# DETECTION AND RECOGNITION OF OBJECTS IN DIGITAL IMAGES USING ELM CLASSIFICATION

**R. Bhuvaneswari* & Dr. Ravi Subban****
* Research Scholar, Mother Teresa Women's University, Kodaikanal, Tamilnadu
** Department of Computer Science, School of Engineering and Technology, Pondicherry University, Pondicherry

**Cite This Article:** R. Bhuvaneswari & Dr. Ravi Subban, "Detection and Recognition of Objects in Digital Images Using ELM Classification", International Journal of Current Research and Modern Education, Volume 3, Issue 1, Page Number 478-486, 2018.



**Abstract:**

Object recognition is one of the hottest research areas, which aims to recognize the objects in digital media, which can be photographs or videos. In order to recognize the objects, the objects should be detected first. The two main considerations about the object recognition system are the accuracy and time consumption rates. Taking this into account, this article presents an effective time conserving object recognition approach based on three important phases. Initially, the points of interest are selected by means of Generalized Kadir Brady (GKB) detector, which considers the geometry and texture pattern of the images. The window size is selected for extracting the contourlet and Gabor Local Vector Pattern (GLVP) features from the window. The feature vector is formed and the Extreme Learning Machine (ELM) classifier is trained, such that the ELM can recognize the objects by means of the knowledge gained in the training process. The performance of the proposed approach is evaluated in four different aspects for proving the efficacy in terms of accuracy, precision, recall, F-measure and time consumption.

**Key Words**: Object Detection, Object Recognition, Feature Extraction & Classification

## 1. Introduction:

Computer vision technology is one of the evergreen research areas of advanced image processing technology. The major application areas of computer vision technology are object tracking, object detection and object recognition based applications [1-3]. All these applications process images for locating or tracking the objects. Due to the over-exploitation of smart phones and wireless technologies, skyrocketing adoption of photographs and videos comes into picture. Most of the applications store these digital images in the database for future reference. The database just provides room to the digital images and videos without any complex operations. However, it would be beneficial to the image analysts, when the objects involved in the images and videos are explicitly specified.

Previously, the involvement of humans is inevitable in the image analytic process. However, the excessive usage of digital photographs and videos makes it impossible to process every single image or video. The major drawback of this technique is the time consumption and inefficiency. In order to address this issue, automated object detection and recognition techniques hit the scene, which can work without the intervention of humans. Hence, efficient automatic object detection and recognition system is attainable, only when the individual objects are correctly detected by employing advanced image processing techniques.

Though it may seem simple, several complex tangles are concealed inside. The objects appear in the images are neither uniform nor follows any standard. Additionally, the same object may appear differently, when it is captured in different lighting, angle and illumination conditions. Numerous objects are involved in a photograph or video and all the objects are to be detected perfectly. Irrespective of all these differences, the object detection system must be capable of identifying the object. Taking all these challenges into account, this work aims to propose an efficient object detection system for photographs by employing advanced image processing techniques. The goal of this work is attained by segregating the complete work into three significant phases, which are points of interest localization, feature extraction, object detection and recognition.

The efficiency of the work is determined by the choice of points of interest and hence, optimal points of interest must be selected. The points of interest must be sharp and crisp, which makes sense that all the significant points of interest are taken into account and are sufficient to achieve a better object detection and recognition system. The points of interest are selected by means of Generalized Kadir-Brady (GKB) detector, which is an enhancement of classical KB detector. As soon as the points of interest are selected, the features of the points are extracted.

Feature extraction is the heart of any object detection and recognition system, as the object recognition accuracy is determined by the efficiency of the features being extracted from the area of the points of interest. Hence, this work gives more importance to feature extraction and is achieved by the combined features of Gabor Local Vector Pattern (GLVP) and contourlet. Finally, Extreme Learning Machine (ELM) is employed as the classifier to differentiate between the objects and the decision of classifier is based on the features used for training the object recognition system. The highlighting points of this work are listed below.

- ✓ The employment of GKB detector helps the object recognition system in selecting the sharp and crispy points of interest. This feature overthrows the need of unnecessary processing, which reduces the computational and space complexity.
- ✓ The combination of GLVP and contourlet features increases the discrimination ability of the object recognition system.
- ✓ The ELM classifier is employed for differentiating and recognizing the objects, as this classifier is fast learning and efficient.

The remainder of this paper is organized in the following way. Section 2 reviews the related works with respect to object detection and recognition system. The proposed object detection and recognition system is elaborated in section 3. The performance of the proposed object detection and recognition system is analysed in section 4. The conclusions of this article are presented in section 5.

**2. Review of Literature:**

This section presents the related state-of-the-art techniques with respect to object detection and recognition systems.

In [4], a material based salient object detection scheme is proposed for hyperspectral images. This work extracts the explicit attributes of the materials by employing the hyperspectral unmixing model. This model pre-processes the hyperspectral images, such that this scheme is able to work with different spectral response and resolutions. Both the local and global features are extracted from the images. The performance of this work is tested over 45 images. An object detection system based on depth information of images is presented in [5]. The depth information of the images is used in combination with RGB upon the Convolutional Neural Network (CNN). The depth information of images contains several visual properties, which depicts the objects of interest. The colour and the depth features are utilized for training the CNN. However, this work involves computational overhead.

A contextual model based on deep features is proposed for detecting objects in [6]. This work utilizes the appearance and the contextual information of an image. The contextual information of this work considers the relationship between the objects and the CNN builds the context based features. A completely connected Conditional Random Field (CRF) is designed for representing the object detection system. This system suffers from computational complexity and the precision rates can still be improved. In [7], a contour segment grouping technique is proposed for detection objects. This work incorporates three different pre-processing techniques for enhancing the object detection accuracy rates. A shape descriptor is employed for performing matching operation between the edges and the contour. The location of the target object is then identified. The performance of this work is tested upon four datasets, yet this work suffers from time complexity.

In [8], a multi-scale contrast based salient object detection scheme is proposed. This work employs the colour and texture features in varying scales for improving the salient point detection rate. A multi-scale feature correction scheme is utilized to enhance the features with respect to pixels. The drawback of this work is that it involves computational complexity. In [9], the image contours are represented by v4 shape features and the objects are detected. The V4 neurons extract the shape features, which can represent the contours of the objects. A self-organizing map based neural network is presented with respect to the V4 features. However, this work focuses only on shape feature.

In [10], the salient objects are detected by means of high-level background priors. This work employs a background estimation model for detecting the objects. The background information of an image is obtained by means of bounding boxes and three priors are taken into account. The three priors are background connectivity, background contrast and spatial distribution. This work consumes more time and computational power as well. A proposal selection based object detection system is presented in [11]. Initially, certain good region proposals are selected with respect to objectness. The initial saliency result is generated by considering the top ranking proposals. A structural ranker is then trained, such that the proposals can be ordered and the saliency map is obtained for the images. This work consumes more time and suffers from computational overhead.

A robust sparse representation model is proposed for object detection in [12], which considers the sparse errors. This type of saliency detection improves the robustness of the system. The representation coefficients and reconstruction errors are utilized for constructing the saliency measure. The salient object is identified by combining the local consistency prior to the robust sparse representation model. However, this approach does not work for images with Gaussian noise. A deep convolutional neural network based salient object detection technique is proposed in [13]. This work generates the coarse saliency map of images by means of fully convolutional network. This fully convolutional network predicts the salient objects globally and then the superpixel segmentation is performed for decomposing the image. The identified salient objects are finally enhanced. As this work involves two steps, it takes much time to identify the salient objects.

In [14], a fully connected conditional random field is employed for detecting objects. This work exploits the co-occurrence and geometric relationship between the objects. The co-occurrence, spatial interaction and scaling information of the objects are combined together. The feature embedding technique is utilised for detecting the geometry of the object. The abnormal objects are detected by employing the mean field approximation. An object detection technique based on boosted local binaries is proposed in [15]. The discrimination ability and generalized power of this descriptor is maintained by the structure-aware framework. This descriptor considers several local neighbouring regions in different scales and locations. The regions are selected by means of Adaboost algorithm associated with a penalty term. This work claims that the object detection accuracy is improved however, certain objects are not detected by the descriptor. The reason is that the efficiency of the descriptor relies on the localization of the image regions.

In [16], an efficient graph based search for object detection in images is proposed. This work pre-processes the images with five different techniques for eliminating the unwanted edges in the real images. The graphs are constructed by taking the edges into account and the adjacent nodes are linked with respect to the similarity degree. This work involves so much of computational overhead and the results can still be improved in terms of accuracy. In [17], an object recognition system based on probability density estimation is proposed for unmanned aerial vehicle application. Probabilistic Bayesian technique is utilized for recognizing the objects present in the images. The probability density estimation is developed by means of regression model.

A graph based image representation model is presented in [15]. This model represents the image patches as nodes and the nearby nodes are linked together. Initially, the image graph is generated and the image objects are matched by means of seed expansion strategy. The features are extracted from the image patches and they are utilized for image representation.

Most of the existing works utilize neural networks for detecting objects being present in the images. Taking this into account, this work intends to present an object detection and recognition system by detecting the points of interest, followed by which the features are extracted Finally, the objects are recognized by means of ELM classifier. The proposed approach is elaborated in section 3.

**3. Proposed Object Detection and Recognition System:**

This section describes the proposed object detection and recognition system in addition to the overall flow of the approach.

**3.1 Overall Flow of the Approach:**

The main goal of this work is to detect and recognize all the objects being present in an image. This kind of object detection is helpful in the fields of image analysis and decision making applications. As the applicability of the object detection system is more, the demand for the system is raising. Taking this into consideration, this work proposes a novel object detection and recognition system that relies on points of interest selection, feature extraction and recognition. The overall flow of the work is depicted in figure 1.

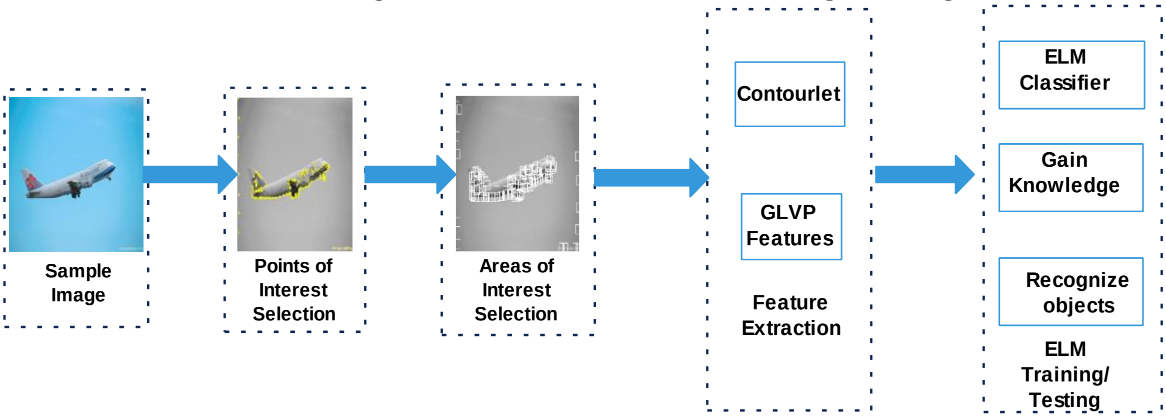



Figure 1: Overall flow of the work

The points of interest of an image are selected by means of GKB detector, which works well with all kinds of images. The reason for selecting the points of interest is the increased efficiency and reduced time complexity. Yet, the efficiency of the object detection and recognition system depends on the choice of points of interest. More number of points of interest increases the computational complexity and excluding important points of interest decreases the accuracy rates. Hence, optimal points of interest are chosen by GKB detector. The reason for the choice of GKB detector is that it considers the geometry and the texture of the image for capturing the points of interest. The GLVP and contourlet features are then extracted by considering the points of interest. These two feature extractors perform well in object detection and hence, the features are utilized. Additionally, the GLVP and contourlet features can sustain rotation and worst illumination conditions. The feature set is constructed and the ELM is trained with the feature vector. During the process of testing, when an input image is passed the classifier can distinguish between the objects present in the image. This is possible only when the classifier is trained effectively. The following sub-sections explain all the phases involved in the proposed object detection and recognition system.

**3.2 Points of Interest Capturing Process:**

The points of interest selection are the important process that attempts to capture potential points being present in the image. The underlying reason for the selection of points of interest is that the elimination of the need to process the whole image, which in turn improves the efficiency and reduces the computational power and time consumption. The major challenge involved in this phase is the optimal choice of interest points, as inclusion of unimportant points and exclusion of important points impacts over the results of the work. Hence, this work shows keen attention in the choice of points of interest and is achieved by means of GKB detector, which is an enhanced version of KB detector.

The classic KB detector selects the points of interest by focussing more on the complex areas of the image and hence, there are chances to miss out the important points. Usually, the KB detector misses the corner side pixels of an image, which may cause serious impact. The GKB detector beats the KB detector by considering the texture and the geometrical properties of the image. The working principle of GKB is presented as follows.

Consider an image with $\wp$ elements, such that the pixel $p$ belongs to $\wp$ and it can be represented by the following equation.

$$o_R(CP) = \{p \in \wp; dis(CP,p) \le R\} \quad (1)$$

In the above equation, $o$ is the object, $R$ is the radius, $CP$ is the centre pixel and $dis(CP,p)$ is the distance between the centre pixel and the pixel. For each and every element, a mapping function $MF$ is derived. The probability mass function of $p \in \wp$ at scale $sc$ is computed by considering the sum of mappings and the probability mass function is computed by

$$pmf_{i,sc} = \frac{\sum_{p1 \in o_R sc(CP)} w(p1,CP) MF_i(p1)}{\sum_{j=1}^{k} \sum_{p1 \in o_R sc(CP)} w(p1,CP) MF_j(p1)} \quad (2)$$

In the above equation, $w$ is the Gaussian weight and is calculated as follows.

$$w(p1,CP) = e^{\frac{-||p1-CP||}{sc^2}} \quad (3)$$

Hence, the entropy of a point at scale $sc$ is computed by

$$H(CP,R_{sc}) = -\sum_{i=1}^{k} pmf_{i,R_{sc}} \log(pmf_{i,R_{sc}}) \quad (4)$$

The importance of a point is determined by calculating the product of $H(CP,R_{sc})$ and $w(CP,R_{sc})$. These points are grouped by taking the point with greater importance into account and the remaining points are disregarded from consideration. Hence, this way of point of interest selection focuses on geometry and the texture of the image, rather than complex regions. By following this way, the points of interest are selected and the features are extracted as follows.

**3.3 GLVP and Contourlet Feature Extraction:**

As soon as the points of interest are selected, the area surrounding the points of interest is manipulated for the purpose of feature extraction. The window size that encloses the point of interest is varied and the experiments are performed. The window sizes being considered by this work are $8 \times 8$, $16 \times 16$ and $32 \times 32$. The features are extracted from the pixels of the corresponding window and the feature vector is formed. Initially, the contourlet transform is applied followed by which the GLVP features are extracted.

Contourlet is a transform that satisfies the essential properties of multiresolution and directionality [19]. Multiresolution is a facility that can represent the digital images in all the resolutions ranging from coarse to fine texture. The directionality feature should consider all the directional orientation of the image. Though the functionality of wavelet considers the multiresolution feature, it does not follow the directionality feature. Curvelets can provide these facilities, yet it operates well on continuous images rather than discrete images. This is the point where contourlet stand strong with multidirection and multiresoultion capability and it operates on the discrete domain. As contourlet can focus on multiple directions at a single scale with the rich set of filter banks, this work employs contourlet as well.

The contourlet clubs the Laplacian Pyramid (LP) and Directional Filter Bank (DFB) for attaining its goal. The image components with great frequency are captured by DKB, such that the minimal frequency components are disregarded. Let $a_0[n]$ be an input image, which is fed into the LP. The LP then generates $K$ bandpass images as denoted by the following equation

$$b_k[n]; k = (1,2, \dots K) and\ a_K[n] \quad (5)$$

In the above equation, $b_k[n]; k = (1,2, \dots K)$ denotes the image from fine to coarse order, where the lowpass image is represented by $a_K[n]$. Here, the $k^{th}$ level of LP decomposes the input image $a_{k-1}[n]$ into two, which are coarse $a_k[n]$ and fine $b_k[n]$ image respectively. Each single bandpass image $b_k[n]$ is decomposed to the degree of $d_k$ to $2^{d_k}$ bandpass directional images as denoted by

$$a_{k,l}^{(d_k)}[n]; \quad k = (0,1, \dots, 2^{d_k} - 1) \quad (6)$$

This work employs a bi-level contourlet, where the initial and the second level contourlet provides two and four subbands respectively. The initial level and the second level matrices possess $4 \times 120$ and $16 \times 120$ elements respectively. Both these matrices are clubbed together, which returns the matrix with $20 \times 120$ elements. When the bilevel contourlet is applied, the GLVP features are extracted as follows.

The GLVP feature extraction technique relies on the operation of gabor filter and LVP. The gabor filter is known for its better texture feature extraction capability while detecting the edges as well. The gabor filter $(GF)$ is generated by

$$GF = \frac{1}{2\pi\sigma_a\sigma_b} exp\left[\frac{-1}{2}\left(\frac{a}{\sigma_a}\right)^2 + \left(\frac{b}{\sigma_b}\right)^2 + kF(acos\theta + bcos\theta)\right] \quad (7)$$

In equation 7, $\sigma_a$ and $\sigma_b$ are the spatial width of the pixels $a\ and\ b$, $F$ is the frequency rate and $\theta$ is the orientation. The LVP considers the orientations of $0°, 45°, 60°, 90°$. The features are extracted by varying the window size, as stated earlier. The feature vector is constructed as follows.

$$LVP_{or,d}(p_i) = \{LVP_{or,d}\ |o = 0°, 45°, 60°, 90°\} \quad (8)$$

$$LVP_{or,d}(p_i) = \{LVP_{or,d}\ |dis = 1,2,3\} \quad (9)$$

In the equations 8 and 9, *or* is the orientation and distance is $dis$ are the orientation and distance between the pixels respectively. The GLVP is figured out as follows

$$GLVP(p_i) = Gabor\ (p_i) \cup LVP_{or,d}(p_i) \quad (10)$$

By following this way, the GLVP features are extracted and the feature vector is formed. The so formed feature vector is utilized for training the classifier and the classification process is described in the forthcoming section.

**3.4 Object Recognition by ELM:**

ELM is the fast learning and efficient classifier [23]. In order to recognize between different objects, the ELM is trained by means of train samples and the knowledge is rendered to it. The process of knowledge gaining is possible through the constructed feature vectors. During the process of testing, the classifier applies the knowledge gained in the training process to recognize the objects. Let $X$ be the training samples represented by $(a_j, b_j)$, where $a_j = [a_{j1}, a_{j2}, \dots, a_{jn}]^T \in L^n$ and $a_j$ denotes the training entity $j$ in dimension $n$. $b_j = [b_{j1}, b_{j2}, \dots, b_{jm}]^T \in L^m$ denotes the training class $j$ with dimension $m$, where $m$ is the total number of classes being computed.

A Single hidden Layer Feed-Forward Neural Network (SLFN) is constructed by an activation function $act(x)$ and $NR$ neurons, as shown below.

$$\sum_{i=1}^{NR} \gamma_i act(wt_i.act_j + bs_i) = b_j; j = 1,2, \dots, n \quad (11)$$

In equation 11, $wt_i$ is the weights shown as weight vectors, such that $wt_i = [wt_{i1}, wt_{i2}, \dots, wt_{in}]^T$, which interlinks the $i^{th}$ hidden neuron with the input neurons as represented by $i = [i_1, i_2, \dots, i_m]^T$. The weight vector interconnects the $i^{th}$ hidden neuron to the output neurons and the bias of the $i^{th}$ hidden neuron is represented by $bs_i$. The $w_i$ and $bs_i$ are chosen in a random fashion. Now, the SLFN can be written as

Consider $HDL$ as the ELM's hidden layer output matrix and the $i^{th}$ column of $HDL$ represents the $i^{th}$ hidden neurons output vector with respect to the inputs $a_{j1}, a_{j2}, \dots, a_{jn}$.

$$HDL = \begin{bmatrix} act(wt_1.a_1 + bs_i) & \dots & act(wt_{NR}.a_1 + bs_{NR}) \\ \vdots & \vdots & \vdots \\ act(wt_1.a_n + bs_i) & \dots & act(wt_{NR}.a_n + bs_{NR}) \end{bmatrix} \quad (12)$$

$$\gamma = \begin{bmatrix} \gamma_1^T \\ \vdots \\ \gamma_G^T \end{bmatrix} \quad (13)$$

$$B = \begin{bmatrix} b_1^T \\ \vdots \\ b_n^T \end{bmatrix} \quad (14)$$

The matrix form is represented as

$$HDL\gamma = B \quad (15)$$

The output weight is computed by the norm least-square solution as represented below.

$$\gamma = HDL^{\dagger}B \quad (16)$$

$HDL^{\dagger}$ is the $HDL$'s Moore-Penrose generalized inverse. The ELM is trained with the paramenters such as number of classes $m$, the activation function $act(x)$, number of hidden neurons $NR$ and the number of ELM in ensemble $E$. During training, the ELM is passed with the training set $TS = \{(a_j, b_j) | a_j \in L^n, b_j \in L^m; j = 1,2, \dots, N\}$. The ELM is trained by computing $\gamma$ for all $TS$ using eqn.16.

The classifier recognizes the objects by means of the knowledge gained from the training process. The training process imparts knowledge to the classifier by means of feature vector. In the testing phase, the input image is treated with the points of interest selection and the feature extraction phases. The so formed feature vector is matched with the feature vector of the trained set and the objects are recognized.

**4. Results and Discussion:**

The proposed object recognition approach is implemented on Matlab 8.1 version, over a stand alone system with 8 GB RAM. The experimental analysis utilizes Caltech dataset [21] for testing the efficiency of the proposed approach. The Caltech5 dataset contains a total of 3691 images, out of which 1846 images are used for training the proposed object recognition approach and the remaining 1844 images are utilized for testing the proposed approach. The performance of the proposed approach is evaluated by means of four standard performance metrics such as precision, recall, accuracy, F-measure and error rates. All the mentioned performance metrics rely on four basic measures such as true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates.

Precision rate is the ratio of TP and the summation of TP and FP. The value of precision is indirectly proportional to the FP rates. It is beneficial for an object recognition system to have greater precision rates. Similarly, the recall rates depend on the FN rates. The lesser the FN rates, the greater is the recall rates. Based on the precision ad recall rates, the F-measure is computed. An object recognition system is considered to be reliable, when it achieves the maximum precision, recall and F-measure rates. The formulae for computing the precision (P), recall ($\mathscr{R}$), F-measure ($\mathscr{F}$), accuracy (A) and error rates (E) are presented below.

$$P = \frac{TP}{TP+FP} \tag{17}$$

$$\mathscr{R} = \frac{\text{TP}}{\text{TP+FN}} \tag{18}$$

$$\mathscr{F} = \frac{2(P \times \mathscr{R})}{P + \mathscr{R}} \tag{19}$$

$$\text{A} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{20}$$

$$\text{E} = 100 - \text{A} \tag{21}$$

A novel object recognition system must prove greater precision, recall, f-measure and accuracy rates. Conversely, the error rates and the time consumption must be as minimal as possible. The following subsections evaluate the performance of the proposed approach by varying the points of interest selection, feature extraction and classification techniques. Additionally, the performance of the proposed approach is compared against the analogous object recognition approaches.

The performance of the proposed approach is tested by varying only the points of interest selection technique and the results are evaluated. The points of interest techniques being taken into consideration are KB and DKB. The experimental results are tabulated in table 1.

Table 1: Performance evaluation by varying points of interest selection techniques

| Image Category | Precision (%) | | | Recall (%) | | | F-measure (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | KB | DKB | GKB | KB | DKB | GKB | KB | DKB | GKB |
| Aeroplane | 90.2 | 97.9 | 98.6 | 87.6 | 96.2 | 97.4 | 88.88 | 97.04 | 97.9 |
| Motorbikes | 90.8 | 96.7 | 97.4 | 86.9 | 95.9 | 96.3 | 88.8 | 96.29 | 96.84 |
| Human Faces | 92.6 | 97.9 | 98.2 | 84.3 | 94.9 | 96.1 | 88.26 | 96.37 | 97.13 |
| Car | 91.8 | 98.1 | 98.4 | 82.3 | 95.3 | 96.2 | 86.79 | 96.67 | 97.28 |
| Leaves | 90.3 | 96.9 | 97.3 | 81.8 | 94.6 | 96.1 | 85.4 | 95.73 | 96.69 |
| Average | 91.14 | 97.5 | 97.9 | 84.58 | 95.3 | 96.4 | 87.62 | 96.42 | 97.16 |

| Image Category | Accuracy (%) | | | Error rate (%) | | |
|---|---|---|---|---|---|---|
| | KB | DKB | GKB | KB | DKB | GKB |
| Aeroplane | 93.4 | 95.2 | 96.8 | 6.6 | 4.8 | 3.2 |
| Motorbikes | 91.6 | 94.8 | 96.4 | 8.4 | 5.2 | 3.6 |
| Human Faces | 89.4 | 96.3 | 96.8 | 10.6 | 3.7 | 3.2 |
| Car | 90.3 | 96.1 | 96.4 | 9.7 | 3.9 | 3.6 |
| Leaves | 92.4 | 93.8 | 95.9 | 7.6 | 6.2 | 4.1 |
| Average | 91.42 | 95.24 | 96.4 | 8.58 | 4.76 | 3.54 |

From the experimental results, it is observed that the GKB performs better than the DKB and KB detector. The major reason for the better performance of GKB is that it considers the geometry and the texture of the images. The DKB detector works by computing the entropy histogram. The KB detector misses out the geometrical points and considers only the complex regions of an image. Hence, the choice of GKB detector is justified.

Feature extraction technique decides the efficiency of the object recognition system and is always recommended to extract rich set of features that can support in achieving effective object recognition. The choice of the combination of contourlet and GLVP features is justified by checking the individual performances of GLVP and contourlet. The experimental results are tabulated as follows.

Table 2: Performance analysis by varying feature extraction techniques

| Image Category | Precision (%) | | | Recall (%) | | | F-measure (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | C | GLVP | GLVP+C | C | GLVP | GLVP+C | C | GLVP | GLVP+C |
| Aeroplane | 80.2 | 83.9 | 98.6 | 73.2 | 82.6 | 97.4 | 76.54 | 83.24 | 97.9 |
| Motorbike | 78.6 | 82.6 | 97.4 | 71.8 | 83.2 | 96.3 | 75.04 | 82.89 | 96.84 |
| Human Face | 79.2 | 80.9 | 98.2 | 74.3 | 83.8 | 96.1 | 76.67 | 82.32 | 97.13 |
| Car | 77.8 | 83.4 | 98.4 | 74.8 | 81.9 | 96.2 | 76.27 | 82.64 | 97.28 |
| Leaf | 78.3 | 80.7 | 97.3 | 70.4 | 82.6 | 96.1 | 74.14 | 81.63 | 96.69 |
| Average | 78.8 | 82.3 | 97.9 | 72.9 | 82.82 | 96.4 | 75.7 | 82.5 | 97.16 |

| Image Category | Accuracy (%) | | | Error rate (%) | | |
|---|---|---|---|---|---|---|
| | C | GLVP | GLVP+C | C | GLVP | GLVP+C |
| Aeroplane | 78.2 | 82.3 | 96.8 | 21.8 | 17.7 | 3.2 |
| Motorbike | 76.9 | 86.4 | 96.4 | 23.1 | 13.6 | 3.6 |
| Human Face | 79.4 | 84.3 | 96.8 | 20.6 | 15.7 | 3.2 |
| Car | 78.9 | 86.1 | 96.4 | 21.1 | 13.9 | 3.6 |
| Leaf | 76.8 | 84.8 | 95.9 | 23.2 | 15.2 | 4.1 |
| Average | 78.04 | 84.78 | 96.4 | 21.96 | 15.2 | 3.54 |

From the experimental analysis, the effectiveness of the combination of contourlet and GLVP is justified. The performance of the contourlet is not upto the mark and so is the GLVP. However, when both these feature extraction techniques are combined the performance is increased considerably. The main reason for the better performance is that multidirectional and multiresolutional ability at different scales. The following section compares the proposed approach by varying the window sizes with reference to the point of interest.

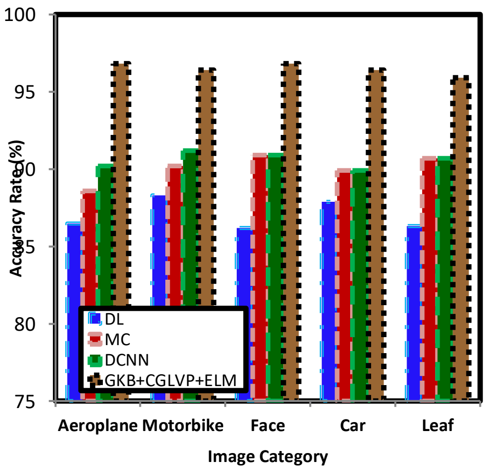
The window sizes of the processing area is varied as $8 \times 8$, $16 \times 16$, $32 \times 32$ and the best window size that suits the proposed approach is selected. On analysis, it is found that $16 \times 16$ works optimal for the proposed approach and the experimental results are presented as follows.

Table 3: Performance analysis by varying the window size

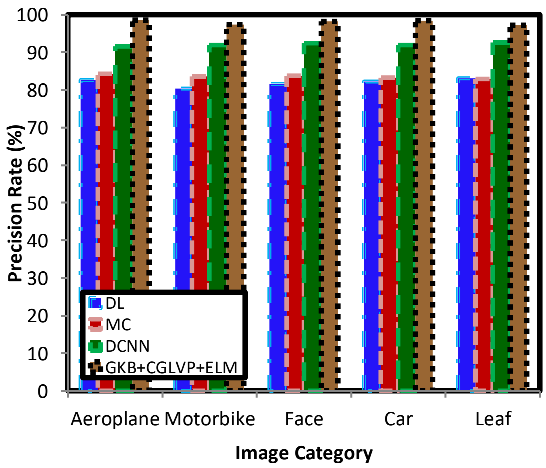
| Image Category | Precision (%) | | | Recall (%) | | | F-measure (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 8 × 8 | 16 × 16 | 32 × 32 | 8 × 8 | 16 × 16 | 32 × 32 | 8 × 8 | 16 × 16 | 32 × 32 |
| Aeroplane | 84.6 | 98.6 | 89.3 | 82.6 | 97.4 | 84.3 | 83.58 | 97.9 | 86.72 |
| Motorbike | 86.7 | 97.4 | 89.9 | 83.7 | 96.3 | 84.9 | 85.17 | 96.84 | 87.32 |
| Human Face | 84.3 | 98.2 | 88.7 | 82.9 | 96.1 | 86.7 | 83.59 | 97.13 | 87.7 |
| Car | 88.4 | 98.4 | 89.8 | 84.3 | 96.2 | 89.2 | 86.3 | 97.28 | 89.49 |
| Leaf | 87.3 | 97.3 | 90.2 | 82.4 | 96.1 | 89.8 | 84.77 | 96.69 | 89.99 |
| Average | 86.26 | 97.9 | 89.5 | 83.18 | 96.4 | 86.98 | 84.68 | 97.16 | 88.24 |

| Image Category | Accuracy (%) | | | Error rate (%) | | |
|---|---|---|---|---|---|---|
| | 8 × 8 | 16 × 16 | 32 × 32 | 8 × 8 | 16 × 16 | 32 × 32 |
| Aeroplane | 88.4 | 96.8 | 84.3 | 11.6 | 3.2 | 15.7 |
| Motorbike | 87.9 | 96.4 | 85.9 | 12.1 | 3.6 | 14.1 |
| Human Face | 87.4 | 96.8 | 86.2 | 12.6 | 3.2 | 13.8 |
| Car | 89.2 | 96.4 | 87.9 | 10.8 | 3.6 | 12.1 |
| Leaf | 88.3 | 95.9 | 87.3 | 11.7 | 4.1 | 12.7 |
| Average | 88.24 | 96.4 | 86.32 | 11.76 | 3.54 | 13.68 |

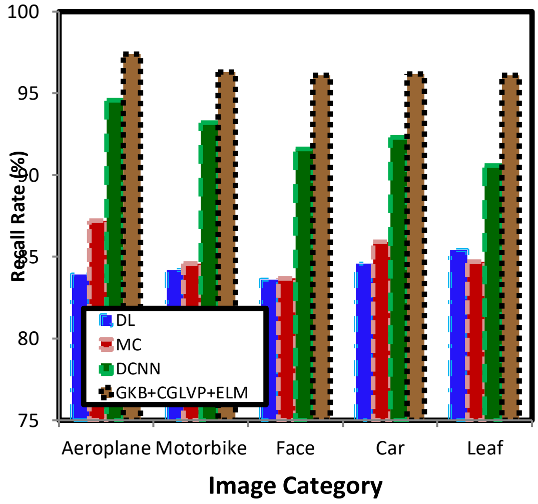
From the above table, it is clearly evident that the proposed approach works well with the window size of $16 \times 16$, rather than the window sizes of $8 \times 8$ and $32 \times 32$. Hence, the proposed object recognition approach works well with the window size of $16 \times 16$. The following section compares the performance of the proposed approach with the state-of-the-art object recognition techniques. The techniques employed for comparison are deep learning (DL) [2], multiscale contrast (MC) [8], DCNN [13].



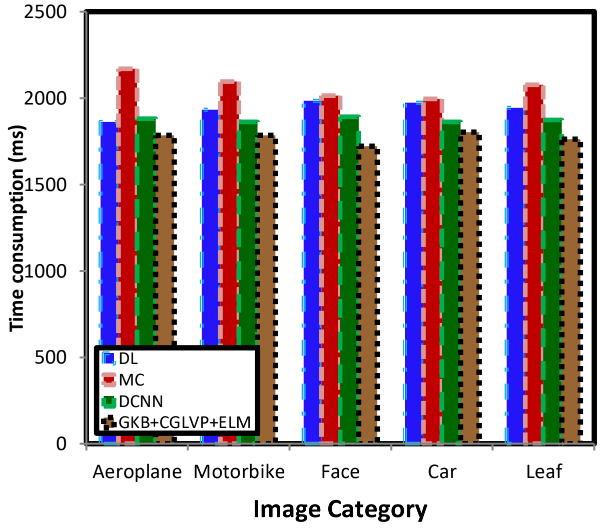

(a)




(b)




(c)




(d)

Figure (a) Accuracy rate analysis, (b) precision rate analysis (c) recall rate analysis (d) time consumption analysis

From the experimental analysis, it is evident that the proposed approach outperforms the existing approaches in terms of accuracy, precision, recall and time consumption. The proposed approach proves greater accuracy, precision and recall rates by consuming minimal amount of time. Hence, the objective of object recognition approach is attained.

**5. Conclusion:**

This article presents an object detection and recognition approach that relies on three significant phases, which are points of interest selection, feature extraction and recognition phases. The points of interest are selected by means of GKB detector and then the surrounding regions of window size $16 \times 16$. The contourlet and GLVP features are extracted from the window size and the feature vector is formed. The ELM classifier is then trained with the constructed feature vector for gaining the knowledge. In the testing process, the ELM is prompted to classify between the objects. The performance of the proposed approach is tested upon Caltech5 dataset in different aspects, such as by varying the points of interest selection, feature extraction and recognition. Additionally, the proposed approach is compared with the recent object recognition approaches and the experimental results are presented. In future, this work is planned to be extended to detect and recognize objects in videos.

**6. References:**

1. Min, W., Zhang, Y., Li, J., & Xu, S. (2018). Recognition of pedestrian activity based on dropped-object detection. Signal Processing, 144, 238-252.
2. Bresilla, K., Manfrini, L., Morandi, B., Boini, A., Perulli, G., & Grappadelli, L. C. (2018). Using Deep Learning Real-Time Object Detection Convolution Neural Networks for Fast Fruit Recognition in the Tree. World Academy of Science, Engineering and Technology, International Journal of Agricultural and Biosystems Engineering, 5(8).
3. Seidenari, L., Baecchi, C., Uricchio, T., Ferracani, A., Bertini, M., & Del Bimbo, A. (2018, January). Object Recognition and Tracking for Smart Audio Guides. In Italian Research Conference on Digital Libraries (pp. 163-168). Springer, Cham.
4. Jie Liang, Jun Zhou, Lei Tong, Xiao Bai, Bin Wang, "Material based salient object detection from hyperspectral images", Pattern Recognition, Vol.76, pp.476-490, 2018.
5. Saihui Hou, Zilei Wang, Feng Wu, "Object detection via deeply exploiting depth information", Neurocomputing, Vol. 286, pp.58-66, 2018.
6. Wenqing Chu, Deng Cai, "Deep feature based contextual model for object detection", Neurocomputing, Vol. 275, pp. 1035-1042, 2018.
7. Hui Wei, ChengzhuanYang, Qian Yu, "Contour segment grouping for object detection", Journal of Visual Communication and Image Representation, Vol.48, pp. 292-309, 2017.
8. Hai Wang, Lei Dai, Yingfeng Cai, Xiaoqiang Sun Long Chen, "Salient object detection based on multi-scale contrast", Neural Networks, Vol.101, pp.47-56, 2018.
9. Hui Wei, Zheng Dong, Luping Wang, "V4 shape features for contour representation and object detection", Neural Networks, Vol.97, pp.46-61, 2018.
10. Gang Wang, Yongdong Zhang, Jintao Li, "High-level background prior based salient object detection", Journal of Visual Communication and Image Representation, Vol.48, pp.432-441, 2017.

11. Lihe Zhang, Qin Zhou, "Salient object detection via proposal selection", Neurocomputing, In Press, 2018. DoI: 10.1016/j.neucom.2018.01.050.
12. Liu Yi, Zhang Qiang, Han Jungong, Wang Long, "Salient object detection employing robust sparse representation and local consistency", Image and Vision Computing, Vol.69, pp.155-167, 2018.
13. Ying Li, Fan Cui, Xizhe Xue, Jonathan Cheung-Wai Chan, "Coarse-to-fine salient object detection based on deep convolutional neural networks", Signal Processing: Image Communication, Vol.64, pp.21-32, 2018.
14. Jongsuk Oh, Hong-In Kim, Rae-Hong Park, "Context-based abnormal object detection using the fully-connected conditional random fields", Pattern Recognition Letters, Vol. 98, pp.16-25, 2018.
15. Haoyu Ren, Ze-Nian Li, "Object detection using boosted local binaries", Pattern Recognition, Vol.60, pp. 793-801, 2016.
16. Hui Wei, Chengzhuan Yang, Qian Yu, "Efficient graph-based search for object detection", Information Sciences, Vol. 385, pp.395-414, 2017.
17. V. P. Kharchenko, A. G. Kukush, N. S. Kuzmenko, I. V. Ostroumov, "Probability density estimation for object recognition in unmanned aerial vehicle application", IEEE International Conference on Actual Problems of Unmanned Aerial Vehicles Developments, 17-19 Oct, Kiev, Ukraine, 2017.
18. Jie Zhu, Shufang Wu, Xizhao Wang, Guoqing Yang, Liyan Ma, "Multi-image matching for object recognition", IET Computer Vision, Vol.12, No.3, pp.350-356, 2018.
19. Do, Minh N., and Martin Vetterli. "The contourlet transform: an efficient directional multiresolution image representation." IEEE Transactions on image processing 14.12 (2005): 2091-2106.
20. Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang, Extreme Learning Machine for Regression and Multiclass Classification, IEEE Transactions on systems, Man and Cybernetics - Part B, Vol.42, No.2, pp.513-529, 2012.