Image and Vision Computing 28 (2010) 677–695

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Gradient histogram: Thresholding in a region of interest for edge detection

Debashis Sen *, Sankar K. Pal

Center for Soft Computing Research, Indian Statistical Institute, 203 B.T. Road, Kolkata, West Bengal 700108, India

ARTICLE INFO

Article history:
Received 28 September 2007
Received in revised form 1 October 2009
Accepted 26 October 2009

Keywords:
Edge detection
Gradient operators
Random process
Gradient histogram
Skewness and kurtosis
Threshold determination
Non-maximum suppression
Hysteresis thresholding

ABSTRACT

Selecting a threshold from the gradient histogram, a histogram of gradient magnitudes, of an image plays a crucial role in a gradient based edge detection system. This paper presents a methodology to determine the threshold from a gradient histogram generated using any kind of linear gradient operator on an image. We consider the image as a random process with dependent samples, model the gradient histogram using theories of random process and random input to a system, and determine a region of interest in the gradient histogram using certain properties of a probability density function. Standard histogram thresholding techniques are then used within the region of interest to get the threshold value. To obtain the edges, this threshold value is then used as the upper threshold of the hysteresis thresholding technique that follows the non-maximum suppression operation applied on the gradient magnitude image. The proposed methodology of determining a threshold in a gradient histogram is deduced through rigorous analysis and hence it helps in achieving consistently appreciable edge detection performance. Experimental results using different real-life and benchmark images are shown to demonstrate the effectiveness of the proposed technique.

1. Introduction

Edge detection is a very important low-level image processing operation, which is essential in order to carry out various higher level tasks such as motion and feature analysis, understanding, recognition and retrieval from databases. Over four decades of research work on edge detection have resulted in the development of a plethora of simple to complex techniques. However, as stated in [1–6], the problem of finding the edges has not been solved in its entirety and no universally accepted technique exists.

Considering the intensity of an image as a two dimensional function, it is generally accepted that the edges in the image are "meaningful" discontinuities (changes) of this function [6–8]. "Meaningful" discontinuities refer to the edges due to various regions in an intensity image, whereas, the "non-meaningful" discontinuities are those due to inherent texture and various noise. It may be noted that we shall henceforth refer 'intensity image' as 'image' for simplicity.

The problem of edge detection has been tackled using various different paradigms such as surface fitting [9,10], optimization of a criterion [11–18], statistical testing [19,20] and soft computing [21,22]. Among the various methods in all the paradigms, the popular are the ones based on finding the "amount" of edge at each pixel of the image. Authors in [11,15,23] have, respectively, suggested the use of first-, second- and fourth-order derivatives to obtain the "amount" of edge at a pixel location. As the use of higher-order derivatives makes the system susceptible to high frequency noise and also results in poorer localization of the edges [6,24], we shall consider the first-order derivative, which is used to calculate the gradient at each pixel in the image.

Authors in [11–14] have optimized various criteria in order to find operators for computing the gradient, which represents the edge at every pixel in the image. It is interesting to note that all these gradient operators may be considered as a smoothing system followed by the first-order derivative operation. In addition to an optimal gradient operator, Canny [11,25] proposed a technique called the non-maximum suppression (NMS), where the fact that an edge at a pixel is legitimate only when the gradient magnitude at that pixel assumes a maximum in the direction of the gradient. He suggested that the NMS be used as a post-processing operation along with the gradient operator for edge detection in order to obtain edges of single pixel width.

Most of the optimal gradient operators reported in the literature are based on the three performance criteria, namely, good detection, good localization and only one response to a single edge, given by Canny [11]. As discussed in [11,13], the optimizations of these three criteria are contradictory and hence a trade-off is required. Such a trade-off implies that the smoothing operation cannot remove the "false" edges due to the "non-meaningful" discontinuities completely. Hence, after all the above mentioned processes of smoothing, first-order derivative and NMS, we still are at a stage where we need a decision making system in order to distin-

* Corresponding author. Tel.: +91 33 25752048.
E-mail addresses: dsen_t@isical.ac.in (D. Sen), sankar@isical.ac.in (S.K. Pal).

0262-8856/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2009.10.010

guish the edges which represent the "meaningful" discontinuities from the "false" edges.

As mentioned in [11,15,1,3,10], the gradient magnitude values representing the "meaningful" discontinuities in an image are in general greater than those representing the "non-meaningful" ones. Assuming that the above condition is always true, a global histogram (gradient histogram) of the gradient magnitudes at all the pixels in the image can be constructed and then a thresholding algorithm can be used to determine a threshold in order to remove the unwanted "false" edges. Ensuring that the above assumption holds, is a job to be taken care of, when the gradient operator is designed. However, when the above assumption is not met, that is, an image has certain "non-meaningful" discontinuities with higher value of gradient magnitude than those of certain "meaningful" ones, some edges would be eliminated and some "false" edges would remain. Now, it could be considered that the gradient magnitudes of the "meaningful" discontinuities are always greater than those of the "non-meaningful" ones within local neighborhoods and hence local thresholding [26–28] could be applied to eliminate the "false" edges. But it is well known that local thresholding techniques are highly sensitive to noise and the size of the local neighborhood considered [28]. Moreover, local thresholding techniques using non-overlapping blocks [28] could give chaotic edge detection results when there is no constraint considered regarding the connectivity of the edges detected in neighboring blocks. In practice, it is inevitable that an image would have certain "non-meaningful" discontinuities with higher value of gradient magnitude than those of certain "meaningful" ones, and hence perfect separation of the edges from the "false" edges would be impossible.

Commenting on the determination of a threshold from the gradient histogram to eliminate the "false" edges, Canny in [11] reported an associated problem of edge streaking and suggested that a technique called hysteresis thresholding, which employs two thresholds, be used to overcome the problem. Edge streaking arises when portions of edge contours are eliminated mainly because of the usage of a single threshold value, which is determined from the gradient histogram, on an image of gradient magnitudes (gradient image) that has certain "non-meaningful" discontinuities with higher gradient magnitude value than certain "meaningful" ones. However, the threshold determined from the gradient histogram can be considered as the upper threshold of the hysteresis process and the lower threshold can be used to capture some of the edges due to "meaningful" discontinuities that were missed by the upper threshold. The lower threshold is usually determined by multiplying a constant (<1) with the upper threshold. But, as the analysis in [29] proves, in order to determine the ratio of the upper threshold to the lower threshold, one needs to know in advance the number of edge pixels in the image under consideration.

Our aim in this paper is to study and develop a gradient histogram thresholding methodology required to determine an appropriate threshold value that can be used as the upper threshold of the hysteresis process, when any arbitrary gradient image is encountered. Note that the design of the gradient operator and determination of various other parameters of gradient based edge detection systems, which themselves are separate areas of study, are not our prime concern.

In [11], Canny states that the performance of his edge detection system is critically dependent on the process of gradient histogram thresholding and this is true for most gradient based edge detection systems [1,2,6]. The usual practice has been to apply any conventional histogram thresholding algorithm [30–34] to the gradient histogram in order to determine the threshold. Such approaches do not yield acceptable result over a wide range of images as the thresholding algorithms are not specific to gradient histogram. Considering the criticality of the thresholding process in the overall edge detection system, it is surprising that except for a few researchers [10,3,6,29,35], no one has addressed this problem of gradient histogram thresholding to the best knowledge of the authors.

In [3], the gradient histogram was assumed to be a weighted sum of two gamma densities in order to identify a threshold by parameter estimation. However, no justification was given for such an assumption. The authors in [6] standardize the edge magnitude over the whole image under consideration and claim by experimentation that such a process stabilizes the edge detection results over a variety of images. But the stabilization achieved neither implies that the thresholds considered will be satisfactory over a wide range of images nor ensures that it will help automatic threshold selection. In [10], a cubic facet model for image and a Gaussian model for noise were considered and various parameters were estimated in order to determine the threshold using the Bayesian decision theory. Such a method is very complex and suffers from the errors due to the facet modeling, parameter estimation and decision making. The author in [35] modeled the distribution of the gradient magnitudes due to the "non-meaningful" discontinuities using a Rayleigh density function in the quest to obtain an appropriate threshold. We shall see in this paper, that such a model implies that the whole image is almost homogeneous having noise and texture which follow a Gaussian model. Considering a gradient histogram model similar to the one suggested in [35], the authors in [29] proposed a technique to determine the thresholds required in the hysteresis process assuming that certain prior information was available.

In this paper, we carry out a rigorous analysis of the processing of an image in a gradient based edge detection system and then propose a methodology to determine an appropriate threshold from the gradient histogram. The threshold determined is used as the upper threshold in the hysteresis process in order to extract the actual (desired) edges in an image. In the proposed methodology, a region of interest (ROI) in the gradient histogram is first obtained by computing certain skewness and kurtosis values, and comparing them to certain deduced constants. These constants are determined considering images as random processes and then by approximately modeling a gradient histogram. Once the ROI in the gradient histogram of the image under consideration has been obtained, any general histogram thresholding algorithm could be applied to the ROI in order to determine the threshold value. We consider, for example, the histogram thresholding algorithms given in [30,34]. The effectiveness of the proposed technique is demonstrated using some real-life and benchmark images through qualitative and ground-truth based quantitative results. A very interesting aspect of the proposed method of threshold determination is that it is applicable to gradient histograms obtained after the application of any linear gradient operator on the image.

The organization of the paper is as follows. In Section 2, we derive approximate expressions for the histogram and the gradient histogram of an image considering it as a random process. The procedure to obtain the ROI in a gradient histogram and the threshold determination process is explained in Section 3. Brief descriptions of the NMS and the hysteresis processes used are also given in this section. Experimental results on a few real-life and benchmark images are shown in Section 4 and the superiority of the proposed method over the other existing ones is demonstrated. The paper concludes with Section 5 by providing an overview of the contributions made.

2. Representation of images using random processes and gradient histogram modeling

Representing images by random processes fields is intuitively appealing as the gray values (or intensity values) at various locations in images depend on many aspects and can hardly be un-

iquely determined using a mathematical expression, rule or table. Representation of images by random processes implies that the gray values at different locations in an image are generated by random variables. Now, any arbitrary image is of spatially changing nature and therefore the presence of edges. Hence, we may realistically consider that different regions in an image can be represented by different stationary random processes [36], preserving the spatially changing nature of the overall image. Therefore, the gray values at different pixels within a region of an image are generated by random variables having the same probability density function (PDF) with identical characteristic parameters. We assume that the PDF is short-tailed. The PDF cannot be long-tailed, as such a PDF would imply the presence of abrupt values within a region. Such an abrupt value in an image should be associated with a pixel of an edge and not with a pixel within a region. Note that by region we are referring to homogeneous region that might contain texture and noise.

Random processes are characterized by the joint PDFs of the various random variables representing the states of the random processes at various times [36]. In this paper, a random process whose underlying finite-dimensional joint PDFs are given by a specific family of multivariate functions shall be referred using the corresponding family name. For example, we shall refer a random process characterized by multivariate Gaussian functions (PDFs) as a Gaussian random process. Let us now consider the following two definitions.

Definition 1. A random process is called wide sense stationary (WSS) if it has a constant mean (that is, identical first-order densities) and its autocorrelation depends only on time interval (time separation between samples/random variables) [36].

Definition 2. A random process is ζ-dependent when various samples (random variables) separated by a time greater than a finite constant ζ are mutually independent [36].

The concept behind Definition 2 has been arrived from the fact that in many cases the mutual dependence of the samples of a random process decreases with the increase in time separation between them [36]. Assuming that dependency between pixels in an image dies down with reduction in proximity between them, we consider that each region of an image is represented by a WSS ζ-dependent Gaussian random process. A WSS Gaussian process implies a strict sense stationary (or stationary) Gaussian process as a Gaussian process is completely characterized by its mean and autocovariance [36]. Therefore, in effect we have considered the short-tailed PDF of the random variables generating the gray values at different pixels within a region of an image as a Gaussian PDF. It is also considered that the identical mean and variance of the Gaussian random variables corresponding to the Gaussian process equals the sample mean and sample variance of the region under consideration and hence the process is ergodic [36]. Note that such a representation of the regions in an image by different Gaussian processes is analogous to the non-stationary framework used in [37].

As mentioned earlier, the gray values in different regions of an image are generated by random variables having Gaussian PDFs with different parameters. Now, when we consider any arbitrary image as a whole, the gray value at a particular pixel would correspond to any one of various Gaussian functions. Therefore, in effect we may consider that a selection process is taking place that decides the region to which a pixel belongs, before the gray value at the pixel is generated. Hence, a random variable representing the gray value at a pixel in an image should also reflect the process of selection. On incorporation of the selection process, it is straightforward from probability theory [36] that the random variables generating the gray values in an image should follow a mixed Gaussian (weighted sum of Gaussians) PDF. Such a representation allows us to consider an image as a single random process rather than representing each region in the image using a random process and this makes our analysis easier.

A random variable, say Λ_1, having a PDF ($f(\cdot)$) given by a weighted sum of Gaussian functions is referred to as a mixed Gaussian random variable. A mixed Gaussian PDF is expressed as

$$f(\Lambda_1) = \sum_r \frac{W_r}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(\Lambda_1 - I_r)^2}{2\sigma_r}\right) \tag{1}$$

where the weight W_r is the ratio of the area of the rth region to the whole area of the image, I_r and σ_r are, respectively, the mean and the standard deviation of the rth region. Now, it can be easily shown that a multivariate mixed Gaussian PDF is in fact mixed multivariate Gaussian PDF. Let Λ_1 and Λ_2 be mixed Gaussian random variables. Then the corresponding bivariate mixed Gaussian PDF is of the form

$$f(\Lambda_1, \Lambda_2) = \sum_{\acute{r}} \frac{\acute{W}_{\acute{r}}}{2\pi\sigma_{1\acute{r}}\sigma_{2\acute{r}}\sqrt{1-\rho_{\acute{r}}^2}} \exp\left(-\frac{1}{2(1-\rho_{\acute{r}}^2)}\left(\frac{(\Lambda_1 - I_{1\acute{r}})^2}{\sigma_{1\acute{r}}^2} + \frac{(\Lambda_2 - I_{2\acute{r}})^2}{\sigma_{2\acute{r}}^2} - 2\rho_{\acute{r}}\frac{(\Lambda_1 - I_{1\acute{r}})(\Lambda_2 - I_{2\acute{r}})}{\sigma_{1\acute{r}}\sigma_{2\acute{r}}}\right)\right) \tag{2}$$

For simplicity, we have used a single notation $\acute{r}$ to represent all the regions corresponding to Λ_1 and Λ_2. This implies, if r_a and r_b, respectively, denote the regions in Λ_1 and Λ_2, then $\acute{r}$ represents all combinations of r_a and r_b.

Now, as deduced earlier, the random variables generating the gray values at different pixels in an image have the same mixed Gaussian PDF. Hence, from the earlier analysis, we get that an image can be represented as a mixed Gaussian random process when different regions in the image are considered to have been generated by different WSS ζ-dependent Gaussian random processes. We shall now assume that the value of ζ is same for all the Gaussian random processes and that the mixed Gaussian random process is also a ζ-dependent process. We shall also assume that the dependence between various samples/random variables of the ζ-dependent mixed Gaussian random process decreases in a similar manner with the time separation between them and hence the autocorrelation of the process depends on the time interval alone. Therefore, according to Definition 1, the ζ-dependent mixed Gaussian random process is also WSS. It is important to note that the assumptions above does not compromise on the changing nature of an image unlike some of the existing models [3,29,35], which shall be evident later in the paper.

Therefore, it stands that we represent a whole image as a WSS ζ-dependent mixed Gaussian random process, without loosing the edge details (changing nature) in the image. However, as suggested in [38–40], for certain kinds of images a right-skewed PDF may represent the short-tailed PDF mentioned earlier better than a Gaussian PDF. Such a situation especially arise with intensity images captured using coherent imaging systems, where the right-skewed PDF may be approximately considered as a lognormal PDF [39,40]. Hence, in such cases, when we consider the short-tailed PDF as a lognormal PDF, we essentially represent the regions in such images as WSS ζ-dependent lognormal random processes. Therefore, the images obtained from coherent imaging system are represented by WSS ζ-dependent mixed lognormal random processes.

The representation of an image by a WSS ζ-dependent mixed Gaussian or mixed lognormal random process does not cause much of a change to our development of a gradient histogram thresholding methodology. A trivial transformation would be required as a

pre-processing step to the gradient based edge detection system, when images represented by mixed lognormal processes are under consideration.

Let us denote an image represented by a WSS ζ-dependent mixed Gaussian process using X and an image represented by a WSS ζ-dependent mixed lognormal process using $\overline{X}$. As mentioned in Section 1, smoothing is the first operation carried out for edge detection. However, if an image is represented by a WSS ζ-dependent mixed lognormal process, then we first apply a logarithmic transformation on the pixels of the image ($\overline{X}_{ij}$) before smoothing in order to obtain an image that can be represented by a WSS ζ-dependent mixed Gaussian process. Therefore, we have

$$X_{ij} = \ln(\overline{X}_{ij}) \tag{3}$$

where X_{ij} and $\overline{X}_{ij}$ denote the gray values at ith row and jth column in the images X and $\overline{X}$, respectively. Note that X_{ij} and $\overline{X}_{ij}$ are time samples of random processes and hence they are random variables. The deduction that X given in (3) is indeed a WSS ζ-dependent mixed Gaussian process shall be given later.

2.1. Linear gradient operators

The gradient operators for edge detection obtained by various authors [11–14] after optimizing various criteria are linear in nature or can be very closely approximated by a linear operator. The application of any linear gradient operator (H) on a digital image (X) can be represented by a window operation such as

$$Y_{ijg} = \sum_p \sum_q H_{pqg} X_{i+p\ j+q} \tag{4}$$

where

$$p = -k, \ldots, 0, \ldots, +k$$
$$q = -k, \ldots, 0, \ldots, +k$$

In the above, $(2k+1) \times (2k+1)$ gives the size of the window, otherwise referred to as the width of the operator. The symbol g indicates the direction of operation on an image, that is, the horizontal (R) and vertical (C) directions in which the gradient is determined. Hence, the vertical component $Y_{ij\mathrm{R}}$ and the horizontal component $Y_{ij\mathrm{C}}$ of the gradient are orthogonal to each other. The magnitude of the gradient at each pixel in the image is obtained by

$$G_{ij} = \sqrt{(Y_{ij\mathrm{R}})^2 + (Y_{ij\mathrm{C}})^2} \tag{5}$$

Let there be D regions in an image X and let us express X as

$$X_{ij} = I_r + T_{ij} = I_r + \sigma_r V_{ij} \tag{6}$$

where r takes a value from the set of integers $\{1, 2, \ldots, D\}$ depending on the values of the indices of V (zero-mean unit-variance noise), I_r is the mean of the rth region, T is the overall noise (zero-mean noise) and σ_r is variance of the overall noise in the rth region. Note that T_{ij} can be expressed as

$$T_{ij} = N_{ij} + U_{ij} \tag{7}$$

where N_{ij} is a noise corrupting the image, U_{ij} is the inherent texture in the uncorrupted image. Using (6) in (4), we get

$$Y_{ijg} = \sum_p \sum_q H_{pqg} (I_r + \sigma_r V_{i+p\ j+q}) \tag{8}$$

Most of the gradient operators reported in literature are obtained based on the three performance criteria given by Canny [11]. These three criteria are the maximization of signal to noise ratio (SNR), localization criterion and unique response to a single edge criterion. However, as mentioned in [11,13], the simultaneous optimization of these criteria are not possible and hence a trade-off is required. This substantiates the necessity of a decision making system to eliminate the "false" edges due to the overall noise T (see (7)), which is never completely suppressed by H.

From (8), we see that apart from the type of the operator, the width of the operator plays an important role in computing the gradient at each pixel of the image. As suggested in [11], the width of the operator may be obtained based on a local estimate of noise energy. Furthermore, the author of [11] suggests the integration of information from operators working at different scales (different widths) in order to mark the appropriate edges in an image. In general, such a technique requires that the decision making process be applied for each of the operators with different widths.

2.2. The histogram of an image

We shall now obtain the histogram of an image when it is represented by a random process. As mentioned earlier, we shall represent a WSS ζ-dependent mixed Gaussian random process (image) using X and thus X_{ij}, which is the sample of X at $(i,j)th$ position, will be a random variable. Let us first put forth a formal definition of the histogram of an image, which is a random process. Note that, as the probability of occurrences of gray values in an image can be depicted as their number of occurrences in the image normalized by the total number of pixels, we consider that the image histogram gives the probability of occurrences of gray values.

Let $A = \{a_t, \forall\ t\}$ be a finite set of samples (random variables) generated, respectively, at times $t = 1, 2, \ldots, n$ by a random process, say Υ. Let β_n be a random variable representing an experiment whose result is an outcome of a random experiment represented by any one randomly chosen equally likely sample in A (see Fig. 1).

Definition 3. The histogram ($h(\Upsilon)$) of an image represented by a random process Υ is given by the probability density function of the random variable β_n, that is, $h(\Upsilon) = f(\beta_n)$, with n as the number of pixels in the image.

As mentioned earlier, $h(\Upsilon)$ gives the probability of occurrences of gray values in the underlying image. From the explanation of the random variable β_n, it is evident that the PDF of β_n would also give the probability of occurrence of gray value in the image, which is a random process and hence the above definition.

Theorem 1. *If Υ is a ζ-dependent random process with identical first-order probability density functions denoted by $f(a_t)$, then $f(\beta_n) \simeq f(a_t)$ as $n \to \infty$.*

The proof of the aforesaid theorem is given in Appendix A.

Corollary 1. *If Υ is an independent identically distributed (i.i.d.) random process, then $f(\beta_n) \simeq f(a_t)$ for any value of n.*

Corollary 2. *If Υ is a Gaussian or mixed Gaussian wide sense stationary ζ-dependent random process, then the random variable β_n will have a Gaussian or mixed Gaussian probability density function, respectively.*

Corollary 3. *If Υ is a lognormal or mixed lognormal wide sense stationary ζ-dependent random process, then the random variable β_n will have a lognormal or mixed lognormal probability density function, respectively.*

The deductions of the aforesaid three corollaries are discussed in Appendix B.

From (A.11)–(A.13) in Appendix A, we find that when $n \gg \zeta, h(\Upsilon) \approx f(a_t)$. We consider such a case with images as it can be realistically assumed that the span of dependent pixels in an image is considerably small compared to the size of the image. As we have considered the image (X) as a WSS ζ-dependent mixed Gaussian random process, using Theorem 1 and Corollary 2 we get the histogram of an image as

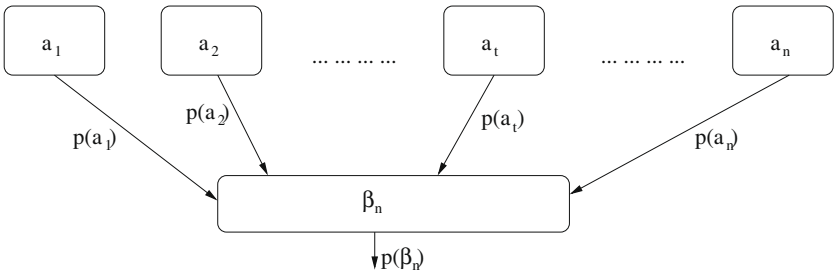


Fig. 1. Pictorial representation of the random experiment given by β_n.

$$h(X) \approx f(X_{ij}) = \sum_{r=1}^{D} \frac{W_r}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(X_{ij} - I_r)^2}{2\sigma_r}\right) \tag{9}$$

where I_r and σ_r are the mean and standard deviation corresponding to the rth region in the image. The symbol W_r represents the weights as explained for (1). Using Theorem 1 and Corollary 3 we get the histogram of an image ($\overline{X}$) represented by WSS ζ-dependent mixed lognormal random process as

$$h(\overline{X}) \approx f(\overline{X}_{ij}) = \sum_{r=1}^{D} \frac{W_r}{\sqrt{2\pi}s_r\overline{X}_{ij}} \exp\left(-\frac{(\ln\overline{X}_{ij} - m_r)^2}{2s_r}\right) \tag{10}$$

where m_r and s_r are the mean and standard deviation of the natural logarithm of the samples in the regions, respectively. As given in (3), the natural logarithmic operation is applied on $\overline{X}_{ij}$ in order to obtain an image that can be represented by a WSS ζ-dependent mixed Gaussian process. We may visualize this operation as a WSS ζ-dependent mixed lognormal random process ($\overline{X}$) being passed through a memoryless non-linear system. We shall now put forth the following proposition about the application of a ζ-dependent process to a memoryless system.

Proposition 1. *If a ζ-dependent process is applied as an input to a memoryless linear or non-linear system, the output shall also be a ζ-dependent process with the same value of ζ.*

The proof of the aforesaid proposition is given in Appendix C.

As given in [41], if the input random process to a non-linear memoryless system is WSS, then the output random process is also WSS. Hence, using Proposition 1 and the fact that the logarithm of a mixed lognormal random variable is mixed Gaussian distributed, we may conclude that we shall have a WSS ζ-dependent mixed Gaussian random process at the output of the non-linear memoryless system, that is, the natural logarithmic operation. Therefore, it is true that the X obtained in (3) is a WSS ζ-dependent mixed Gaussian random process and we shall have a $h(X)$ of the form given in (9).

In literature pertaining to image processing, assumption of an arbitrary model for image histogram has been used many times for various purposes. In this paper, we have used random processes to capture some general characteristics, like spatially changing nature, inter-pixel dependency and noise, of an image and hence arrived at a model for an image histogram through rigorous theoretical analysis.

2.3. Approximate model for a gradient histogram

We shall now develop an approximate model for gradient histograms of gradient magnitude images obtained using any linear gradient operator. As mentioned earlier, in a gradient based edge detecting system, an optimal linear gradient operator is first applied on the image X. Let us now rewrite (4) below

$$Y_{ijg} = \sum_{p} \sum_{q} H_{pqg} X_{i+p\ j+q} \tag{11}$$

The above equation shows the application of a linear system on a digital image in the form of a window operator. Let H_{imp} be the impulse response of the linear system given by H in (11). We shall now put forth the following proposition about the application of a ζ-dependent process to a linear system.

Proposition 2. *If a WSS ζ-dependent process applied as an input to a linear system generates an output process whose samples when uncorrelated are also independent, then the output is also a WSS ζ-dependent process.*

The proof of the aforesaid proposition is given in Appendix C.

If the image X is a mixed Gaussian process, then the output Y given in (11), which represents a linear system, will also be a mixed Gaussian process [36]. Note that, in [36], the above result has been shown for Gaussian random processes. The extension to mixed Gaussian processes is trivial. Now, using Proposition 2, we get that when a WSS ζ-dependent mixed Gaussian process is given as an input to a linear system, the output process will also be WSS ζ-dependent mixed Gaussian process. Therefore, both Y_R and Y_C (see (4)) are WSS ζ-dependent mixed Gaussian random processes.

We shall now rewrite (5), which gives the calculation of the magnitude of the gradient at each pixel in the image X.

$$G_{ij} = \sqrt{(Y_{ijR})^2 + (Y_{ijC})^2} \tag{12}$$

In the above, using Proposition 1, we find that G is also a ζ-dependent random process as the G_{ij} values are obtained working on the same pixels in both Y_R and Y_C. From (12), we also see that the process G will have identical first-order densities as both Y_R and Y_C are WSS random processes. Hence, Theorem 1 can be applied to the gradient magnitude image G in order to find $h(G)$. However, we shall first determine the PDF of G_{ij}. In order to find the PDF of G_{ij}, we require the joint PDF of Y_{ijR} and Y_{ijC}. We know that Y_{ijR} and Y_{ijC} are individually mixed Gaussian distributed and as they are obtained by applying the same mixed Gaussian random process X to different linear systems, they are mutually dependent. Let us now consider a random variable A, such that

$$A = aY_{ijR} + bY_{ijC} \tag{13}$$

where a and b are arbitrary real-valued constants. The random variable A is mixed Gaussian distributed as both Y_{ijR} and Y_{ijC} have been obtained by a linear operation (see (11)) on the same set of various random variables those are individually and jointly mixed Gaussian distributed. The random variable A being mixed Gaussian distributed implies that the joint PDF of Y_{ijR} and Y_{ijC} is a mixed bivariate Gaussian PDF [36]. This property has been shown for Gaussian distributed random variable in [36] and the extension to mixed Gaussian distributed random variable is trivial. Now, using (2) we have

$$f(Y_{ij\mathrm{R}}, Y_{ij\mathrm{C}}) = \sum_{r_G} \frac{\omega_{r_G}}{2\pi v_{1r_G} v_{2r_G}\sqrt{1-\rho_{r_G}^2}} \times \exp\left(-\frac{1}{2(1-\rho_{r_G}^2)}\left(\frac{(Y_{ij\mathrm{R}}-\mu_{1r_G})^2}{v_{1r_G}^2} + \frac{(Y_{ij\mathrm{C}}-\mu_{2r_G})^2}{v_{2r_G}^2} - 2\rho_{r_G}\frac{(Y_{ij\mathrm{R}}-\mu_{1r_G})(Y_{ij\mathrm{C}}-\mu_{2r_G})}{v_{1r_G}v_{2r_G}}\right)\right) \quad (14)$$

where r_G represents all the regions corresponding to $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$, μ_{1r_G} and v_{1r_G} are the means and standard deviations corresponding to $Y_{ij\mathrm{R}}$, and μ_{2r_G} and v_{2r_G} are the means and standard deviations corresponding to $Y_{ij\mathrm{C}}$. The ω_{r_G}s are certain constant weights. Now, as mentioned earlier, $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are orthogonal to each other. Note that the expression in (12) can be used to represent the gradient magnitude only when $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are orthogonal. Therefore, the correlation between $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ is zero, that is, $R_{Y_{ij\mathrm{R}}Y_{ij\mathrm{C}}} = 0$. Hence, the correlation co-efficient between $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ is given by

$$\rho = -\frac{\mu_1\mu_2}{v_1 v_2} = -\frac{\mu_1\mu_2}{\sqrt{(M_1-\mu_1^2)(M_2-\mu_2^2)}} \quad (15)$$

where μ_1 and μ_2 are the first-order moments of $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$, respectively, M_1 and M_2 are the second-order moments and v_1 and v_2 are the standard deviations. We have

$$\mu_1 = \sum_{r_1} \upsilon_{1r_1}\mu_{1r_1}, \quad M_1 = \sum_{r_1} \upsilon_{1r_1} M_{1r_1}$$
$$\mu_2 = \sum_{r_2} \upsilon_{2r_2}\mu_{2r_2}, \quad M_2 = \sum_{r_2} \upsilon_{2r_2} M_{2r_2} \quad (16)$$

where r_1 and r_2, respectively, represent the regions corresponding to $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$, and υ_{1r_1} and υ_{2r_2} are the weights corresponding to the random variables $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$, respectively. The constants ω_{r_G}s in (14) are formed from the various products of υ_{1r_1} and υ_{2r_2}. From (16), we see that the first (mean) and second-order moments of $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are the weighted sums of the first (mean) and second-order moments of the constituent Gaussian functions.

We shall now consider a few aspects of an image, the modeling of each region in an image using a Gaussian process and then the whole image as a mixed Gaussian process in order to explore the nature of the joint PDF given in (14).

First, let us consider a few realistic aspects of an image. In general, the edges in an image cover much less area in the scene than the homogeneous regions. Hence, the weighting constants υ_{1r_1} and υ_{2r_2} corresponding to those values of r_1 and r_2 when $\mu_{1r_1} \approx 0$ and $\mu_{2r_2} \approx 0$ would be much larger than the others. It can also be realistically assumed that among μ_{1r_1} and μ_{2r_2} for other values of r_1 and r_2, some would have negative values and some would have positive. Hence, the values of μ_1 and μ_2 would be very near to zero. On the other hand, as the values of M_1 and M_2 indicates the presence of texture and noise in various regions, their values would not be small compared to μ_1 and μ_2. Hence, we may conclude from the expression in (15) that the value of ρ would be very small and might be considered negligible. Now, as $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are mixed Gaussian distributed, a small value of ρ suggests that the random variables $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are weakly dependent on each other. Now, as the weighting constants ω_{r_G}s are products of various υ_{1r_1} and υ_{2r_2}, they would be large for those values of r_G when $\mu_{1r_G} \approx 0$ and $\mu_{2r_G} \approx 0$ compared to others. Hence, we may consider that the values of ρ_{r_G} are small in general. Note that r_G denotes all combinations of r_1 and r_2.

Secondly, the values of the random variables $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are obtained by performing a weighted summation operation (see (11)) on the same set of pixels of the input image and hence they would have the same value of variance [36], that is, $v_1 = v_2$. Now, we have considered that each region of an image, which is subjected to the gradient operator, is generated by a Gaussian process and thus the whole image is represented by a mixed Gaussian process. Therefore, the expression in (14) would not have any component such that $v_{1r_G} \neq v_{2r_G}$ and we shall always have $v_{1r_G} = v_{2r_G} = v_{r_G}$.

As we have now deduced that the correlation co-efficient between $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ is very small, we may approximately consider that $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are mutually independent. However, rather than assuming independence of $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$, we come up with a better approximation here. Considering mathematical tractability, we suggest the following approximation of (14)

$$f(Y_{ij\mathrm{R}}, Y_{ij\mathrm{C}}) \approx f_a(Y_{ij\mathrm{R}}, Y_{ij\mathrm{C}}) \quad (17)$$

and

$$f_a(Y_{ij\mathrm{R}}, Y_{ij\mathrm{C}}) = \begin{cases} \sum_{r_G} \frac{\omega_{r_G} N_1}{2\pi v_{r_G}^2\left(\sqrt{1-\rho_{r_G}^2}\right)} \exp\left(-\frac{1}{2(1-\rho_{r_G}^2)}\left[\Psi_{\mathrm{R}}^2 + \Psi_{\mathrm{C}}^2 - \rho_{r_G}\left(\Psi_{\mathrm{R}}^2 + \Psi_{\mathrm{C}}^2\right)\right]\right), & \Psi_{\mathrm{R}} > 0 \text{ and } \Psi_{\mathrm{C}} > 0 \\ & \Psi_{\mathrm{R}} < 0 \text{ and } \Psi_{\mathrm{C}} < 0 \\ \sum_{r_G} \frac{\omega_{r_G} N_{-1}}{2\pi v_{r_G}^2\left(\sqrt{1-\rho_{r_G}^2}\right)} \exp\left(-\frac{1}{2(1-\rho_{r_G}^2)}\left[\Psi_{\mathrm{R}}^2 + \Psi_{\mathrm{C}}^2 + \rho_{r_G}\left(\Psi_{\mathrm{R}}^2 + \Psi_{\mathrm{C}}^2\right)\right]\right), & \Psi_{\mathrm{R}} > 0 \text{ and } \Psi_{\mathrm{C}} < 0 \\ & \Psi_{\mathrm{R}} < 0 \text{ and } \Psi_{\mathrm{C}} > 0 \\ \sum_{r_G} \frac{\omega_{r_G} N_0}{2\pi v_{r_G}^2\left(\sqrt{1-\rho_{r_G}^2}\right)} \exp\left(-\frac{1}{2(1-\rho_{r_G}^2)}\left[\Psi_{\mathrm{R}}^2 + \Psi_{\mathrm{C}}^2\right]\right), & \Psi_{\mathrm{R}} = 0 \\ & \Psi_{\mathrm{C}} = 0 \end{cases} \quad (18)$$

where

$$\Psi_{\mathrm{R}} = \frac{(Y_{ij\mathrm{R}} - \mu_{1r_G})}{v_{r_G}}, \quad \Psi_{\mathrm{C}} = \frac{(Y_{ij\mathrm{C}} - \mu_{2r_G})}{v_{r_G}}$$

and N_0, N_1, and N_{-1} are constants such that the area under the curve f_a is unity. In other words, we suggest that the suitability of the relation given in (17) implies that $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are weakly dependent. An analysis on the aforementioned approximation in given in Appendix D.

Now considering $f_a(Y_{ij\mathrm{R}}, Y_{ij\mathrm{C}})$, we shall find the PDF of G_{ij}. According to [36], the PDF of G_{ij} is given by

$$f(G_{ij}) \approx \int_{-G_{ij}}^{G_{ij}} \frac{G_{ij}}{\sqrt{G_{ij}^2 - Y_{ij\mathrm{C}}^2}} \left[f_a\left(\sqrt{G_{ij}^2 - Y_{ij\mathrm{C}}^2}, Y_{ij\mathrm{C}}\right) + f_a\left(-\sqrt{G_{ij}^2 - Y_{ij\mathrm{C}}^2}, Y_{ij\mathrm{C}}\right)\right] dY_{ij\mathrm{C}} \quad (19)$$

Substituting the expression of f_a in (18), we arrive at:

$$f(G_{ij}) \approx \sum_{r_G} \omega_{r_G} \left[N_1 \times \sqrt{\frac{1+\rho_{r_G}}{1-\rho_{r_G}}} \times \frac{G_{ij}\exp\left(-\frac{G_{ij}^2 + \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}}{2v_{r_G}^2(1+\rho_{r_G})}\right)}{v_{r_G}^2(1+\rho_{r_G})} \times I_0\left(\frac{G_{ij}\sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}}{v_{r_G}^2(1+\rho_{r_G})}\right) + N_{-1} \times \sqrt{\frac{1-\rho_{r_G}}{1+\rho_{r_G}}} \times \frac{G_{ij}\exp\left(-\frac{G_{ij}^2 + \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}}{2v_{r_G}^2(1-\rho_{r_G})}\right)}{v_{r_G}^2(1-\rho_{r_G})} \times I_0\left(\frac{G_{ij}\sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}}{v_{r_G}^2(1-\rho_{r_G})}\right) + N_0 \times \sqrt{1-\rho_{r_G}^2} \times \frac{G_{ij}\exp\left(-\frac{G_{ij}^2 + \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}}{2v_{r_G}^2\left(1-\rho_{r_G}^2\right)}\right)}{v_{r_G}^2(1-\rho_{r_G}^2)} \times I_0\left(\frac{G_{ij}\sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}}{v_{r_G}^2\left(1-\rho_{r_G}^2\right)}\right)\right] \quad (20)$$

where I_0 is the zeroth-order modified Bessel function of the first kind. The random variable G_{ij} represents the magnitude of the gradient present at the (i,j)th position of the image under consideration. As we can see from (20) the PDF of G_{ij} is approximately a weighted sum of Rician [42] PDFs. It is easily evident that this is true even when $Y_{ij\mathrm{R}}$ and $Y_{ij\mathrm{C}}$ are considered mutually independent. Now, consider the case when $\mu_{1r_G} \approx 0$ and $\mu_{2r_G} \approx 0$, that is, the homogeneous areas (A_H) in the image. Then, we have

$$f(G_{ij})_{A_H} \approx \sum_{r_G \in A_H} \omega_{r_G} \left[N_1 \times \sqrt{\frac{1+\rho_{r_G}}{1-\rho_{r_G}}} \times \frac{G_{ij} \exp\left(-\frac{G_{ij}^2}{2v_{r_G}^2(1+\rho_{r_G})}\right)}{v_{r_G}^2(1+\rho_{r_G})} + N_{-1} \times \sqrt{\frac{1-\rho_{r_G}}{1+\rho_{r_G}}} \times \frac{G_{ij} \exp\left(-\frac{G_{ij}^2}{2v_{r_G}^2(1-\rho_{r_G})}\right)}{v_{r_G}^2(1-\rho_{r_G})} + N_0 \times \sqrt{1-\rho_{r_G}^2} \times \frac{G_{ij} \exp\left(-\frac{G_{ij}^2}{2v_{r_G}^2(1-\rho_{r_G}^2)}\right)}{v_{r_G}^2\left(1-\rho_{r_G}^2\right)} \right] \quad (21)$$

Hence, the PDF of G_{ij} representing the magnitude of gradient at locations in the homogeneous areas of the image is a weighted sum of Rayleigh PDFs. Note that the Rayleigh PDF is a special case of the Rician PDF.

The value of v_{r_G} indicates the variation of the overall noise in a region. In an image, this variation is considerably less than the difference in the means of any two regions separated by an edge. Therefore, in the areas of the image containing edges (A_E), we shall have $\sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2} \gg v_{r_G}$. Referring to [43], we find that in such cases, the PDF of G_{ij} is given by a weighted sum of Gaussian PDFs as shown below:

$$f(G_{ij})_{A_E} \approx \sum_{r_G \in A_E} \omega_{r_G} \left[N_1 \times \sqrt{\frac{1+\rho_{r_G}}{1-\rho_{r_G}}} \times \exp\left(-\frac{(G_{ij} - \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2})^2}{2v_{r_G}^2(1+\rho_{r_G})}\right) + N_{-1} \times \sqrt{\frac{1-\rho_{r_G}}{1+\rho_{r_G}}} \times \exp\left(-\frac{\left(G_{ij} - \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}\right)^2}{2v_{r_G}^2(1-\rho_{r_G})}\right) + N_0 \times \sqrt{1-\rho_{r_G}^2} \times \exp\left(-\frac{\left(G_{ij} - \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}\right)^2}{2v_{r_G}^2\left(1-\rho_{r_G}^2\right)}\right) \right] \quad (22)$$

Hence, we see that every Rician PDF in (20) reduces either to a Rayleigh PDF or to a Gaussian PDF in our analysis. Now, the whole image is given by the accumulation of A_H and A_E. Hence, we have $f(G_{ij}) = f(G_{ij})_{A_H} + f(G_{ij})_{A_E}$. Therefore, we get

$$f(G_{ij}) \approx \sum_{r_G \in A_H} \omega_{r_G} \left[N_1 \times \sqrt{\frac{1+\rho_{r_G}}{1-\rho_{r_G}}} \times \frac{G_{ij} \exp\left(-\frac{G_{ij}^2}{2v_{r_G}^2(1+\rho_{r_G})}\right)}{v_{r_G}^2(1+\rho_{r_G})} + N_{-1} \times \sqrt{\frac{1-\rho_{r_G}}{1+\rho_{r_G}}} \times \frac{G_{ij} \exp\left(-\frac{G_{ij}^2}{2v_{r_G}^2(1-\rho_{r_G})}\right)}{v_{r_G}^2(1-\rho_{r_G})} + N_0 \times \sqrt{1-\rho_{r_G}^2} \times \frac{G_{ij} \exp\left(-\frac{G_{ij}^2}{2v_{r_G}^2\left(1-\rho_{r_G}^2\right)}\right)}{v_{r_G}^2\left(1-\rho_{r_G}^2\right)} \right] + \sum_{r_G \in A_E} \omega_{r_G} \left[N_1 \times \sqrt{\frac{1+\rho_{r_G}}{1-\rho_{r_G}}} \times \exp\left(-\frac{(G_{ij} - \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2})^2}{2v_{r_G}^2(1+\rho_{r_G})}\right) + N_{-1} \times \sqrt{\frac{1-\rho_{r_G}}{1+\rho_{r_G}}} \times \exp\left(-\frac{\left(G_{ij} - \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}\right)^2}{2v_{r_G}^2(1-\rho_{r_G})}\right) + N_0 \times \sqrt{1-\rho_{r_G}^2} \times \exp\left(-\frac{\left(G_{ij} - \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2}\right)^2}{2v_{r_G}^2\left(1-\rho_{r_G}^2\right)}\right) \right] \quad (23)$$

As mentioned earlier, the ω_{r_G} values for $r_G \in A_E$ will be much smaller than those for $r_G \in A_H$. Now using Theorem 1, we have $h(G) \approx f(G_{ij})$. That is, the approximate model for a gradient (magnitude) histogram of an image is given by the expression in (23).

The presence of more than one Rayleigh function in (21) and more than one Gaussian function in (22) reflects the various regions and hence the changing nature of an image. Therefore, the existing gradient histogram modeling techniques [3,35,29], which consider that the distribution of the gradient magnitudes due to the "non-meaningful" discontinuities may be represented by a single density function of known type, clearly compromise on the changing nature of images. For example, as mentioned in Section 1, the authors in [35,29] use a Rayleigh density function. It is evident from (21), that it is a specific case of our model and use of a single Rayleigh density function implies that an image consists of predominantly one region having noise and texture which follow a Gaussian model.

3. Threshold determination, post-processing and the hysteresis process

Once we have the histogram of gradient magnitudes $h(G)$ (see (23)), our aim would be to determine a suitable threshold value that separates $f(G_{ij})_{A_H}$ from $f(G_{ij})_{A_E}$. As suggested in [35,44], we may consider the well known Bayesian technique [45], which requires the estimation of the underlying parameters, in order to carry out the threshold determination. However, as our gradient histogram modeling is unique in capturing the spatially changing nature and inter-pixel dependency in images, it results in a complicated expression (see (23)) compared to the existing models [3,35,29]. Therefore, theoretically, parameter estimation would be a non-trivial task and its implementation would require a lot of computation undermining the simplicity and effectiveness of a thresholding technique.

In this section, we propose a novel technique to determine an appropriate threshold from the gradient histogram $h(G)$ that would be used as the upper threshold of the hysteresis process (Sections 1 and 3.4) in order to separate the actual (desired) edges from the "false" edges. The proposed technique is designed based on some theoretical analysis and certain empirical observations. Although the proposed technique is not an optimal one as the Bayesian technique (minimizes the classification error [45]), it does not require the estimation of the underlying parameters. It will be evident from the experimental results reported later in this paper (Section 4) that the use of the proposed gradient histogram model and the proposed threshold determination technique is much preferable than an unsuitable model and the Bayesian thresholding technique. We shall now explain the proposed threshold determination technique from a gradient histogram $h(G)$.

3.1. The region of interest (ROI) in the gradient histogram

As mentioned earlier, in order to obtain the edges in an image, we need a threshold value to distinguish and separate $f(G_{ij})_{A_H}$ ($h(G)_{A_H}$) and $f(G_{ij})_{A_E}$ ($h(G)_{A_E}$) given the $h(G)$ of a particular image. In other words, we need to distinguish between the components (bins indicating the number of occurrences) of the gradient histogram ($h(G)$) given by the expression of the weighted sum of Rayleigh PDFs and the weighted sum of Gaussian PDFs (see (23)).

We leave the problem of distinguishing the components from weighted sum of Rayleigh PDFs and weighted sum of Gaussian PDFs for later and now consider another histogram, say, $h^1(G)$, where we need to distinguish between the components given by just one Rayleigh PDF and one Gaussian PDF. We shall now explore certain properties of Rayleigh and Gaussian PDFs those might be

useful in distinguishing the components given by one from those by the other. The skewness and kurtosis values of the density functions may be used for this purpose, as for a Rayleigh and a Gaussian PDF, skewness and kurtosis are constants which do not vary with the characteristic parameters of the density function. The skewness and kurtosis values are also not affected when the density function is multiplied by a scalar constant. This conveys that in certain cases, the properties such as skewness and kurtosis may be used when the characteristic parameters of Rayleigh and Gaussian PDFs are not known. The expressions of skewness (S) and kurtosis (K) [36] are

$$S = \frac{\Theta_3}{\Theta_2^{3/2}} \tag{24}$$

$$K = \frac{\Theta_4}{\Theta_2^2} \tag{25}$$

where Θ_n stands for nth-order central moment corresponding to the PDF of the random variable under consideration. For a Rayleigh PDF, we have $S \approx .63$ and $K \approx 3.25$ and for a Gaussian PDF, we have $S = 0$ and $K = 3$.

Note that we shall consider the finite range approximations of the histograms $h(G)$ and $h^1(G)$, but retain the same notations. In (23), we have deduced $h(G)$ as a mixture of a weighted sum of Rayleigh and Gaussian PDFs. We have considered $h^1(G)$ as a mixture of a Rayleigh and a Gaussian PDF. It is well known that for a given image the G_{ij} values will have a finite maximum. Hence, we consider that certain finite range approximations of the infinite range PDFs represented by $h(G)$ and $h^1(G)$ are acceptable and the area under an infinite range density function is not appreciably different from the area under the corresponding finite range approximation. For $h^1(G)$, we consider suitable finite range approximations of a Rayleigh and a Gaussian PDF as follows:

- The area under a Rayleigh PDF in the finite range $[0 \;\; 4 \times l_H]$, where l_H is the mode of the density function, is more than 99.9% of the total area under the corresponding infinite range Rayleigh PDF. Hence, in this case, we consider a finite range larger than or equal to $[0 \;\; 4 \times l_H]$ as appropriate.
- The area under a Gaussian PDF in the finite range $[l_E - 3 \times sd \;\; l_E + 3 \times sd]$, where l_E is the mode of the density function and sd represents its standard deviation, is more than 99.7% of the total area under the corresponding infinite range Gaussian density function. Hence, in this case, we consider a finite range larger than or equal to $[l_E - 3 \times sd \;\; l_E + 3 \times sd]$ as appropriate.

As mentioned in Section 1, determination of a single threshold from the gradient histogram has an underlying assumption that the gradient magnitude values representing the "meaningful" discontinuities in an image are in general greater than those representing the "non-meaningful" ones. Therefore, for the gradient histogram $h(G)$ (see (23)), any technique determining a single global threshold considers

$$\min_{\forall r_G \in A_E} \sqrt{\mu_{1r_G}^2 + \mu_{2r_G}^2} > \max\left(\left[\max_{\forall r_G \in A_H} v_{r_G}(1+\rho_{r_G}) \;\; \max_{\forall r_G \in A_H} v_{r_G}(1-\rho_{r_G}) \;\; \max_{\forall r_G \in A_H} v_{r_G}(1-\rho_{r_G}^2)\right]\right) \tag{26}$$

For $h^1(G)$ we shall have $l_E > l_H$. Therefore, in $h^1(G)$, which is a mixture of a Rayleigh and a Gaussian PDF, the Rayleigh PDF contributes the components at the smaller values those the random variables G_{ij} can take, whereas the Gaussian PDFs contribute the components at the larger values (see Fig. 2). Hereafter, we shall represent the values those the random variables G_{ij} (or the random process G) can take by G_v. As can be seen from Fig. 2, there is an overlapping region in the histogram where both the density functions have significant contributions to the components.

In order to get a suitable threshold value, say T^1, from $h^1(G)$ to separate the components due to a Rayleigh PDF and a Gaussian PDF, we calculate the instantaneous skewness ($S(G_v)$) or kurtosis ($K(G_v)$) of $h^1(G)$ progressing along the increasing values of G_v (see Fig. 2). That is, $S(G_v)$ and $K(G_v)$ for $h^1(G)$, respectively, give the skewness and kurtosis values for the data represented by all those bins of $h^1(G)$ where the gradient magnitude value is less than or equal to G_v. The first G_v value at which the instantaneous skewness or kurtosis, respectively, equals 0.63 or 3.25 is considered as the threshold value.

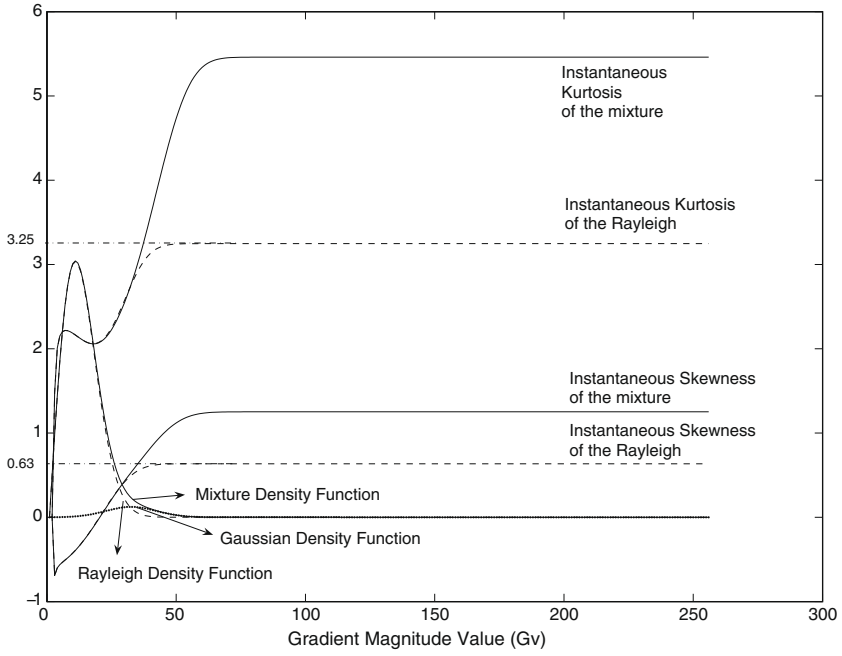


Fig. 2. Separation of components due to a Rayleigh and a Gaussian PDF using instantaneous skewness and kurtosis values of their mixture. (The PDFs shown have been multiplied with a constant to represent them in the given scale.)

A few empirical observations and related analysis regarding the above proposed method of threshold determination for $h^1(G)$ reveals certain important facts about the method. The empirical observations we find interesting are:

1. Let us consider a histogram $h^{ray}(G)$, which is given by only an acceptable finite range Rayleigh PDF and not by any mixture density function. It is empirically observed that for $h^{ray}(G)$ the instantaneous skewness and kurtosis at G_v values less than four times the mode of the Rayleigh PDF is always smaller than 0.63 and 3.25, respectively (see Fig. 2).
2. Rigorous empirical analysis, which was performed considering different parameter values of the underlying PDFs, has also revealed that the instantaneous skewness and kurtosis values corresponding to $h^1(G)$ is never less than those corresponding to $h^{ray}(G)$ for any value G_v (for an example, see Fig. 2).

The above two observations suggest that we may set a bound on the threshold determined from $h^1(G)$ using the proposed approach. The bound may be represented as

$$T^1\begin{cases} \geqslant & max[l_E - 3\times sd \quad 0], & l_E - 3\times sd < 4\times l_H \\ = & 4\times l_H, & l_E - 3\times sd \geqslant 4\times l_H \end{cases} \tag{27}$$

It can be easily deduced that when we have $l_E - 3\times sd \geqslant 4\times l_H$, the threshold obtained using the proposed approach (skewness or kurtosis based) will be same as the threshold value that would have been obtained using the Bayesian thresholding technique [45], which is optimal in terms of minimizing the classification error [45], if the parameters of the density functions were known. Note that when $l_E - 3\times sd < 4\times l_H$, the threshold determination using the skewness and kurtosis based approaches may yield different T^1 values.

We shall now consider the actual problem of distinguishing the components from a weighted sum of Rayleigh PDFs and a weighted sum of Gaussian PDFs in order to determine a suitable threshold, say T, from the gradient histogram $h(G)$. We have shown that the components of $h^1(G)$ corresponding to a Rayleigh and a Gaussian PDF may be distinguished using the skewness and kurtosis values of a Rayleigh PDF. In a similar manner, in order to distinguish the components of $h(G)$ corresponding to a weighted sum of Rayleigh PDFs from the components corresponding to a weighted sum of Gaussian PDFs, we may calculate the instantaneous skewness $(S(G_v))$ or kurtosis $(K(G_v))$ and set the value as threshold where the skewness or kurtosis first equals certain constants. From previous analysis, we infer that these constants should be the skewness and kurtosis values for a weighted sum of Rayleigh PDFs provided that they do not vary with the weights and the characteristic parameters of the Rayleigh PDFs. In order to calculate the skewness and kurtosis for a weighted sum of Rayleigh PDFs, let us consider the following expression of the sum

$$V(\varpi) = \int_0^{\imath} \frac{\varkappa\varpi}{\vartheta^2}\exp\left(-\frac{\varpi^2}{2\vartheta^2}\right)d\vartheta \tag{28}$$

where $\varkappa$ (not a deterministic function of ϑ) represents the weights, which may be different for the various Rayleigh PDFs and $\int_0^{\imath}\varkappa\, d\vartheta = 1$ as the total area under $V(\varpi)$ is unity. The symbol ϑ corresponds to the mode of a Rayleigh PDF. It is to be noted that the expression of the sum in (28) may be used to approximately represent a weighted sum of a large number of Rayleigh PDFs and not that of a few, as (28) suggests that there are infinite Rayleigh PDFs with their modes between 0 and $\imath$. It is evident that calculating the skewness and kurtosis of $V(\varpi)$ is a non-trivial problem. Let us now drop the weights $\varkappa$ in $V(\varpi)$ to get, say, $V^1(\varpi)$ from (28) as given below

$$V^1(\varpi) = \int_0^{\imath} \frac{\varpi}{\vartheta^2}\exp\left(-\frac{\varpi^2}{2\vartheta^2}\right)d\vartheta \tag{29}$$

By carrying out the variable substitution $\frac{\varpi^2}{\vartheta^2} = \varepsilon^2$ in the above expression we have

$$V^1(\varpi) = \int_{\frac{\varpi}{\imath}}^{\infty}\exp\left(-\frac{\varepsilon^2}{2}\right)d\varepsilon = \sqrt{2\pi}\left(\frac{1}{2} - \mathrm{erf}\left(\frac{\varpi}{\imath}\right)\right) \tag{30}$$

We calculate the skewness S (24) and kurtosis K (25) of $V^1(\varpi)$ given by (30) and obtain $S \approx 1.25$ and $K \approx 4.7$. Hence, we see that the skewness and kurtosis values of $V^1(\varpi)$ do not vary with the underlying parameters. Now, it is empirically observed that when the weights $\varkappa$ are reintroduced, that is, $V(\varpi)$ is considered and also when the number of Rayleigh PDFs in the weighted sum is not large, we have $S \lesssim 1.25$ and $K \lesssim 4.7$ (for an example, see Fig. 3).

In the above we have found that contrary to our requirements, the skewness and kurtosis values of an arbitrary weighted sum of Rayleigh PDFs vary with the underlying parameters. For example, the skewness and kurtosis of an arbitrary weighted sum of Rayleigh PDFs, respectively, take the values 0.63 and 3.25 when there is only one Rayleigh PDF in the expression of weighted sum, and take the values 1.25 and 4.7 when there are a large (infinite) number of Rayleigh PDFs in the expression of weighted sum. It is also found that an arbitrary weighted sum of Rayleigh PDFs will have a skewness value between 0.63 and 1.25, and a kurtosis value between 3.25 and 4.7. Therefore, as we do not have particular constants to which the instantaneous skewness $(S(G_v))$ and kurtosis $K(G_v)$ would be compared, we cannot determine a particular threshold T from the gradient histogram $h(G)$. However, as ranges of values are known instead of particular constants, we can obtain a region of interest (ROI) in the gradient histogram $h(G)$ that contains T. We shall discuss this a little later.

Now, similar to the case of $h^1(G)$, we shall put forth some empirical observations and related analysis regarding the use of instantaneous skewness and kurtosis to determine a threshold from the gradient histogram $h(G)$. Some interesting empirical observations are:

1. Let us consider a histogram $h^{sray}(G)$, which is given by an arbitrary weighted sum of Rayleigh PDFs. It is empirically observed that for $h^{sray}(G)$ the instantaneous skewness and kurtosis at G_v values less than four times the largest of all the modes of the various Rayleigh PDFs is always less than 1.25 and 4.7, respectively (see Fig. 4). It is also empirically observed that instantaneous skewness and kurtosis at G_v values greater than four times the largest of all the modes of the various Rayleigh PDFs is always greater than 0.63 and 3.25, respectively (see Fig. 4).
2. Rigorous empirical analysis, which was performed considering different parameter values of the underlying sums of PDFs, has also revealed that the instantaneous skewness and kurtosis values corresponding to $h(G)$ is never less than those corresponding to $h^{sray}(G)$ for any value G_v (For an example, see Fig. 4). It is also observed that the skewness and kurtosis values corresponding to a gradient histogram $h(G)$ is always greater than 1.25 and 4.7, respectively (for an example, see Fig. 4).

The above observations guarantee that we can obtain a ROI given any arbitrary gradient histogram. We have considered 235 different real-life images to test the observation that the skewness and kurtosis values corresponding to an arbitrary gradient histogram should, respectively, be greater than 1.25 and 4.7, and have not found a single image where the observation does not hold. Therefore, as the observation is true for our gradient histogram model in (23), the indication is that our model of gradient histogram in is very much valid for a wide range of images.

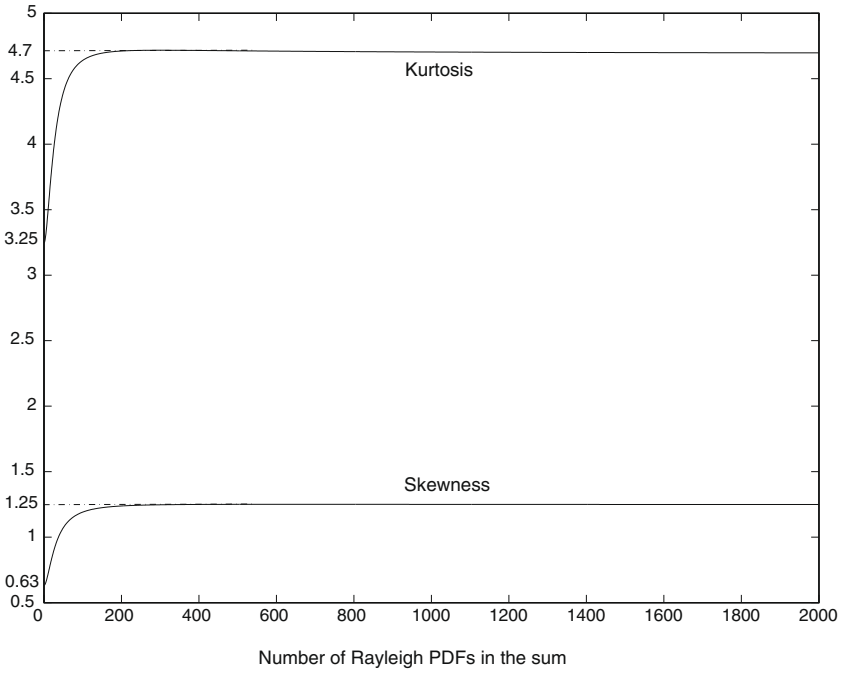


Fig. 3. Empirical calculation of the skewness and kurtosis of an arbitrary weighted sum of Rayleigh PDFs.

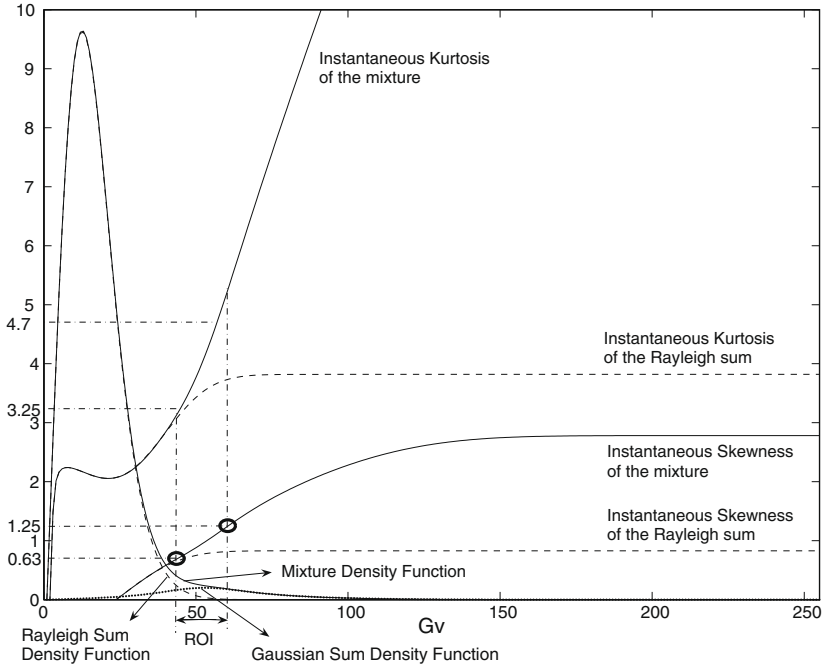


Fig. 4. Separation of a weighted sum of Rayleigh and a weighted sum of Gaussian PDFs using instantaneous skewness and kurtosis values of their mixture and the determination of the ROI. (The PDFs shown have been multiplied with a constant to represent them in the given scale.)

3.1.1. The region of interest (ROI)

Consider the expression in (23). We see that if we come across an image for which all ρ_{r_G} equals 0 and all v_{r_G} are equal or there is predominantly one homogeneous region in the image, then the weighted sum of the Rayleigh PDFs will reduce to a single Rayleigh PDF. Hence, in such a case, we should consider that value of G_v as the threshold where we get $S(G_v) = .63$ or $K(G_v) = 1.25$. On the other hand, in some other image we might have a large number of objects with noise such that all v_{r_G} are different and several ρ_{r_G} do not equal 0. In such a case, the weighted sum of Rayleigh PDFs will contain a large number of Rayleigh PDFs and *hence we should consider that value of* G_v *as the threshold, where* $S(G_v) = 3.25$ *or* $K(G_v) = 4.7$.

Therefore, given a gradient histogram $h(G)$, we may define a ROI within which the threshold value T would lie. This ROI contains all the components of $h(G)$ starting from the G_v when $S(G_v) = 0.63$ or $K(G_v) = 1.25$ until the value of G_v when $S(G_v) = 1.25$ or $K(G_v) = 4.7$. Note that use of skewness or kurtosis may produce

different ROIs. However, if there is no overlapping region in the gradient histogram, then they would produce the same ROI. *In this paper, we shall use both the skewness and kurtosis based approaches, and define the ROI from the smallest G_v to the largest G_v among the four values obtained (see Fig. 4).*

3.2. Selecting a threshold from the region of interest (ROI)

From the analysis presented till now, we have obtained a ROI in the gradient histogram $h(G)$ where the required threshold would lie. In order to find a threshold value T that would be used as the upper threshold of the hysteresis process to eliminate the "false" edges, we now need to consider only the components in the ROI and not the overall gradient histogram $h(G)$.

After the ROI in the gradient histogram is obtained, we suggest that any conventional histogram thresholding algorithm may be used on the components in the ROI in order to determine the threshold T. In our experiment (see Section 4), we consider the two histogram thresholding algorithms, namely, Otsu's method [30] and beam theory based method [34], as examples, to be used on the components in the ROI. It will be evident from the experimental results reported later in Section 4 that although the conventional histogram thresholding algorithms are not suited for a gradient histogram, they can readily be used without compromising on performance when restricted to only the components within ROI.

3.3. Non-maximum suppression (NMS)

Canny [11,25] pointed out that there should only be one response of an edge detection system to a single edge in the image and to ensure this, he presented a technique referred to as the non-maximum suppression (NMS) operation to be applied as a post-processing operation on the gradient magnitude image. The NMS operation considers the fact that an edge at a pixel is legitimate only when the gradient magnitude at that pixel assumes a maximum in the gradient direction within a local surrounding.

We calculate the gradient direction at each location in the image under consideration using the following expression:

$$Z_{ij} = \tan^{-1}\left(\frac{Y_{ij\mathrm{R}}}{Y_{ij\mathrm{C}}}\right) \tag{31}$$

For simplicity, the values of the directions obtained are then approximated $\left(Z_{ij}^{a}\right)$ to the closest among the following set, [0° 45° 90° 135°]. We then retain only those G_{ij} which are greater than the other gradient values in a 3×3 local surrounding and in the corresponding gradient directions Z_{ij}^{a}. We may represent the NMS operation explained above as

$$G_{ij}^{NMS} = \begin{cases} G_{ij} & G_{ij} \geqslant \kappa_{ij} \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

where G^{NMS} represents the image obtained at the output of the NMS operation. Note that κ_{ij} is a random variable. Now, it can be easily deduced that the PDF of G_{ij}^{NMS} can be represented as

$$f(G_{ij}^{NMS}) \simeq \varrho_{ij}\delta(G_{ij}) + u(G_{ij} - \kappa_{ij})f(G_{ij}) \tag{33}$$

where ϱ_{ij} is a weight which is dependent on the location (i,j). The symbols δ and u stand for unit impulse and step functions, respectively. It is evident from (33), that the image G^{NMS} does not have identical first-order densities. In such a case, the determination of the histogram $h(G^{NMS})$ becomes a non-trivial task as Theorem 1 is not applicable. However, as can be seen from (33), the splitting of $f(G_{ij}^{NMS})$ into parts implies the splitting of $f(G_{ij})$, as other parts in the expression cannot be split. Hence, we may use the threshold T determined from the gradient histogram $h(G)$ $(f(G_{ij}))$ on the post-processed gradient image G^{NMS}.

3.4. Hysteresis thresholding

Hysteresis thresholding is an algorithm proposed by Canny [11] to be applied on the post-processed gradient image G^{NMS} in order to mark the edges of the underlying image. This algorithm is designed to tackle the problem of edge streaking [11], which is a very common one when a single threshold is used for the thresholding operation on G^{NMS}.

The hysteresis thresholding algorithm employs two different thresholds called the upper threshold T_u and the lower threshold T_l. The following statements below describe the process.

(1) If the gradient magnitude value at any pixel is above T_u, that pixel is immediately marked as a part of an edge.
(2) If the gradient magnitude value at any pixel is below T_l, that pixel is immediately rejected to be a part of an edge.
(3) If the gradient magnitude value at any pixel is above T_l and below T_u, that pixel is considered to be a part of an edge only when it is connected (lies within 8-neighborhood) to a pixel already marked as a part of an edge. This step is repeated until no new pixel is marked as a part of an edge.

As mentioned earlier, we consider $T_u = T$, where T is the threshold determined from the ROI in a gradient histogram and $T_l = \frac{1}{\mathsf{F}} \times T_u$ with $\mathsf{F} > 1$. As mentioned in Section 1, in order to determine F one needs to know in advance the number of edge pixels in the image under consideration [29]. As knowing the number of edge pixels accurately in advance is not possible, the value of F determined will be as appropriate as one's guess. To the best knowledge to authors, there exist no other work on the determination of the ratio F in related literature. However, it is empirically observed that, as suggested in [11], a value of F between 2 and 3 is quite acceptable. As our prime focus is on the determination the upper threshold T_u from a gradient histogram, we fix $\mathsf{F} = 2.5$ and do not change it over various images or gradient operators.

4. Experimental results and discussion

In this section, we provide experimental results using a few real-life and benchmark images in order to demonstrate the superiority of the proposed method over the others. As mentioned in Section 3.2, we have considered, for example, the following two conventional thresholding algorithm in order to get the threshold from a ROI determined using the proposed methodology.

(1) Maximization of separability (Otsu's method) [30]This method presented by Otsu is a non-parametric technique maximizing the within-region homogeneity and minimizing the between-region homogeneity to get the threshold.
(2) Beam theory and minimization of ambiguity [34]This method given by Sen and Pal uses the beam theory from structural mechanics to carry out a histogram modification process and then minimization of the ambiguity in the information given by the modified histogram to get the threshold.

Among the few existing techniques to perform gradient histogram thresholding, we consider the ones proposed in [3,29,10] for comparison. We also consider the technique given in [4], proposed to find a threshold from an unimodal histogram and the histogram thresholding techniques proposed in [34,30].

To the best knowledge of the authors, no globally accepted unsupervised (not requiring ground truth) objective measure is

available for quantitative evaluation of edge detection performance. Moreover, unsupervised objective performance measures are often misleading especially when used with the underlying images having complex scenes [1,46]. Edge detection evaluation based on the ground truth of synthetically generated images could also be misleading or biased, as every synthetically generated image is obtained by obeying certain rules and definitions, and often contain only simple geometric patterns [1,46]. The authors in [1] suggested that only visual method and then statistical analysis be used to judge the performance of an edge detector. Later, the authors in [46,47], respectively, used receiver operating characteristics (ROC) and precision-recall characteristics (PRC) from the pattern classification literature [45], and human labeled ground truth in numerous real-life images for edge detection evaluation. The evaluation methods in [46,47], which strive to imitate human observers, represent the state-of-the-art in edge detection performance evaluation.

In this paper, we consider qualitative evaluation and human labeled ground-truth based quantitative evaluation of performance. Note that, our goal in this section is to evaluate the performance of the proposed and existing gradient histogram thresholding methodologies, and not that of an entire edge detection system. We use the various gradient histogram thresholding methodologies to determine threshold values from the gradient histogram, and as mentioned earlier, use the thresholds obtained as the upper threshold values in the hysteresis process applied to the post-processed gradient image in order to get the edges. In the process of evaluation of gradient histogram thresholding methodologies, we choose a particular gradient operator and assign specific values to various other parameters of a gradient based edge detection system.

In most of the results we present in this section, we shall use the linear gradient operator given by Canny [11]. As discussed in Section 2.1, the width of the operator $((2k+1)\times(2k+1))$ is an important parameter. For the Canny's gradient operator, we have

$$H_{pq\mathrm{R}} = \frac{q}{\chi}\exp\left(-\frac{p^2+q^2}{2\Sigma^2}\right) \tag{34}$$

$$H_{pq\mathrm{C}} = \frac{p}{\chi}\exp\left(-\frac{p^2+q^2}{2\Sigma^2}\right) \tag{35}$$

Hence, the operator is the negative of the first derivative of the Gaussian function, Σ gives its spread and χ is a constant. We take $k = 3\Sigma$. Now in order to determine Σ, let us consider the application of the Canny's operator in the vertical direction. Using (34) in (8), we have

$$Y_{ij\mathrm{R}} = \frac{\Sigma}{\chi}\sum_P\sum_Q P\exp\left(-\frac{P^2+Q^2}{2}\right)I_r + \frac{\Sigma}{\chi}\sum_P\sum_Q P \times \exp\left(-\frac{P^2+Q^2}{2}\right)\sigma_r \times V_{i+P\Sigma\ j+Q\Sigma} \tag{36}$$

where $P = p/\Sigma, Q = q/\Sigma$. When $k = 3\Sigma$, both P and Q take values from -3 to 3. From (36), we see that a large Σ would result in good suppression of the overall noise, but the localization of the edges would be poor and a small Σ would give the opposite results. As suggested in [11], we use an estimated value of standard deviation of the overall noise in the image, σ, to get the value of Σ using the expression $\Sigma = \alpha\sigma$. We empirically find that it is appropriate for α to have a value in the range [10 20], when σ is normalized by the maximum grayscale value of 255. The value of σ is obtained as the average of local (5×5) standard deviation measures of the image. We take $\alpha = 15$ and do not change it over various images. Note that similar analysis can also be shown for the application of the Canny's operator in the horizontal direction.

We shall now consider the qualitative evaluation of the performance of the various gradient histogram thresholding methods using a few real-life images. Fig. 5 shows the edges obtained using gradient based edge detection systems containing Canny's gradient operator, the NMS operation, the various gradient histogram thresholding techniques and the hysteresis process. The image considered was captured using a non-coherent imaging system and hence we represent it by a mixed Gaussian random process. We see that the use of the proposed technique (Fig. 5e and f) of finding an ROI and then determining the upper threshold of the hysteresis process considering the components within the ROI results in the extraction of the appropriate edges and satisfactory elimination of the unwanted due to inherent texture and noise. It is evident from the figure that the results obtained when the other thresholding techniques are used to get the upper threshold, are totally unacceptable, except when the methods proposed in [4,29] are used. Note that the gradient histogram model assumed in [29] is a specific case of the model that we have deduced earlier in this paper. It is also interesting to note in Fig. 5 that when the method in [4] that does not consider any specific gradient histogram model is used, satisfactory result is obtained. On the contrary, the use of many other existing gradient histogram modeling based thresholding techniques results in unsatisfactory edge detection performance. This shows the significance of considering a proper model for an image. Fig. 5c and d, respectively, show the edges obtained when the smallest and the largest gradient magnitude values within the ROI are considered as the upper threshold of the hysteresis process. As can be seen, the results in Fig. 5c and d are fairly satisfactory and this demonstrates the effectiveness of the proposed method of determining the ROI.

An image obtained using a synthetic aperture radar system, which is a coherent imaging system, is considered in Fig. 6. Hence, the image is represented by a mixed lognormal random process. Therefore, the Canny's gradient operator is used on an image obtained applying the natural logarithmic transformation on the image shown in Fig. 6a. The gradient histogram shown in Fig. 6b, which is considered for the determination of the threshold to be used as the upper threshold in the hysteresis process, is that of the transformed image. We see that similar to the case in Fig. 5, the edge detection system using the technique proposed by [3] fails to remove the "false" edges due to texture and noise. Although edge detection systems using the other existing thresholding techniques give satisfactory results, the ones using the proposed thresholding methodology (Fig. 6e and f) outperform all the others in reducing "false" edges and also in extracting the appropriate edges. Similar to the case in Fig. 5, the fairly satisfactory results (Fig. 6c and d) obtained considering the smallest and the largest gradient magnitude values within the ROI as the upper threshold of the hysteresis process demonstrate the effectiveness of the proposed method of determining the ROI.

The consistency of a fully automatic thresholding technique in providing acceptable results is extremely important. In Fig. 7, we have considered six images of widely different types and applied a gradient based edge detection system, which uses our methodology of threshold determination, on them. From visual inspection we find that the edge detection results obtained in all of them are highly satisfactory. The wideness in the types of the images considered is indicated using their homogeneity measures (angular second moment [48]) given below the images, with a higher value indicating larger homogeneity.

In the above qualitative evaluation of performance, we have demonstrated the superiority of the proposed gradient histogram thresholding methodology over the other existing ones and also shown the consistency of the proposed methodology. Now we shall substantiate our claims of superiority and consistency of the proposed methodology with the help of human labeled ground-truth

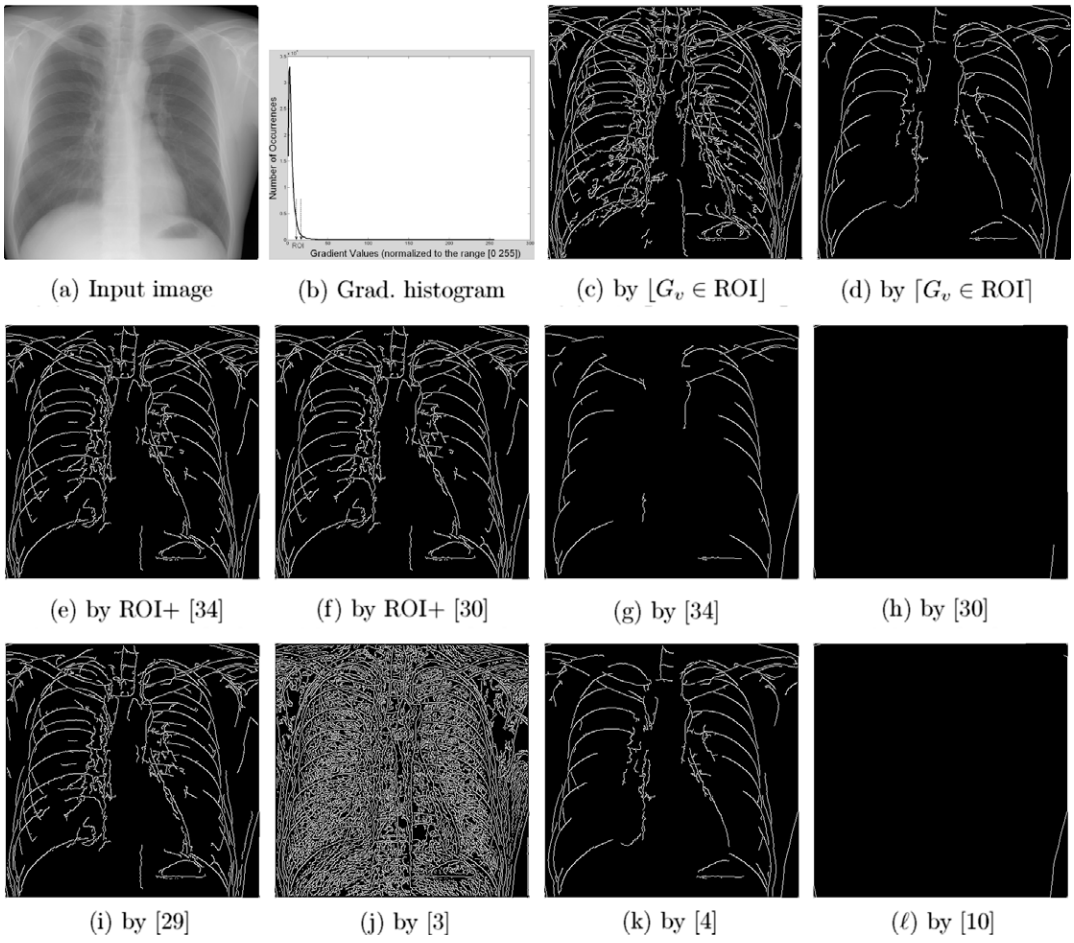


Fig. 5. Edges in a non-coherent X-ray image obtained using Canny's gradient operator, the various gradient histogram thresholding techniques, the NMS operation and the hysteresis process. [*Note:* (e and f) → proposed.]

based quantitative evaluation of performance. Let us consider the ROC plot, which is a plot of percentage unmatched ground truth pixels versus percentage false positive edge pixels, and the PRC plot, which is a plot of the fraction of true positives that are detected rather than missed (recall) versus the fraction of detections that are true positives rather than false positives (precision), for the quantitative analysis. The dataset of benchmark images considered are the 60 images used in [46] (available at http://figment.csee.usf.edu/edge/roc/). Among the 60 images, 50 represent the general domain of generic object recognition from grayscale images, and 10 represent the domain of aerial image analysis [46]. The ground truth for these 60 images has been created manually, where each pixel in an image has been labeled to be in one of three classes, namely, edge, non-edge and "don't care".

Fig. 8 shows the ROC and PRC plots for the edges obtained using gradient based edge detection systems containing Canny's gradient operator, the NMS operation, various gradient histogram thresholding techniques and the hysteresis process on the 60 benchmark images. The gradient histogram thresholding methods considered for comparison in Fig. 8 are the proposed technique of applying the method in [30] to the components within ROI (circles in black), the techniques in [29] (circles in magenta), [4] (circles in blue), [3] (circles in green) and [10] (circles in red). Note that, each value marked by a colored circle in the plots correspond to the edges obtained in an image by the edge detection system that uses any one of the various gradient histogram thresholding technique. The 5 squares in relevant colors represent the two dimensional arithmetic mean (values given in the figure) of the values (colored circle) corresponding to the edge detection system using the five gradient histogram thresholding techniques considered. Note that, if a result obtained by an edge detection system is free from any error with respect to the underlying ground truth, it will produce a value at (0,0) of the ROC plot and at (1,1) of the PRC plot. Hereafter, we shall refer the points (0,0) of the ROC plot and (1,1) of the PRC plot as the ideal. It is evident from the plots in the figure that the means corresponding to the edge detection system using the proposed methodology are the nearest to the ideal in terms of Euclidean distance. It is also evident that the values corresponding to the edge detection systems using the methods in [29,4] and the proposed methodology are distributed near (in terms of Euclidean distance) to the ideal in the plots. Therefore, we may say that the use of the proposed methodology and the methods in [29,4] consistently result in appreciable performance over the 60 images considered with the edge detection system using the proposed gradient histogram thresholding methodology doing slightly better than the other two. The 60 images, corresponding ground truths and the various edge detection results obtained for the 60 images can be found at http://dsen.cscr.isi.googlepages.com/dsenskpivcres for qualitative evaluation of performance.

We shall now take the help of statistical hypothesis testing in order to analyze the observed superiority of the edge detection system using the proposed methodology (ROI + [30]) over the other existing ones. One way of judging the goodness of the values obtained in the ROC and PRC plots could be their Euclidean distance from the (0,0) and (1,1) points, respectively. Let us assume that the

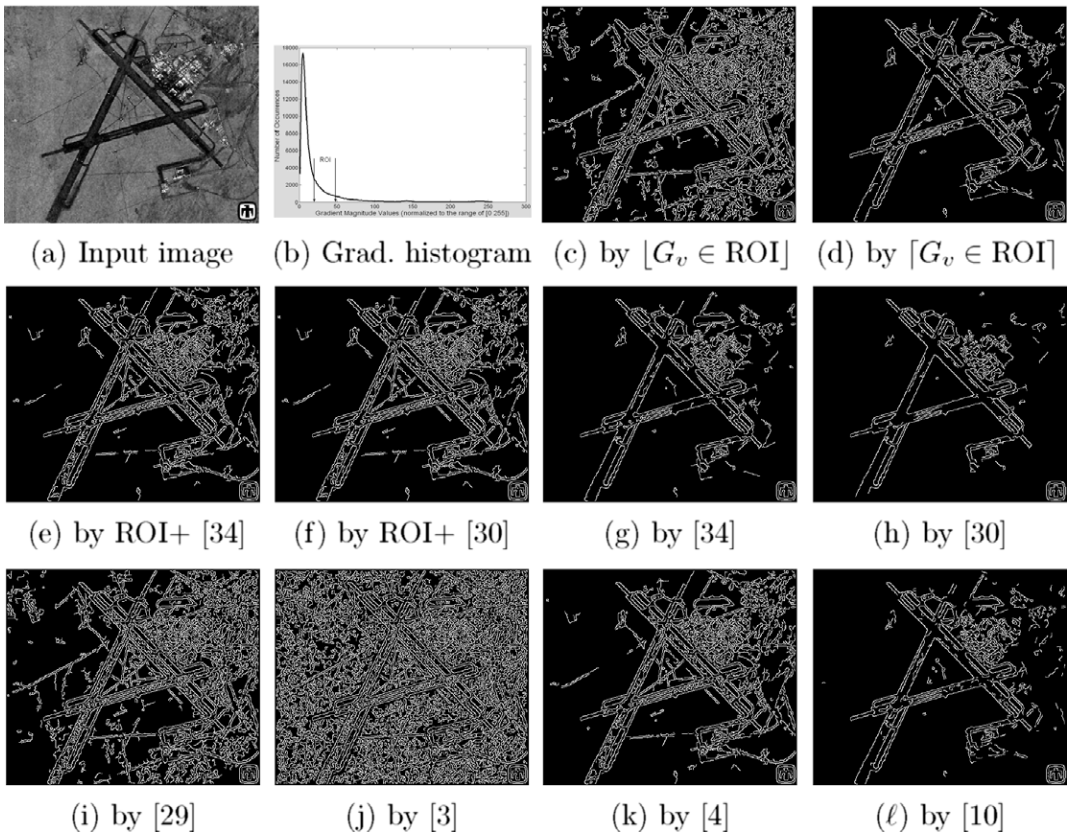


Fig. 6. Edges in a coherent synthetic aperture radar image obtained Canny's gradient operator, the various gradient histogram thresholding techniques, the NMS operation and the hysteresis process. [*Note:* (e and f) → proposed.]

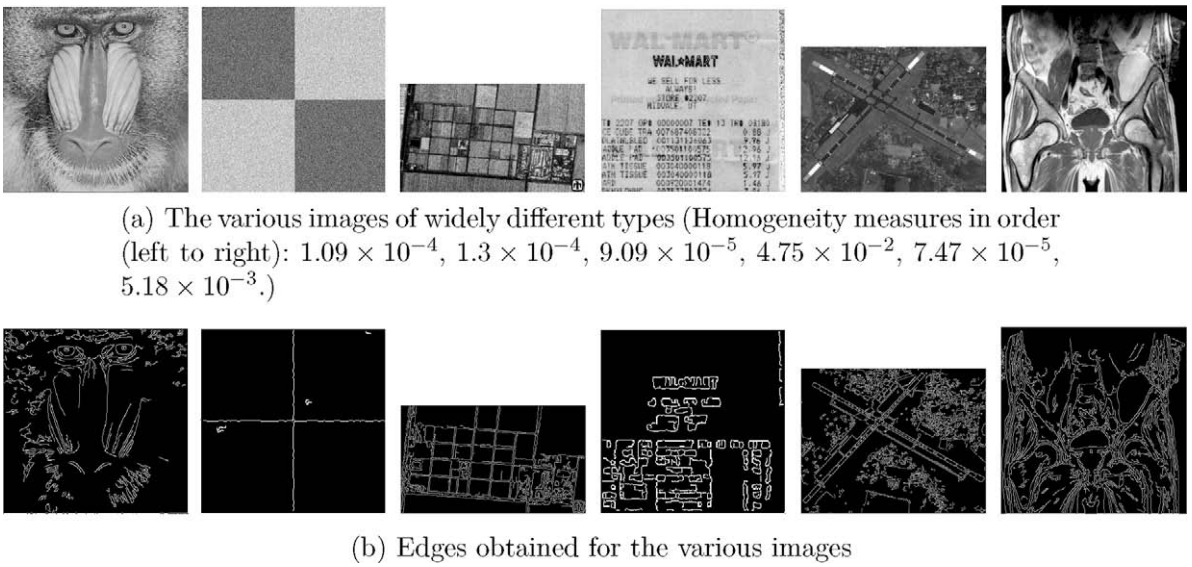
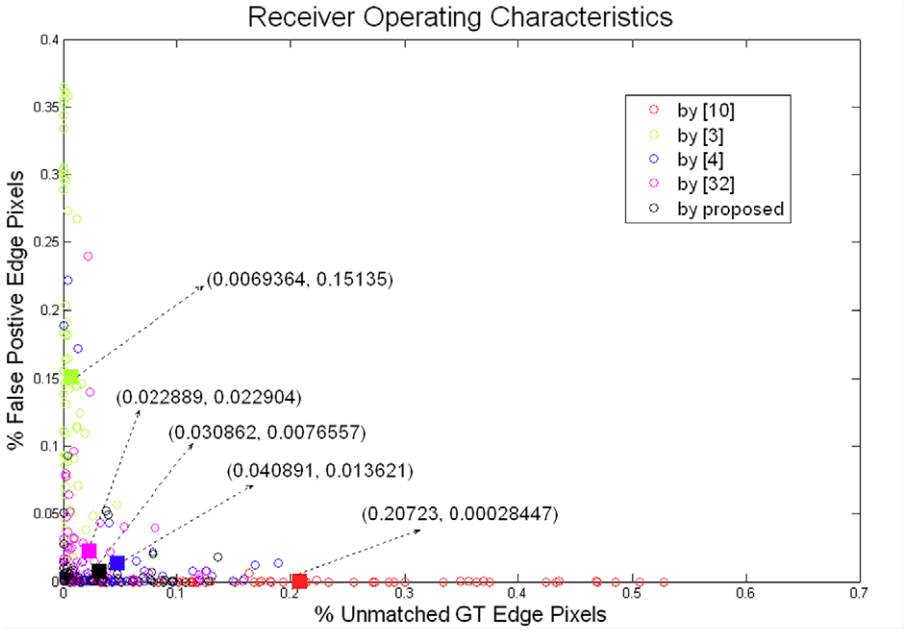
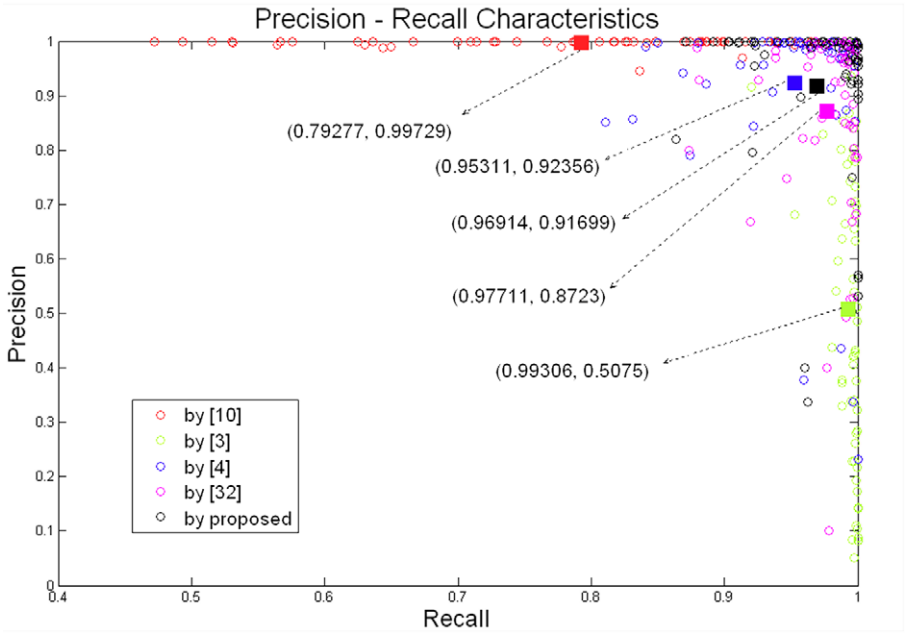


Fig. 7. Edges obtained in various images using Canny's gradient operator based edge detection system that contains the proposed thresholding technique of determining an ROI and then using [34].

Euclidean distance of a value corresponding to an edge detection system using any one gradient histogram thresholding technique from the ideal is random in nature and taken from a normally distributed population. We shall now use the *t*-test [49] for comparing the performance of the system using the proposed method with a system using an existing method at a time. Note that, the *t*-test, which we shall use, considers that the samples (Euclidean distances) corresponding to the two systems to be compared have come from normal distributions with unknown and possibly unequal variances (Behrens–Fisher problem [49]). Let us consider the alternative hypothesis (H_a) that 'the average Euclidean distance of the values obtained using a system with an existing thresholding method from the ideal is greater than the average Euclidean distance of the values obtained using the system with



(a) ROC plot



(b) PRC plot

Fig. 8. ROC and PRC plots containing values corresponding to the edges obtained using Canny's gradient operator, the various gradient histogram thresholding techniques, the NMS operation and the hysteresis process.

the proposed thresholding method from the ideal'. Table 1 gives the *p*-values [49] obtained using one-sided (right-tailed) *t*-test considering the alternative hypothesis H_a. In the table, a *p*-value gives the probability that a superior performance obtained, when the proposed method is used, would have been due to chance alone. Therefore, in Table 1, smaller the *p*-value compared to 0.5, more better is the system with the proposed method compared to the system with the considered existing method. As can be seen from the table, when we consider the *p*-values corresponding to ROC, the system using the proposed method is way ahead of those

Table 1
p-value based comparison of the proposed gradient histogram thresholding methodology with the other existing ones.

t-test results Proposed compared to ↓	ROC *p*-value	PRC *p*-value
[29]	0.27254	0.14382
[4]	0.008125	0.47289
[3]	1.25×10^{-11}	4.23×10^{-16}
[10]	1.25×10^{-13}	1.1×10^{-4}

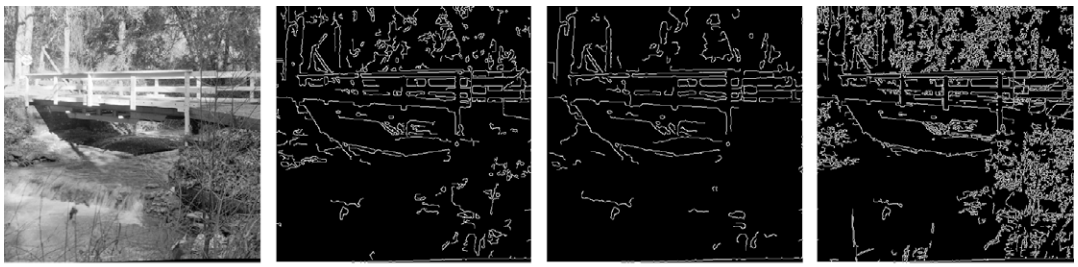


Fig. 9. Edges obtained when different linear gradient operators are used along with the proposed thresholding technique of determining an ROI + [30].

using the methods in [4,3,10]. When the system using the proposed method is compared to the one using the method in [29], we see that the probability of obtaining a superior result by chance alone, when the proposed method is used, is only about 0.27. Hence, we may say that the system using the proposed method is best among the ones considered for comparison in terms of ROC. It can also be seen from the table that when we consider the p-values corresponding to PRC, the system using the proposed method is ahead of those using the methods in [29,3,10]. However, when the system using the proposed method is compared to the one using the method in [4], we see that the probability of obtaining a superior result by chance alone, when the proposed method is used, is about 0.47. Therefore, we may say that the system using the proposed method is better than the ones considered for comparison in terms of PRC, except for the system using the method in [4] where the results obtained could be equally good.

Fig. 9 demonstrates the usage of the proposed thresholding technique with different linear gradient operators, namely, Canny's step edge detecting operator [11] (width determined using overall noise standard deviation σ), Petrou's ramp edge detecting operator [14] (width: 6×6) and Sobel's operator [50] (width: 3×3). Note that Sobel's operator is not an optimal operator unlike the other two. It is evident from the figure that although the design and usage of an appropriate linear gradient operator is crucial for edge detection, the proposed thresholding technique could be used with any linear gradient operator.

As mentioned in Section 2.1, integration of information from gradient operators working at different scales has been suggested in literature to mark the appropriate edges in an image. In such systems, it is required that the gradient histogram thresholding process is applicable at the various operator scales. To demonstrate such a feasibility of our thresholding methodology, in Fig. 10, we have considered the Canny's gradient operator with different widths, essentially implying different scales ($k \propto \Sigma$) and it is evident that appreciable results are obtained in all. Note that the proposed methodology is based on Theorem 1, where the number of pixels in the image $n \gg \zeta$. It is understood that the value of ζ increases linearly with the increase in the operator scale (width). Hence, we conclude that the proposed methodology of threshold determination is suitable for edge detection systems using gradient operators working at different scales as long as $n \gg \zeta$.

5. Conclusion

Detecting the edges in an image using gradient value based methodologies is heavily dependent on a decision process that uses threshold values determined from the gradient histogram of the image. Compared to the designing of optimal gradient operators and methods to thin the edges to single pixel width, the specific problem of threshold determination from a gradient histogram has received very less attention in the edge detection community. This ignorance could be attributed to the fact that there exist numerous automatic histogram thresholding techniques those may be applied. However, it is well known that such general techniques are not reliable as they are not designed considering any property of a gradient histogram.

The novel methodology described in this article, which determines a threshold from a gradient histogram to be used as the upper threshold of the hysteresis thresholding process, has been systematically obtained based on rigorous theoretical analysis and empirical observations of certain aspects about an image. Theories of random process and random inputs to a system have been considered to deduce a general model for a gradient histogram of an image. Certain properties of a probability density function have been used to determine a region of interest (ROI) in the gradient histogram, from which the threshold value has been obtained using standard histogram thresholding techniques. Experiments have been carried out using different real-life and benchmark images, and different gradient operators. It has been observed, both qualitatively and quantitatively, that the use of the proposed methodology of determining an ROI in a gradient histogram consistently results in appreciable edge detection performance and hence the proposed methodology of threshold determination could be considered to be robust over the type of image. Moreover, the proposed methodology has been seen to score over the few existing techniques of the same paradigm and some standard histogram

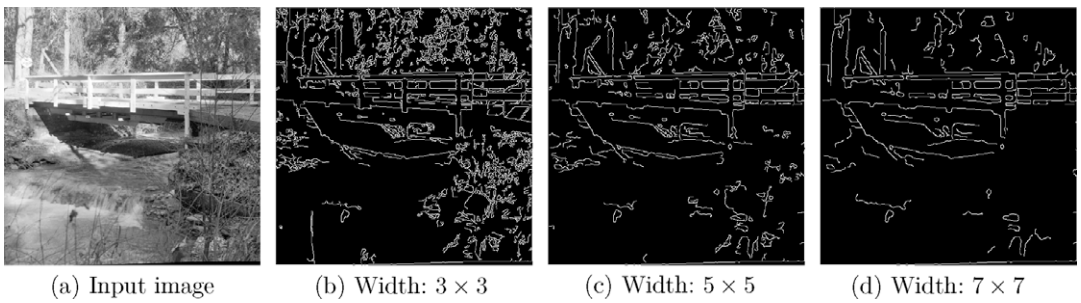


Fig. 10. Edges obtained when Canny's gradient operator with different widths are used along with the proposed thresholding technique of determining an ROI + [30].

thresholding methods in simultaneous extraction of the appropriate edges from images and removal of the unwanted due to inherent texture and noise.

Acknowledgements

D. Sen thanks Mr. Suman Saha for his comments and suggestions. S.K. Pal thanks the Government of India for the J.C. Bose National Fellowship.

Appendix A. Proof of Theorem 1

Let us first consider Υ as any arbitrary random process. Let $p(\Phi)$ represent the probability that the random variable Φ takes a particular value, say ϕ. Therefore, we have

$$f(\Phi) \Rightarrow p(\Phi = \phi)\ \forall\ \phi \tag{A.1}$$

That is, $f(\Phi)$ is a curve representing the $p(\Phi)$ values for all ϕ and the area under the curve is 1. Now, for $n = 2$ we have $A = \{a_1, a_2\}$ and we may obtain $p(\beta_2)$ as

$$p(\beta_2) = \frac{1}{2}\left(\frac{1}{2}\times p(a_1) + \frac{1}{2}\times p(a_2/a_1)\right) + \frac{1}{2}\left(\frac{1}{2}\times p(a_1/a_2) + \frac{1}{2}\times p(a_2)\right) \tag{A.2}$$

When Υ has identical first-order densities we get $f(a_1) \simeq f(a_2)$ and hence all $p(a_1) = p(a_2)$. Let us now consider the following two cases.

Case 1: The samples or random variables a_1 and a_2 are mutually independent. In such a case, (A.2) reduces to

$$p(\beta_2) = \frac{1}{2}\times p(a_1) + \frac{1}{2}\times p(a_2) = p(a_1) = p(a_2) \tag{A.3}$$

Case 2: The samples or random variables a_1 and a_2 are dependent, such that, either $p(a_1/a_2) = p(a_2/a_1) = 1$ or $p(a_1/a_2) = p(a_2/a_1) = 0$. When $p(a_1/a_2) = p(a_2/a_1) = 1$ (A.2) reduces to

$$p(\beta_2) = \frac{1}{2}\left(\frac{1}{2}\times p(a_1) + \frac{1}{2}\right) + \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2}\times p(a_2)\right) = \frac{1}{2}p(a_1) + \frac{1}{2} = \frac{1}{2}p(a_2) + \frac{1}{2} \tag{A.4}$$

and when $p(a_1/a_2) = p(a_2/a_1) = 0$, (A.2) reduces to

$$p(\beta_2) = \frac{1}{2}\left(\frac{1}{2}\times p(a_1)\right) + \frac{1}{2}\left(\frac{1}{2}\times p(a_2)\right) = \frac{1}{2}p(a_1) = \frac{1}{2}p(a_2) \tag{A.5}$$

From careful analysis of Cases 1 and 2, we find that in general the expression for $p(\beta_2)$ can be written as

$$p(\beta_2) = \frac{1}{2}p(a_1) + \frac{1}{2}\times \acute{C} = \frac{1}{2}p(a_2) + \frac{1}{2}\times \acute{C} \tag{A.6}$$

where

$$\acute{C} = \frac{1}{2}p(a_1/a_2) + \frac{1}{2}p(a_2/a_1) \tag{A.7}$$

and hence $0 \leqslant \acute{C} \leqslant 1$. Note that the value of $\acute{C}$ varies with the value (event) of β_2 under consideration. Let us now consider $A = \{a_t, \forall\ t\}$, where $t = 1, 2, \ldots, n$. In this case, we have the union of $n!$ different mutually exclusive combinations in the expression of $p(\beta_n)$. Thus, we have

$$p(\beta_n) = \frac{1}{n!}\left(\frac{1}{n}\times p(a_1) + \frac{1}{n}\times p(a_2/a_1) + \frac{1}{n}\times p(a_3/(a_1 \cap a_2)) + \cdots + \frac{1}{n}\times p(a_n/(\cap_1^{n-1} a_t))\right) + \cdots \text{ other } (n! - 1) \text{ different combinations} \tag{A.8}$$

When we consider Υ as a ζ-dependent random process, we have

$$p(\beta_n) = \frac{1}{n!}\left(\frac{1}{n}\times p(a_1) + \frac{1}{n}\times p(a_2/a_1) + \frac{1}{n}\times p(a_3/(a_1 \cap a_2)) + \cdots + \frac{1}{n}\times p(a_{\zeta+1}/(\cap_1^{\zeta} a_t)) + \frac{1}{n}\times p(a_{\zeta+2}/(\cap_2^{\zeta+1} a_t)) + \frac{1}{n}\times p(a_{\zeta+3}/(\cap_3^{\zeta+2} a_t)) + \cdots + \frac{1}{n}\times p(a_{2\zeta+2}/(\cap_{\zeta+2}^{2\zeta+1} a_t)) + \cdots + \frac{1}{n}\times p(a_n/(\cap_{n-\zeta}^{n-1} a_t))\right) + \cdots \text{ other } (n! - 1) \text{ different combinations} \tag{A.9}$$

We observe in (A.9), that a random variable, say, $a_{\zeta+2}$ depends only on the random variables a_2 to $a_{\zeta+1}$, that is, the previous ζ samples. In a similar manner, as shown graphically in Fig. 11, we infer that there are ζ samples in A those are dependent on any a_t and $n - \zeta$ samples those are independent of that a_t. Therefore, referring to (A.2)–(A.6), we obtain the expression of $p(\beta_n)$ as

$$p(\beta_n) = \frac{1}{n!}\left(\frac{n-\zeta}{n}p(a_t) + \frac{\zeta}{n}\times C_1\right) + \frac{1}{n!}\left(\frac{n-\zeta}{n}p(a_t) + \frac{\zeta}{n}\times C_2\right) + \cdots + \frac{1}{n!}\left(\frac{n-\zeta}{n}p(a_t) + \frac{\zeta}{n}\times C_{n!}\right)$$

where we have $p(a_1) = p(a_2) = \cdots = p(a_n) = p(a_t)$, considering that Υ has identical first-order densities and $0 \leqslant C_1, C_2, \ldots, C_{n!} \leqslant 1$. Simplifying the expression in (A.10) we have

$$p(\beta_n) = \frac{n-\zeta}{n}p(a_t) + \frac{\zeta}{n}\times C \tag{A.11}$$

where $n! \times C = C_1 + C_2 + \cdots + C_{n!}$ and hence $0 \leqslant C \leqslant 1$. Note that the value of C varies with the value (event) of β_n under consideration. From (A.11), we have

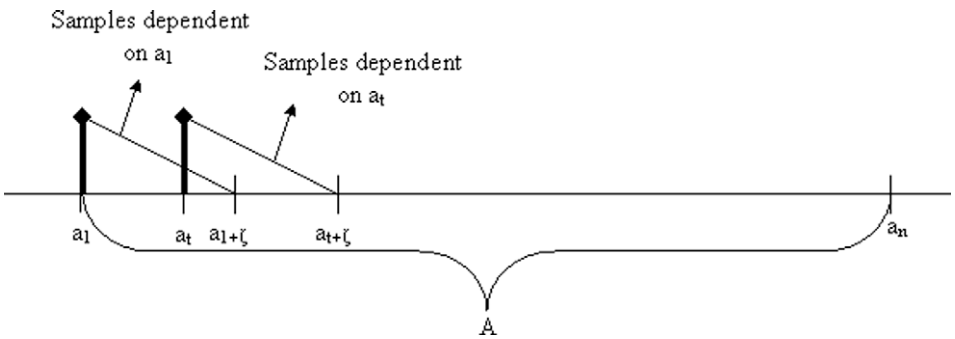
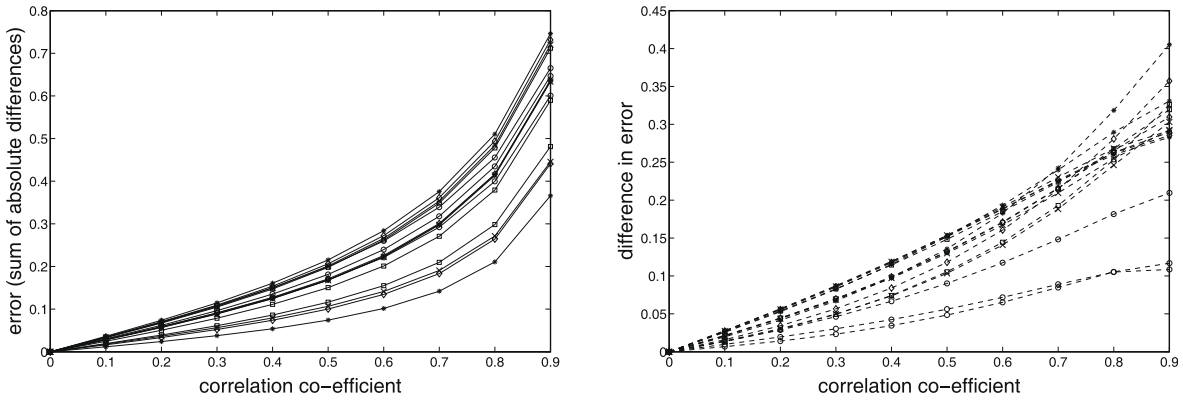


Fig. 11. Dependence and independence of various samples in A on any $a_t \in A$.



(a) Plots of Ω versus ρ_1 (b) Plots of $(\Omega_i - \Omega_a)$ versus ρ_1

Fig. 12. The error of the proposed approximation of joint Gaussian PDF and its supremacy over the independence assumption.

$$p(\beta_n) = p(a_t), \text{ as } n \to \infty \tag{A.12}$$

From (A.1), we find that (A.12) implies

$$f(\beta_n) \simeq f(a_t), \text{ as } n \to \infty \tag{A.13}$$

Appendix B. Deductions of Corollaries 1–3

Corollary 1. *This corollary is evident from (A.11), as $\zeta = 0$ when Υ is an i.i.d. random process.*

Corollaries 2 and 3:

These corollaries can be easily deduced using the following two points:

1. From Definition 1, all finite-dimensional joint probability density function of the various random variables of a Gaussian, lognormal, mixed Gaussian and mixed lognormal random processes are multivariate Gaussian, lognormal, mixed Gaussian and mixed lognormal functions, respectively.
2. As given in [36], jointly Gaussian, lognormal, mixed Gaussian and mixed lognormal random variables are always marginally Gaussian, lognormal, mixed Gaussian and mixed lognormal distributed, respectively. Note that in [36], the above property has been shown only for jointly Gaussian random variables. The extensions to the other mentioned random variables are trivial.

Appendix C. Proofs of Propositions 1 and 2

Proposition 1. *This proposition can be easily deduced from the fact that if two random variables are mutually dependent then their functions involving only that random variable shall also be mutually dependent [36]. As each individual sample of the output process from a memoryless system can be considered as a function of each individual sample of the input process, the output process would be ζ-dependent with the same value of ζ when the input process is ζ-dependent.*

Proposition 2. *As given in [36], a WSS random process to a linear system produces another WSS random process. We know that a WSS ζ-dependent process will have an autocorrelation function (ACF) which is zero outside the time interval ζ [36]. Let the ACF of a WSS ζ-dependent process applied to a linear system as in (11) be expressed as $R(\tau)(1 - u(\tau - \zeta))$, where the variable τ denotes time interval and u stands for the unit step function. According to [36], the ACF of the WSS random process at the output of the linear system would be $H_{\text{imp}}(\tau) * H_{\text{imp}}(-\tau) * R(\tau)(1 - u(\tau - \zeta))$. Hence, the ACF of the output is also zero outside the time interval ζ, which implies that the samples of the output separated by a time greater than ζ are uncorrelated. Now, if the output process is such that its samples when uncorrelated are also independent, then from Definition 2, we get that the output is a WSS ζ-dependent process. Gaussian, lognormal, mixed Gaussian and mixed lognormal random processes are examples of processes whose samples when uncorrelated are also independent.*

Appendix D. An analysis on the approximation of joint mixed Gaussian PDF of two weakly dependent random variables

From (14) and (18), we see that the approximation used is

$$2|\Psi_R \Psi_C| \approx \begin{cases} \left(\Psi_R^2 + \Psi_C^2\right), & \Psi_R, \Psi_C \neq 0 \\ 0, & \Psi_R, \Psi_C = 0 \end{cases} \tag{D.1}$$

As deduced earlier, we have $|\rho_{r_G}| \ll 1$. We see that when $|\rho_{r_G}| \ll 1$, the approximation is carried out in a term which has an insignificant contribution in the whole expression. We shall now empirically calculate an error of the approximation in (17), when r_G takes a solitary value, say $r_G = 1$. Note that when r_G takes only one value, the expression in (14) reduces to a joint Gaussian PDF and the expression in (18) gives the proposed approximation for a joint Gaussian PDF. We consider the following expression of a sum of absolute differences as the error measure

$$\Omega = \sum_{Y_{ijR}} \sum_{Y_{ijC}} |f(Y_{ijR}, Y_{ijC}) - f_a(Y_{ijR}, Y_{ijC})| \tag{D.2}$$

We find that, in general $\Omega \lesssim 0.2$ for $\rho_1 \leqslant 0.5$, that is, considering that ρ_1 is small, we find that the sum of absolute difference between the joint Gaussian PDF and our approximation is in general less than about 20% of the total area under the joint Gaussian PDF. Fig. 12a shows the plot of Ω against various values of ρ_1 for different sets of parameters of the PDF. We also find that the approximation in (17) always gives a smaller Ω than the case when Y_{ijR} and Y_{ijC} are considered independent. Fig. 12b shows the difference in the error of the proposed approximation (Ω_a) and that of the independence assumption (say, Ω_i), with the former being subtracted from the latter for different sets of parameters of the PDF. As can be seen this difference of error is always positive for all values of ρ_1 and hence our approximation for the joint Gaussian PDF is better than the independence assumption in terms of Ω.

References

[1] M. Heath, S. Sarkar, T. Sanocki, K. Bowyer, A robust visual method for assessing the relative performance of edge-detection algorithms, IEEE Trans. Pattern Anal. Machine Intell. 19 (12) (1997) 1338–1359.
[2] M. Heath, S. Sarkar, T. Sanocki, K. Bowyer, Comparison of edge detectors: a methodology and initial study, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 1996, pp. 143–148.
[3] P.V. Henstock, D.M. Chelberg, Automatic gradient threshold determination for edge detection, IEEE Trans. Image Process. 5 (5) (1996) 784–787.
[4] P.L. Rosin, Unimodal thresholding, Pattern Recognit. 34 (11) (2001) 2083–2096.
[5] P.L. Rosin, Edges: saliency measures and automatic thresholding, Mach. Vis. Appl. 9 (4) (1997).
[6] R.R. Rakesh, P. Chaudhuri, C.A. Murthy, Thresholding in edge detection: a statistical approach, IEEE Trans. Image Process. 13 (7) (2004) 927–936.
[7] A. Koschan, M. Abidi, Detection and classification of edges in color images, IEEE Signal Process. Mag. 22 (1) (2005) 64–73.
[8] M. Basu, Gaussian-based edge-detection methods – a survey, IEEE Trans. Syst., Man, Cybern. C 32 (3) (2002) 252–260.
[9] V.S. Nalwa, T.O. Binford, On detecting edges, IEEE Trans. Pattern Anal. Mach. Intell. 8 (6) (1986) 699–714.
[10] O.A. Zuniga, R.M. Haralick, Gradient threshold selection using the facet model, Pattern Recognit. 21 (5) (1988) 493–503.
[11] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. 8 (6) (1986) 679–698.
[12] J. Shen, S. Castan, An optimal linear operator for step edge detection, Graph. Models Image Process. 54 (2) (1992) 112–133.
[13] S. Sarkar, K.L. Boyer, On optimal infinite impulse response edge detection filters, IEEE Trans. Pattern Anal. Mach. Intell. 13 (11) (1991) 1154–1171.
[14] M. Petrou, J. Kittler, Optimal edge detectors for ramp edges, IEEE Trans. Pattern Anal. Mach. Intell. 13 (5) (1991) 483–491.
[15] D. Marr, E. Hildreth, Theory of edge detection, Proc. Roy. Soc. Lond. B207 (1980) 187–217.
[16] V. Torre, T.A. Poggio, On edge detection, IEEE Trans. Pattern Anal. Mach. Intell. 8 (2) (1986) 147–163.
[17] S.M. Smith, J.M. Brady, SUSAN – a new approach to low level image processing, Int. J. Comput. Vis. 23 (1) (1986) 45–78.
[18] S. Alkaabi, F. Deravi, Variation of ISEF edge detector, IEE Electron. Lett. 39 (16) (2003) 1174–1175.
[19] P. Qiu, S.M. Bhandarkar, An edge detection technique using local smoothing and statistical hypothesis testing, Pattern Recognit. Lett. 17 (8) (1996) 849–872.
[20] P. deSouza, Edge detection using sliding statistical tests, Comput. Vis. Graph. Image Process. 23 (1) (1983) 1–14.
[21] J.H. Han, T.Y. Kim, Ambiguity distance: an edge evaluation measure using fuzziness of edges, Fuzzy Sets Syst. 126 (3) (2002) 311–324.
[22] S.K. Pal, R.A. King, On edge detection of X-ray images using fuzzy set, IEEE Trans. Pattern Anal. Mach. Intell. 5 (1) (1983) 69–77.
[23] K. Ghosh, S. Sarkar, K. Bhaumik, A possible mechanism of zero-crossing detection using the concept of the extended classical receptive field of retinal ganglion cells, Biol. Cybern. 93 (1) (2005) 1–5.
[24] R.C. Gonzalez, R.E. Woods, Digital Image Processing, second ed., Pearson Education, India, 2002.
[25] J. Canny, Finding edges and lines in images, Technical Report, Massachusetts Institute of Technology, Cambridge, MA, USA, 1983.
[26] M.K. Kundu, S.K. Pal, Thresholding for edge detection using human psychovisual phenomena, Pattern Recognit. Lett. 4 (6) (1986) 433–441.
[27] G.S. Jung, R.H. Park, Automatic edge extraction using locally adaptive threshold, Electron. Lett. 24 (11) (1988) 711–712.
[28] A. Elinabrouk, A. Aggoun, Edge detection using local histogram analysis, Electron. Lett. 34 (12) (1998) 1216–1217.
[29] E.R. Hancock, J. Kittler, Adaptive estimation of hysteresis thresholds, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 1991, pp. 196–201.
[30] N. Otsu, A threshold selection method from gray-level histogram, IEEE Trans. Syst., Man, Cybern. 9 (1) (1979) 62–66.
[31] J.N. Kapur, P.K. Sahoo, A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision, Graphics, and Image Processing 29 (1985) 273–285.
[32] W.-H. Tsai, Moment-preserving thresholding: a new approach, Comput. Vis. Graph. Image Process. 29 (1985) 377–393.
[33] J. Kittler, J. Illingworth, Minimum error thresholding, Pattern Recognit. 19 (1) (1986) 41–47.
[34] D. Sen, S.K. Pal, Histogram thresholding using beam theory and ambiguity measures, Fundam. Inform. 75 (1–4) (2007) 483–504.
[35] J.F. Haddon, Generalised threshold selection for edge detection, Pattern Recognit. 21 (3) (1988) 195–203.
[36] A. Papoulis, S.U. Pillai, Probability, Random Variable and Stochastic Processes, fourth ed., McGraw-Hill, 2001.
[37] X. Descombes, M. Sigelle, F. Préteux, Estimating Gaussian Markov random field parameters in a nonstationary framework: application to remote sensing imaging, IEEE Trans. Image Process. 8 (4) (1999) 490–503.
[38] Y. Dong, B.C. Foster, A.K. Milne, Comparison of radar image segmentation by Gaussian and gamma Markov random field models, Int. J. Remote Sens. 24 (4) (2003) 711–722.
[39] D. Sen, M.N.S. Swamy, M.O. Ahmad, Unbiased homomorphic system and its application in reducing multiplicative noise, IEE Proc.: Vis. Image Signal Process. 153 (5) (2006) 521–537.
[40] J.W. Goodman, Speckle Phenomena in Optics: Theory and Applications, Roberts and Company Publishers, USA, 2006.
[41] A. Zayezdny, D. Tabak, D. Wulich, Engineering Applications of Stochastic Processes: Theory, Problems and Solutions, Research Studies Press and John Wiley & Sons, England, 1989.
[42] S.O. Rice, Mathematical analysis of random noise, Bell Syst. Tech. J. 24 (1945) 46–156.
[43] S. Haykin, Communication Systems, second ed., Wiley Eastern, India, 1994.
[44] I.E. Abdou, W.K. Pratt, Quantitative design and evaluation of enhancement/thresholding edge detectors, Proc. IEEE 67 (5) (1979) 753–763.
[45] R.O. Duda, P.E. Hart, D.G. Stock, Pattern Classification, second ed., Wiley Interscience, USA, 2000.
[46] K. Bowyer, C. Kranenburg, S. Dougherty, Edge detector evaluation using empirical roc curves, Comput. Vis. Image Understand. 84 (1) (2001) 77–103.
[47] D.R. Martin, C.C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 530–549.
[48] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst., Man, Cybern. 3 (6) (1973) 610–621.
[49] E.L. Lehmann, J.P. Romano, Testing Statistical Hypothesis, third ed., Springer, USA, 2005.
[50] I.E. Sobel, Camera model and machine perception, Ph.D. thesis, Standford University, 1970.