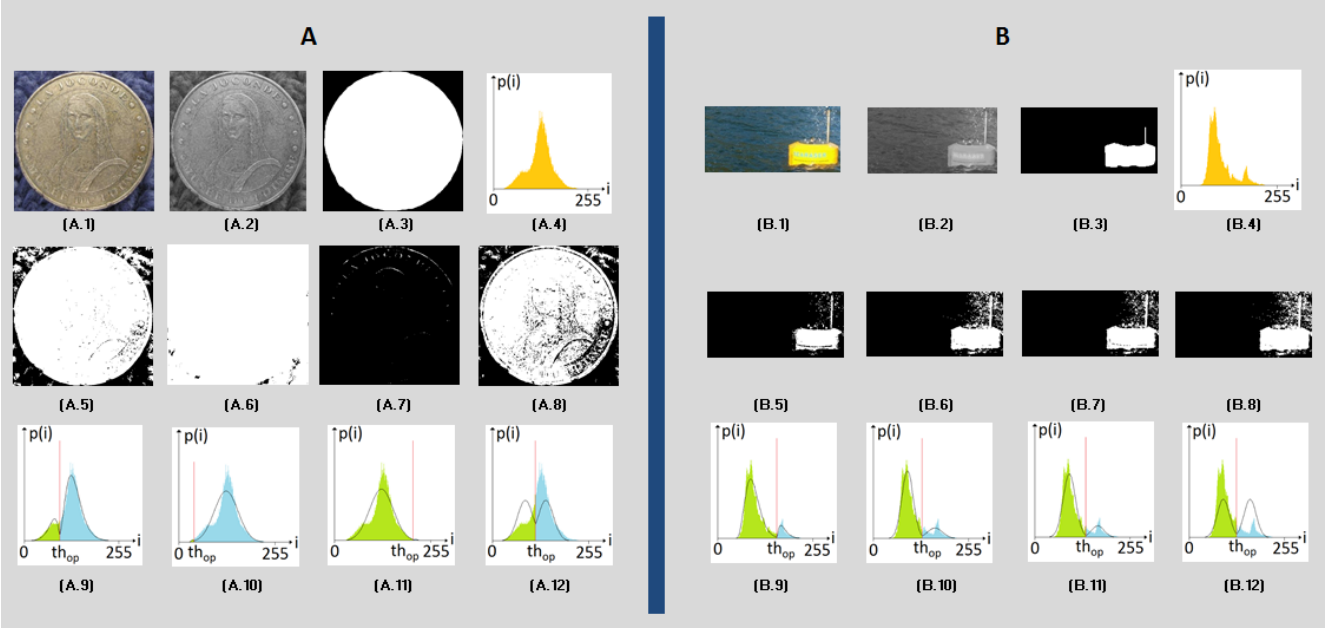

Fig. 2. Parts A and B show the segmentation results for an image depicting a coin and an image depicting a buoy on the sea surface, respectively. (A.1/B.1) original color images; (A.2/B.2) original intensity images; (A.3/B.3) manually segmented images; (A.4/B.4) intensity probability density functions PDF^I obtained from the histogram of the original image; (A.5/B.5), (A.6/B.6), (A.7/B.7) and (A.8/B.8) segmented images obtained by the proposed, Kittler's, Kurita's and Otsu's algorithms, respectively; (A.9/B.9), (A.10/B.10), (A.11/B.11) and (A.12/B.12) optimal models $PDF^{S,op}$ obtained by the proposed, Kittler's, Kurita's and Otsu's algorithms, respectively, superimposed on the PDF^I obtained from the histogram of the original image.

This results allow us to conclude that the proposed algorithm in average delivered slightly more precise segmentation results than those obtained by the other three. This was also confirmed by a subjective evaluation of the segmented images. In addition, since the average F1-score of the proposed algorithm for low contrast images was 16%, 18% and 10% higher than those obtained by Kittler's, Kurita's and Otsu's algorithms, respectively, the proposed algorithm is undoubtedly the best option when segmenting low contrast images.

The approximation quality of the model of the intensity probability density function obtained by each one of the four algorithms was also evaluated. For quality evaluation, the mean square error between the found optimal model and the intensity probability density function computed from the intensity image histogram was compared. In this regard, the approximation quality of the proposed algorithm was 27%, 60% and 78% better than Kittler's, Kurita's and Otsu's algorithms, respectively. This confirms our hypothesis that the new model is capable of describing more complex shapes of intensity probability density functions than the currently used models consisting of the sum of two weighted two-parameter Gaussian distributions, which resulted in a performance increase of the maximum likelihood thresholding approach.

The additional parameters and complexity of the proposed algorithm likelihood function resulted in an increase of the processing time by a factor of 3 with respect to Kittler's, Kurita's or Otsu's algorithms. However, since the processing times are very low, this increase is not very significant even for real-time applications.

REFERENCES

[1] A. El-Zaart and D. Ziou. Statistical modelling of multimodal sar images. *International Journal of Remote Sensing*, 28(10):2277–2294, 2007.

[2] A. El-Zaart. Images thresholding using isodata technique with gamma distribution. *Pattern Recognition and Image Analysis*, 20(1):29–41, 2010.

[3] D.H. AlSaeed, A. Bouridane, A. ElZaart, and R. Sammouda. Two modified otsu image segmentation methods based on lognormal and gamma distribution models. In *Information Technology and e-Services (ICITeS), 2012 International Conference on*, pages 1–5, March 2012.

[4] A. Al-Hussain and A. El-Zaart. Moment-preserving thresholding using gamma distribution. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, volume 6, pages V6–323–V6–325, April 2010.

[5] E. Assidan and A. El-Zaart. Fast optimal multimodal thresholding based on between-class variance using a mixture of gamma distributions. In *Systems, Man and Cybernetics (SMC), IEEE 2009 International Conference on*, pages 2599–2602, Oct 2009.

[6] A.A. Ghosn, A. El-Zaart, and E. Assidan. Mammogram images thresholding based on between-class variance using a mixture of gamma distributions. In *Advances in Computational Tools for Engineering Applications (ACTEA), 2012 International Conference on*, pages 75–79, Dec 2012.

[7] A.A. Amory, A. El Zaart, A.O. Rokabi, H. Mathkour, and R. Sammouda. Fast optimal thresholding based on between-class variance using mixture of log-normal distribution. In *Information Technology and e-Services (ICITeS), 2012 International Conference on*, pages 1–4, March 2012.

[8] E. Assidan and A. El-Zaart. Fast optimal multimodal thresholding based on between-class variance using a mixture of gamma distributions. In *Systems, Man and Cybernetics (SMC), IEEE 2009 International Conference on*, pages 2599–2602, Oct 2009.

[9] Ali El Zaart, Djemel Ziou, Shengrui Wang, and Qingshan Jiang. Segmentation of {SAR} images. *Pattern Recognition*, 35(3):713 – 724, 2002.

[10] Z. Tao, H. D. Tagare, and J. D. Beaty. Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images. *IEEE Transactions on Medical Imaging*, 25(11):1483–1491, 2006.

[11] Y. Bazi, L. Bruzzone, and F. Melgani. Image thresholding based on

the em algorithm and the generalized gaussian distribution. *Pattern Recognition*, 40(2):619–634, 2007.

[12] R. C. H. Cheng and T. C. Iles. Embedded models in three-parameter distributions and their estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1):135–149, 1990.

[13] H. Hirose. Maximum likelihood parameter estimation in the three-parameter gamma distribution. *Computational Statistics and Data Analysis*, 20(4):343–354, 1995.

[14] W. T. Vetterling. *Numerical Recipes: Example Book (C)*. Numerical recipes in C / William H. Press. Cambridge University Press, 1992.

[15] S. C. Choi and R. Wette. Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics*, 11(4):683–690, 1969.

[16] J. Fan. Notes on poisson distribution-based minimum error thresholding. *Pattern Recognition Letters*, 19(56):425–431, 1998.

[17] J. Fan and W. Xie. Minimum error thresholding: A note. *Pattern Recognition Letters*, 18(8):705–709, 1997.

[18] W. Feller. A direct proof of stirling's formula. *The American Mathematical Monthly*, 74(10):1223–1225, 1967.

[19] C. Goutte and E. Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In D. Losada and J. M. Fernández-Luna, editors, *Advances in Information Retrieval*, volume 3408 of *Lecture Notes in Computer Science*, pages 345–359. Springer Berlin Heidelberg, 2005.

[20] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, 19(1):41–47, 1986.

[21] D. Kundu and R. D. Gupta. A convenient way of generating gamma random variables using generalized exponential distribution. *Computational Statistics and Data Analysis*, 6(51):2796–2802, 2007.

[22] T. Kurita, N. Otsu, and N. Abdelmalek. Maximum likelihood thresholding based on population mixture models. *Pattern Recognition*, 25(10):1231–1240, 1992.

[23] L. LeCam. Maximum likelihood: An introduction. *International Statistical Review / Revue Internationale de Statistique*, 58(2):153–171, 1990.

[24] G. Martinez. Criterion for automatic selection of the most suitable maximum-likelihood thresholding algorithm for extracting object from their background in a still image. In *Machine Vision Applications (MVA), 2005 International Conference on*, pages 10–13, 2005.

[25] J. M. Mendel. *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Pearson Education, 1995.

[26] I. J. Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100, 2003.

[27] J. A. Noble and D. Boukerroui. Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging*, 25(8):987–1010, 2006.

[28] R. H. Norden. A survey of maximum likelihood estimation. *International Statistical Review / Revue Internationale de Statistique*, 40(3):329–354, 1972.

[29] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, SMS-9(1):62–66, 1977.

[30] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.

[31] S. Raut, M. Raghuvanshi, R. Dharaskar, and A. Raut. Image segmentation no.150; a state-of-art survey for prediction. In *Advanced Computer Control (ICACC), 2009 International Conference on*, pages 420–424, 2009.

[32] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–165, 2004.