

RESEARCH ARTICLE

Efficient Multi-Object Recognition Using GMM Segmentation Feature Fusion Approach

AYSHA NASEER¹, HAMDAN A. ALZHRANI², NOUF ABDULLAH ALMUJALLY³,
KHALED AL NOWAISER⁴, NAIF AL MUDAWI⁵, ASAAD ALGARNI⁶,
AND JEONGMIN PARK⁷

¹Department of Computer Science, Air University, Islamabad 44000, Pakistan

²College of Computing and Informatics, Saudi Electronic University, Riyadh 11673, Saudi Arabia

³Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

⁴Department of Computer Engineering, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

⁵Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia

⁶Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Saudi Arabia

⁷Department of Computer Engineering, Tech University of Korea, Siheung-si, Gyeonggi-do 15073, South Korea

Corresponding author: Jeongmin Park (jmpark@tukorea.ac.kr)


This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2024-RS-2022-00156326) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation). This research is supported and funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Group Funding program grant code (NU/RG/SERC/12/40). The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FFR-2024-231-03.”

ABSTRACT Machines need to be able to recognize and understand complex visual surroundings to function at their best in a variety of contexts. Here, we address the difficult problem of multi-object recognition to obtain a sophisticated knowledge of complex visual environments, tackling issues such as size, occlusion, fluctuations in object traits, and complicated backdrops. Our contribution is to provide novel methods (Gaussian mixture model and mean-shift algorithms) for inferring multiple object segmentation in complicated visuals, introducing a unique multiclass object classification strategy utilizing benchmark datasets. Notably, by utilizing local appearance, texture, and geometry characteristics, our technique considerably improves classification accuracy by integrating a Multi-Layer Perceptron (MLP) with area signatures and local descriptors. By facilitating accurate object matching and identification based on local appearance, texture, and geometric features, local descriptors are essential for collecting particular information and regions of interest in images. When compared to state-of-the-art methods, empirical validation on MSRC and Corel 10k datasets shows better performance, especially when managing object occlusion problems. With an accuracy of 90.6% and 89.69%, respectively, our suggested system performs better than industry standards for multi-object classification on both datasets, highlighting the significant progress our method makes to the area of multiclass object classification in challenging visual contexts.

INDEX TERMS Gaussian mixture model, mean-shift segmentation, feature fusion, saliency map, multi-object categorization, local descriptors.

I. INTRODUCTION

The Multi-object detection is essential for granting machines the ability to comprehend and experience complicated visual scenes. Multiple objects of interest must be located and

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

identified simultaneously inside images or video streams to complete the task. Machines must be able to discern various objects in visual input in order to interact intelligently with their surroundings. Real-world scenes usually contain a range of objects with different sizes, shapes, orientations, and spatial relationships [1], [2]. Traditional object detection methods centered around locating a single instance of

an object within an image, but there is now a fast-growing need for systems that can handle numerous objects at once. Numerous applications, such as autonomous driving, surveillance [3], [4], medical imaging, and robotics [5], [6], depend heavily on these activities. Machines can comprehend their surroundings, make wise decisions, and communicate with the environment when they can detect and segment objects [7]. One of the primary issues regarding multi-object detection is coping, with the complexity and diversity of object appearances [8], [9]. Objects may display differences because of varying illumination, perspectives, occlusions, and background clutter [8]. The inclusion of interesting objects with varying scales and aspect ratios may also increase the task's complexity. It is vital to develop reliable algorithms that can consistently locate numerous objects in a range of situations in order to deal with these challenges effectively. By giving each unique object in an image or video a pixel-level label, object segmentation accomplishes the task.

Segmentation offers a more precise understanding of object borders and their geographic extent than object detection, which detects the bounding box of objects [9]. However, handling variations in object appearance, occlusions, scale, and crowded backgrounds presents challenges for object detection. It is not easy to get precise and consistent detection in many environments [10]. On the other hand, there are issues with precisely delineating object boundaries during object segmentation, mainly when there are complicated shapes, overlapping objects, and partial occlusions. The precise segmentation of objects at the pixel level remains a research challenge [11]. Finally, difficulties with object categorization include deviations between object appearance, class imbalance, a lack of labeled data, and generalizing to unseen objects [12]. It is not easy to create models that can categorize objects accurately under multiple circumstances and adjust to new object classification. In this article, we outline a method for categorizing multiple objects to perform multi-object detection using benchmark datasets to address the difficulties in classifying more than objects in images. The proposed system preprocesses the images in the first phase, we efficiently segment images using two segmentation methods, Mean Shift Segmentation (MSS) and Gaussian Mixture Model (GMM). The output of the two algorithms is compared and examined in the second stage. The final stage involves analyzing the detection of different regions and matching the local descriptors and signatures [13] of the regions of the images to classify various objects [14]. A vital aspect of computer vision and artificial intelligence is multi-object categorization. It adheres to the needs of a wide range of industries and applications, including computer vision tasks like object recognition and detection, vehicular autonomy for secure navigation and avoiding hazards, robotics manipulation, security, and surveillance needs to identify potential threats, augmented reality and virtual reality for seamless incorporation of virtual elements, medical imaging for

diagnosis and treatment planning, automated content moderation and more [15]. Overall, multi-object categorization is essential for automating processes, increasing productivity, and improving user experiences across a wide range of sectors and domains.

In this research, we strive to overcome the shortcomings of existing approaches by incorporating new segmentation techniques and a unique multiclass object classification strategy. A novel methodology is suggested for object classification that may result in improved performance over current techniques by combining MLP with area signatures and local descriptors. This novel approach encourages the pursuit of novel directions in the domain of object segmentation and categorization. The highlights and crucial contributions made in this work are presented below.

In this article, feature detectors, local descriptors, and multilayer perceptron have been employed for object categorization to produce more precise findings. Two segmentation techniques, the Gaussian Mixture Model (GMM) and the Mean Shift Segmentation (MSS), are used to segment data effectively. The result obtained from the two algorithms is compared and analyzed. A feature detector called Maximally Stable Extremal Regions (MSER) is used to identify areas in an image that display consistent and recognizable intensity variations across several scales. Robust and effective feature matching across several images or scenes is made possible by merging feature detectors and descriptors. The distribution of gradient orientations in various local image regions is determined by the Histogram of Oriented Gradients (HOG) descriptor. A multilayer Perceptron (MLP) is employed for multi-object categorization in an image. We compare two benchmark datasets that are publicly accessible: MSRC and Corel 10k datasets for multi-object detection and categorization:

- We use the Mean Shift Segmentation (MSS) and the Gaussian Mixture Model (GMM) as segmentation strategies for effective data segmentation. The outputs from both methods are then compared in the next phase.
- We generate a saliency map by deriving the color, texture, and spatial features of an image.
- Next, we employed a variety of feature extractors and MSER to compute local descriptors that were all generated from the saliency map. Finally, we merged all of the features we had acquired.
- The assessments and experimental findings are performed based on the GMM algorithm and MLP approach on two different datasets.

The rest of the paper is structured as follows: Section II is a full architecture discussion of the proposed multi-object categorization paradigm. Section III discusses previous systems reviews and methodologies. Section IV discusses the evaluations and experimental results based on MLP methodology and GMM algorithm. Lastly, Section V unveils the paper's conclusion, parameters, and considerations for future study.

II. LITERATURE REVIEW

According to references [17], [18], researchers have recently introduced a variety of computer vision algorithms for image segmentation and recognition. The first part of the pertinent study discusses aspects like identifying objects and image segmentation, while the second part focuses on multi-object classification.

A. OBJECT SEGMENTATION

Different researchers have used numerous models to segment and detect objects [19]. Hu et al. [20] suggested an improved RCNN network that combines Region Regression and RCNN to accurately recognize different objects in pictures. The potential region's size must be modified, though. Region mean pooling was employed by Kuan et al. [21] to identify objects in the context; however, this approach involves computing complexity and expense, particularly for big images. A three-stage video tracking approach, which includes detection, tracking, and assessment using MoAG segmentation, was developed by Mahalingam and Subramoniam [22]. However, it might not sufficiently handle occlusions, complicated backdrops, and shifting illumination conditions. Owing to its substantial processing demands, this approach might not be appropriate for scenarios requiring limited resources or real-time applications. Chahyati and Arymurthy [23] make use of mean shift as the primary tracker when the target object was not hidden. To enhance the tracking results, the particle Kalman filter was selected as the primary tracker. The system was not effective when occlusion is present. RetinaNet is used by Zhang and Zhang [24] to determine the various target objects, and the Hungarian method is then applied to detect them. RetinaNet is well known to govern object instances accurately at varying scales, which makes it relevant to a variety of applications. By methodically assigning and associating identified objects, the Hungarian method's inclusion improves accuracy and fortifies the system's overall efficacy. The system's accuracy is constrained by the notion that a complex crowd lessens the detection accuracy. Accurate object recognition [25] can be hampered by complex crowd situations' occlusions, overlapping objects, and complicated patterns. This means that in cluttered situations, the system's performance may be less dependable, indicating a possible disadvantage when used in scenarios with a high object density or complexity. A. Jalal et. al in their research employed region based segmentation with decision trees to detect the object and got 86.01 accuracy. Naseer and Jalal [26] used fuzzy c-mean to segment the object from the image with 86.61 accuracy.

B. MULTI-OBJECT CATEGORIZATION

Many academics have invested their time and efforts in developing systems for object categorization using various methods. A thorough summary of studies pertinent to these models is revealed below. The fluid force concept and scene perception are used to find anomalies by Di et al. [27]. One

class SVM made a choice based on the collected features after extracting fluid and appearance features. The proposed model was ineffective in various situations, therefore, the method's ability to adapt to different circumstances needs to be improved. Lecumberry et al. [28] developed a measure of form similarity and used the steepest descent minimization technique to simulate the iterative shaping of each object. They automatically classified a variety of things using energy optimization. When working with many objects or complicated scenarios, energy optimization strategies may have scaling problems. The optimization challenge becomes increasingly challenging and time-consuming as the number of objects rises. In addition to labeled object categorization algorithms, Shi et al. [29] suggested the MVFL-VC approach. The approach's potential in classifying labeled objects and taking label correlations into account is acknowledged, but its practical application is noted as perhaps limited. Its reliance on a particular dataset for training has drawn criticism as it could make it less flexible to adapt to different datasets or situations. Due to its limits in terms of generality and adaptation to different settings, the technique may not be as useful as it may be, which highlights the need for careful thought in real-world applications. A context-based technique for identifying saliency zones in images has been offered by Prabu et al. [30]. They generated featurepoints deploying a CNN model and tested them on five datasets (Label Me, UIUC-Sports, Scene-15, MIT67, and SUN) before using them, but they only performed well indoors. In an outdoor context, the system accuracy will be significantly reduced, when dealing with complex environments with cluttered backgrounds, occlusions, or numerous salient objects. The method of Shi et al. aids in identifying study gaps in outdoor settings with complicated surroundings, crowded backdrops, occlusions, or plenty of conspicuous things. This lays the groundwork for the significance of our actions in resolving these problems. Dong et al. [46] compare two distinct segmentation approaches then for object categorization they used multi kernel learning.

Comparing the methodology employed in this work to cutting-edge object recognition and segmentation techniques, there are many positive aspects. By combining classic techniques like the Gaussian Mixture Model (GMM) for segmentation, the Histogram of Oriented Gradients (HOG), the Maximally Stable Extremal Regions (MSER), and Accelerated-KAZE (AKAZE) for feature extraction with the use of a Multilayer Perceptron (MLP) for detection, our approach offers a distinctive blend of conventional and contemporary methods. This hybrid technique may make segmentation and detection assessments less complicated to understand due to its interpretability. Because of the inherent adaptability and fine-tuning aspects of our methodology, it is possible to create a framework that is customized to specific dataset properties and even to gradually progress towards more sophisticated techniques.

Having been considered, the method offers an acceptable alternative for accurate and resource-efficient object

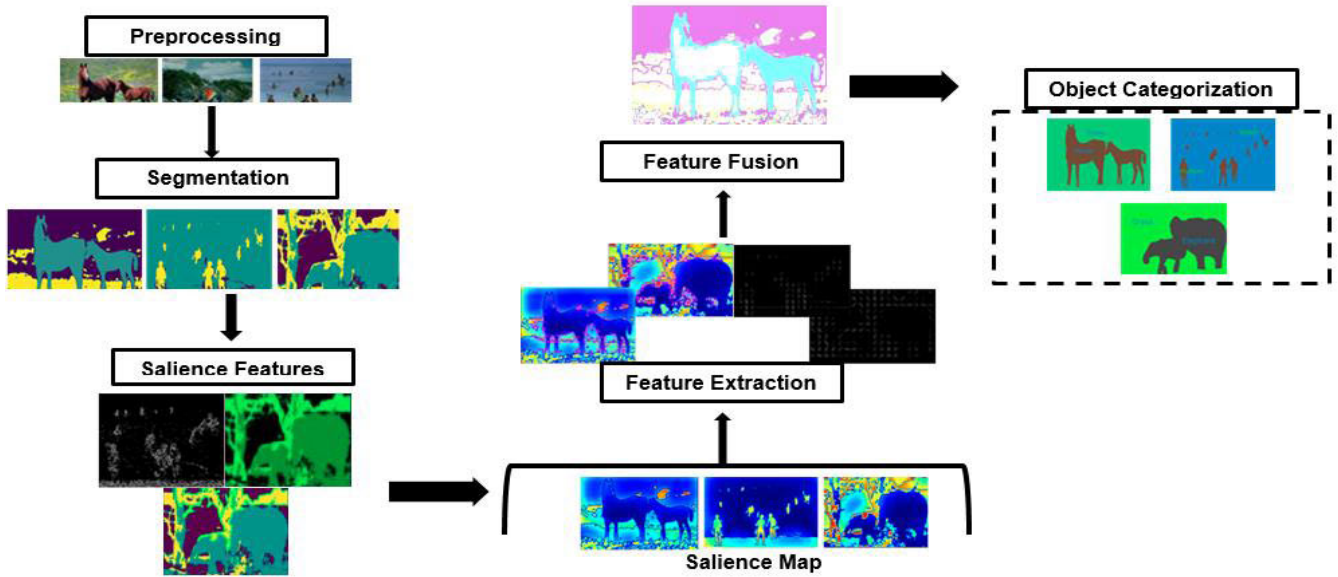


FIGURE 1. Block diagram of the proposed multi-object categorization system.

identification and segmentation by balancing the advantages of modern and conventional techniques.

III. THE PROPOSED MULTI-OBJECT DETECTION AND CATEGORIZATION

In this article, we present an innovative approach for categorizing objects that correctly identifies and labels every target object shown in the image. The first steps of the proposed system include preprocessing, eliminating unwanted details, such as noise elements, and adjusting object dimensions across all data sets' images. Then retrieved data are applied to precise image segmentation using two different segmentation algorithms –the Gaussian Mixture Model(GMM) and mean shift segmentation(MSS). We are able to detect different objects of various sizes for all of the images in the datasets by conducting a comparative analysis of signatures and local descriptors inside the images and evaluating the effectiveness of the multiple areas detector. In this paper, we used the MSER feature detector for evaluating multiple different regions. In the end, Multilayer Perceptron is used to accomplish multi-object categorization. Time complexity is also shown for both datasets using two distinct segmentation methods. The architecture flow diagram for this research work is shown in more depth in Figure 1. The comprehensive schema shows step by step visual description of the proposed methodology. The next subsections provide in-depth explanations of the functions and processes of each of the aforementioned modules. Figure 1 displays an overview at a glance of the proposed model.

A. PRE-PROCESSING

The n Images are taken under various types of circumstances during preprocessing, including variable lighting and

surroundings, which can cause the images to have noise and high-intensity levels (see Figure 2a). In the adaptive weighted median filter, neighboring pixels are given adaptive weights based on weights based on their similarity to the center pixel rather than applying a fixed-weight median filter to all pixels evenly. The Adaptive Weighted Median Filter [31] reduces noise while crucial image details are maintained. It offers a versatile approach to denoising [32], either preserving or suppressing image information [33] as needed by adaptively assigning weights based on similarity. These filters employ a P x Q sliding window that moves across each image. The filtering procedure takes advantage of the image's local statistic weights. using Eq. (1).

$$Z_{(a,b)} = Z_o - \frac{c(Dis)var_{ij}^2}{Mn_{ij}} \quad (1)$$

where “c” is the scaling factor applied to the scale of the filter frame (i.e., 3 × 3), “Z_o” denotes the weight of the central pixel of the filter frame, and “Dis” denotes the Euclidean distance between pixels. The P x Q sliding window's mean and variance are represented correspondingly by the values Mn_{i,j} in Eq. (2) and var_{i,j}² in Eq. (3). Figure. 2 is representing



FIGURE 2. Preprocessed images after removal of noise using AWMF with filter frame 3 × 3.

the filtered images after applying AWMF.

$$Mn_{i,j} = \frac{1}{PQ} \sum_{a=0}^{P-1} \sum_{b=0}^{Q-1} i_{a,b} \quad (2)$$

$$var_{i,j}^2 = \frac{1}{PQ} - 1 \sum_{a=0}^{P-1} \sum_{b=0}^{Q-1} i_{a,b} - Mn_{i,j} \quad (3)$$

B. IMAGE SEGMENTATION

After the pre-processing stage, segmentation—the division of an image into multiple sections or segments—is performed, each of which corresponds to distinct objects or areas of interest in the image, it is known as image segmentation [34] for multi-object. Many techniques and algorithms are available for segmenting images with numerous objects. This paper analyzes two robust methods known for their efficacy in object segmentation: Mean Shift Segmentation (MSS) and Gaussian Mixture Model Image Segmentation (GMM) in this section.

1) MEAN SHIFT SEGMENTATION

The Mean Shift Segmentation [30] technique is used in the proposed system to segment an image into multiple areas. The MSS algorithm estimates the local pixel density and seeks the part of the image’s sample with the largest concentration of similar pixels. The MSS algorithm performs density estimation iteratively to get the minimal local value for density and then quickly shifts all pixels with a local density close to the local minimum density to clusters with comparable attributes [35]. This non-parametric clustering method doesn’t rely on prior information about the objects or image components. As a result, it can swiftly locate cluster centers and carry out effective image segmentation. In the meantime, the suggested technique calculates the minimal local density value using kernel density estimation. At a position of x , $K(z)$ is estimated in D -dimensional space S^D for n pixels z_i , where $i = 1, 2, 3, \dots, n$. The formula is as follows in Eq. (4).

$$K(z) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n D} K\left(\frac{z - z_i}{h_{z_i}}\right) \quad (4)$$

where h_{z_i} is the window function’s (kernel density’s) width, which is measurable by using Eq. (5) and Eq. (6).

$$h_1(z_i) = (1 - d(z_i)) \quad (5)$$

$$h_1(z_i) = (h * h_1(z_i)) \quad (6)$$

where h is a constant and $d(z_i)$ is the probability density function of the specified pixel space. The given requirement is met by the kernel density (window function) $K(z)$ as given in Eq. (7) and Eq. (8). Results from using the suggested MSS technique on the MSRC dataset are shown in Figure. 3.

$$\int K(z) dz = 1 \quad (7)$$

$$\int zK(z) dz = 0 \quad (8)$$



FIGURE 3. Image segmentation achieved through the mean shift segmentation.

2) GAUSSIAN MIXTURE MODEL(GMM)

A function called a Gaussian Mixture is composed of several Gaussians [36], each denoted by the $c \in 1, \dots, K$, where K denotes how many clusters there are in our dataset. The following parameters are the components of each Gaussian “ c ” in the mixture shown in Eq. (9).

- μ defines the mean of the distribution
- \sum determines the width of the covariance. This would be equal to an ellipsoid’s dimensions in a multivariate setting.
- π is a mixing probability that determines the size of the Gaussian function? The formula that follows is the probability density function of a Gaussian distribution.

$$y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (9)$$

Together, the Gaussian distributions in the GMM reflect various data components or clusters, forming a mixed model. Methods such as the Expectation-Maximization (EM) method are used to estimate the parameters of each Gaussian distribution, such as variance and mean, using the obtained data. Figure. 4 shows the pictorial representation of the pdf plot Gaussian distribution for multi-objects.

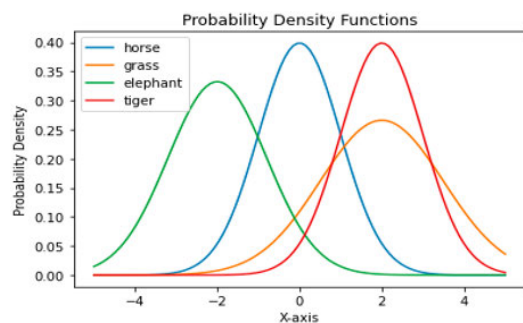


FIGURE 4. Spatial distribution of multiple objects using Gaussian.

Component weights and variances/covariance are the two values that parameterize the Gaussian Mixture Model. “ μ ” is the mean of the k th component of a Gaussian mixture model, which has “ k ” components. The mixture component weights are defined by “ k ” and are given for each component

in Eq. (10).

$$p(x) = \sum_{i=1}^k \pi_i * y \tag{10}$$

where π_i , is the mixing probability that illustrates how big or small the Gaussian function is and y is the one-dimensional pdf of a Gaussian distribution. The mixing coefficients are probabilities and must satisfy the constraint that the total probability [37] must be 1 [38] after normalization as shown in Eq. (11). The output of the suggested GMM method system’s Corel-10k dataset is shown in Figure 5.

$$\sum_{i=1}^k \pi_i = 1 \tag{11}$$



FIGURE 5. Statistical patterns of segmented images by GMM.

Object segmentation accuracy done by using pixel-wise accuracy. It is a metric that calculates the proportion of pixels [39] in an image that are properly identified to all of the pixels in the entire image. Pixel wise accuracy is calculated using Eq. (12). It is simple and easy to understand. It offers an overall segmentation accuracy score for the complete image.

$$PWA = \frac{\text{Number of accurately identified pixels}}{\text{Total number of pixels}} \tag{12}$$

As a result, the suggested system evaluates the segmentation accuracy and computation time efficiency of the output of the GMM and MSS algorithms. When compared to MSS, GMM requires less computing time and yields more lucid results. We used the GMM results for additional analyses because they were more significant and superior to MSS’s performance. Saliency maps will be produced in the next phase by using different characteristics that were taken from the segmented visuals.

C. SALIENCY MAP

A saliency map is a graphic depiction of an image where specific areas are highlighted according to their prominence or significance. These highlighted areas, also known as “saliency regions,” are those that stand out from the surrounding area and most successfully draw attention [40] from the public. Various low-level features of the image, such as A saliency map is a graphic depiction of an image where specific areas are highlighted according to their prominence or significance [41]. These highlighted areas, also known as “saliency regions,” are those that stand out from the surrounding area and most successfully draw attention from the

public. Various low-level features of the image, such as color, texture, and spatial location, are examined via a saliency map to identify the salient regions.

1) COLOR FEATURES

When deciding [42] whether areas of an image are more important than others, color characteristics are extremely important. Colors that stand out and contrast with one another are more likely to catch the eye. As a result, areas with these color qualities [43] frequently stand out on the saliency map. Features of a color might include things like hue, saturation, and intensity [44]. In the saliency map, areas with distinctive color qualities are highlighted.

2) TEXTURE FEATURES

The visual patterns [45] and characteristics included in an image are referred to as texture features. Salient areas are those that stand out from their surroundings with distinctive textures [46]. To locate areas with distinctive textures, texture analysis techniques [47] can be used, such as Gabor filters or Haralick texture features. For instance, the textured surface would be highlighted in the saliency map in an image with a smooth, textured surface and a consistently plain backdrop.

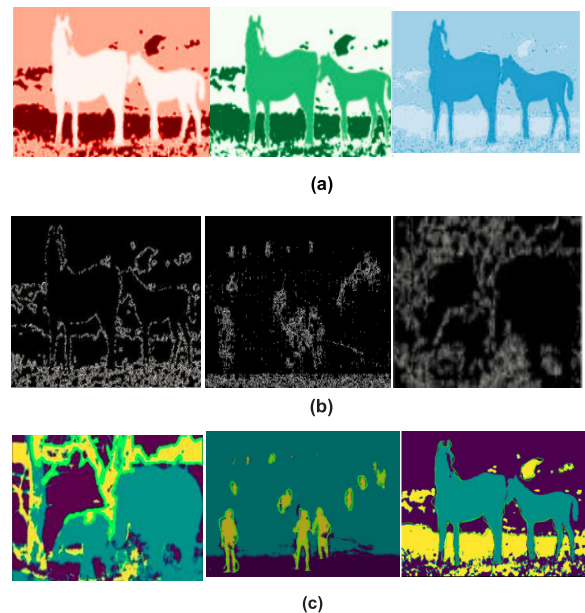


FIGURE 6. (a) Color channels extracted from color features.(b) Texture analysis of segmented images.(c) Spatial patterns depicted from spatial features.

3) SPATIAL FEATURES

The positioning [48] and arrangement [49] of areas or objects inside an image are referred to as spatial characteristics [50]. The centroid [51], bounding box [52], and geometric characteristics [53] of objects are notable spatial attributes. Objects that are positioned differently from the norm or that are at the center of the visual frequently develop alienation. Saliency

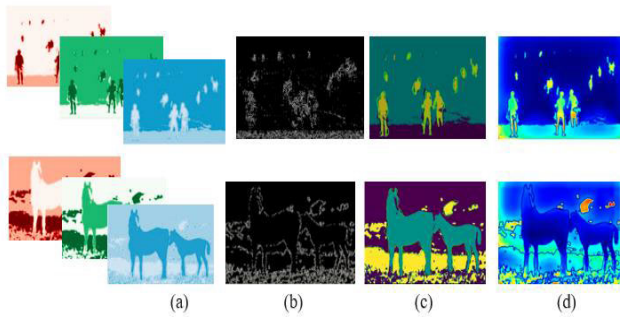


FIGURE 7. Saliency maps (a) Color channels (b) Texture analysis (c) Spatial analysis (d) Saliency features.

maps of the segmented images by combining all feature maps (color + texture + spatial) are shown in Figure. 7.

D. FEATURE DETECTOR AND LOCAL DESCRIPTORS

To get better results, two distinct types of features are extracted and integrated. we applied a feature detector to segmented images. Feature detectors pinpoint certain areas of interest in images or data. These detectors have been designed to endure various situations, including variations in scale, rotation, and lighting. They seek distinguishing characteristics that can serve as reference points [54] for additional research, such as corners [55], edges [56], and blobs [57].

1) FEATURE DETECTOR: MAXIMALLY STABLE EXTREMAL REGIONS

MSER works by identifying areas in an image that keep steady across various thresholds [38]. These areas frequently correspond to blobs or regions with consistent gradients in texture or intensity. Based on the variance in the region’s intensity or gray-level values under various threshold levels, the algorithm below defines stability.

Detected features from saliency maps by maximally stable extremal regions are shown in Figure. 8. The scale parameter is time t, and bigger values result in simpler visual representations. Fig. 8 shows the results of KAZE features.

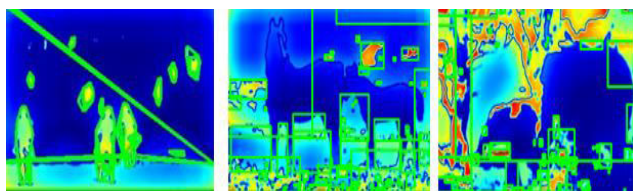


FIGURE 8. Stable regions are depicted by MSER feature detector.

2) HISTOGRAM OF ORIENTED GRADIENTS (HOG)

The HOG (Histogram of Oriented Gradients) local descriptor [39] is a feature descriptor in the vision field. By dividing images into small, overlapping cells and producing histograms of gradient orientations inside these cells. It extracts local gradient information from an image by breaking it

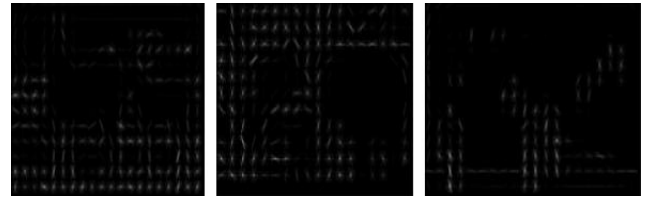


FIGURE 9. Gradient Orientations depicted using HOG.

up into tiny, overlapping cells and producing histograms of gradient orientations inside them. The horizontal and vertical gradients are used to determine the HOG Descriptor. This technique requires a single feature kernel, which can count the “h” gradient from the image’s (H_h) (see Eq. 13) and (H_v) (see Eq. 14) gradients using a Sobel operation. Features extracted using HOG have been shown in Figure. 9.

$$H_h = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad (13)$$

$$H_v = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (14)$$

$$H = \sqrt{H_h^2 + H_v^2} \quad (15)$$

where H_h Sobel horizontal matrix and H_v , Sobel vertical matrix. We determined by applying the square root of the gradients H_h and H_v . The value of the gradient direction H can be computed using Equation (15). To capture the varying intensities in the image, gradients are produced in the “h” and “v” dimensions. Based on the “h” and “v” gradients, each pixel’s gradient orientation (angle) and magnitude are computed. The value act as a representation of the local gradient data as define in Eq.16 and Eq. 17:

$$||h, v|| = \sqrt{(H_{h,v} - H_{h+1,v})^2 + (H_{h,v} - H_{h,v+1})^2} \quad (16)$$

$$\emptyset(h, v) = atan2[(H_{h,v} - H_{h+1,v}) \cdot (H_{h,v+1} - H_{h,v})] \quad (17)$$

3) ACCELERATED-KAZE(AKAZE)

The feature extraction technique known as AKAZE, or Accelerated-KAZE, is used in computer vision to find and characterize key points or interest points in digital images. AKAZE stands out for its effectiveness and adaptability. It is appropriate for tasks demanding swift and precise feature extraction since it is built to operate effectively in real-time applications. AKAZE has scale and rotation invariance, which allows it to identify and characterize key points in an image regardless of their size or orientation. Additionally, it is effective in circumstances involving object tracking since it is resistant to affine transformations like shearing and perspective distortions. AKAZE computes descriptors that store details about the immediate picture region surrounding each key point in addition to identifying key points. For applications like image matching, object identification, and

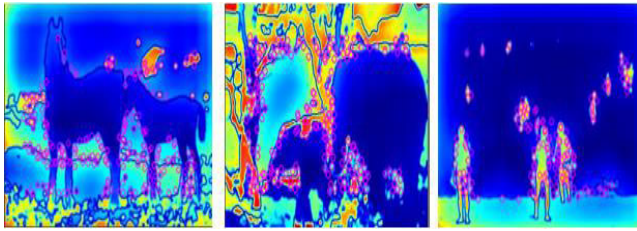


FIGURE 10. Corners, edges and blobs extracted by AKAZE.

image stitching, this makes it indispensable. The output showcases the detected features by AKAZE in Figure 10.

4) FEATURE FUSION

In this section, we combine the previously acknowledged qualities, including attributes obtained from HOG (F_{hog}), MSER (F_{mser}), and AKAZE (F_{akaze}) characteristics. According to reference [40], the feature vectors are first normalized to provide uniformity within the merged feature vector. After being averaged, the HOG, AKAZE, and MSER features are joined together to create a full feature vector by using Eq. 18. Resultant images after fusion have been shown in Figure 11. Fused images passed to multi kernels for object categorization.

$$F_{fused} = F_{hog} + F_{akaze} + F_{mser} \quad (18)$$

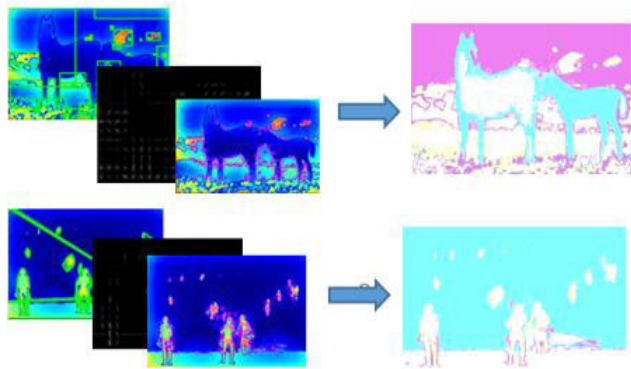


FIGURE 11. Integration of distinctive features extracted via HOG, AKAZE and MSER.

IV. OBJECT CATEGORIZATION

This section describes categorization that doesn't call on previously determined classes. Instead, it looks for naturally occurring cluster groups within a dataset based on similarities or shared characteristics. In this paper, we define how the suggested system achieves various object categorizations based on numerous areas and signatures of the regions in complex images using the Multilayer Perceptron (MLP) classifier. When categorizing objects, a region R of an image "i" (which contains clusters k of numerous items represented by different colors) is initially set for the local descriptor D_i (i.e., HOG, AKAZE) and establishes the region R of

image "i". Using a function F_R from local descriptors D_i as $F_R: D_i \leftarrow S_i$, the signature S_i is calculated. The following mathematical formula (Eq. 19,20,21) is used to derive this F_R conversion.

$$C_k = \frac{1}{|k|} \sum_{i=1}^n \sum_{j=1}^n D_{ijk} \quad (19)$$

$$\mu_k = \frac{1}{|k|} \sum_i \sum_j (D_{ijk} - C_k)(D_{ijk} - C_k)^T \quad (20)$$

$$\mu_{i,k} = \sum_j (D_{ijk} - C_k)(D_{ijk} - C_k)^T - \mu_k \quad (21)$$

where D_{ijk} stands for the descriptors of image "i" that belong to cluster k , C_k denotes the cluster's center, $|k|$ for the total number of descriptors in the clusters, "k" of each image in a class, and " μ_k " for the average of clusters k 's centered descriptors. The computation of an image's signature is represented by $\mu_{i,k}$. Then, $\mu_{i,k}$ is turned into a vector called $V_{i,k}$. The concatenation of all the signature vectors $V_{i,k}$ for each cluster in an image "i" results in the signature vector $S_i = \{V_{i,1}, V_{i,2}, \dots, V_{i,k}\}$.

A multilayer perceptron (MLP) classifier comprises multiple layers of interconnected nodes, also referred to as artificial neurons or perceptron. It is a feedforward neural network model; it denotes a single direction of data flow from the input layer to the output layer via the hidden layers [41]. Each node in an MLP classifier receives input signals from the nodes in the layer above (as shown in Figure 12), processes the weighted sum of the inputs using an activation function, and then generates an output $\hat{y}(x)$. The network can learn complicated patterns and make non-linear decisions due to the activation function, which brings non-linearities into the system. Representing objects as feature vectors and training [42] an MLP (as shown in algorithm 1) to organize them into several categories are the steps involved in using a multilayer perceptron (MLP) for categorizing objects. Any perceptron's procedure could be described as shown in Figure 12.

Algorithm 1 Multilayer Perceptron

Initialize input, output, and hidden layer, weights and biases

1. Input $x_{1,2,\dots,n}$, $y(x)$, weights $\leftarrow c$ randomly
2. DO compute $(x_{1,2,\dots,n}, \hat{y}(x))$ where $x_{1,2,\dots,n} = \{\text{set of features}\}$, $y(x)$ is original class $\hat{y}(x)$ is a label for the predicted class.

- a. Calculate output $\leftarrow \hat{y}(x)$
- b. measure error, $\emptyset \leftarrow \hat{y}(x) - y(x)$
- c. Update w by using \emptyset as given below:

$$\Delta c = c_{new} - c_{old}$$

3. Repeat 2 Until error $\delta = 0$ or start converging

$$c_{new} = c_{old} + a * \emptyset * x$$

$$a \leftarrow \text{learning rate}$$

$$\emptyset \leftarrow \text{error}$$

Return/Output: $\hat{y}(x)$

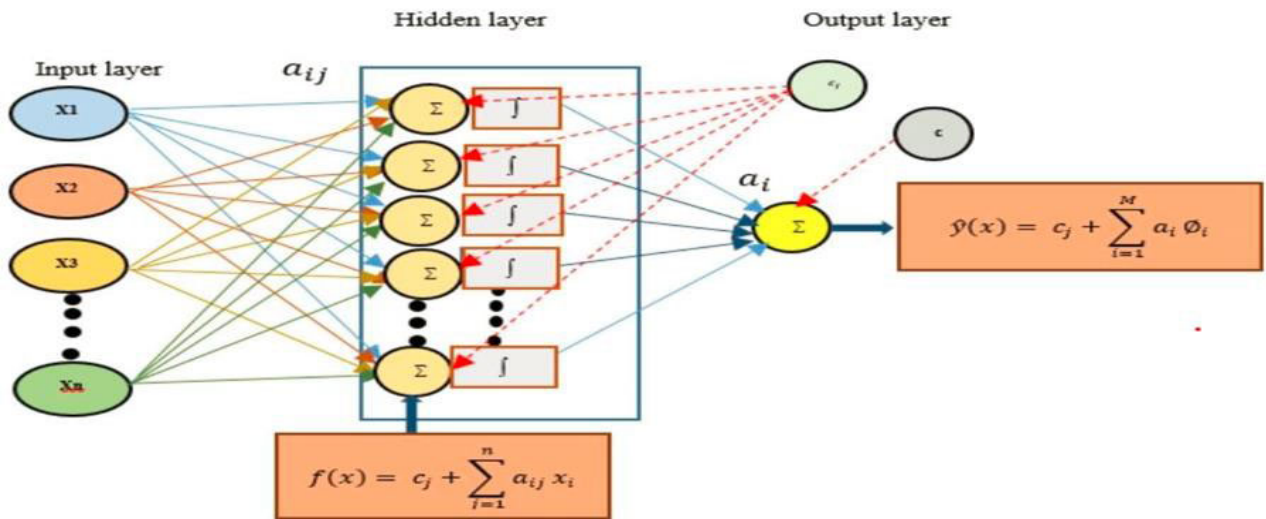


FIGURE 12. The architecture of MLP to object recognition.



FIGURE 13. Multi-objects classified and categorized based on learned features and patterns by MLP classification.

In order to attain similarity throughout the entire image, an image, however, contains many regions. Therefore, Figure. 13 illustrates the object categorization method using MLP.

V. EXPERIMENTAL SETUP AND ANALYSIS

This section provides thorough experiments that back up the proposed paradigm. Tools from MATLAB and Google Colab (Python) were applied for processing and experimentation. A 2.4GHz Intel Core i7 processor, 16GB of RAM, and Windows 10 Pro comprised the hardware setup. The tests are broken down into three categories: feature identification, local descriptor extraction from images, and segmentation evaluation using MSRC and Corel 10k datasets. In the last section, the model we recommend is contrasted with cutting-edge techniques. In the final three subsections, there are summaries of the dataset, performance metrics and findings, and corresponding remarks.

A. DATASETS DESCRIPTION

1) MSRC DATASET

The MSRC-v2 dataset [42], [52] included 591 different kinds of objects in dynamic contexts such as city structures, hilly terrain, traffic signs, and beaches. The dataset consists of 12 distinct classes, such as bike, car, cow, chair, bird, flower,

house, plane, signboard, tree, sheep, book, and building. The images in the collection have a 213×320 resolution and each image has a complex background.

2) COREL 10K DATASET

10,000 challenging images in 10 classes, with various sizes and backgrounds, are part of the Corel-10k dataset. Twenty classes, including a horse, deer, car, water, building, elephant, plane, tree, tiger, bike, wolf, dog, boat, flower, bear, sky, land, cat, bird, and fish, were the subjects of our experimental evaluations.

3) EXPERIMENT 1: EXPERIMENTAL RESULTS USING PROPOSED APPROACH

The suggested methodology has been used over the MSRC dataset. Table 1 in the form of the confusion matrix showing the accuracy for object categorization over MSRC dataset. The Corel-10k dataset is utilized to test the experiments over the proposed model and displayed using Table 2 in the form of a confusion matrix [43]. It is evident from the table that the offered methodology was able to attain an accuracy rate of 89.69% in this experiment. Table 3 gives object segmentation accuracy of Corel 10k dataset while Tables 4 and 5 give us computational time of both datasets MSRC and Corel 10k respectively.

4) EXPERIMENT 2: EVALUATION USING OTHER CONVENTIONAL

We have examined the suggested MLP-based multi-object categorization method with other existing, related systems using the Corel 10k datasets. Tables 6 and 7 provide a comparison of our method with traditional state-of-the-art procedures. When compared to other traditional models, the suggested approach has improved the mean accuracy rate to 89.69%.

TABLE 1. Confusion matrix for object categorization using MLP accuracy on the MSRC-v2 [67] dataset.

Objects	Bi	Cw	Ca	Sp	Dg	Gr	By	Du	Wr	Fl
Bi	0.91	0.0	0.0	0.3	0.0	0.0	0.2	.02	.01	0.0
Cw	0.0	0.93	0.3	0.0	.0	0.2	0.0	0.0	0.0	0.0
Ca	0.0	0.4	0.84	0.0	.04	0.1	0.0	0.0	0.0	0.0
Sp	0.3	0.0	0.0	0.92	0.0	0.0	0.2	0.0	0.0	0.0
Dg	0.0	0.2	0.3	0.0	0.89	0.0	0.0	0.0	0.0	0.0
Gr	0.0	0.2	0.4	0.0	0.3	0.91	0.0	0.0	0.0	0.0
By	0.0	0.0	0.0	0.0	0.0	0.0	0.89	0.0	0.6	0.0
Du	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.92	0.1	0.5
Wr	0.1	0.0	0.0	0.0	0.0	0.0	0.5	0.1	0.90	0
Fl	0.2	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.95
Mean accuracy = 90.60%										

*Bi = Bird, Cw = Cow, Ca = Car, Sp= Sheep, Dg = Dog, Gr = Grass, By= Bicycle, Du = Duck, Wr = Water, Fl = Flower

TABLE 2. Confusion matrix for object categorization using MLP accuracy on the Corel-10K [60] dataset.

Objects	Ho	El	Bu	Ap	Tr	Tg	Bi	Bd	Fs	Ld
Ho	.92	0.0	0.0	0.3	0.0	0.0	0.1	0.2	0.0	0.0
El	0.0	.88	0.2	0.0	0.1	0.2	0.0	0.0	0.0	0.0
Bu	0.0	0.1	.84	0.0	0.4	0.2	0.0	0.0	0.0	0.0
Ap	.03	0.0	0.0	.90	0.0	0.0	0.2	0.0	0.0	0.0
Tr	0.0	0.1	0.4	0.0	.91	0.3	0.0	0.0	0.0	0.0
Tg	0.0	0.0	0.4	0.0	0.1	.89	0.0	0.0	0.0	.02
Bi	0.0	0.0	0.0	0.0	0.0	0.0	.86	0.0	0.6	0.0
Bd	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.83	.01	0.9
Fs	0.2	0.0	0.0	0.0	0.0	0.0	0.5	.01	.89	0.0
Ld	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.0	.01	.90
Mean accuracy = 89.69%										

* Ho = Horse, El= Elephant, Bu= Building, Ap = Airplane, Tr = Tree, Tg= Tiger, Bi = Bike, Bd = Bird, Fs = Fish, Ld= Land

TABLE 3. Object segmentation using GMM accuracy on the Corel-10k dataset.

Objects	Accuracy	Objects	Accuracy
Ho	94.0	Tg	82.4
El	97.5	Bi	87.0
Bu	84.8	Bd	86.8
Ap	88.9	Fs	92.3
Tr	84.4	Ld	83.3
Mean Segmentation Accuracy = 88.1%			

TABLE 4. Comparison of object segmentation algorithms’ computing times on the MSRC dataset.

Objects	GMM	MSS	Objects	GMM	MSS
Bi	41.5	43.7	Gr	51.4	52.2
Cw	97.5	101.5	By	36.2	41.8
Ca	45.8	46.1	Du	172.9	201.5
Sp	65.9	71.2	Wr	29.8	33.3
Dg	35.2	43.5	Fl	76.5	78.2
GMM's average calculation time is 55.27s MSS's average calculation time is 71.30 s					

Figures 14 and 15 are showing ANOVA test to compare the accuracies of both datasets Corel-10k and MSRC, respectively.

Abbreviations used in this article are shown with their descriptions in Table 8. This table contains a comprehensive overview of all the abbreviations used in the text,

together with the full forms or descriptions that relate to each one.

VI. RESULTS AND DISCUSSIONS

In this research, we present an innovative way of segmenting images and processing them by combining two different

TABLE 5. Comparison of object segmentation algorithms’ computing times on the Corel-10k dataset.

Objects	GMM	MSS	Objects	GMM	MSS
Ho	112.0	131.2	Tg	133.2	156.3
El	96.5	114.2	Bi	170.9	199.2
Bu	171.0	188.9	Bd	131.2	157.0
Ap	150.2	170.3	Fs	135.0	162.7
Tr	94.1	105.9	Ld	97.5	113.2

GMM's average calculation time is 129.16s
MSS's average calculation time is 149.89s

TABLE 6. Accuracy comparison for object detection of the proposed approach to SOTA using Corel-10k dataset.

Procedures	Overall Accuracy (%)
A. Ahmad et. al [60]	86.10%
N. Kahyan et. al[61]	56.31%
M. Bansal et. al[62]	85.90%
A. Ahmad et. al [67]	88.75%
Proposed system	90.6%

TABLE 7. Accuracy comparison for object detection of the proposed approach to SOTA using MSRC dataset.

Procedures	Overall Accuracy (%)
X. Long et. al[63]	81.0%
C. Cheng et. al [64]	83.70%
R. Raja et. al.[65]	87.33%
Y. Zhao et. al [66]	79.20%
A. Ahmad et. al [67]	85.75%
Proposed Model	89.69%

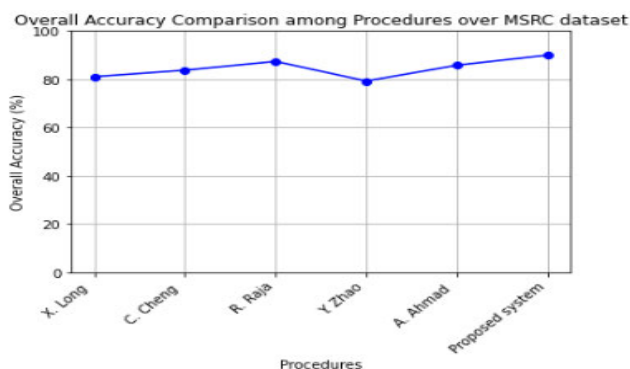


FIGURE 15. ANOVA TEST among accuracies over MSRC dataset.

step, the use of both approaches enables a thorough assessment of their relative outputs, exposing possible benefits and drawbacks. We also provide a comprehensive approach that extracts color, texture, and spatial information for the creation of a saliency map. This special combination captures a wide variety of visual features, improving the saliency map’s accuracy. The saliency map is then used to generate a variety of feature extractors and Maximally Stable Extremal Regions (MSER) for the purpose of computing local descriptors. It is noteworthy that these descriptions have been integrate.

VII. CONCLUSION

Even though our suggested approach provides novel segmentation methods and a distinctive classification scheme, it’s essential to acknowledge its limitations. These include challenges with generalizing across different datasets, possible sensitivity to parameter choices, and computational complexity. Further studies have to concentrate on enhancing these constraints in order to augment the technique’s suitability and efficiency in real-world applications.

VIII. FUTURE WORK

Our next work will focus on enhancing scene recognition and object identification. We are concentrating on improving algorithms to accurately and sensitively analyze scenes, optimizing for use in real-time, and adjusting to various environments. In order to ensure optimal model performance in dynamic contexts, we also seek to handle temporal fluctuations and dynamic situations.

Overall Accuracy Comparison using ANOVA TEST among Procedures over Corel-10k dataset

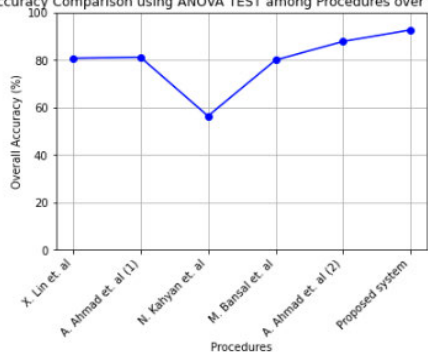


FIGURE 14. ANOVA TEST among accuracies over Corel 10k dataset.

TABLE 8. Abbreviations with their descriptions.

Abbreviations	Descriptions
MSS	Mean Shift Segmentation
GMM	Gaussian Mixture Model
AKAZE	Accelerated
HOG	Histogram of Gradient
MSER	Maximally Stable Extremal Regions
RCNN	R Convolution Neural Network
MLP	Multi-Layer Perceptron

segmentation strategies: The Gaussian Mixture Model (GMM) and the Mean Shift Segmentation (MSS). In the next

ACKNOWLEDGMENT

This research is supported and funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

- [1] H. Xu, L. Yao, Z. Li, X. Liang, and W. Zhang, "Auto-FPN: Automatic network architecture adaptation for object detection beyond classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6649–6658.
- [2] J. Huang, H. Ma, F. Sedano, P. Lewis, S. Liang, Q. Wu, W. Su, X. Zhang, and D. Zhu, "Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST-PROSAIL model," *Eur. J. Agronomy*, vol. 102, pp. 1–13, Jan. 2019, doi: 10.1016/j.eja.2018.10.008.
- [3] H. Tian, N. Huang, Z. Niu, Y. Qin, J. Pei, and J. Wang, "Mapping winter crops in China with multi-source satellite imagery and phenology-based algorithm," *Remote Sens.*, vol. 11, no. 7, p. 820, Apr. 2019, doi: 10.3390/rs11070820.
- [4] H. Tian, J. Pei, J. Huang, X. Li, J. Wang, B. Zhou, Y. Qin, and L. Wang, "Garlic and winter wheat identification based on active and passive satellite imagery and the Google Earth engine in northern China," *Remote Sens.*, vol. 12, no. 21, p. 3539, Oct. 2020, doi: 10.3390/rs12213539.
- [5] M. Qi, S. Cui, X. Chang, Y. Xu, H. Meng, Y. Wang, and T. Yin, "Multi-region nonuniform brightness correction algorithm based on L-channel gamma transform," *Secur. Commun. Netw.*, vol. 2022, pp. 1–9, Apr. 2022, doi: 10.1155/2022/2675950.
- [6] W. Zheng, S. Lu, Y. Yang, Z. Yin, and L. Yin, "Lightweight transformer image feature extraction network," *PeerJ Comput. Sci.*, vol. 10, p. e1755, Jan. 2024, doi: 10.7717/peerj-cs.1755.
- [7] W. Zheng, S. Lu, Z. Cai, R. Wang, L. Wang, and L. Yin, "PAL-BERT: An improved question answering model," *Comput. Model. Eng. Sci.*, vol. 55, pp. 1–10, Dec. 2023, doi: 10.32604/cmescs.2023.046692.
- [8] P. Zhou, J. Qi, A. Duan, S. Huo, Z. Wu, and D. Navarro-Alarcon, "Imitating tool-based garment folding from a single visual observation using hand-object graph dynamics," *IEEE Trans. Ind. Informat.*, early access, Jan. 1, 2024, doi: 10.1109/TII.2023.3342895.
- [9] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, "Towards open world object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5826–5836.
- [10] Q. She, R. Hu, J. Xu, M. Liu, K. Xu, and H. Huang, "Learning high-DOF reaching-and-grasping via dynamic representation of gripper-object interaction," *ACM Trans. Graph.*, vol. 41, no. 4, pp. 1–14, Jul. 2022, doi: 10.1145/3528223.3530091.
- [11] H. Zhang, H. Liu, and C. Kim, "Semantic and instance segmentation in coastal urban spatial perception: A multi-task learning framework with an attention mechanism," *Sustainability*, vol. 16, no. 2, p. 833, Jan. 2024, doi: 10.3390/su16020833.
- [12] Y. Xu, E. Wang, Y. Yang, and Y. Chang, "A unified collaborative representation learning for neural-network based recommender systems," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 11, pp. 5126–5139, Nov. 2022, doi: 10.1109/TKDE.2021.3054782.
- [13] Y. Han, B. Wang, T. Guan, D. Tian, G. Yang, W. Wei, H. Tang, and J. H. Chuah, "Research on road environmental sense method of intelligent vehicle based on tracking check," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 1261–1275, Jan. 2023, doi: 10.1109/TITS.2022.3183893.
- [14] D. Yang, T. Zhu, S. Wang, S. Wang, and Z. Xiong, "LFRSNet: A robust light field semantic segmentation network combining contextual and geometric features," *Frontiers Environ. Sci.*, vol. 10, p. 1443, Oct. 2022, doi: 10.3389/fenvs.2022.996513.
- [15] Y. Fang, H. Min, X. Wu, W. Wang, X. Zhao, and G. Mao, "On-ramp merging strategies of connected and automated vehicles considering communication delay," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15298–15312, Sep. 2022, doi: 10.1109/TITS.2022.3140219.
- [16] Z. Xiao, J. Shu, H. Jiang, J. C. S. Lui, G. Min, J. Liu, and S. Dustdar, "Multi-objective parallel task offloading and content caching in D2D-aided MEC networks," *IEEE Trans. Mobile Comput.*, vol. 22, no. 11, pp. 6599–6615, 2022, doi: 10.1109/TMC.2022.3199876.
- [17] M. Yang, H. Wang, K. Hu, G. Yin, and Z. Wei, "IA-Net: An inception-attention-module-based network for classifying underwater images from others," *IEEE J. Ocean. Eng.*, vol. 47, no. 3, pp. 704–717, Jul. 2022, doi: 10.1109/JOE.2021.3126090.
- [18] C. Fu, H. Yuan, H. Xu, H. Zhang, and L. Shen, "TMSO-Net: Texture adaptive multi-scale observation for light field image depth estimation," *J. Vis. Commun. Image Represent.*, vol. 90, Feb. 2023, Art. no. 103731, doi: 10.1016/j.jvcir.2022.103731.
- [19] X. Xu and Z. Wei, "Dynamic pickup and delivery problem with transshipments and LIFO constraints," *Comput. Ind. Eng.*, vol. 175, Jan. 2023, Art. no. 108835, doi: 10.1016/j.cie.2022.108835.
- [20] X. Hu, T. Tang, L. Tan, and H. Zhang, "Fault detection for point machines: A review, challenges, and perspectives," *Actuators*, vol. 12, no. 10, p. 391, Oct. 2023, doi: 10.3390/act12100391.
- [21] K. Kuan, G. Manek, J. Lin, Y. Fang, and V. Chandrasekhar, "Region average pooling for context-aware object detection," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1347–1351.
- [22] T. Mahalingam and M. Subramoniam, "A robust single and multiple moving object detection, tracking and classification," *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 2–18, Jan. 2021.
- [23] D. Chahyati and A. M. Arymurythi, "Multiple human tracking using retinanet features, Siamese neural network, and Hungarian algorithm," in *Proc. IAEME*, vol. 10, 2020, pp. 465–475.
- [24] X. Zhang and L. Zhang, "Real-time crowd counting with human detection and human tracking," in *Proc. Int. Conf. Neural Inf. Process.*, 2014, pp. 1–8.
- [25] J. Zhang, G. Ye, Z. Tu, Y. Qin, Q. Qin, J. Zhang, and J. Liu, "A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 46–55, Mar. 2022.
- [26] A. Naseer and A. Jalal, "Pixels to precision: Features fusion and random forests over labelled-based segmentation," in *Proc. IEEE Conf. Bhurban*, Bhurban, Pakistan, Jan. 2023, pp. 1–6.
- [27] Y. Di, R. Li, H. Tian, J. Guo, B. Shi, Z. Wang, and Y. Liu, "A maneuvering target tracking based on fastIMM-extended Viterbi algorithm," *Neural Comput. Appl.*, vol. 35, pp. 1–10, Oct. 2023, doi: 10.1007/s00521-023-09039.
- [28] F. Lecumberry, A. Pardo, and G. Sapiro, "Multiple shape models for simultaneous object classification and segmentation," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 3001–3004.
- [29] J. Shi, H. Zhu, S. Yu, W. Wu, and H. Shi, "Scene categorization model using deep visually sensitive features," *IEEE Access*, vol. 7, pp. 45230–45239, 2019.
- [30] S. Prabu, V. Balamurugan, and K. Vengatesan, "Design of cognitive image filters for suppression of noise level in medical images," *Measurement*, vol. 141, pp. 296–301, Jul. 2019.
- [31] M. Yan, J. Cai, J. Gao, and L. Luo, "K-means cluster algorithm based on color image enhancement for cell segmentation," in *Proc. 5th Int. Conf. Biomed. Eng. Informat.*, Oct. 2012, pp. 295–299.
- [32] M. Zheng, K. Zhi, J. Zeng, C. Tian and L. You, "A hybrid CNN for image denoising," *J. Artif. Intell. Technol.*, vol. 2, no. 3, pp. 93–99, 2022.
- [33] J. Meng, Y. Li, H. Liang, and Y. Ma, "Single image dehazing based on two-stream convolutional neural network," *J. Artif. Intell. Technol.*, vol. 2, pp. 100–110, Jun. 2022.
- [34] D. Li, K. D. Ortigas, and M. White, "Exploring the computational effects of advanced deep neural networks on logical and activity learning for enhanced thinking skills," *Systems*, vol. 11, no. 7, p. 319, Jun. 2023, doi: 10.3390/systems11070319.
- [35] S. Lu, J. Yang, B. Yang, Z. Yin, M. Liu, L. Yin, and W. Zheng, "Analysis and design of surgical instrument localization algorithm," *Comput. Model. Eng. Sci.*, vol. 137, no. 1, pp. 669–685, 2023, doi: 10.32604/cmescs.2023.027417.
- [36] F. S. Hassan and A. Gutub, "Improving data hiding within colour images using hue component of HSV colour space," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 56–68, Mar. 2022.
- [37] Y. Shi, J. Xi, D. Hu, Z. Cai, and K. Xu, "RayMVSNet++: Learning ray-based 1D implicit fields for accurate multi-view stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13666–13682, Jul. 2023, doi: 10.1109/tpami.2023.3296163.
- [38] X. Hu, Q. Kuang, Q. Cai, Y. Xue, W. Zhou, and Y. Li, "A coherent pattern mining algorithm based on all contiguous column Biclust," *J. Artif. Intell. Technol.*, vol. 2, pp. 80–92, May 2022.

- [39] G. Zhou, S. Su, J. Xu, Z. Tian, and Q. Cao, "Bathymetry retrieval from spaceborne multispectral subsurface reflectance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2547–2558, 2023, doi: [10.1109/JSTARS.2023.3249789](https://doi.org/10.1109/JSTARS.2023.3249789).
- [40] G. Zhou, G. Lin, Z. Liu, X. Zhou, W. Li, X. Li, and R. Deng, "An optical system for suppression of laser echo energy from the water surface on single-band bathymetric LiDAR," *Opt. Lasers Eng.*, vol. 163, Apr. 2023, Art. no. 107468, doi: [10.1016/j.optlaseng.2022.107468](https://doi.org/10.1016/j.optlaseng.2022.107468).
- [41] G. Zhou, G. Wu, X. Zhou, C. Xu, D. Zhao, J. Lin, Z. Liu, H. Zhang, Q. Wang, J. Xu, B. Song, and L. Zhang, "Adaptive model for the water depth bias correction of bathymetric LiDAR point cloud data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, Apr. 2023, Art. no. 103253, doi: [10.1016/j.jag.2023.103253](https://doi.org/10.1016/j.jag.2023.103253).
- [42] J. Chen, Y. Song, D. Li, X. Lin, S. Zhou, and W. Xu, "Specular removal of industrial metal objects without changing lighting configuration," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 3144–3153, 2024, doi: [10.1109/TII.2023.3297613](https://doi.org/10.1109/TII.2023.3297613).
- [43] Y. Dong, B. Xu, T. Liao, C. Yin, and Z. Tan, "Application of local-feature-based 3-D point cloud stitching method of low-overlap point cloud to aero-engine blade measurement," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023, doi: [10.1109/tim.2023.3309384](https://doi.org/10.1109/tim.2023.3309384).
- [44] D. Jiang, G. Li, Y. Sun, J. Hu, J. Yun, and Y. Liu, "Manipulator grabbing position detection with information fusion of color image and depth image using deep learning," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 12, pp. 10809–10822, Dec. 2021.
- [45] S. Li, J. Chen, W. Peng, X. Shi, and W. Bu, "A vehicle detection method based on disparity segmentation," *Multimedia Tools Appl.*, vol. 82, no. 13, pp. 19643–19655, May 2023, doi: [10.1007/s11042-023-14360-x](https://doi.org/10.1007/s11042-023-14360-x).
- [46] Y. Tao, J. Shi, W. Guo, and J. Zheng, "Convolutional neural network based defect recognition model for phased array ultrasonic testing images of electrofusion joints," *J. Pressure Vessel Technol.*, vol. 145, no. 2, Apr. 2023, Art. no. 024502, doi: [10.1115/1.4056836](https://doi.org/10.1115/1.4056836).
- [47] R. Zhang, L. Li, Q. Zhang, J. Zhang, L. Xu, B. Zhang, and B. Wang, "Differential feature awareness network within antagonistic learning for infrared-visible object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 8, pp. 2456–2464, 2023, doi: [10.1109/TCSVT.2023.3289142](https://doi.org/10.1109/TCSVT.2023.3289142).
- [48] W. Dong, J. Zhao, J. Qu, S. Xiao, N. Li, S. Hou, and Y. Li, "Abundance matrix correlation analysis network based on hierarchical multihead self-cross-hybrid attention for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 4, 2023, doi: [10.1109/TGRS.2023.3235401](https://doi.org/10.1109/TGRS.2023.3235401).
- [49] W. Dong, Y. Yang, J. Qu, S. Xiao, and Y. Li, "Local information-enhanced graph-transformer for hyperspectral image change detection with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 9, 2023, doi: [10.1109/TGRS.2023.3269892](https://doi.org/10.1109/TGRS.2023.3269892).
- [50] B. Cheng, D. Zhu, S. Zhao, and J. Chen, "Situation-aware IoT service coordination using the event-driven SOA paradigm," *IEEE Trans. Netw. Service Manage.*, vol. 13, no. 2, pp. 349–361, Jun. 2016, doi: [10.1109/TNSM.2016.2541171](https://doi.org/10.1109/TNSM.2016.2541171).
- [51] S. Yang, Q. Li, W. Li, X. Li, and A.-A. Liu, "Dual-level representation enhancement on characteristic and context for image-text retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 8037–8050, Nov. 2022, doi: [10.1109/TCSVT.2022.3182426](https://doi.org/10.1109/TCSVT.2022.3182426).
- [52] H. Liu, H. Yuan, Q. Liu, J. Hou, H. Zeng, and S. Kwong, "A hybrid compression framework for color attributes of static 3D point clouds," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1564–1577, Mar. 2022, doi: [10.1109/TCSVT.2021.3069838](https://doi.org/10.1109/TCSVT.2021.3069838).
- [53] Q. Liu, H. Yuan, R. Hamzaoui, H. Su, J. Hou, and H. Yang, "Reduced reference perceptual quality model with application to rate control for video-based point cloud compression," *IEEE Trans. Image Process.*, vol. 30, pp. 6623–6636, 2021, doi: [10.1109/TIP.2021.3096060](https://doi.org/10.1109/TIP.2021.3096060).
- [54] Z. Cui, H. Sheng, D. Yang, S. Wang, R. Chen, and W. Ke, "Light field depth estimation for non-lambertian objects via adaptive cross operator," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 2, pp. 1199–1211, 2023, doi: [10.1109/TCSVT.2023.3292884](https://doi.org/10.1109/TCSVT.2023.3292884).
- [55] J. Li, N. Zhou, J. Sun, S. Zhou, Z. Bai, L. Lu, Q. Chen, and C. Zuo, "Transport of intensity diffraction tomography with non-interferometric synthetic aperture for three-dimensional label-free microscopy," *Light. Sci. Appl.*, vol. 11, no. 1, p. 154, Jun. 2022, doi: [10.1038/s41377-022-00815-7](https://doi.org/10.1038/s41377-022-00815-7).
- [56] D. Cheng, L. Chen, C. Lv, L. Guo, and Q. Kou, "Light-guided and cross-fusion U-Net for anti-illumination image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8436–8449, Dec. 2022, doi: [10.1109/TCSVT.2022.3194169](https://doi.org/10.1109/TCSVT.2022.3194169).
- [57] Y. Liu, G. Li, and L. Lin, "Cross-modal causal relational reasoning for event-level visual question answering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11624–11641, Jun. 2023, doi: [10.1109/TPAMI.2023.3284038](https://doi.org/10.1109/TPAMI.2023.3284038).
- [58] H. Min, Y. Li, X. Wu, W. Wang, L. Chen, and X. Zhao, "A measurement scheduling method for multi-vehicle cooperative localization considering state correlation," *Veh. Commun.*, vol. 44, Dec. 2023, Art. no. 100682, doi: [10.1016/j.vehcom.2023.100682](https://doi.org/10.1016/j.vehcom.2023.100682).
- [59] J. Li, C. Zhang, Z. Liu, R. Hong, and H. Hu, "Optimal volumetric video streaming with hybrid saliency based tiling," *IEEE Trans. Multimedia*, vol. 25, pp. 2939–2953, 2022, doi: [10.1109/TMM.2022.3153208](https://doi.org/10.1109/TMM.2022.3153208).
- [60] J. Zhang, J. Ren, Y. Cui, D. Fu, and J. Cong, "Multi-USV task planning method based on improved deep reinforcement learning," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 3676–3689, 2024, doi: [10.1109/jiot.2024.3363044](https://doi.org/10.1109/jiot.2024.3363044).
- [61] H. Xu, J. Chen, and Q. Li, "Highlight removal from a single grayscale image using attentive GAN," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, Art. no. 1988441, doi: [10.1080/08839514.2021.198844](https://doi.org/10.1080/08839514.2021.198844).
- [62] A. Ahmed, A. Jalal, and K. Kim, "Region and decision tree-based segmentations for multi-objects detection and classification in outdoor scenes," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Islamabad, Pakistan, Dec. 2019, pp. 209–2095.
- [63] N. Kayhan and S. Fekri-Ershad, "Content based image retrieval based on weighted fusion of texture and color features derived from modified local binary patterns and local neighborhood difference patterns," *Multimedia Tools Appl.*, vol. 80, nos. 21–23, pp. 32763–32790, Sep. 2021.
- [64] M. Bansal, M. Kumar, M. Kumar, and K. Kumar, "An efficient technique for object recognition using Shi-Tomasi corner detection algorithm," *Soft Comput.*, vol. 25, no. 6, pp. 4423–4432, Mar. 2021.
- [65] X. Long, H. Lu, Y. Peng, X. Wang, and S. Feng, "Image classification based on improved VLAD," *Multimedia Tools Appl.*, vol. 75, no. 10, pp. 5533–5555, May 2016.
- [66] C. Cheng, X. Long, and Y. Li, "VLAD encoding based on LLC for image classification," in *Proc. 11th Int. Conf. Mach. Learn. Comput.*, Feb. 2019, pp. 417–422.
- [67] R. Raja, S. Kumar, and M. R. Mahmood, "Color object detection based image retrieval using ROI segmentation with multi-feature method," *Wireless Pers. Commun.*, vol. 112, no. 1, pp. 169–192, May 2020.
- [68] Y. Zhao, X. Guo, and Y. Lu, "Semantic-aligned fusion transformer for one-shot object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7601–7611.
- [69] A. Ahmed, A. Jalal, and K. Kim, "A novel statistical method for scene classification based on multi-object categorization and logistic regression," *Sensors*, vol. 20, no. 14, p. 3871, Jul. 2020.

AYSHA NASEER received the M.S. degree in computer engineering from the Center for Advanced Studies in Engineering (CASE), Islamabad, Pakistan. She is currently pursuing the Ph.D. degree in computer science with Air University, Islamabad. Her research interests include artificial intelligence, computer vision, machine learning algorithms, deep learning, image and video processing, and intelligent systems.



HAMDAN A. ALZAHIRANI received the master's degree in information technology from The University of Sydney, Australia, and the Ph.D. degree in security from the University of Colorado at Colorado Springs, Colorado Springs, USA, in 2016. He joined Saudi Electronic University, in 2017. He is currently an Assistant Professor with the College of Computing and Informatics. His current research interests include e-commerce technologies, privacy and security, identity management, cryptography, cyber security, and biometrics security.

NOUF ABDULLAH ALMUJALLY received the Ph.D. degree in computer science from the University of Warwick, U.K. She is currently an Assistant Professor of computer science with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University (PNU), Riyadh, Saudi Arabia. Her research interests include human-computer interaction (HCI), artificial intelligence (AI), machine learning, deep learning, and computer-based applications.



KHALED AL NOWAISER received the Ph.D. degree in computer science from Glasgow University, Scotland. He is currently an Assistant Professor with the Computer Engineering Department, Prince Sattam bin Abdulaziz University, Saudi Arabia. His research interests include computer vision, optimization techniques, and performance enhancement.



ASAAD ALGARNI received the Ph.D. degree in software engineering from North Dakota State University, USA. He is currently an Assistant Professor with the Department of Computer Sciences, College of Computing and Information Technology, Northern Borders University, Saudi Arabia. His research interests include software engineering, computer vision applications, and machine learning.



NAIF AL MUDAWI received the master's degree in computer science from Australian La Trobe University, in 2011, and the Ph.D. degree from the College of Engineering and Informatics, University of Sussex, Brighton, U.K., in 2018. He is currently an Assistant Professor with the Department of Computer Science and Information Systems, Najran University. He has published many research and scientific articles in many prestigious journals in various disciplines of computer science. During the master's degree, he was a member of the Australian Computer Science Committee.



JEONGMIN PARK received the Ph.D. degree from the College of Information and Communication Engineering, Sungkyunkwan University, in 2009. He is currently an Associate Professor with the Department of Computer Engineering, Tech University of Korea, South Korea. Before joining Tech University of Korea, in 2014, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI) and a Research Professor with Sungkyunkwan University, South Korea. His research interests include high-reliable autonomous computing mechanisms and human-oriented interaction systems.

...