Received May 7, 2021, accepted June 10, 2021, date of publication June 14, 2021, date of current version June 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089210

Quality Assessment Methods to Evaluate the Performance of Edge Detection Algorithms for Digital Image: A Systematic Literature Review

NAZISH TARIQ[1], ROSTAM AFFENDI HAMZAH[2], THEAM FOO NG[3], SHIR LI WANG[4], AND HAIDI IBRAHIM[1], (Senior Member, IEEE)

[1]School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia
[2]Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka, Durian Tunggal 76100, Malaysia
[3]Centre of Global Sustainability Studies (CGSS), Level 5, Hamzah Sendut Library, Universiti Sains Malaysia, Minden 11800, Malaysia
[4]Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjung Malim 35900, Malaysia

Corresponding author: Haidi Ibrahim (haidi_ibrahim@ieee.org)

This work was supported by the Ministry of Higher Education (MoHE) Malaysia through the Fundamental Research Grant Scheme (FRGS) under Grant 203\PELECT\6071421.

ABSTRACT A segmentation process is usually required in order to analyze an image. One of the available segmentation approaches is by detecting the edges on the image. Up to now, there are many edge detection algorithms that researchers have proposed. Thus, the purpose of this systematic literature review is to investigate the available quality assessment methods that researchers have utilized to evaluate the performance of the edge detection algorithms. Due to the vast number of available literature in this area, we limit our search to only open-access publications. A systematic search in five publisher websites (i.e., IEEExplore, IET digital library, Wiley, MDPI, and Hindawi) and Scopus database was carried out to gather resources that are related to the edge detection algorithms. Seventy-three publications that are about developing or comparing edge detection algorithms have been chosen. From these publication samples, we have identified 17 quality assessment methods used by researchers. Among the popular quality assessment methods are visual inspection, processing time, confusion-matrix based measures, mean square error (MSE)-based measures, and figure of merit (FOM). This survey also indicates that although most of the researchers only use a small number of test images (i.e., less than 10 test images), there are available datasets with a larger number of images for digital image segmentation that researchers can utilize.

INDEX TERMS Digital image processing, edge detection algorithm, image segmentation, assessment, validation, quality measures, reviews.

I. INTRODUCTION

In some computer vision applications, image segmentation is required to identify objects and analyze images automatically or help humans find the region of interest [1]–[3]. Edge detection is one of the significant branches available for image segmentation. The applications of edge detection are not limited. Many applications can benefit from edge detection. One can use edge detection to detect cracks on a surface captured on a photograph [4]–[6]. Edge detection is also commonly used to assess the shape of an object. For example, edge detection is used to find an object's perimeter, which is useful to find other shape's characteristics, such as centroid and circularity [7]. Edge detection can also be used for art purposes. The natural picture can be transformed into a cartoon-type picture with the help of edge detection [8].

Many researchers have come out with various algorithms for edge detection. Edges can be identified by recognizing the locations where there are drastic changes in intensity levels. This identification can be made by inspecting the gradient or derivative of the image. Thus, many algorithms are based on the first derivative or second derivative of the image. Popular edge detection algorithms, such as the Sobel, Prewitt, Robert and Canny edge detectors are using this concept [9], [10]. In addition to the spatial domain approaches, some algorithms are developed in the transformed domains, such as wavelet

The associate editor coordinating the review of this manuscript and approving it for publication was Fahmi Khalifa.

and curvelet [11], [12]. Recently, machine learning and deep learning approaches have also been used to develop edge detection algorithm [13]–[15].

The edge detection output is the edge or contour constructed by the edgels (edgel is the term used to present the edge element). When developing an edge detection algorithm, researchers need to evaluate their method's performance in terms of the edge or contour generated.

Up to now, various quality assessment methods have been used by researchers to evaluate the edges generated by their approaches. Therefore, this observation motivates us to conduct a systematic literature review to see how the previous studies related to the development or comparison of edge detection algorithms evaluated their works. This review aims to see whether there is already a standard procedure to evaluate the edge detection algorithms or not. Four main aspects will be investigated in this work. They are: (1) the assessment method used by researchers, (2) the number of test images used, (3) how their ground truth is obtained, and (4) the number of benchmark methods utilized.

This review hopefully will bring benefits to several parties. First, for researchers developing a new edge detection algorithm, this review can become a guideline for selecting suitable evaluation approaches. This review can give a basic idea to them to decide their test images and the number of benchmark methods. On the other hand, this review can hopefully help researchers develop a new quality assessment measure to evaluate edges. Based on this systematic literature review findings, we also will provide some suggestions for future works.

This paper is divided into five sections. Section I, which is this section, introduces this literature review. Then, Section II will explain our methodology in executing this literature review. Next, Section III will present our result, followed by Section IV that will discuss the results. The final section, which is Section V, is for our conclusion.

II. METHODOLOGY

In this work, the systematic literature review was executed based on the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [16], [17]. As there are many works related to edge detection, we scope our search only to open-access articles. The literature search has been carried out on four publisher websites (i.e., IEEExplore, IET digital library, Wiley, MDPI, and Hindawi) and the Scopus database.

A. SEARCH STRATEGY

The search was performed in February 2021. The search was performed by using the keyword: "edge detection algorithm". The related flowchart is shown in Figure 1.

The search returned us with 1,149 results from the IEEE website (https://ieeexplore.ieee.org). Therefore, we limited the search for the open-access publications only, which returned us with only 17 results.

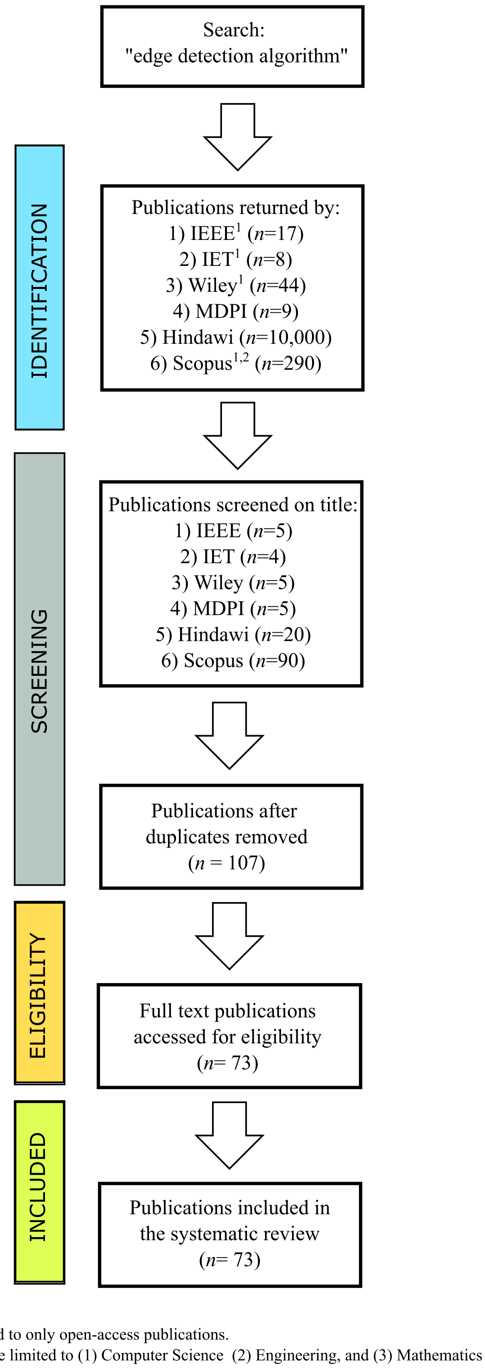

FIGURE 1. Flowchart of the publication selection process.

For the IET (https://digital-library.theiet.org/), 144 articles have been given as the search results. Then, we limited our search to only open-access publications. This search returns only eight articles.

Wiley's website (https://onlinelibrary.wiley.com/) returns 921 publications. The results have been reduced to 44 publications when we choose only open-access publications.

MPDI is an open-access publisher. We have executed the search on www.mpdi.com. This website presented us with nine publications as a result.

We did the search in https://www.hindawi.com/. Hindawi is another well-known open-access publisher. This website suggested 10,000 titles of publications.

Scopus (https://www.scopus.com/) has given us 3,197 publications as the initial result of the search. Therefore, we have limited the search to only open-access publications (i.e., ''All Open Access'') and limit the subject areas to ''Computer Science'', ''Engineering'', and ''Mathematics''. With these settings, we obtained 290 articles.

B. STUDY SELECTION

Due to the vast number of suggested publications that we have obtained from our searches, as presented in Subsection II-A, we first screened the publications' titles. The publications were included as the candidates of our samples if their title meets the following criteria:

- The title is showing that the work is related to the digital image processing field.
- The title should reflect that the publication is about the development, usage, or evaluation of edge detection algorithm.

For Hindawi's website, most of the suggested publications are not related to this literature review. Therefore, we arranged the results based on the relevancy. Then, we inspect only a few search pages of this website, which are really related to our work. Based on this approach, we kept five publications from IEEE, four publications from IET, five publications from Wiley, five publications from MDPI, 20 publications from Hindawi, and 90 publications from Scopus (please refer to Figure 1).

After this first screening, the publications from these six sources were combined, and duplicates were removed. Then, based on the full-text publications, the publications were selected for this systematic literature review if they fulfill the following criteria:

- The text in the publication is written in English.
- The publication is on digital image processing.
- The publication is about the development of a new edge detection algorithm, improvement to edge detection algorithm, evaluation of edge detection algorithm, or the usage of the edge detection algorithm.
- The evaluation used in the publication should not be too specific towards certain applications.

The full-text publications are excluded from this systematic literature review if any of the above condition is not met. The process of selecting the full-text publications for inclusion in this work was distributed independently among the authors of this paper.

C. DATA EXTRACTION AND ANALYSIS APPROACH

For this systematic literature review, data extraction was concentrated on the following four aspects: (i) quality assessment method, (ii) test images, (iii) ground-truth, and (iv) the number of benchmark methods. Data extracted from each publication were: (a) authors, (b) year of publication, (c) quality assessment method(s) used, (d) the number of test images used, (e) information about the ground truth, and (f) the number of benchmark methods used.

III. RESULTS

This section is divided into five subsections. Subsection III-A gives the search results for the publication selection process. Then, Subsection III-B presents the quality assessment methods that are used by our samples. Next, information about the test images used by the researchers are provided in Subsection III-C. After that, we explain our finding related to the ground-truth in Subsection III-D. Finally, Subsection III-E presents the gathered information about the number of benchmark methods.

A. SEARCH RESULTS

After the removal of duplicates, we identified a total of 107 publications (i.e., $n = 107$) from six sources (i.e., IEEE, IET, Wiley, MDPI, Hindawi, and Scopus) as the candidates for our literature review. Following the full-text publications screening (i.e., to assess the publications' eligibility), 73 publications (i.e., $n = 73$) have met the inclusion criteria. This publication selection process is shown in Figure 1.

B. QUALITY ASSESSMENT METHODS

There are 17 quality assessment methods have been used in these 73 publications. Each of this method is explained by the following subsections.

1) VISUAL INSPECTION

Visual inspection is the most used quality assessment method found in these 73 publication samples. All of these publication samples are using visual inspection to discuss the performance of their edge detection algorithm. Output images are shown for the evaluation. This quality assessment method is not an objective assessment method. The ground truth is not required for this assessment. The outputs are judged subjectively by the human, based on their preference. However, there are variations in how the authors presented their outputs. Some of the examples are shown in Figure 2.

Lasserre *et al.* [18] have used five images to evaluate their edge detection algorithm's performance. Two images have been used to evaluate the directional gradient magnitudes (i.e., approach as shown in Figure 2(b)). These gradient magnitude images are used to show that the proposed method has successfully removed strong gradient values that are not related to the important edges, according to this work. The other three images have been used to evaluate the appearance of the areas that are enclosed by the generated contours (i.e., approach as shown in Figure 2(d)). These images have also been used to discuss how close the generated contours with the expected contour by humans.

Sun *et al.* [19] have used six hyperspectral images, consisting of artificial images and real natural scene images, for their evaluation. The artificial images were generated using a freeware called HYper-spectral data viewer for

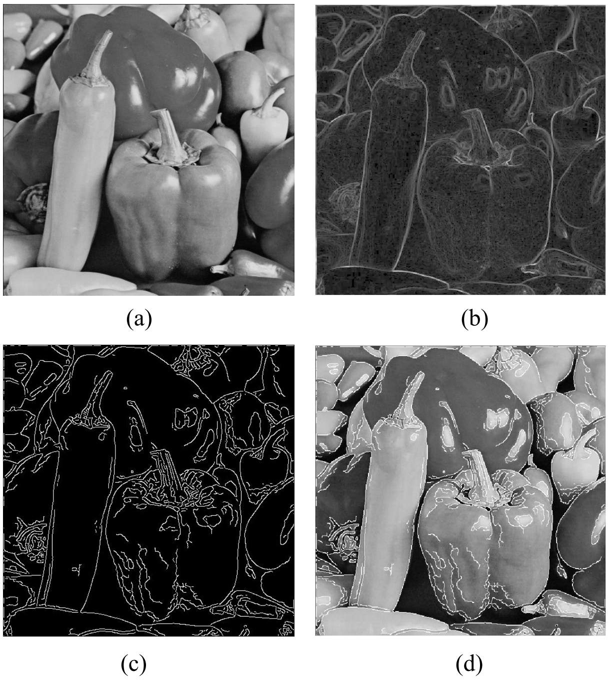

FIGURE 2. **Examples of visual inspection. (a) Input image. (b) Gradient magnitude. (c) Detected edges. (d). Detected edges overlaid with the input image.**

Development of Research Applications (HYDRA) [20]. They have commented on the edges generated by the algorithms (i.e., approach as shown in Figure 2(c)). Poor methods usually generate many small, spurious, or grainy edges.

Chen and Chiu [21] compared the edges generated by their method with four other methods. Yahya *et al.* [22] evaluated the edges' appearance from their method and compared them with five other methods.

Researchers also have compared the images of edges obtained from different image resolutions. An example is a work on ore image edge detection by Wang and Zhang [23]. They have evaluated the edges that are obtained by the binarization process. Three input images with different resolutions have been used for this purpose.

To evaluate the segmentation results of their work, Xu and Xiao [24], have utilized two images for this purpose. The images used are showing the wind turbine blade. They evaluate the performance in terms of edges detected and also the segmented region.

2) PERCENTAGE OF ACCEPTABLE SEGMENTATION CONTOUR

In work by Lasserre *et al.* [18], they use the percentage of acceptable segmentation contour as one of their evaluations. This measure can be considered as an extension of the visual inspection assessments. No ground truth is needed for this measure. This percentage is defined as the percentage of output images that have generated contours similar to those expected by a human.

3) NUMBER OF DETECTED EDGELS

In work by Sundani *et al.* [25], they have measured the number of edgels as one of their evaluations. They assume that a good edge detection algorithm will produce more edges. As they developed an improvement to the Canny algorithm, they have also calculated the increment of edgels with respect to the edges detected by the Canny algorithm. This measure is defined as:

$$\text{Increment} = \frac{N_q - N_C}{S_i} \times 100\% \tag{1}$$

where N_q is the number of edgels detected by their method, N_C is the number of edgel from the Canny algorithm, and S_i is the image's size.

4) CONFUSION MATRIX-BASED MEASURES

In this reference-based evaluation, in addition to the detected edges by the algorithm, the researchers also need to provide the ground truth for each image tested. The ground truth contains the targeted edgels; how the edges or contour should be. By comparing the detected edges with the ground truth, four classes are generated, which are true positive (TP), true negative (TN), false negative (FN), and false positive (FP). This is defined by the confusion matrix shown in Figure 3.

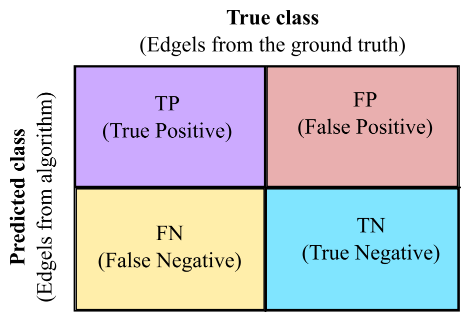

FIGURE 3. **A confusion matrix to define TP, TN, FP and FN.**

The true positive TP is when the edgel is both defined by the algorithm and also by the ground truth. The true negative TN is when the non-edge pixel defined by the ground truth is correctly identified as the non-edge pixel by the algorithm. The false-positive FP is when the non-edge pixel defined by the ground truth is wrongly identified as the edgel by the algorithm. The false-negative FN is when the actual edgel defined by the ground truth is wrongly identified as a non-edge pixel by the algorithm.

From these four classes, several other quality measures can be defined. Precision is defined as:

$$\text{Precision} = \frac{\#(\text{TP})}{\#(\text{TP+FP})} \tag{2}$$

where #(.) is the number of elements in a set. This measure is also called as positive predictive value (PPV).

Recall is expressed as:

$$\text{Recall} = \frac{\#(\text{TP})}{\#(\text{TP+FN})} \tag{3}$$

This measure is also known as true positive rate (TPR), sensitivity, or hit rate.

The false positive rate (FPR), or fall-out is given as:

$$\text{FPR} = \frac{\#(\text{FP})}{\#(\text{FP+TN})} \tag{4}$$

The true negative rate (TNR) is described as:

$$\text{TNR} = \frac{\#(\text{TN})}{\#(\text{TN+FP})} \tag{5}$$

The negative predictive value (NPV) is presented by the following equation:

$$\text{NPV} = \frac{\#(\text{TN})}{\#(\text{TN+FN})} \tag{6}$$

The error rate (Err) is defined as:

$$\text{Err} = \frac{\#(\text{FP})}{\#(\text{TP+FN})} \tag{7}$$

The accuracy (Acc) is given as:

$$\text{Acc} = \frac{\#(\text{TP+TN})}{\#(\text{TP+TN+FP+FN})} \tag{8}$$

The F1-score is defined as:

$$\text{F1-score} = \frac{\#(\text{2TP})}{\#(\text{2TP+FN+FP})} \tag{9}$$

The researchers in our samples did not use all the formulas listed above but use selected evaluations for their work. For example, Dong *et al.* [26] used only the true positive rate (i.e., Recall) and the false positive rate (FPR) in their work. Huang *et al.* [27] utilized the sensitivity measure (i.e., Recall). Dorafshan *et al.* [6] evaluated the methods by using the positive predictive value (i.e., Precision), true positive rate (i.e., Recall), accuracy, negative predictive value, and F1-score. Luo *et al.* [28] have used the true positive rate (i.e., Recall), false-positive rate (FPR), and accuracy (Acc). They also have added the versions of Recall, FPR, and Acc that still accepted detected edgels by the algorithms, which deviate a few pixels from the ground truth. Bogdan *et al.* [29] utilized Recall, Precision, and F1-score.

Khunteta and Ghosh [30] and Xu *et al.* [31] used a measure called P-value (or precision value) for their evaluation purpose. This measure is defined as:

$$\text{P-value} = \frac{\#(\text{TP})}{\#(\text{TP+FP+FN})} \tag{10}$$

Sun *et al.* [19] have used F-measure as their quantitative evaluation. Their definition of F-measure (F_α) is:

$$F_\alpha = \frac{\text{Precision} \times \text{Recall}}{\alpha.\text{Precision} + (1-\alpha).\text{Recall}} \tag{11}$$

with $\alpha \in [0, 1]$ is the weighting parameter.

Qu *et al.* [15] and Zheng *et al.* [32] have also used F-measure in their work, but their definition for the F-measure is:

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{12}$$

Related to the F-measure are two other quality measure, which are optimal dataset scale (ODS), and optimal image scale (OIS). The ODS is the value of F-measure obtained from a fix contour threshold, while the OIS is when the optimal threshold is applied individually to each test image. Qu *et al.* [15] have used F-measure, ODS and OIS in their work. Besides, they are also have discussed the plot of precision versus recall (i.e., PR curve) to compare the performance of their method with other methods.

5) NUMBER OF THE FALSE EDGES

Al-Jarrah *et al.* [33] compared the number of false edges obtained from several edge detection approaches. A low number of false edges indicates a good edge detection method. Unfortunately, the definition of the false edges is not mentioned in this work, whether it is false-positive edges, false-negative edges, or both.

6) RECEIVER OPERATING CHARACTERISTIC (ROC)

The receiver operating characteristic (ROC) is obtained by plotting the true positive rate (TPR) versus the false positive rate (FPR). If the plot shows that the curve is nearer to the left-hand border and the ROC space's top border, then the method is considered to produce accurate results. On the other hand, if the curve is around 45° of the ROC space, this shows that the results are less accurate. Zheng *et al.* [32] used ROC as one of the evaluation criteria in their work.

7) EDGE DETECTION ERROR, ϵ

The measure of edge detection error ϵ was introduced by Ma and Staunton [34]. It compares the edge detected by the algorithm with the ground truth, as defined as:

$$\epsilon = 1 - \frac{\#(G \cap E)}{\#(G)} \tag{13}$$

where G is the ground truth, E is the detected edges, and #(.) is the number of elements in a set. A good edge detection algorithm will give a small value of ϵ. Guo and Sengur [35] have used this measure for their work.

8) INTERSECTION OVER UNION (IoU)

The measure of intersection over union (IoU) can be used to measure the correlation between the edges on the ground truth, with the edges detected by the algorithm. A high IoU value indicates a high correlation of edges between these two edge images. Zheng *et al.* [32] have used IoU as one of the performance measures in their work.

Yoon [36] has used IoU to measure the edge similarity. In this work, IoU is called as an edge similarity or an overlapping rate. The similarity is defined as:

$$S(G, E) = \frac{G \cap E}{G \cup E} \tag{14}$$

where E is the result from the segmentation, and G is the ground truth.

9) MISCLASSIFICATION RATE (MCR)

Misclassification rate (MCR) is defined as:

$$\mathrm{MCR} = \frac{\sum |B_G \cap B_E| + \sum |F_G \cap F_E|}{\sum (B_G + F_E)} \times 100\% \tag{15}$$

where G and E are the ground truth and the detected edges, respectively. Then, F_i and B_i (with $i = G$, or E) stand for the foreground pixels, and background pixels, respectively. A low MCR value indicates a good edge detection result. This measure has been used by the work by Ergen [37].

10) BUFFER ANALYSIS METHOD

In work by Zhang *et al.* [14], one template SAR image has been used in their buffer analysis method. The buffer analysis method works by counting the number of detected edgels within a 3×3 buffer around the correct edgel (i.e., the ground truth). There are two types of analysis presented; the percentage B_i ($i = 0, 1, 2, 3$), and the cumulative percentage S_i ($i = 0, 1, 2, 3$), where i is the radius of the mask used for the evaluation.

11) STRUCTURAL SIMILARITY INDEX (SSIM)

Mittal *et al.* [38] and Arulpandy and Pricilla [39] have use SSIM to evaluate their work. SSIM can be defined as:

$$\mathrm{SSIM}(G, E) = [l(G, E), c(G, E), s(G, E)] \tag{16}$$

which means that this measure is based on the luminance (l), contrast (c), and structure (s).

12) EDGE BASED SSIM (ESSIM)

Arulpandy and Pricilla [39] have use edge based structural similarity index (ESSIM) to evaluate their method. This measure can be defined as:

$$\mathrm{ESSIM}(G, E) = [l(G, E), c(G, E), e(G, E)] \tag{17}$$

which means that this measure is based on the luminance (l), contrast (c), and edges (e).

13) CONTINUITY EVALUATION

Zheng *et al.* [40] have used a measure called a continuity evaluation index in their work. The continuity index ρ is defined as:

$$\rho = \frac{\epsilon_1 n_s}{\epsilon_s n_1} \tag{18}$$

where ϵ_1 and ϵ_s are the number of edge points from the output, and the ground truth, respectively, whereas n_1 and n_s are the number of edge lines on the output, and the ground truth, respectively. Higher value of ρ indicates better performance.

Yang *et al.* [41] used a similar measure to ρ. However, this measure does not need any ground truth. This measure, denoted as R, is defined as:

$$R = \frac{\mathrm{TEN}}{\mathrm{CEN}} \tag{19}$$

where TEN is the number of the detected edgels, while CEN is the number of edge segments. A higher value of R indicates that the segment is longer (i.e., more continuous).

The work by Yu *et al.* [42] used a similar but more detailed equation. They define A as the number of edgels detected by the algorithms, B is the number of segments detected by 4-connected component analysis, and C is the number of segments detected by 8-connected component analysis. They measured two values, which are C/A (to present the edge continuity) and C/B (to present the edge width). A good edge detection method is expected to give a smaller value of these ratios.

14) MEAN SQUARE ERROR (MSE), ROOT-MEAN-SQUARE-ERROR (RMSE), SIGNAL-TO-NOISE RATIO (SNR), OR PEAK SIGNAL-TO-NOISE RATIO (PSNR)

Mittal *et al.* [38], Yang *et al.* [41] use the mean square error (MSE) as one of their evaluation method. If both the resultant edge image E and the ground truth image G are both in size of $M \times N$ pixels, MSE is defined as:

$$\mathrm{MSE}(E, G) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} ||E_{x,y} - G_{x,y}||^2 \tag{20}$$

where x, y are the spatial coordinates. Smaller magnitude of MSE indicates a better performance.

A similar measure to MSE is the root-mean-square-error (RMSE), which is defined as:

$$\mathrm{RMSE}(E, G) = \sqrt{\mathrm{MSE}(E, G)} \tag{21}$$

RMSE has been used in the work by Mendhurwar *et al.* [43].

Other option is by using peak signal-to-noise ratio (PSNR) to evaluate the performance of edge detection algorithms. PSNR is defined as:

$$\begin{aligned}\mathrm{PSNR}(E, G) &= 10 \log_{10} \frac{255^2 MN}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (G_{x,y} - E_{x,y})^2} \\ &= 10 \log_{10} \frac{255^2}{\mathrm{MSE}(E, G)}\end{aligned} \tag{22}$$

In contrast to the MSE, a higher magnitude of PSNR indicates a better performance. Researchers such as Yahya *et al.* [22], Fu [44], Chi and Gao [45], Mendhurwar *et al.* [43], Ren *et al.* [46], Sert and Avci [47], and Tang *et al.* [48] have used PSNR as one of the evaluation measures for their work. El Araby *et al.* [49], Hu *et al.* [50], and Sudhakara and Meena [51] have used both MSE and PSNR.

Mendhurwar *et al.* [43] also have used signal-to-noise ratio (SNR) to evaluate their method. This measure is defined as:

$$\mathrm{SNR}(E, G) = \sqrt{\frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (E_{x,y})^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (G_{x,y} - E_{x,y})^2}}$$

$$= \sqrt{\frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}(E_{x,y})^2}{MN \times \mathrm{MSE}(E,G)}} \tag{23}$$

Kurdi *et al.* [52] have also used SNR in their work. However, their SNR is reported in decibel (dB). Zhao [53] have utilized a slightly different equation for their SNR measure.

15) ENTROPY

In 1948, E. Shannon had proposed a measure, called entropy, to evaluate the information content [38]. The information entropy is defined as:

$$H(I) = -\sum_{i=0}^{L} p_i \log_2 p_i \tag{24}$$

where I is the image that we want to evaluate, p_i is the rate of recurrence of pixels with intensity i, and $H(I)$ is Shannon's entropy. This measure is also called the average information content (AIC). Researchers such as Mittal *et al.* [38], and Sudhakara and Meena [51] have employed Shannon's entropy to investigate whether their generated edges are meaningful or not.

16) FIGURE OF MERIT, FOM

Figure of merit (FOM) was introduced by Pratt [54]. This measure is defined as:

$$\mathrm{FOM} = \frac{1}{\max(N_E, N_G)}\sum_{k=1}^{N_E}\frac{1}{1+\alpha d^2(k)} \tag{25}$$

where N_G is the numbers of the actual edges, N_E is the number of the detected edges by the algorithm, α is the scaling constant, and $d(k)$ is the displacement of the detected edge from the actual edge. This measure has been used by researchers, including Yahya *et al.* [22], Chen *et al.* [55], Mendhurwar *et al.* [43], Yang *et al.* [41], Elaraby and Moratal [56], Ergen [37], Gonzalez *et al.* [57], Liu *et al.* [58], Ren *et al.* [46], Sadiq *et al.* [59], Sert and Avci [47], Guo and Sengur [35], and Nes [60]. In the work by Liu and Ren [61], they addressed FOM as the quality factor Q.

17) PROCESSING TIME, OR COMPUTATIONAL COMPLEXITY

Processing time, or known as the CPU time, is a measure on how long the segmentation process needs to completely segment the image. They normally measured in seconds, ot milliseconds. Lasserre *et al.* [18] have evaluate the processing time, based on three categories of input, which are easy, medium and complex. Processing time is also has been used by researchers such as Wang and Zhang [23], and Yahya *et al.* [22].

In the work by Liu and Ren [61], in addition to the processing time, they also have reported the memory footprint, which was measured in megabytes (MB). Higher value of memory footprint indirectly indicates that the algorithm is more complex.

Ma *et al.* [62] compared the complexity of the edge detection algorithms. They have evaluated the complexity of their method depending on the cases. On the other hand, Qu *et al.* [15] used frame per second (FPS) to indicates the processing time needed by the algorithms.

C. TEST IMAGES

Not all of the publication samples in this study reported the sources of their input images. Only 13 from 73 publication samples (i.e., 17.81%) have reported their image source. Nevertheless, we have observed that some of the works are using standard images, such as "Lena" and "Cameraman", together with their unique input images. Zhang *et al.* [14] have used one SAR image obtained from the National Library Website. Xu and Xiao [24] used images from work by Heath *et al.* [63]. Sun *et al.* [19] generate artificial images to represent hyperspectral images by using HYDRA [20]. Chen and Chiu used five natural images, with textures, taken from Flickr (www.flickr.com).

Zhang *et al.* [64] used 795 training images and 654 testing images taken from the NYUD2 dataset. Tang *et al.* [48] used images obtained from the UCID test database. Cao *et al.* [65] have used the Pascal VOC2012 database (http://host.robots.ox.ac.uk/pascal/VOC/voc2012/) for their input. This database is available to researchers and has 17,125 images in 20 categories. Six publications have used Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500). These publications are the works by Mittal *et al.* [38], Qu *et al.* [15], Zheng *et al.* [32], Chen *et al.* [55], Bogdan *et al.* [29] and Zheng *et al.* [40]. BSDS500 provides 200 images for training, 100 images for validation, and 200 images for testing.

We also have investigated the number of test images used by the researchers. Some of the publications did not explicitly specify the number of test images used by them. This is shown in Figure 4. From this figure, we can see that the number of test images used by the researchers is varied. It can be observed that 40 from 73 publication samples (i.e., 54.79%) are using five or fewer test images. On the other hand, eight publications have used more than 100 test images. These eight publications are the works by:

1) Alpert *et al.* [66] in year 2010. The work was on detecting weak curved edges in noisy images. The test images consist of 63 binary images and 100 grayscale images.
2) Chen *et al.* [55] in the year 2015, which is using BSDS500. The edge detection is done based on nonsubsampled contourlet transform and edge tracking.
3) Zhang *et al.* [64] in the year 2016, which is using the NYUD2 dataset. The work involves with a machine learning approach, which is based on structured forests.
4) Dorafshan *et al.* [6] in the year 2018. The work is on detecting the crack in the concrete and involves implementing deep convolutional neural networks.
5) Cao *et al.* [65] in the year 2018, which is using Pascal VOC2012. This work implemented a parallel approach

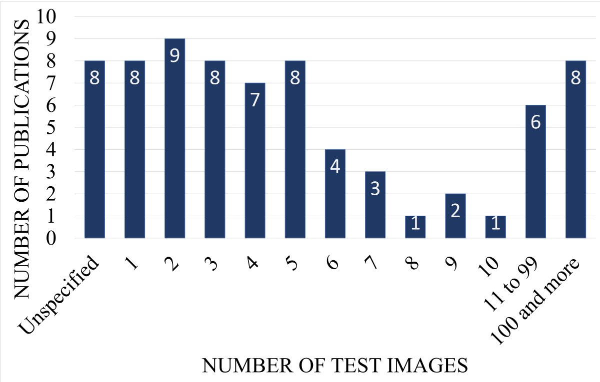

FIGURE 4. The total number of test images used by the publication samples.

to detect edges utilizing the Hadoop platform. Sixty thousand images have been used in this work.

6) Qu *et al.* [15] in the year 2019, which is using BSDS500. This work involves a deep neural network.
7) Mittal *et al.* [38] in the year 2019, which is using BSDS500. The aim is to produce edges with better connectivity.
8) Zheng *et al.* [40] in the year 2020, which is using BSDS500. The work was on developing an adaptive edge detection algorithm.

From the list above, we can see that the works with more than 100 test images are mostly from the year 2018 and above. The majority of them are using the BSDS500 dataset. Some of them are implementing deep neural network, which requires a large number of images.

We also investigated the number of images that the researchers used for their visual inspection assessment. The finding is shown in Figure 5. From this figure, we can see that majority of the publication samples are using less than five images for this purpose. Although the majority of the researchers are using a small number of test images for the visual inspection, it is worth noting that some of the researchers did further extensive assessments. These assessments include inspecting the algorithms' performance under different parameter settings, different image resolutions, or different levels of noise corruption.

FIGURE 5. The total number of test images used for the visual inspection.

D. GROUND TRUTH

Almost all of our publication samples do not discuss their ground truth, although reference-based evaluation methods have been utilized in their works. For researchers using Berkeley Segmentation Data Set (BSDS500), this dataset provides 100 test images with ground truth. The boundaries in these ground truth images are labeled by a human. Researchers such as Mittal *et al.* [38], Qu *et al.* [15], and Zheng *et al.* [32] are using BSDS500 in their research.

In work by Xu and Xiao [24], their own edge detection algorithm develops the ground truth. By changing the regularization parameters α_j, with $j \in \{1, 2, \ldots, N\}$, their method will develop N different edge maps $D_j, \in \{1, 2, \ldots, N\}$. Next, they generate N potential ground truths (PGTs) based on the information from D_j. The Chi-square test was conducted on each PGT_j, and the PGT_j with the best test value is selected as the estimated ground truth (EGT).

E. NUMBER OF BENCHMARK METHODS

We have investigated the number of benchmark methods used by the publication samples. The result is shown in Figure 6. As shown by this figure, four publications do not have any benchmark method. Fifteen publications only have one method to compare. Nevertheless, most of these publications are the extension to the well-known edge detector algorithm, such as Canny or Sobel, where the comparisons were only made to this base method.

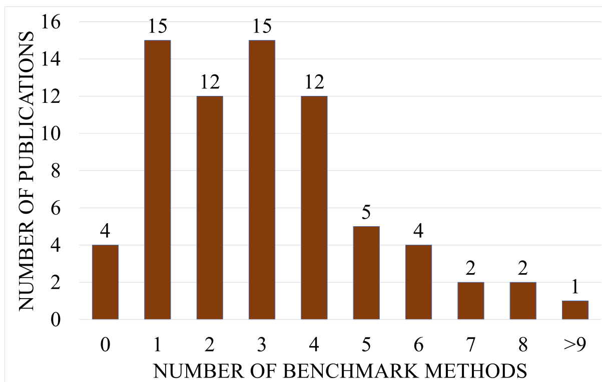

FIGURE 6. The number of benchmark methods.

From Figure 6, it can be seen that the majority of the methods have one to four benchmark methods. Canny, Sobel, and Prewitt are among the popular benchmark methods.

In these samples, only one publication has more than nine methods for the comparisons. This work is by Qu *et al.* [15], where they have around 20 benchmark methods to assess their deep neural network.

IV. DISCUSSIONS

From Section III-B, those 17 assessment methods obtained from 73 publication samples can be classified into the following three general categories:

(I) Subjective evaluation. Under this category, the performance of the edge detection algorithm is judged subjectively by the human(s). Therefore, the reference image or the ground truth is not needed. Methods under this category are:
  1. Visual inspection (all 73 publications).
  2. Percentage of acceptable contour (one publication).

(II) No-reference image quality assessment methods. Methods under this category do not require the reference image or the ground truth. The measurement is done by the information provided solely by the results from the algorithm. Methods under this category are:
  1. Continuity evaluation (two publications).
  2. Processing time, or computational complexity (23 publications).

(III) Full-reference image quality assessment methods. The reference image or the ground truth is needed by this category. The performance of the method is evaluated by comparing the result from the algorithm with the reference image. Methods under this category are:
  1. Number of detected edgels (one publication).
  2. Confusion matrix-based measures (14 publications).
  3. Number of false edges (one publication).
  4. ROC (one publication).
  5. Edge detection error (one publication).
  6. Intersection over union (IoU) (two publications).
  7. Buffer analysis method (one publication).
  8. SSIM (two publications).
  9. Edge based SSIM (ESSIM) (one publication).
  10. Continuity evaluation (one publication).
  11. MSE, RMSE, SNR, or PSNR (14 publications).
  12. Entropy (two publications).
  13. Figure of merit (FOM) (14 publications).

The continuity evaluation methods fall into two categories because Yang *et al.* [41], and Yu *et al.* [42] do the calculation based on the output edges only. In contrast, the method by Zheng *et al.* [40] needs the ground truth.

Researchers also combined assessment methods to evaluate the performance of edge detection algorithm. Table 1 presents the summary of the assessment used in these 73 publications. There are 25 combinations in total, ranging from one evaluation method to the combination of four methods. From this literature review, we found that there is still no standard assessment method available for evaluating the edge detection algorithm. There is no common assessment combination that can be found in Table 1. Each assessment method has its advantages and disadvantages, where the researcher uses or develops the measure based on their research's aim. Most of the assessment methods are also used by only one publication.

We observed that visual inspection is the most popular assessment method used by researchers. All of these 73 publications have utilized visual inspection. Even 20 of them did the evaluation solely based on the visual inspection. Although relatively easy, the visual inspection is essential to verify that the algorithm produces results similar to the one expected from a human. However, the visual inspection used may not be the same with each other. The visual inspection may include checking the edges or contour, displaying the gradient magnitude, or evaluating the edges shown on the input image.

Works by Sze *et al.* [67], Alpert *et al.* [66], for example, use only visual inspection for their assessments because the objects on their test images are simple or having strong contrast, where the edges of interest are relatively easy to be identified. However, some other methods are using more complex test images. For these cases, not all edges on the image will be considered for the evaluation. Judgment of the output quality is depending on the preference of the human evaluator. Thus, the visual inspection is prone to the inter-observer variation, where the result reported by one observer may not be the same as a result by other observers.

To get a more truthful result, the researchers should report who the evaluators are and how many evaluators are employed for their visual inspection procedure. In our opinion, the researchers should avoid being the evaluator, as the preference might be biased towards their own method. It would be much better to recruit people outside the research group to be the evaluator. It is also good to have more than one evaluator for this purpose so that the inter-observer variation effect can be reduced. We also believe that the visual inspection should be paired with other quantitative quality assessment methods.

The second most popular edge detection assessment method is the processing time or complexity analysis, with 23 from 73 publications are utilizing it. This measure is used to evaluate the processing speed of the methods, where a shorter processing time is desired. This measure is crucial, especially for those methods designed for real-time applications. The other popular assessment methods are (i) confusion matrix-based measures, (ii) MSE, RMSE, SNR, or PSNR, and (iii) figure of merit (FOM), where each of them has been reported in 14 publications. The other evaluation methods can be considered as not popular as only three or fewer publications are using them.

From Table 1, we can see that most of the publications using a combination of three or four quality assessment methods are recent publications, which are from the year 2018 and above. This observation indicates that the researchers are still not satisfied with the currently available assessment methods, as a certain assessment only addressed a specified aspect for the evaluation. Therefore, we would like to suggest the researchers have at least four quality assessment methods to cover the following four evaluation aspects:

1) The appearance of the edges. This can be observed by using subjective evaluation methods.
2) Processing time, or complexity analysis. This measure is important as a simple and fast method is generally desired.

TABLE 1. Summary of the quality assessment methods of 73 publications.

No.	Assesement methods used	Publications	Total number of publications
1	Visual inspection only	Sze et al., 1997 [67], Alpert et al., 2010 [66], Yang and Xu, 2011 [68], Liu et al., 2014 [4], Liu, Wang and Duan, 2014 [69], Asghari and Jalali, 2015 [70], Singh, Bajpai and Pandey, 2015 [71], Wei and Qiao, 2015 [72], Mathur, Mathur and Mathur, 2016 [73], Liu and Mao, 2018 [74], Liu et al., 2018 [75], Shi and Luo, 2018 [76], Wang and Wang, 2018 [77], Xu and Xiao, 2018 [24], Xu, Xu and Zuo, 2019 [78], Zhang et al., 2018 [79], Bauganssa, Sbihi and Zaim, 2019 [80], Chen and Chiu, 2019 [21], Kabir and Mondal, 2019 [81], Sekehravani, Babulak and Masoodi, 2020 [82]	20 publications
2	(1) Visual inspection **AND** (2) MSE, RMSE, SNR, or PSNR	Zhao, 2011 [53], Chi and Gao, 2014 [45], Kurdi, Grantner and M Abdel Qader, 2017 [52], Hu et al., 2019 [50]	4 publications
3	(1) Visual inspection **AND** (2) Figure of merit (FOM)	Nes, 2012 [60], Elaraby and Moratal, 2017 [56], Gonzalez, Melin and Castillo, 2017 [57], Liu et al., 2018 [58], Sadiq et al., 2019 [59]	5 publications
4	(1) Visual inspection **AND** (2) Processing time, or complexity	Swargha and Rodrigues, 2012 [83], Shi et al., 2015 [84], Bastan, Bukhari and Breuel, 2017 [85], Gunawan et al., 2017 [86], Wang and Zhang, 2017 [23], Cao et al., 2018 [65], Jing et al., 2018 [87], Zixin et al., 2018 [88], Ezzaaki et al., 2020 [89], Ma, Ma and Chu, 2020 [62]	10 publications
5	(1) Visual inspection **AND** (2) Confusion matrix-based measures	Khunteta and Ghosh, 2014 [30], Dong et al., 2015 [26], Zhang et al., 2016 [64], Huang, Yu and Zuo, 2017 [27], Dorafshan, Thomas and Maguire, 2018 [6], Yang et al., 2018 [13]	6 publications
6	(1) Visual inspection **AND** (2) Number of detected edgels	Sundani et al., 2019 [25]	1 publication
7	(1) Visual inspection **AND** (2) Buffer analysis method	Zhang et al., 2019 [14]	1 publication
8	(1) Visual inspection **AND** (2) MSE, RMSE, SNR, or PSNR **AND** (3) Figure of merit (FOM)	Mendhurwar et al., 2011 [43], Ren, Li and Chen, 2013 [46], Sert and Avci, 2019 [47]	3 publications
9	(1) Visual inspection **AND** (2) Misclassification rate (MCR) **AND** (3) Figure of merit (FOM)	Ergen, 2014 [37]	1 publication
10	(1) Visual inspection **AND** (2) Edge detection error **AND** (3) Figure of merit (FOM)	Guo and Sengur, 2014 [35]	1 publication
11	(1) Visual inspection **AND** (2) Confusion matrix-based measures **AND** (3) Figure of merit (FOM)	Chen et al., 2015 [55], Luo et al., 2017 [28]	2 publications
12	(1) Visual inspection **AND** (2) Figure of merit (FOM) **AND** (3) Processing time, or compexity	Liu and Ren, 2019 [61]	1 publication
13	(1) Visual inspection **AND** (2) Percentage of acceptable contour **AND** (3) Processing time, or complexity	Lasserre, Cutt and Moffat, 2008 [18]	1 publication
14	(1) Visual inspection **AND** (2) Confusion matrix-based measures **AND** (3) Processing time, or complexity	Xu et al., 2014 [31], Sun et al., 2017 [19], Qu et al., 2019 [15], Bogdan, Bonchis and Orhei, 2020 [29]	4 publications
15	(1) Visual inspection **AND** (2) Number of false edges **AND** (3) Processing time, or complexity	Al Jarrah, Al Jarrah and Roth, 2018 [33]	1 publication
16	(1) Visual inspection **AND** (2) Intersection over union (IoU) **AND** (3) Processing time, or complexity	Yoon, 2016 [90]	1 publication
17	(1) Visual inspection **AND** (2) MSE, RMSE, SNR or PSNR **AND** (3) Processing time, or complexity	El Araby et al., 2018 [49], Fu, 2020 [44], Tang et al., 2020 [48]	3 publications
18	(1) Visual inspection **AND** (2) Continuity evaluation **AND** (3) Processing time, or complexity	Yu et al., 2019 [42]	1 publication
19	(1) Visual inspection **AND** (2) Continuity evaluation **AND** (3) Confusion matrix-based measures	Zheng et al., 2020 [40]	1 publication
20	(1) Visual inspection **AND** (2) MSE, RMSE, SNR, or PSNR **AND** (3) Entropy	Sudhakara and Janaki Meena, 2019 [51]	1 publication
21	(1) Visual inspection **AND** (2) Structural Similarity Index (SSIM) **AND** (3) Edge based SSIM (ESSIM)	Arulpandy and Pricilla, 2020 [39]	1 publication
22	(1) Visual inspection **AND** (2) Continuity evaluation **AND** (3) MSE, RMSE, SNR, or PSNR **AND** (4) Figure of merit (FOM)	Yang, Xia and Juan, 2016 [41]	1 publication
23	(1) Visual inspection **AND** (2) Structural Similarity Index (SSIM) **AND** (3) MSE, RMSE, SNR, or PSNR **AND** (4) Entropy	Mittal et al., 2019 [38]	1 publication
24	(1) Visual inspection **AND** (2) MSE, RMSE, SNR, or PSNR **AND** (3) Figure of merit (FOM) **AND** (4) Processing time, or complexity	Yahya et al., 2019 [22]	1 publication
25	(1) Visual inspection **AND** (2) Confusion matrix-based measures **AND** (3) Receiver operating characteristic (ROC) **AND** (4) Intersection over union (IoU)	Zheng et al., 2019 [32]	1 publication

3) Pixel-by-pixel evaluation of the detected edges. This can be obtained by using full-reference image quality assessment methods.
4) Continuity evaluation. This evaluation is important because in most applications, a closed contour or continuous edges are required.

One of the main drawbacks of the reference-based evaluation method for evaluating edge detection algorithms is that the edges detected by the algorithm, and the edges in the ground truth, are usually presented by thin segments. Therefore, it is challenging to find the overlap portion of these segments (i.e., the true positive), especially in segmenting weak edges. As a consequence, some assessment measures that consider the slight deviation of the detected edges from the ground truth, similar to the work by Luo *et al.* [28], should be further investigated in the future.

The development of algorithms, which do not require training or validation image dataset, may continue to use a low number of test images. The number of test images between one to five images is widely accepted, as found from Section III-C of this literature survey. However, we believe that the use of more test images is better as this will indicate that their methods are robust and can work on various types of input images. We want to suggest that there should be at least three test images to evaluate the edge detection algorithm. One image is to present a simple image, one image with moderate complexity, and another one with highly complex structures or edges.

Nevertheless, with the rapid development of deep learning methods for image processing, including for edge detection purposes, researchers may be involved with deep learning, which requires a large number of images for their methods. For this purpose, the researcher can create their dataset or use the available dataset such as BSDS500 or Pascal VOC2012.

Most of the quality assessment methods found in this literature review are full-reference methods. This finding shows the importance of the ground truth. Unfortunately, Section III-D of this literature review shows that most authors did not explain how they obtained their ground truth. Generally, the ground truth is presented by a binary image, where each pixel is labeled either as edgel or as non-edgel. It is worth noting that not all the edges from the image are the desired output edges. Therefore, some explanation on how the edges are classified into useful and non-useful edges should be provided. In the future, we would like to suggest that the ground truth is not binary but should contain information about useful edges, non-useful edges, and non-edges (i.e., three categories).

Furthermore, suppose the ground truth is created by humans. In order to reduce the effect of inter-human variation, more than one human operator is needed to mark the edges for the ground truth. Therefore, in the future, research on how to combine edge segments from multiple human operators should be conducted.

We also found that the number of benchmark methods that are less than five methods is acceptable in Section III-E. Well-known edge detection algorithms such as Canny, Sobel, and Prewitt are among the popular choice for the benchmark methods. We found out that the comparison with the state-of-art method is not a requirement for researches in this area. This observation is maybe because the developed algorithms are specific to a particular type of application only. However, in the future, we would like to encourage researchers to compare the performance of the developed algorithm with state-of-art methods. For example, suppose the work is to introduce a new improvement to Canny edge detection. In that case, the comparison should not only done towards the original Canny edge detector but should also include the recently improved Canny versions.

V. CONCLUSION

To conclude, from these 73 publication samples, we found that there are 17 quality assessment methods used by researchers to evaluate the performance of their edge detection algorithms. These assessment methods can be generally classified into three categories, which are (i) subjective evaluation, (ii) no-reference image quality assessment methods, and (iii) full-reference image quality assessment methods. Most of the assessment methods are under full-reference methods.

This literature review reveals that there is still no standard method to evaluate edge detection algorithm. There is also no standard in terms of the number of test images and the benchmark methods used. Most of the researchers also not explain how they obtain the ground truth.

Based on this review, we have suggested that the evaluation methods combine at least four approaches to cover four main evaluation aspects. The number of the test images should be at least three to cover at least three levels of edge complexity. For the ground truth, we suggested that the ground truth be created by more than one human and contains three classes (i.e., useful edges, non-useful edges, and non-edges). We also encourage the researchers to compare their methods with recent related methods. More works are needed in developing a standard quality assessment method for edge detection algorithms.

REFERENCES

[1] F. Garcia-Lamont, J. Cervantes, A. López, and L. Rodriguez, "Segmentation of images by color features: A survey," *Neurocomputing*, vol. 292, pp. 1–27, May 2018.

[2] M. R. H. M. Adnan, A. M. Zain, H. Haron, M. Z. C. Azemin, and M. Bahari, "Consideration of Canny edge detection for eye redness image processing: A review," in *Proc. IOP Conf., Mater. Sci. Eng.*, 2019, vol. 551, no. 1, Art. no. 012045.

[3] R. Yaacob, C. D. Ooi, H. Ibrahim, N. F. N. Hassan, P. J. Othman, and H. Hadi, "Automatic extraction of two regions of creases from palmprint images for biometric identification," *J. Sensors*, vol. 2019, pp. 1–12, Jan. 2019.

[4] T.-S. Liu, R.-X. Liu, P. Zeng, and S.-W. Pan, "Improved Canny algorithm for edge detection of core image," *Open Autom. Control Syst. J.*, vol. 6, no. 1, pp. 426–432, 2014.

[5] H.-Y. Yin and L. Tang, "Crack detection of highway prestressed concrete bridge based on image recognition," in *Proc. IEEE Int. Conf. Ind. Appl. Artif. Intell. (IAAI)*, Dec. 2020, pp. 82–87.

[6] S. Dorafshan, R. J. Thomas, and M. Maguire, "Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete," *Construct. Building Mater.*, vol. 186, pp. 1031–1045, Oct. 2018.

[7] S. Patel, P. Trivedi, V. Gandhi, and I. G. Prajapati, "2D basic shape detection using region properties," *Int. J. Eng. Res. Technol.*, vol. 2, no. 5, pp. 1147–1153, 2013.

[8] P. Sikchi, N. Beknalkar, and S. Rane, "Real-time cartoonization using Raspberry Pi," *Int. J. Comput. Technol.*, vol. 1, no. 6, pp. 284–287, 2014.

[9] S. K. Katiyar and P. V. Arun, "Comparative analysis of common edge detection techniques in context of object extraction," 2014, *arXiv:1405.6132*. [Online]. Available: http://arxiv.org/abs/1405.6132
[10] A. Shihab, "Comparative study among Sobel, Prewitt and Canny edge detection operators used in image processing," *J. Theor. Appl. Inf. Technol.*, vol. 96, p. 9, Oct. 2018.
[11] K. Kumar, N. Mustafa, J.-P. Li, R. A. Shaikh, S. A. Khan, and A. Khan, "Image edge detection scheme using wavelet transform," in *Proc. 11th Int. Comput. Conf. Wavelet Actiev Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2014, pp. 261–265.
[12] T. Gebäck and P. Koumoutsakos, "Edge detection in microscopy images using curvelets," *BMC Bioinf.*, vol. 10, no. 1, p. 75, Dec. 2009.
[13] X. Yang, Q. Zhang, X. Yang, Q. Peng, Z. Li, and N. Wang, "Edge detection in Cassini astronomy image using extreme learning machine," in *Proc. MATEC Web Conf.*, vol. 189, 2018, Art. no. 06007.
[14] Z. Zhang, Y. Liu, T. Liu, Y. Li, and W. Ye, "Edge detection algorithm of a symmetric difference kernel SAR image based on the GAN network model," *Symmetry*, vol. 11, no. 4, p. 557, Apr. 2019.
[15] Z. Qu, S.-Y. Wang, L. Liu, and D.-Y. Zhou, "Visual cross-image fusion using deep neural networks for image edge detection," *IEEE Access*, vol. 7, pp. 57604–57615, 2019.
[16] S. Morton, A. Berg, L. Levit, and J. Eden, *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC, USA: National Academic Press, 2011.
[17] D. Moher, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Ann. Internal Med.*, vol. 151, no. 4, p. 264, Aug. 2009.
[18] P. Lasserre, B. Cutt, and J. Moffat, "Edge-detection of the radiation field in double exposure portal images using a curve propagation algorithm," *J. Appl. Clin. Med. Phys.*, vol. 9, no. 4, pp. 3–16, Sep. 2008.
[19] G. Sun, A. Zhang, J. Ren, J. Ma, P. Wang, Y. Zhang, and X. Jia, "Gravitation-based edge detection in hyperspectral images," *Remote Sens.*, vol. 9, no. 6, p. 592, Jun. 2017.
[20] T. Rink, W. P. Menzel, P. Antonelli, T. Whittaker, K. Baggett, L. Gumley, and A. Huang, "Introducing HYDRA: A multispectral data analysis toolkit," *Bull. Amer. Meteorol. Soc.*, vol. 88, no. 2, pp. 159–166, Feb. 2007.
[21] S.-C. Chen and C.-C. Chiu, "Texture construction edge detection algorithm," *Appl. Sci.*, vol. 9, no. 5, p. 897, Mar. 2019.
[22] A. A. Yahya, J. Tan, B. Su, K. Liu, and A. N. Hadi, "Image edge detection method based on anisotropic diffusion and total variation models," *J. Eng.*, vol. 2019, no. 2, pp. 455–460, Feb. 2019.
[23] Q. Wang and G. Zhang, "Ore image edge detection using HOG-index dictionary learning approach," *J. Eng.*, vol. 2017, no. 9, pp. 542–543, Sep. 2017.
[24] H. Xu and Y. Xiao, "A novel edge detection method based on the regularized Laplacian operation," *Symmetry*, vol. 10, no. 12, p. 697, Dec. 2018.
[25] D. Sundani, S. Widiyanto, Y. Karyanti, and D. T. Wardani, "Identification of image edge using quantum Canny edge detection algorithm," *J. ICT Res. Appl.*, vol. 13, no. 2, pp. 133–144, 2019.
[26] E. Dong, Y. Zhao, X. Yu, J. Zhu, and C. Chen, "An improved NMS-based adaptive edge detection method and its FPGA implementation," *J. Sensors*, vol. 2016, Dec. 2015, Art. no. 1470312.
[27] L. Huang, X. Yu, and X. Zuo, "Edge detection in UAV remote sensing images using the method integrating Zernike moments with clustering algorithms," *Int. J. Aerosp. Eng.*, vol. 2017, Feb. 2017, Art. no. 1793212.
[28] S. Luo, J. Yang, Q. Gao, S. Zhou, and C. A. Zhan, "The edge detectors suitable for retinal OCT image segmentation," *J. Healthcare Eng.*, vol. 2017, pp. 1–13, Jan. 2017.
[29] V. Bogdan, C. Bonchis, and C. Orhei, "Custom dilated edge detection filters," *J. WSCG*, vol. 2020, no. 2020, pp. 161–168, 2020.
[30] A. Khunteta and D. Ghosh, "Edge detection via edge-strength estimation using fuzzy reasoning and optimal threshold selection using particle swarm optimization," *Adv. Fuzzy Syst.*, vol. 2014, Dec. 2014, Art. no. 365817.
[31] G. Xu, Y. Zhao, R. Guo, B. Wang, Y. Tian, and K. Li, "A salient edges detection algorithm of multi-sensor images and its rapid calculation based on PFCM kernel clustering," *Chin. J. Aeronaut.*, vol. 27, no. 1, pp. 102–109, Feb. 2014.
[32] Z. Zheng, B. Zha, Y. Xuchen, H. Yuan, Y. Gao, and H. Zhang, "Adaptive edge detection algorithm based on grey entropy theory and textural features," *IEEE Access*, vol. 7, pp. 92943–92954, 2019.
[33] R. Al-Jarrah, M. Al-Jarrah, and H. Roth, "A novel edge detection algorithm for mobile robot path planning," *J. Robot.*, vol. 2018, Jan. 2018, Art. no. 1969834.
[34] L. Ma and R. C. Staunton, "A modified fuzzy C-means image segmentation algorithm for use with uneven illumination patterns," *Pattern Recognit.*, vol. 40, no. 11, pp. 3005–3011, Nov. 2007.
[35] Y. Guo and A. Şengür, "A novel image edge detection algorithm based on neutrosophic set," *Comput. Electr. Eng.*, vol. 40, no. 8, pp. 3–25, Nov. 2014.
[36] J. Yoon, "A new Bayesian edge-linking algorithm using single-target tracking techniques," *Symmetry*, vol. 8, no. 12, p. 143, Dec. 2016.
[37] B. Ergen, "A fusion method of Gabor wavelet transform and unsupervised clustering algorithms for tissue edge detection," *Sci. World J.*, vol. 2014, Mar. 2014, Art. no. 964870.
[38] M. Mittal, A. Verma, I. Kaur, B. Kaur, M. Sharma, L. M. Goyal, S. Roy, and T.-H. Kim, "An efficient edge detection approach to provide better edge connectivity for image analysis," *IEEE Access*, vol. 7, pp. 33240–33255, 2019.
[39] P. Arulpandy and M. T. Pricilla, "Salt and pepper noise reduction and edge detection algorithm based on neutrosophic logic," *Comput. Sci.*, vol. 21, no. 2, pp. 193–209, Apr. 2020.
[40] Z. Zheng, B. Zha, H. Yuan, Y. Xuchen, Y. Gao, and H. Zhang, "Adaptive edge detection algorithm based on improved grey prediction model," *IEEE Access*, vol. 8, pp. 102165–102176, 2020.
[41] L. Yang, C. Xia, and C. Juan, "Image edge detection based on Gaussian mixture model in nonsubsampled contourlet domain," *J. Electr. Comput. Eng.*, vol. 2016, Jul. 2016, Art. no. 4125909.
[42] C. Yu, F. Ji, X. Jing, and M. Liu, "Dynamic granularity matrix space based adaptive edge detection method for structured light stripes," *Math. Problems Eng.*, vol. 2019, Mar. 2019, Art. no. 1959671.
[43] K. Mendhurwar, S. Patil, H. Sundani, P. Aggarwal, and V. Devabhaktuni, "Edge-detection in noisy images using independent component analysis," *ISRN Signal Process.*, vol. 2011, Apr. 2011, Art. no. 672353.
[44] R. Fu, "Accurate detection of image edge: Comparison of different algorithms," *Int. J. Mechatronics Appl. Mech.*, vol. 1, no. 7, pp. 119–123, 2020.
[45] C. Chi and F. Gao, "Palm print edge extraction using fractional differential algorithm," *J. Appl. Math.*, vol. 2014, Apr. 2014, Art. no. 896938.
[46] F. Ren, B. Li, and Q. Chen, "Single parameter logarithmic image processing for edge detection," *IEICE Trans. Inf. Syst.*, vol. E96.D, no. 11, pp. 2437–2449, 2013.
[47] E. Sert and D. Avci, "A new edge detection approach via neutrosophy based on maximum norm entropy," *Expert Syst. Appl.*, vol. 115, pp. 499–511, Jan. 2019.
[48] J. Tang, Y. Wang, C. Huang, H. Liu, and N. Al-Nabhan, "Image edge detection based on singular value feature vector and gradient operator," *Math. Biosci. Eng.*, vol. 17, no. 4, pp. 3721–3735, 2020.
[49] W. Araby, A. Madian, M. Ashour, I. Farag, and M. Nassef, "Radiographic images fractional edge detection based on genetic algorithm," *Int. J. Intell. Eng. Syst.*, vol. 11, no. 4, pp. 158–166, Aug. 2018.
[50] Z. Hu, C. Deng, Y. Shao, and C. Wang, "Edge detection method based on lifting B-spline dyadic wavelet," *Int. J. Performability Eng.*, vol. 15, no. 5, p. 1472, 2019.
[51] M. Sudhakara and M. Janaki Meena, "An edge detection mechanism using L*A*B color-based contrast enhancement for underwater images," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, pp. 41–48, 2019.
[52] H. A. Kurdi, L. J. Grantner, and M. I. Abdel-Qader, "Fuzzy logic based hardware accelerator with partially reconfigurable defuzzification stage for image edge detection," *Int. J. Reconfigurable Comput.*, vol. 2017, Mar. 2017, Art. no. 1325493.
[53] Z. Xiaoli, "Edge detection algorithm based on multiscale product with Gaussian function," *Procedia Eng.*, vol. 15, pp. 2650–2654, Jan. 2011.
[54] W. K. Pratt, *Digital Image Processing*, 2 ed. New York, NY, USA: Wiley, 1993.
[55] E. Chen, J. Wang, L. Qi, and W. Lv, "A novel multiscale edge detection approach based on nonsubsampled Contourlet transform and edge tracking," *Math. Problems Eng.*, vol. 2015, Feb. 2015, Art. no. 504725.
[56] A. Elaraby and D. Moratal, "A generalized entropy-based two-phase threshold algorithm for noisy medical image edge detection," *Scientia Iranica*, vol. 24, no. 6, pp. 3247–3256, 2017.
[57] C. Gonzalez, P. Melin, and O. Castillo, "Edge detection method based on general type-2 fuzzy logic applied to color images," *Information*, vol. 8, no. 3, p. 104, Aug. 2017.
[58] S. Liu, X. Geng, Y. Zhang, S. Zhang, J. Zhang, Y. Xiao, C. Huang, and H. Zhang, "An edge detection method based on wavelet transform at arbitrary angles," *IEICE Trans. Inf. Syst.*, vol. E101.D, no. 9, pp. 2392–2400, 2018.

[59] B. O. Sadiq, E. O. Ochia, O. S. Zakariyya, and A. F. Salami, ''On the accuracy of edge detectors in number plate extraction,'' *Baltic J. Modern Comput.*, vol. 7, no. 1, pp. 19–30, 2019.
[60] P. G. Nes, ''Fast multi-scale edge-detection in medical ultrasound signals,'' *Signal Process.*, vol. 92, no. 10, pp. 2394–2408, Oct. 2012.
[61] C. Liu and C. Ren, ''Research on coal-rock fracture image edge detection based on Tikhonov regularization and fractional order differential operator,'' *J. Electr. Comput. Eng.*, vol. 2019, May 2019, Art. no. 9594301.
[62] Y. Ma, H. Ma, and P. Chu, ''Demonstration of quantum image edge extration enhancement through improved Sobel operator,'' *IEEE Access*, vol. 8, pp. 210277–210285, 2020.
[63] M. D. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer, ''A robust visual method for assessing the relative performance of edge-detection algorithms,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 12, pp. 1338–1359, Dec. 1997.
[64] H. Zhang, Z. Wen, Y. Liu, and G. Xu, ''Edge detection from RGB-D image based on structured forests,'' *J. Sensors*, vol. 2016, Jul. 2016, Art. no. 5328130.
[65] J. Cao, L. Chen, M. Wang, and Y. Tian, ''Implementing a parallel image edge detection algorithm based on the otsu-canny operator on the Hadoop platform,'' *Comput. Intell. Neurosci.*, vol. 2018, 2018, Art. no. 3598284.
[66] S. Alpert, M. Galun, B. Nadler, and R. Basri, *Detecting Faint Curved Edges in Noisy Images* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 6314. Berlin, Germany: Springer, 2010, pp. 750–763.
[67] C.-J. Sze, H.-Y. M. Liao, H.-L. Hung, K.-C. Fan, and J.-W. Hsieh, *Multi-scale Edge Detection Via Normal Changes* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 1310. Berlin, Germany: Springer, 1997, pp. 22–29.
[68] G. Yang and F. Xu, ''Research and analysis of image edge detection algorithm based on the MATLAB,'' *Procedia Eng.*, vol. 15, pp. 1313–1318, Jan. 2011.
[69] X. Liu, X. Wang, and Z. Duan, ''Canny edge detection based on iterative algorithm,'' *Int. J. Secur. Appl.*, vol. 8, no. 5, pp. 41–50, 2014.
[70] H. M. Asghari and B. Jalali, ''Edge detection in digital images using dispersive phase stretch transform,'' *Int. J. Biomed. Imag.*, vol. 2015, Mar. 2015, Art. no. 687819.
[71] K. K. Singh, M. K. Bajpai, and R. K. Pandey, ''A novel approach for edge detection of low contrast satellite images,'' *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 40, pp. 211–217, Mar. 2015.
[72] Z. Wei and J. Qiao, ''Industrial flame edge detection algorithm based on gray dominant filter,'' *Int. J. Multimedia Ubiquitous Eng.*, vol. 10, no. 2, pp. 87–96, Feb. 2015.
[73] N. Mathur, S. Mathur, and D. Mathur, ''A novel approach to improve Sobel edge detector,'' *Procedia Comput. Sci.*, vol. 93, pp. 431–438, Jan. 2016.
[74] R. Liu and J. Mao, ''Research on improved Canny edge detection algorithm,'' in *Proc. MATEC Web Conf.*, vol. 232, no. 4, p. 03053, 2018.
[75] L. Liu, F. Liang, J. Zheng, D. He, and J. Huang, ''Ship infrared image edge detection based on an improved adaptive canny algorithm,'' *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 3, Mar. 2018, Art. no. 155014771876463.
[76] C. Shi, ''Edge detection algorithm based on color space variables,'' *Int. J. Performability Eng.*, vol. 14, no. 5, pp. 885–890, 2018.
[77] Z.-X. Wang and W. Wang, ''The research on edge detection algorithm of lane,'' *EURASIP J. Image Video Process.*, vol. 2018, no. 1, pp. 1–9, Dec. 2018.
[78] H. Xu, X. Xu, and Y. Zuo, ''Applying morphology to improve canny operator's image segmentation method,'' *J. Eng.*, vol. 2019, no. 23, pp. 8816–8819, Dec. 2019.
[79] K. Zhang, Y. Zhang, P. Wang, Y. Tian, and J. Yang, ''An improved Sobel edge algorithm and FPGA implementation,'' *Procedia Comput. Sci.*, vol. 131, pp. 243–248, Jan. 2018.
[80] I. Bouganssa, M. Sbihi, and M. Zaim, ''Laplacian edge detection algorithm for road signal images and FPGA implementation,'' *Int. J. Mach. Learn. Comput.*, vol. 9, no. 1, pp. 57–61, 2019.
[81] M. A. Kabir and M. R. H. Mondal, ''Intensity gradient based edge detection for pixelated communication systems,'' *J. Eng.*, vol. 2019, no. 12, pp. 8463–8470, Dec. 2019.
[82] E. A. Sekehravani, E. Babulak, and M. Masoodi, ''Implementing Canny edge detection algorithm for noisy image,'' *Bull. Electr. Eng. Informat.*, vol. 9, no. 4, pp. 1404–1410, Aug. 2020.
[83] K. Swargha and P. Rodrigues, ''Performance amelioration of edge detection algorithms using concurrent programming,'' *Procedia Eng.*, vol. 38, pp. 2824–2831, Jan. 2012.
[84] Y. Shi, Y. Gu, L.-L. Wang, and X.-C. Tai, ''A fast edge detection algorithm using binary labels,'' *Inverse Problems Imag.*, vol. 9, no. 2, pp. 551–578, 2015.
[85] M. Baştan, S. S. Bukhari, and T. Breuel, ''Active Canny: Edge detection and recovery with open active contour models,'' *IET Image Process.*, vol. 11, no. 12, pp. 1325–1332, Dec. 2017.
[86] T. S. Gunawan, I. Z. Yaacob, M. Kartiwi, N. Ismail, N. F. Za'bah, and H. Mansor, ''Artificial neural network based fast edge detection algorithm for MRI medical images,'' *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 7, no. 1, pp. 123–130, 2017.
[87] Y. Jing, J. Liu, Z. Liu, and H. Cao, ''Fast edge detection approach based on global optimization convex model and split bregman algorithm,'' *Diagnostyka*, vol. 19, no. 2, pp. 23–29, 2018.
[88] L. Zixin, L. Jinyue, L. Yang, and W. Liming, ''A fast tool edge detection method based on Zernike moments algorithm,'' *IOP Conf., Mater. Sci. Eng.*, 2018, vol. 439, no. 3, Art. no. 032106.
[89] A. Ezzaki, M. Lhoussaine, M. E. Ansari, F.-A. Moreno, R. Zenouhi, and J. G. Jimenez, ''Edge detection algorithm based on quantum superposition principle and photons arrival probability,'' *Int. J. Electr. Comput. Eng.*, vol. 10, no. 2, pp. 1655–1666, 2020.
[90] I. Yoon, H. Joung, and J. Lee, ''Zynq-based reconfigurable system for real-time edge detection of noisy video sequences,'' *J. Sensors*, vol. 2016, Aug. 2016, Art. no. 2654059.

NAZISH TARIQ received the B.S. degree in biomedical engineering from the Sir Syed University of Engineering and Technology, Pakistan, in 2011, and the M.E. degree in industrial control and automation from Hamdard University, Pakistan, in 2017. She is currently pursuing the Ph.D. degree in electrical and electronic engineering with Universiti Sains Malaysia (USM). She is also a Research Associate with the Office of Research and Innovation and Commercialization, Dow University of Health Sciences, Pakistan. Her research interests include digital image processing, signal processing, and deep learning.

ROSTAM AFFENDI HAMZAH received the B.Eng. degree majoring in electronic engineering from the Universiti Teknologi Malaysia, and the M.Sc. degree majoring in electronic system design engineering and the Ph.D. degree majoring in image processing from Universiti Sains Malaysia. He is currently a Senior Lecturer with the Universiti Teknikal Malaysia Melaka.

THEAM FOO NG received the B.Sc. degree (Hons.) in mathematics and the M.Sc. degree in statistics from Universiti Sains Malaysia (USM), Malaysia, and the Ph.D. degree from the University of New South Wales (UNSW), Australia. He worked with the School of Electrical and Electronic Engineering, Engineering Campus, USM, Nibong Tebal, Malaysia. He was a permanent Representative on behalf of CGSS with the Division of Industry and Community Network (DICN), USM. He is currently with the Centre for Global Sustainability Studies (CGSS), USM, Malaysia, where he is also a Senior Lecturer. He is a Coordinator of the South East Asia Sustainability Network (SEASN). He has more than 30 research articles and book chapters published in established international journals and book chapters. His research interests include machine learning, pattern recognition, computational intelligence, image processing, education for sustainable development, and sustainability. He has experience as a guest editor, an associate editor, a section editor, and a reviewer for a few reputable journals/book series.

SHIR LI WANG received the bachelor's and master's degrees from Universiti Sains Malaysia, and the Ph.D. degree from the University of New South Wales Sydney, Australia. She is currently a Senior Lecturer with the Department of Computing, Faculty of Art, Computing and Industry Creative, Universiti Pendidikan Sultan Idris (UPSI). Her interests and passions in artificial intelligence motivate her to explore the potential of artificial intelligence in other domains. Her research works focus on the improvement of algorithms and problem solving on the basis of evolutionary algorithms, neural networks, and deep learning. Her current research interest includes adaptive parameters in evolutionary algorithms.

HAIDI IBRAHIM (Senior Member, IEEE) received the B.Eng. degree in electrical and electronic engineering from Universiti Sains Malaysia, Malaysia, and the Ph.D. degree in image processing from the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, U.K., in 2005. His research interests include digital image and signal processing, and analysis.