International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.14, No.1 (Jul-23)

http://dx.doi.org/10.12785/ijcds/140125

# Higher Order Textural Statistics for Object Segmentation in Unconstrained Environments

**Fatema A. Albalooshi**[1]

[1]*College of IT, University of Bahrain, Sakheer, Kingdom of Bahrain, P.O. Box 32038*

*Received 27 Nov. 2022, Revised 30 Apr. 2023, Accepted 8 May. 2023, Published 1 Jul. 2023*

**Abstract:** This paper presents an object segmentation technique that builds on the success of Seeded-Region Growing (SRG) segmentation. SRG methods are typically initialized by a single point or patch in the image that represents the object of interest. Unlike previous approaches which utilize patches of the object of interest to obtain first and second-order characteristics, the author explores the potential of higher-order textural statistical descriptors. The proposed unsupervised approach relies on both the homogeneous and heterogeneous textural characteristics of the selected object region to iteratively expand the boundary to encompass the full object. In addition, the research proposes a dynamic selection criterion for determining segmentation parameters based on patch neighborhood features. The presented experiments are conducted in unconstrained environments wherein a textural description of the object of interest is extracted and the proposed algorithm automatically segments it from the background and other captured objects in the scene. The approach is evaluated using various subsets of the PASCAL Visual Object Classes (VOC) challenge imagery. Through quantitative metrics and analysis, the proposed algorithmic framework outperforms state-of-the-art methods for segmenting objects with non-homogeneous textural descriptors from complex real-world environments.

**Keywords:** Object Segmentation, Seeded Region Growing, Textural Statistical Descriptor, Foreground Extraction, High Order Moments, Nonhomogeneous Segmentation

## 1. INTRODUCTION

The human visual interpretation system is intuitive and has an incredible ability to learn and distinguish individual objects apart from one another. Through continuous visual learning, the brain is able to identify shapes and objects in the most complex environments and under drastically varying conditions. Many computer vision algorithms attempt to study and understand the brain's intuitive process, and furthermore, they try to construct mimicking adaptable techniques. Considerable work has been done to create a computer vision algorithm that can segment out and identify objects in complex environments, however, the topic remains an active research area with many unforeseen challenges.

In many image processing and computer vision applications, segmentation is an initial preprocessing step in which the input imagery is transformed and separated into foreground and background regions. The foreground regions contain the objects and areas of interest while the background regions contain the rest of the scene clutter. Image segmentation is a challenging problem since a universal set of object descriptors does not exist. Therefore, segmentation algorithms are developed and geared to specific applications. For example, automatically segmenting bones in medical imagery [1] requires a vastly different set of features and descriptors than segmenting an airplane from the sky. The challenges of the problem are embedded within the object's shape complexity, variations in color and texture, scale and appearance changes, and scene clutter present in the image. Image segmentation plays a vital role in many computer vision applications, and it is an active area of research in the field that often focus on improving extraction accuracy, reliability, and consistency, or providing a computationally efficient result. In many cases, applications that rely on segmentation as a preprocessing step often evaluate segmentation algorithms based on the accuracy of the resultant extracted region of interest. Techniques used by human action and activity recognition algorithms are motivated by the ability to perform feature extraction and classification on the segmented object [2]. In situations where real-time video content analysis is necessary or where the segmented region is input to another real-time system, it becomes more necessary that the segmentation procedure is performed efficiently while maintaining adequate performance standards. The complexity of segmentation procedures also varies. For some applications, simple thresholding can provide adequate separation between the foreground and background, while in other scenarios adaptable parameters must be adjusted to compensate for specific challenges of the environment. In motion detection applications, for example, regions

*E-mail address: falbalooshi@uob.edu.bh*

of interest are identified by comparing consecutive video frames to determine anomalies. In such scenarios, motion-based segmentation methods [3], [4] can provide a reliable separation between the foreground object and the rest of the captured scene. In other applications, object segmentation is performed using the color and spatial information of the region of interest in addition to the motion[5]. Finally, for static scenes where the objects are not characterized by movement or anomalies, texture-based segmentation [6] can be utilized to isolate the objects of interest from the background.

This paper presents a semi-automatic object segmentation technique that utilizes textural descriptors to identify and extract objects of interest from the surrounding area in unconstrained environment. The proposed technique combines the well-established principles of geometric moments with novel multimodal Seeded Region Growing (SRG). By performing higher-order moment analysis, the algorithm is able to extract textural statistics that describe the selected region of interest. Through experimentation on the 2012 PASCAL Visual Object Classes (VOC) database [7], the authors present a quantitative evaluation and analysis that demonstrates an improvement in recall and precision over other state-of-the-art semi-automatic segmentation methods. The contributions of this paper are as follows:

- Providing a methodology to extract objects of interest from the surrounding unconstrained environment. In particular, textural descriptors allow the extraction of information that contributes to the discrimination between objects of similar color.
- Novel addition of a discriminating descriptor that provides homogeneous and heterogeneous characteristics of the object region.
- Patch-based seeded region growing is designed for instances in which a scene contains nonhomogeneous objects.
- Proposing future directions for research on patch-based segmentation.

The remainder of the paper is organized as follows. Section 2 provides a summary of the vast amount of related segmentation work and published literature. In Section 3, the proposed segmentation technique is explained in detail, including the steps involved. Section 4 illustrates the results of the segmentation algorithm along with the analysis and discussion of the findings. Finally, the conclusions and future direction of the work are presented in Section 5.

## 2. Related Works

The notion of seeded region growing was first proposed by Adams and Bischof [8] and has been extensively used to solve segmentation challenges. The region-growing concept entails that a region can be defined by characteristics exhibited by a single pixel or a small patch of pixels that is referred to as a seed.

To automate the seed selection process, Narkhede et al. [9] proposed a seeded region growing segmentation technique that utilizes edge detection to initialize the seeding process automatically. An extension of edge-based seeding utilized color edges which are obtained by integrating an edge detector and an SRG technique. Color edge centroids provide spatial object characteristics that are sensitive to the object's borders. Sun et al. [10] continued the edge-based work, but in contrast to predecessors, this technique penalized regions around edges. Sun et al. used a controllable fusion module (CFM), in which semantic region growth is utilized. To reduce the algorithm's sensitivity to an arbitrarily selected location for the initialization seed, the technique utilizes an adaptive edge loss function (AEL) to select seeds' locations. Shrivastava and Bharti [11] took a different approach to the region's growing problem. By focusing on defining automatic seed selection methods for medical imagery segmentation, they introduced different region-growing algorithms for image segmentation with automatic seed selection.

As a part of the developments that incorporate the edge and color modalities for region growing, Dalvand et al. [12] proposed a tissue-like P system to automatically select the initial seeds in complicated backgrounds. The seeded region growing segmentation method was applied and tested on complicated backgrounds, moreover, CUDA programming language and Graphic Processing Unit (GPU) are utilized to construct the suggested model in parallel.

As an extension to that work, Bangare [13] presented a fuzzy region-growing technique that incorporated approaches based on the usage of a fuzzy min-max neural network approach. To achieve an adaptable segmentation output, they employed four segmentation steps, consisting of pre-processing, dynamic modified region growing for threshold generation, texture feature generation, and region merging..

Other object statistics such as variance, texture, and homogeneity have also been explored. Through statistics, Orozco-del-Castillo et al. [14] presented a novel way to use texture-based SRG that utilized extracted features to automate the volume extraction in seismic data. The technique was tested on a variety of allochthonous salt bodies.

In an attempt to segment out multiple objects at a time, Zhang et al. [15] utilized a similar region-growing approach for multi-object segmentation. The method is based on the Chebyshev inequality and a kernel density estimation method to do background modeling. When an accurate background model is obtained, an adaptive thresholding algorithm is used to classify image pixels as foreground points, background points, or unresolved points.

With the evaluation of sensor technology, many segmentation techniques began to use alternative information sources. In their study, Gómez et al. [16] presented ASRG-IB1, an automated SRG segmentation method that is capable of segmenting both color (RGB) and multi-spectral imagery. The generation of seeds is automated by means of histogram analysis, whereby a pixel in the image is identified as a seed if its intensity values for each band lie within a specific representative range. To carry out the segmentation process,

the algorithm employs instance-based learning as a distance criterion.

In an effort to incorporate the successes of color and texture region growing, Song et al. [17] introduced a color-texture segmentation technique that utilizes a region-level Markov random field model (RMRF). This method calculates a color descriptor from the image's color space. The color distance can be computed between candidate regions. The textural information is extracted from the higher-order statistics of the image intensities. Similarly, Carson et al. [18] present an algorithm that segments images based on texture and color information. These methods have been tested on natural images with complex real-world backgrounds. In the evaluation of the proposed segmentation method, RMRF and Carson's Blobworld method are used for direct performance comparison.

All the discussed seeded region growing (SRG) techniques can be categorized as using one of two types of seeds: (1) SRG segmentation utilizing single-pixel seeds and (2) patch-based SRG segmentation. The following sections provide a detailed discussion of the differences between the two types of SRG methods.

### A. SRG Segmentation Utilizing Single Pixel Seeds

Seeded Region Growing (SRG) carries out image segmentation by considering a collection of a predefined set of seed locations as reference points. The method involves separating the image into regions based on the seed points, creating a segmentation map that begins with small pixel regions associated with the seed points. The method involves examining the pixels neighboring each seed point individually, and checking if they are similar enough to the seed based on a uniformity test. If they pass the test, they are assigned to the seed's growing region and then added to the growing region of the seed. The process of assigning neighboring pixels to the growing regions continues until all the pixels that meet the similarity criteria have been added. The initial regions are created using a seeding scheme, and the criteria for merging regions are determined beforehand based on homogeneity. Lastly, the classification is performed with the pixels in the same region that are labeled by the same symbol and the pixels in variant regions that are labeled by different symbols. All the labeled and unlabeled pixels are often referred to as the allocated and unallocated pixels, respectively [8].

The SRG algorithm can be mathematically described as follows. Let $U$ be the collection of pixels that have not yet been assigned to any region, but are adjacent to at least one of the labeled regions,

$$U = \left\{ (x,y) \notin \bigcup_{i=1}^{n} P_i | N(x,y) \cap \bigcup_{i=1}^{n} P_i \neq \phi \right\} \tag{1}$$

where $P_i$ defines the $n$ seeded regions for $i = 1, 2, 3, \cdots, n$, and $\phi$ is the empty set. Furthermore, let $N(x,y)$ be the set of eight immediate neighbors of the pixel $(x,y)$, as shown in Figure 1. For $(x,y) \in U$, if $N(x,y)$ meets just one of the $P_i$ regions, then the label of the region is defined as the index, $\psi(x,y) \in \{1, 2, \cdots, n\}$. The index is chosen in a way that ensures that the two sets have some common elements $N(x,y) \cap P_{\psi(x,y)} \neq \phi$. Also, let $\zeta(x,y)$ be defined as a distance metric to measure the dissimilarity between the pixel being tested at $(x,y)$ and its adjacent labeled region. This distance, denoted as $\zeta(x,y)$ is computed as:

$$\zeta(x,y) = |g(x,y) - g(X_i^c, Y_i^c)| \tag{2}$$

where the distance measure $\zeta(x,y)$ is calculated using the color values of the testing pixel at location $(x,y)$, denoted by $g(x,y)$, and the average color values of the homogeneous region that satisfies the criteria for the pixel, denoted by $g(X_i^c, Y_i^c)$, where $(X_i^c, Y_i^c)$ is the centroid of the region $P_i$. When the neighborhood $N(x,y)$ of a pixel meets two or more labeled regions (e.g. $P_i$ and $P_{i+1}$), the value of $\psi(x,y)$ is chosen such that $N(x,y)$ intersects with $P_i$ and the distance $\zeta(x,y)$ is minimized. The labeling is determined based on the following conditions:

$$\psi(x,y) = \min_{(x,y) \in U} \{\zeta(x,y) | j \in \{1, 2, \cdots, n\}\} \tag{3}$$

where $j$ represents the newly allocated pixel label.

The SRG process outlined is iterated until all pixels in the image have been assigned to their respective regions. The definitions presented in Equations (1) and (3) ensure that the final partition of the image is divided into a set of regions as homogeneous as possible on the basis of the given constraints. Overall, the SRG algorithms that use a single pixel seed have been proven robust and rapid in various experimental environments. However, the most significant downside to using the single pixel seeds SRG algorithm is the lack of discriminative information embedded into the content of the seed.

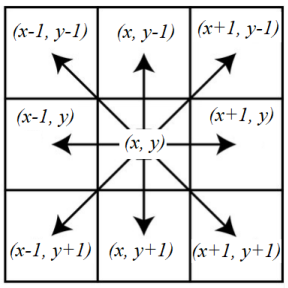



Figure 1. The SRG segmentation considers the 8-connected neighbors of a pixel $(x,y)$, denoted by $N(x,y)$.

### B. Patch-based SRG Segmentation

In many scenarios, the intensity and color variations that describe the object of interest are non-homogeneous. For these non-homogeneous objects, extracting a patch of seed points provides a more complete description of the distinguishing object's appearance. Various statistics on the selected patch can provide useful discriminatory information on the object. Previous research, as cited in [19], has investigated how effective the average brightness level of the pixels in the starting patch is. When taken

as an identifier, the mean provides a stable noise-resilient statistic that is capable of segmenting sufficiently well for selected objects and in non-cluttered environments. The patch-based SRG technique provides improved performance over the single seed regions in scenes containing various non-homogeneous objects.

## 3. Proposed Method

Considering the vast amount of research conducted, it is chosen to focus on the textural descriptor component of the SRG segmentation framework. The proposed approach introduces a method for segmenting objects that employ SRG and distinctive textural characteristics to identify the region of interest for the object in a challenging environment. A complex environment is defined as a captured scene that contains numerous objects and non-uniform backgrounds that result in challenging segmentation conditions. The objective of the proposed technique is to accurately and precisely segment out user-selected objects from the scene. The output of the proposed method can be used as an adaptive input into a classification algorithm. The effectiveness of classification using the proposed segmentation method is not explored in this paper.

The presented semi-automatic algorithm, illustrated in Figure 2, is initialized by the human operator. The selected locations are used for determining the object of interest. The locations provide single-pixel seed points or regional patch centers that are employed in the SRG segmentation. Once a location on the object of interest has been selected, the proposed segmentation method works in an automatic and unsupervised manner to complete the segmentation process.

The presented contribution to the SRG segmentation work is a novel addition of a discriminating descriptor that provides homogeneous and heterogeneous characteristics of the object region. As mentioned before, the novelty lies in uniquely utilizing the color and texture attributes for characterizing the object. Moreover, the introduced concept of utilizing seeded region growing based on patches is intended for situations where a scene consists of non-uniform objects. Utilizing a patch provides important statistics for defining the object's color characteristics. Secondly, the patch allows the extraction of texture information that allows for the discrimination between objects of similar color. The technique is evaluated and qualitatively compared to other SRG methods that use color and texture statistics. In addition to the novel SRG texture discrimination descriptor, the proposed method introduces a dynamic parameter selection scheme. Specifically, the thresholds that control the growing region expansion are derived based on local feature analysis. In the following sections, the algorithmic methodology is described.

### A. Object of Interest Seed Selection

The setup discussed in this section provides a brief overview of how the segmentation experiments and user interface were conducted. In application, the operator would be displayed an image and select an object of interest in the captured scene. The proposed technique utilizes the areas around the selected location as seed points for the SRG segmentation. Throughout the conducted experiments various operator selection methods were used to test usability and system intuitiveness. The default application uses a standard computer mouse setting where a point-and-click procedure allows the user to select the object. Interestingly, a more modern and user-friendly approach to get the operator's selection is by using an electronic touchscreen visual display that the users can control by touching the screen with a special stylus, pen, or by finger, as shown in Figure 3. The experiments confirmed earlier findings [20] that touchscreens are an intuitive way to enable the user to interact directly with what is displayed.

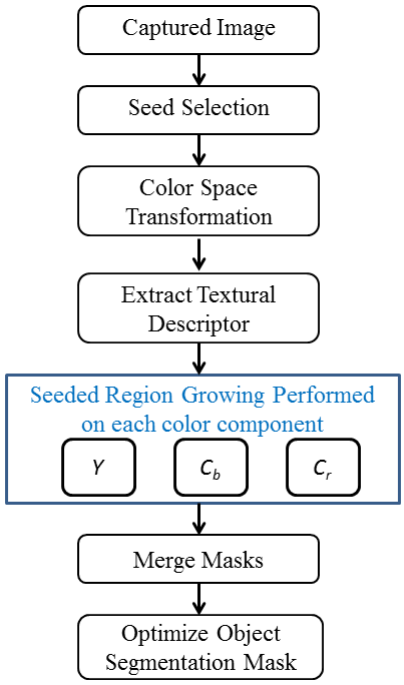



Figure 2. Algorithmic flow of the proposed object segmentation technique.

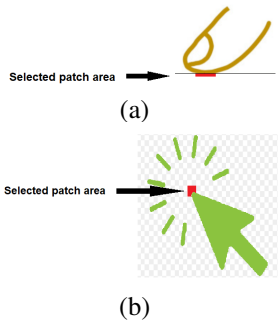



Figure 3. Illustration of an object of interest seed selection.(a) selection by electronic touchscreen visual display with a finger. (b) selection by mouse point-and-click procedure.

Technology-inclined users who interface with computers on a regular basis tended to utilize the computer mouse, while others preferred the smartphone-like interface of the touchscreen display. Thus, the system introduced the option to select the technique for identifying the object of interest to be segmented from the scene. This paper does not address the psychological analysis needed to draw conclusions on the user experience, however, it is important to note for the purpose of repeatability that the experimental setup was aimed to provide the users with the most convenient operation method.

### B. Color Space Transformation

Leveraging the success of previous multichannel segmentation techniques [21], the proposed algorithm extends the monochrome SRG technique to three-channel color images. The empirically selected $YC_BC_R$ color model is a common color representation for digital video systems [22] and it is used here to provide improved color and contrast differentiation. In this scheme, the ($Y$) element denotes the luminance, whereas the $C_b$ and $C_r$ components indicate the color difference signals. For overtness, the following values and operations were used to convert the captured RGB image to the improved $YC_BC_R$ color space,

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4}$$

The color space transformation is a key conceptual ingredient to increasing image contrast and aiding the discriminative textural descriptor.

### C. Textural descriptor

The contribution of this paper is embedded within the textual descriptor and metric described in this section. Once the captured image has been transformed to produce a $YC_BC_R$ image, each channel is processed independently. For each channel, a patch of the neighborhood surrounding the seed location is considered for determining the texture characteristics. Geometric moment theory and patch moments have a proven record and have extensively been used as textural descriptors for visual pattern recognition [23], [24]. For any particular patch the $(p+q)th$ order geometric moment, $m_{pq}$, of a single channel image $f(x;y)$ is defined as

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \tag{5}$$

Moreover, central moment $\mu_{pq}$ and normalized moment $v_{pq}$ are defined as

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-x_c)^p (y-y_c)^q f(x,y)dxdy \tag{6}$$

$$v_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\omega}} \tag{7}$$

respectively, where $(x_c, y_c)$ characterize the the coordinates of the centroid of $f(x,y)$, and $\omega = (p+q+2)/2$. on the other hand, complex moment $c_{pq}$ of image $f(x,y)$ can be defined as

$$c_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+iy)^p (x-iy)^q f(x,y)dxdy \tag{8}$$

where $i$ represents imaginary unit. The equations defining central and normalized complex moments are given in Equations (6) and (7).

It is possible to represent the complex moment using geometric moments, as follows:

$$c_{pq} = \sum_{k=0}^{p}\sum_{j=0}^{q} \binom{p}{k}\binom{q}{j} (-1)^{q-j} . i^{p+q-k-j} . m_{k+j,p+q-k-j} \tag{9}$$

and vice versa,

$$m_{pq} = \frac{1}{2^p + qi^q} \sum_{k=0}^{p}\sum_{j=0}^{q} \binom{p}{k}\binom{q}{j} (-1)^{q-j} . c_{k+j,p+q-k-j} \tag{10}$$

In contrast to prior SRG methods that use patch statistics by comparing the average intensity level of the seed's neighborhood, the proposed SRG algorithm employs moment data as a textural descriptor of the surrounding region. The moment statistic provides a more accurate interpretation of the region of interest and provides a more discriminating contrast with the non-object regions.

Once the textural statistics are computed, they are fed into the SRG framework through Equation 1. For each pixel that neighbors the seeded region, the iterative algorithm computes a distance metric between the textural descriptor of the seed patch and the textural descriptor of the area centered about the neighboring pixel. The distance measure is evaluated and classified using hard thresholds. If the difference falls below the specified threshold then the neighboring pixel is added to the seed point region. This process is repeated until there are no more pixels added to the seed regions. Algorithmic parameters, such as the initial patch size and descriptor distance thresholds, remain unchanged through the algorithmic procedure. It is important to note that in order to perform a fair algorithmic comparison, the algorithmic parameter set values must remain the same throughout the experiments. Adjusting parameters for specific targets can lead to higher and more accurate rates, but does not support the automatic or semi-automatic requirements of the presented segmentation technique.

### D. Selection of SRG Thresholds

The autonomy of the segmentation technique, after the operator has selected the target or targets of interest, is a key component of the novelty of this work. In addition to the altered patch descriptor, the proposed segmentation method presents a parameter automation scheme. The selection criterion of SRG thresholds is obtained from the selected seeded region patches. The automation scheme utilizes the independently computed textural values from each of the three channels for every selected patch. For each channel, the mean and the standard deviation value of the patch

texture are computed. The standard deviation is used as the textural threshold for the corresponding image channel. Through empirical experimentation, this property has proven to give the algorithm the ability to perform well for a wide range of objects. The automation scheme allows the segmentation algorithm to learn the variations in the texture of the selected object. The standard deviation provides a wide range of expectable variations of the object's texture without including unwanted objects or backgrounds. The range in variation is intended to capture pixels of the same texture, but whose texture descriptor may have been affected by an image artifact. This threshold selection scheme has proven to obtain great performance without the need to constantly alter thresholding parameters.

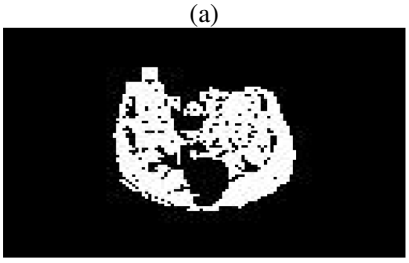
### *E. Segmentation Mask Merging*

In order to produce an optimized segmentation mask, the algorithm merges the outputs of each channel to create the final output. For each color component $YC_BC_R$ the resulting thresholded and labeled mask is considered. The proposed method utilizes median filter to merge the components into a single binary mask. That is, when two of the three channels indicate the same label for a pixel, then the merged output is assigned the majority label. The filtering results in the removal of certain falsely segmented regions as well as the removal of object pixels that have been mislabeled in the previous algorithmic stages.

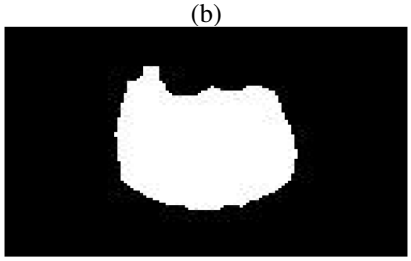
In many cases, the result of the merger appears sporadic and the object of interest appears incomplete. An illustration of the resulting merger is depicted in Figure 4. In Figure 4(a), an original image from the PASCAL VOC 2012 segmentation dataset [7] is shown, wherein a patch on the boat is selected as the object of interest. In Figure 4(b), the resulting merged mask is illustrated. The merging scheme performs well by eliminating false detections but removes several significant portions of the boat, or the object of interest, as well. To compensate for potential losses, the segmentation algorithm is assisted by the usage of morphological operations. Morphology most commonly operates on binary images. Morphological operations, such as intersection, union, and inclusion allow the segmentation algorithm to fill in or close holes in the binary mask. The closing and opening processes create a cleaner and more complete representation of the selected object mask.

The filtering processes utilize a structuring element, or kernel, to perform the operations. The closing operation is defined as an image dilation operation followed by an image erosion operation. Similarly, the opening operation is vice versa and is defined as an erosion followed by dilation. The closing and opening processes use the same structuring element for both operations. The shape and size of the structuring element play an important role in the resulting mask [25]. The segmentation algorithm presented in this paper uses a 3 x 3 structuring element. As illustrated in Figure 4(c), the morphological closing process eliminates minor gaps and concave angles in the combined mask while maintaining the shape and dimensions of the targeted object. The final binary mask is used as the algorithmic output. The following section describes how the output mask from the proposed algorithm compares to the several state-of-the-art segmentation methods.

(a)



(b)



(c)

Figure 4. The outcomes of employing morphological operations. (a) original image. (b) before morphology.(c) after morphology.

## 4. Results and Discussion

The evaluation of the proposed segmentation method is presented visually and quantitatively. A standardized image dataset was selected for testing to provide a fair comparison with other segmentation techniques. The challenging Pattern Analysis Statistical modeling and Computational Learning (PASCAL) Visual Object Classes (VOC) 2012 dataset [7] provides numerous real-world images with specified objects of interest in each image. The dataset is often used for segmentation algorithm evaluation due to the available truthed segmentation masks provided with the PASCAL VOC imagery.

The segmentation problem is evaluated as a classification algorithm, wherein each pixel is classified as being either part of the object of interest or part of the background clutter. Utilizing standard classification metrics that compare the computed results to the truthed mask, the following measures have been calculated. For each test image, the True Positive (TP) metric indicates the number of pixels that belong to the confirmed object area and were correctly identified as the object of interest by the algorithm. Secondly, the True Negative (TN) measure

represents the total number of pixels that were correctly detected as background by the algorithm. Thirdly, the False Positive (FP) measure indicates the total number of pixels that have been incorrectly detected as the object. And lastly, False Negative (FN) is the total number of pixels that are incorrectly detected as background region. As a validation of this binary classification problem, the following must be true,

$$TP + FP + TN + FN = I_M I_N \tag{11}$$

where $I_M$ and $I_N$ represent the number of rows and columns in the test image, and the product, $I_M I_N$, is the total number of pixels in the tested image and thus is the total number of classifications by the algorithm.
Although trivial, these computed metrics are necessary to derive the standard classification rates used for algorithmic comparisons. In addition to deriving the true positive, true negative, false positive, and false negative rates, additional statistics provide a more complete understanding of the segmentation algorithm's performance. Recall (RE) is a metric that evaluates the accuracy of correctly classified positive cases relative to the actual number of positive cases. It quantifies the percentage of true positive cases that the model correctly identifies out of all positive cases that are present in the dataset.

$$RE = \frac{TP}{(TP + FN)} \tag{12}$$

In the same manner, the Specificity (SP) evaluates the precision of the system in identifying negative cases and is calculated as:,

$$SP = \frac{TN}{(TN + FP)} \tag{13}$$

False negative rate (FNR)is characterized as:

$$FNR = \frac{FP}{(FP + TN)} \tag{14}$$

False positive rate (FNR), described as:

$$FPR = \frac{FN}{(FN + TP)} \tag{15}$$

Lastly, a commonly used metric of Precision (P) measures the ability of a model to correctly identify positive instances out of all instances predicted as positive, defined as

$$P = \frac{TP}{(TP + FP)} \tag{16}$$

As the metrics indicate in the following sections, the most accurate segmentation outputs result in high recall (RE), specificity (SP), and precision (P) values, while low false positive rate (FPR) and false negative rate(FNR) values.

### A. *Visual Assessment of Segmentation*

The visual assessment of the segmentation algorithm provides a performance metric that directly translates to the user's experience. This assessment allows the user to visualize the strengths and weaknesses of the segmentation. Strengths are considered as the algorithm's robustness to illumination or shadows, while weaknesses are visualized as objects missing prominent information. In Figure 5, images from the PASCAL VOC were selected to demonstrate the performance of the proposed technique. Figures 5(a) - (b) depict the original images with the user selected patched marked in green. Each image represents a type of challenging segmentation scenario. In Figure 5(a), the object of interest is the weaved basket. The texture of the basket distinguishes it from the background clutter, however, the shadows cast within the basket cause segmentation challenges. On the contrary, in Figure 5(b), the object of interest is a single train section and takes up a small portion of the image. Visually, the output presented in Figures

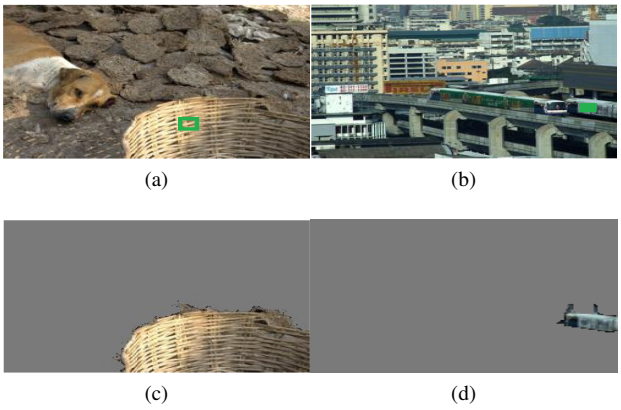
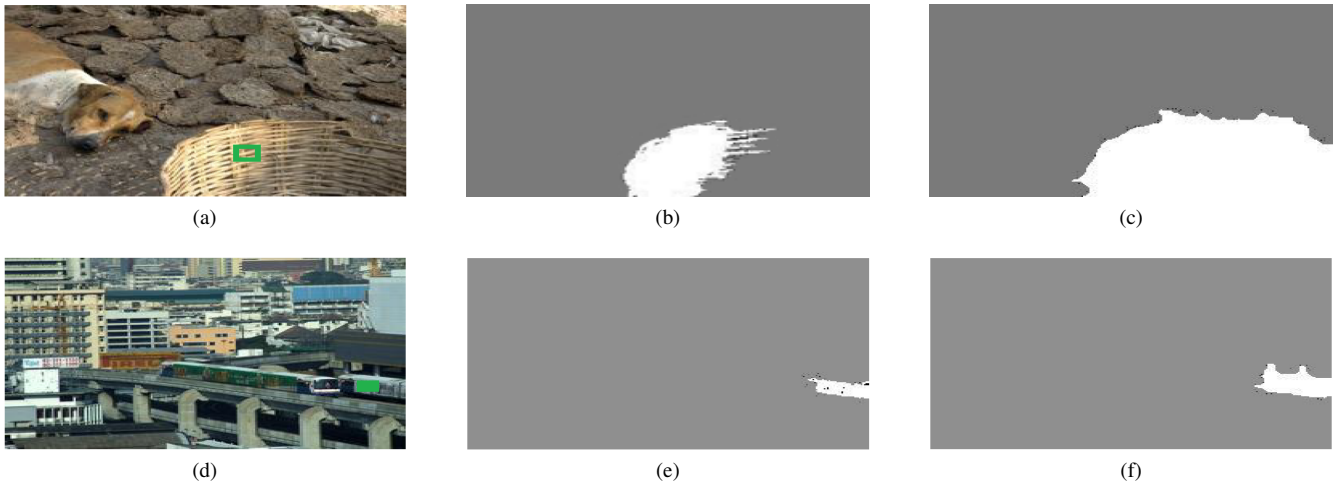



Figure 5. Visual assessment of the performance of the proposed segmentation technique. For each scene, the user's seed selection is marked in green. (a), (b) Challenging segmentation environment. (c), (d) The resulting segmentation output.

5(c) - (d) can be interpreted as successful with limited amounts of false positives and false negatives visible in the segmented outputs. The performance of the proposed method is compared to the original SRG method [8] in Figure 6, where the same images and same seed locations are used to compute the illustrated masks. Figures 6(a)& (d) present input images with the selected seed patched marked in green. In Figures 6(b) & (e) the computed segmentation output mask of the traditional SRG approach [8] is depicted. In comparison, Figures 6(c) & (f) represent the segmentation masks computed using the proposed technique. Further visual analysis of the segmentation masks reveals that the traditional SRG approach underperforms the proposed technique in all scenarios. In the case of the weaved basket, the traditional SRG technique fails to segment portions of the basket in the shadow. Moreover, the traditional SRG segmentation method omits several parts of the train. The following section will provide quantifiable metrics to describe the performance of the proposed technique.

### B. *Quantitative Evaluation*

The visual assessment highlights the algorithm's abilities to adapt to various scenes. In order to get a quantifiable



(a) (b) (c)

(d) (e) (f)

Figure 6. Comparison of the segmentation mask outputs. (a)& (d) The input image with the selected seed patch is marked in green. (b) & (e) The output mask generated using a traditional SRG approach [8]. (c)& (f) The output mask computed using the proposed segmentation technique.

comparison, the algorithms must be independently compared to a ground truth mask. As mentioned previously, the images utilized in the quantitative evaluation are obtained from the PASCAL VOC dataset.

In Figures 7 and 8 the proposed technique outperforms the SRG technique in all metrics. Each scene is quantified using the recall (RE), specificity (SP), and precision (P) values. Moreover, false positive rate (FPR), and false negative rate (FNR) assessment is also being made. For all scenarios, the proposed method is compared to the traditional SRG approach [8]. Extension of visual assessment is presented in Table I. Metrics in which the proposed technique outperformed the traditional SRG algorithm are marked in bold.

Interestingly, the SRG traditional method produces a slightly better specificity false negative rate (FNR). Note, the proposed technique significantly outperforms the traditional SRG approach in all other metrics.

The calculated metrics for a particular scene may be misleading due to irregularities in the algorithm's performance. To obtain an objective assessment of the proposed technique, the evaluation involved executing the segmentation algorithms on a group of 100 images from the PASCAL VOC database. Each image in the test set was initialized with the same seed region. The output segmentation masks generated by each technique were compared to a manually segmented reference mask. In all the quantified metric categories (with the exception of FNR), the proposed technique outperformed the traditional SRG algorithm. Note that on average the large differences between the metrics occur in the specificity (SP), false positive rate (FPR), and precision (P). Those metrics translate to high levels of false alarms.

The next evaluation compares the proposed algorithm to several state-of-the-art texture-based segmentation algorithms. The RMRF [17] technique and Carson's Blobworld algorithm [18] are breakthrough segmentation methods that have been tested on the PASCAL VOC database. Both RMRF and Carson's Blobworld works are seeded region growing techniques that use the color and texture information from the selected target. A secondary test subset of 80 images was selected from the PASCAL VOC data. The four algorithms evaluated are the traditional SRG [8], RMRF [17], Blobworld [18], and the proposed segmentation technique. All techniques utilized identical initialized seed patches. Table II shows the mean recall (RE) and mean precision (P) metrics computed from the masks generated by the four segmentation techniques on the test dataset. As expected, the traditional SRG approach shows the lowest recall and lowest precision of the compared methods. The RMRF edged out Blobworld in the mean recall with significantly higher mean precision. The proposed segmentation technique proved to have the highest recall of 70% with the highest precision of 88%, indicated in bold. Further analysis indicates that Blobworld, RMRF, and the proposed method had a similar number of misses or false negatives. However, the proposed method on average had a significantly small number of false positives. It is the belief of the authors that this phenomenon occurred due to the discriminatory and distinguishing properties of the higher-order textural descriptors.

In Figure 9, a region of convergence (ROC) graph shows the performance of each technique. Represented as a dashed magenta, the traditional SRG technique hits a 58% false alarm rate before completely detecting the object of interest. Similarly, the Blobworld technique, in green, reaches a high detection rate quick, but hovers at 95%. The RMRF technique, in red, and the proposed segmentation method, in blue, performed similarly with the proposed edging out RMRF. The strength of the proposed method is that it is able to reach high detection rates of 73% with a minimal number of false alarms. The ROC curves confirm the recall

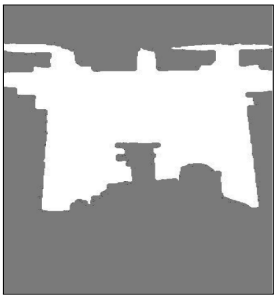
TABLE I. Quantitative performance analysis of the proposed technique for a Subset of 100 images.

| Item | Method | RE | SP | FPR | FNR | P |
|---|---|---|---|---|---|---|
| PASCAL VOC | SRG [8] | 0.73 | 0.44 | 0.60 | **0.27** | 0.45 |
| | Proposed | **0.78** | **0.92** | **0.08** | 0.28 | **0.90** |

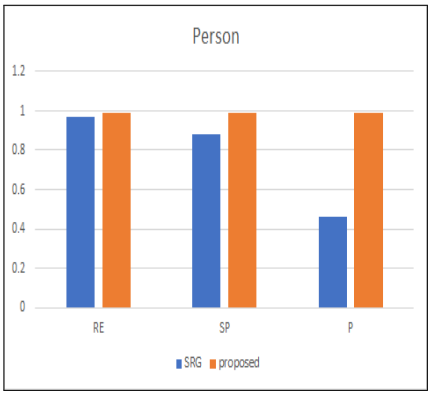
(a)

(b)

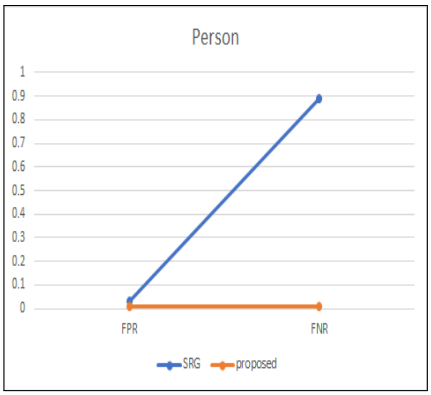
(a)



(b)

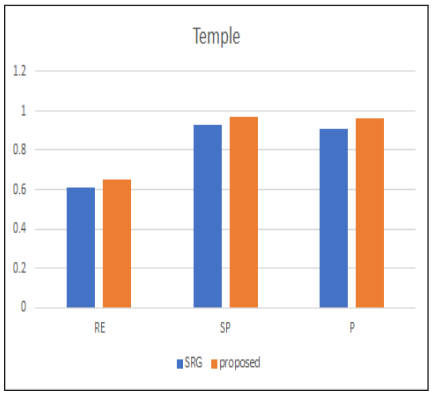



(c)

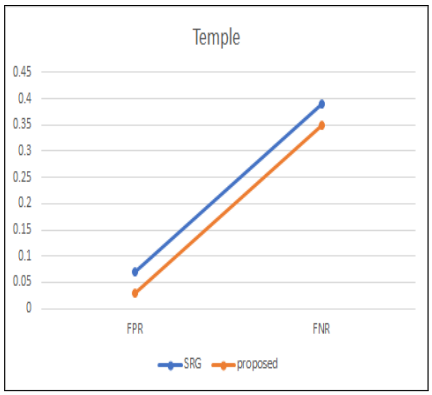



(d)

Figure 7. Comparison of the quantitative evaluation for a person image.(a) The input image with the selected seed patch is marked in green. (b) The desired mask. (c) Quantitative results of both SRG [8] and proposed method. (d) Error rate comparison of both SRG [8] and proposed method.




(c)




(d)

Figure 8. Comparison of the quantitative evaluation for a temple image.(a) The input image with the selected seed patch is marked in green. (b) The desired mask. (c) Quantitative results of both SRG [8] and proposed method. (d) Error rate comparison of both SRG [8] and proposed method.

and precision analysis presented in Table II.

TABLE II. Statistical performance comparison. Techniques were tested on a selected subset of 80 images from the PASCAL VOC database.

| | Recall mean | Precision mean |
|---|---|---|
| SRG [8] | 0.61 | 0.45 |
| Blobworld [18] | 0.68 | 0.63 |
| RMRF [17] | 0.69 | 0.81 |
| Proposed | **0.70** | **0.88** |

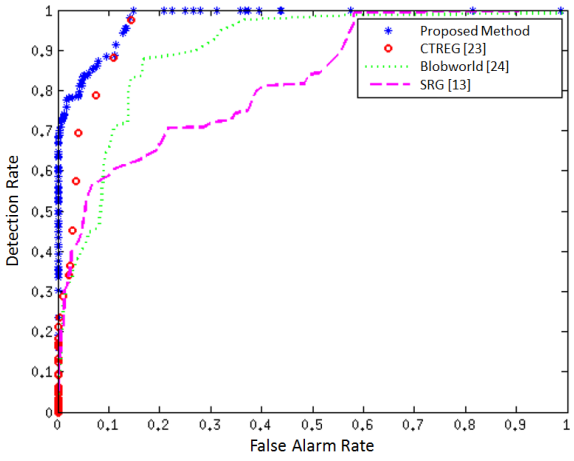



Figure 9. The region of convergence (ROC) graph evaluation and performance comparison.

## 5. Conclusion

In this paper, an efficient higher-order textural-based region-growing algorithm for object segmentation has been proposed. The driving force behind this study was the need for diverse surveillance applications, such as object recognition, change detection, and target tracking. The system introduced in this study is constructed on the accomplishments of seeded region-growing procedures, and it utilizes a user-selected seed patch to extract textural information associated with the object of interest, which can work perfectly when a scene contains a nonhomogeneous object. Following that, a conversion of color space and geometrical analysis of higher order on the patch enables the identification of distinct boundaries of the targeted object. Finally, an adaptive thresholding scheme and morphological filtering allow the proposed method to accurately segment out a complex object from a complex environment. The proposed method was evaluated on the PASCAL visual object classes challenge image dataset. Through visual assessment and quantitative analysis, the proposed technique has proven superior to the traditional SRG algorithm as well as state-of-the-art enhanced SRG techniques.

Upcoming research is focused on examining a distinct set of images, particularly those that query new attributes from the over-segmented areas. Moreover, to enhance segmentation outcomes through the inclusion of additional textural feature descriptors, leveraging a neural network to choose texture features and employing other similarity criteria while carrying out the region growing process.

## References

[1] L. Bonaldi, A. Pretto, C. Pirri, F. Uccheddu, C. G. Fontanella, and C. Stecco, "Deep learning-based medical images segmentation of musculoskeletal anatomical structures: A survey of bottlenecks and strategies," *Bioengineering*, vol. 10, no. 2, 2023. [Online]. Available: https://www.mdpi.com/2306-5354/10/2/137

[2] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions," *Knowledge-Based Systems*, vol. 223, p. 106970, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705121002331

[3] F. A. Albalooshi and V. K. Asari, "Adaptive segmentation technique for automatic object region and boundary extraction for activity recognition," in *Proceedings of the SPIE Conference on Defense, Security, and Sensing: Visual Information Processing XXI*, vol. 8399, 2012, pp. 839 907–839 917.

[4] S. Almeer, F. Albalooshi, and A. Alhajeri, "Oil spill detection system in the arabian gulf region: An azure machine-learning approach," in *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2021, pp. 418–422.

[5] K. Lee and J. Jeong, "Multi-color space network for salient object detection," *Sensors*, vol. 22, no. 9, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/9/3588

[6] X. Zhao, G. Wang, Z. He, and H. Jiang, "A survey of moving object detection methods: A practical perspective," *Neurocomputing*, vol. 503, pp. 28–48, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222008359

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/ // challenges/VOC/voc2012/workshop/index.html. (accessed:01.09.2022).

[8] R. Adams and L. Bischof, "Seeded region growing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 6, pp. 641–647, 1994.

[9] P. R. Narkhede and A. V. Gokhale, "Color image segmentation using edge detection and seeded region growing approach for cielab and hsv color spaces," in *Industrial Instrumentation and Control (ICIC), 2015 International Conference on*. IEEE, 2015, pp. 1214–1218.

[10] X. Sun, M. Xia, and T. Dai, "Controllable fused semantic segmentation with adaptive edge loss for remote sensing parsing," *Remote Sensing*, vol. 14, no. 1, p. 207, 2022.

[11] N. Shrivastava and J. Bharti, "Automatic seeded region growing image segmentation for medical image segmentation: a brief review," *International Journal of Image and Graphics*, vol. 20, no. 03, p. 2050018, 2020.

[12] M. Dalvand, A. Fathi, and A. Kamran, "Flooding region growing: a new parallel image segmentation model based on membrane computing," *Journal of Real-Time Image Processing*, vol. 18, pp. 37–55, 2020.

[13] S. L. Bangare, "Classification of optimal brain tissue using dynamic region growing and fuzzy min-max neural network in brain magnetic resonance images," *Neuroscience Informatics*, vol. 2, no. 3, p. 100019, 2022, multimedia-based Emerging

Technologies and Data Analytics for Neuroscience as a Service (NaaS). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772528621000194

[14] M. Orozco-del Castillo, M. Cárdenas-Soto, C. Ortiz-Alemán, C. Couder-Castañeda, J. Urrutia-Fucugauchi, and A. Trujillo-Alcántara, "A texture-based region growing algorithm for volume extraction in seismic data," *Geophysical Prospecting*, vol. 65, no. 1, pp. 97–105, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2478.12381

[15] K. Zhang, C. Wang, and B. Wang, "A multi-object segmentation algorithm based on background modeling and region growing," in *Advances in Neural Networks – ISNN 2012*, ser. Lecture Notes in Computer Science, J. Wang, G. Yen, and M. Polycarpou, Eds. Springer Berlin Heidelberg, 2012, vol. 7367, pp. 106–115.

[16] O. Gómez, J. González, and E. Morales, "Image segmentation using automatic seeded region growing and instance-based learning," in *Progress in Pattern Recognition, Image Analysis and Applications*, ser. Lecture Notes in Computer Science, L. Rueda, D. Mery, and J. Kittler, Eds. Springer Berlin Heidelberg, 2007, vol. 4756, pp. 192–201.

[17] X. Song, L. Wu, and G. Liu, "Unsupervised color texture segmentation based on multi-scale region-level markov random field models," , vol. 43, no. 2, pp. 264–269, 2019.

[18] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 8, pp. 1026–1038, Aug 2002.

[19] P. Bansal, A. Singhal, and K. Gehlot, "Osteosarcoma detection from whole slide images using multi-feature non-seed-based region growing segmentation and feature extraction," *Neural Processing Letters*, pp. 1–23, 2022.

[20] H. Nam, K.-H. Seol, J. Lee, H. Cho, and S. W. Jung, "Review of capacitive touchscreen technologies: Overview, research trends, and machine learning approaches," *Sensors*, vol. 21, no. 14, p. 4776, 2021.

[21] D. A. Abdelsadek, M. N. Al-Berry, H. M. Ebied, and M. Hassaan, "Impact of using different color spaces on the image segmentation," in *International Conference on Advanced Machine Learning Technologies and Applications*. Springer, 2022, pp. 456–471.

[22] T.-Y. Sun, S.-J. Tsai, and V. Chan, "Hsi color model based lane-marking detection," in *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*. IEEE, 2006, pp. 1168–1172.

[23] A. Hjouji, J. EL-Mekkaoui, M. Jourhmane, and B. Bouikhalene, "New set of non-separable orthogonal invariant moments for image recognition," *Journal of Mathematical Imaging and Vision*, vol. 62, no. 4, pp. 606–624, 2020.

[24] R. Benouini, I. Batioua, K. Zenkouar, and S. Najah, "Fractional-order generalized laguerre moments and moment invariants for grey-scale image analysis," *IET Image Processing*, vol. 15, no. 2, pp. 523–541, 2021.

[25] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2007.

**Fatema A. Albalooshi** short biography Dr. Albalooshi is chairperson at the Computer Engineering Department, College of IT, University of Bahrain. She pursued her Ph.D degree from the University of Dayton (USA). Dr. Albalooshi always contributed as an active member of the Computer Vision and Wide Area Surveillance Laboratory (UD Vision Lab) at the University of Dayton. Dr. Albalooshi had completed her Master of Science degree at the University of Nottingham (United Kingdom) in Electronic Communications and Computer Engineering. Specializing in computer vision research, her primary research focus is on the field of image processing and object segmentation where she has completed several publications in the related areas. Dr. Albalooshi's other research interests include electronic noses where she had published a book entitled "Electronic Nose Technologies and Advances in Machine Olfaction". Moreover, her research interests are in cybersecurity and authentication, object recognition and tracking, image enhancement, neural networks, medical imagery segmentation, e-noses, and 3D reconstruction.