Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 52 (2018) 985–994

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Novel object detection and recognition system based on points of interest selection and SVM classification

R. Bhuvaneswari [a,*], Ravi Subban [b]

[a] Mother Teresa Women's University, Kodaikanal, Tamil Nadu, India

[b] Dept. of Computer Science, School of Engineering and Technology, Pondicherry University, Pondicherry, India

Received 19 June 2018; received in revised form 10 September 2018; accepted 20 September 2018
Available online 9 October 2018

Abstract

Due to the progression in computer vision technology, object recognition systems have gained considerable research interest. Though there are numerous object recognition systems in the literature, there is always a constant demand for better object recognition systems. Taking this as a challenge, this work proposes a novel object recognition system based on points of interest and feature extraction. Initially, the points of interest of the image are selected by means of Derivative Kadir-Brady (DKB) detector and the neighbourhood pixels of a particular window size are selected for further processing. The gabor and curvelet features are extracted from the area of interest, followed by the Support Vector Machine (SVM) classification. The performance of the proposed object recognition system is evaluated against three analogous techniques in terms of accuracy, precision, recall and F-measure. On experimental analysis, it is proven that the proposed approach outperforms the existing approaches and the performance of the proposed work is satisfactory.

Keywords: Object detection; Object recognition; Classification; Points of interest extraction

1. Introduction

Due to the advent of computer technology, the exploitation of images and videos is inevitable. As time progresses, computer vision technology has gained substantial research interest with respect to object recognition. It is easy for humans to recognize and distinguish between different objects. However, machines find it difficult to recognize the objects and an effective algorithm is needed to be incorporated for knowledge building process. The efficiency of object recognition algorithm relies on the image processing concepts being employed. The applications of object recognition are widespread and some of the noteworthy applications are surveillance, robotics, access control systems and so on.

Object recognition is a tougher process, as the image quality may vary from image to image. For instance, most of the images suffer from noise, illumination issues, orientation and so on. The object recognition system must be able to recognize the objects being present in the image, irrespective of all the shortcomings. An image contains several objects and the object recognition system must be capable of differentiating between several objects in the image.

Although there are several object recognition techniques, there is constant demand for reliable and efficient object recognition systems. Most of the existing techniques encounter complexities in terms of computation, memory and time. Taking this issue into account, this work proposes an object recognition system based on the local feature set. Local features render more promising results rather than geometric or appearance based methods (Mikolajczyk and Schmid, 2005; Yokono and Poggio,

* Corresponding author.
E-mail address: bhuvaneswari.mtu@gmail.com (R. Bhuvaneswari).

https://doi.org/10.1016/j.cogsys.2018.09.022
1389-0417/

2006). The proposed work is organized into three phases and they are points of interest extraction, feature extraction and object recognition.

The interest points being present in an image are extracted by means of Derivative based Kadir-Brady detector (DKB). DKB is the enhancement of traditional KB saliency detector, which is proposed by Kadir and Brady in the year 2001 (Kadir and Brady, 2001). Traditional interest point detection schemes work on assumptions and demand perfect image alignments. These requirements degrade the performance of the interest point detection. This issue is addressed by Kadir-Brady (KB) detector, which is based on histograms. The KB detector does not consider the quality of images and the interest points are selected with respect to the information content of the pixels. The information content of the pixels is calculated by the entropy histogram. As this approach relies on histograms, KB detector can work for different kinds of images, irrespective of the image quality. However, the KB detector computes the histogram of pixel intensity in a circular area, which results in the repetitive process. This introduces computational overhead.

Several enhancements of KB detector are proposed in the literature and this work utilizes the Derivative based Kadir-Brady (DKB) detector proposed in (Brown, Windridge, & Guillemaut, 2017). The DKB detector overcomes the shortcoming of KB detector by building the histogram of eigen values of the second moment matrix. This idea eliminates the repetitive processing of image pixels, which in turn saves time and computational power. Additionally, the DKB takes both the geometrical and texture features of the image into account. This is the major reason for the choice of DKB detector.

The features of the area surrounding the interest points are extracted by means of gabor and curvelet transform and the object recognition is attained by the Support Vector Machine (SVM) classifier. SVM is a promising classifier, which can distinguish between the objects effectively. The noteworthy contributions of this work are as follows.

- This work incorporates DKB detector for detecting and extracting interest points, which overthrows the curse of repetitive processing of image pixels.
- DKB detector is efficient as the interest points are selected based on the information content.
- DKB detector considers both the geometrical and texture features of the image pixels.
- The combination of gabor and curvelet features land at better set of features, which are sufficient to recognize the objects.
- With the help of the built knowledge base, the SVM can distinguish between the objects being present in the image effectively.

The remainder of this article is arranged as follows. The related review of literature with respect to object recognition is presented in Section 2. The proposed object recognition technique based on DKB and SVM is presented in Section 3. The performance of the proposed approach is evaluated in Section 4. Section 5 presents the conclusive idea of this article.

2. Review of literature

This section presents the related state-of-the-art literature of object detection and recognition systems and the motivation of the research.

A technique to recognize objects based on shape growth pattern in (Cheddad, Kusetogullari, & Grahn, 2017). This work applies dilations to the shape, so as to figure out an updated dimension of the image with important features. This work is proven to be resilient against noise. However, this work involves computational overhead and the object recognition capability can be improved further. A system for automatic brightness adjustment for recognising the objects in a better way is presented in (Kim, Lee, Kim, & Cho, 2017). This work adjusts the brightness of the image by taking the RGB and ‘L’ component of the CIE LAB space. In the second step, the fuzzy inference system computes the adjustment coefficients of every pixel and the brightness of the image is adjusted. This work claims that the automatic brightness adjustment of images helps in recognising the objects in a better way and can be used in the pre-processing phase of the object recognition system.

A fine grained object recognition technique is proposed in (Srinivas, Lin, & Liao, 2017), which is based on Convolutional Neural Networks (CNN) and dictionary learning. Incorporation of CNN alone suffers from inefficiency and expensiveness. Hence, this work merges online dictionary with the CNN. The dictionaries are built in an incremental fashion and the experimental data are represented sparsely. The results of this work are better in terms of object recognition and efficiency. In (Hsu, 2017), an object recognition system based on adaboost algorithm is proposed, which works on the real time environment. To collect the real time data, this work exploits laser range finder and a camera. The adaboost algorithm is exploited to detect the objects in front of the car and trigger alarm.

In Alom, Alam, Taha and Iftekharuddin (2017), CNN based object recognition technique is proposed in which the initial filters are produced by Cellular Simultaneous Recurrent Networks (CSRN) and the features are extracted. Regularized Extreme Learning Machine (RELM) is employed to perform classification. Though this work is proven to show better performance than CNN, the time complexity involved is more. An online object recognition system is proposed in Sebastien Wong et al. (2017), which can detect and recognize multiple objects being present in a single image. Initially, the objects present in the image are tracked and are classified by means of CNN. This work renders better results at the cost of computational overhead.

In Tang, Song and Chen (2017), an object recognition system for 3D objects is proposed that is based on three important steps. Initially, the remarkable features are collected by a geometric centroid descriptor. This step is followed by the computation of the geometric compatibility between all the objects and the similarity between them. Though this work provides better results, this is meant for recognizing 3D objects. An object recognition system by employing deep convolutional features is presented in Bui, Lech, Cheng, Neville and Burnett (2016). This work extracts image features from deep convolutional neural network and the recursive neural network is applied over it. The accuracy rate of this work is proven to be greater at the cost of minimal computational overhead. However, the time complexity involved in the work can still be reduced.

A system to detect salient object for object recognition and classification is presented in Nikhila and Rawat (2016). This work states that the local contrast and the compactness of the image are complement to each other. The salient objects are detected by means of compactness and the local contrast map, followed by which the objects are recognised. Hence, this work depends on geometrical features which are complex to compute. In Hannat, Zrira, Raoui and Bouyakhf (2016), a real time object recognition system is proposed by means of visual bag of words. Initially, the SURF feature points are computed and the visual bag-of-words are constructed by means of k means clustering algorithm. The Support Vector Machine (SVM) is trained by means the bag of features and hence the objects are recognized.

An object recognition system based on multiple instance learning CNN is proposed in Sun, Han, Liu and Khodayari-Rostamabad (2016). This work overcomes the problems faced by traditional CNN, which requires perfect labels. Attaining perfect labels is a time consuming process. Though this work is an improvisation of CNN, it involves computational complexity. In Yörük, Öner and Akgül (2016), a system for object detection, recognition and pose estimation is proposed by employing hough transform. The SURF features of the images are extracted and the six degrees of freedom are computed. This work relies on geometrical relationship between the objects, which is complex to compute.

Most of the existing works utilize CNN, in order to present an object recognition system. CNN requires the activity of perfect labelling, which introduces computational and time complexity. Additionally, several object recognition systems in the existing literature utilize geometrical relationship between the pixels. Feature based object recognition systems are more promising than geometrical and appearance based systems Happy and Routray (2015). Taking this statement into account, this work proposes a feature based object recognition system, which employs the combination of gabor and curvelet features and SVM.

3. Proposed object detection and recognition system

The intention of this section is to present the proposed object recognition system in an elaborated fashion, in addition to the overview of the work.

3.1. Overview of the work

The main goal of this work is to present a reliable and efficient object recognition system which demands reasonable computational overhead. In order to ensure modularity of the proposed work, the entire work is broken down into three important phases, which are points of interest extraction, feature extraction and object recognition. Each and every phase performs their task effectively, so as to reach the work goal in a better way.

The point of interest extraction phase is attained by the application of DKB detector. The important point to select the DKB detector is that the points of interest are computed by taking the information content of the pixels into account. The information content of the pixels is computed by the entropy histogram. This ensures that no points of interest are left out from further processing. As soon as the points of interest are detected, certain specific area surrounding the point of interest is chosen.

In the second phase, the features are extracted from the compact regions around the interest points and this phase is the heart of the entire object recognition system. The reason for the choice of curvelet is that it can handle the entire frequency spectrum of an image and hence the entire spectral domain is treated. The reason for employing the gabor filter is that it can recognize the distinct objects easily and the gabor features are rich in texture. The SVM classifier is trained with the constructed knowledge base, which is built by the features. The overall flow of the proposed object recognition approach is presented in Fig. 1.

The SVM classifier is then capable of differentiating and recognizing the objects effectively. Hence, this is the final decisive stage and is the brain of the object recognition system. The phase-wise explanation of the proposed object recognition system is as follows.

3.2. Points of interest selection

The points of interest are selected by employing DKB detector (Brown et al., 2017), which is an enhanced version of KB detector (Kadir and Brady, 2001). The summary of original KB detector is presented, followed by which the DKB is described. The KB detector works by considering that the complex areas of an image are important. The complexity is measured in various scales, which is termed as scale-invariance and the pixels with more complexity are chosen as interest points. Additionally, the image pixel must show a reasonable degree of dissimilarity when compared to its neighbouring image scales. This is known as inter-scale saliency. Hence, the KB detector chooses the points of interest by taking the complexity and inter-scale saliency.

In the final stage, all the detected points are clubbed together to combat against noise. The major drawbacks of the standard KB detector are as follows. The KB detector pays more attention towards complex regions and so the probability of missing the geometrical points is greater.

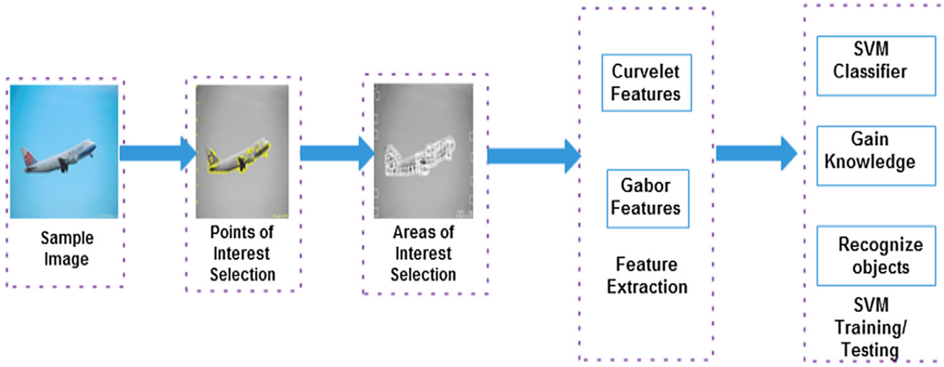

Fig. 1. Overall flow of the proposed object recognition system.

Mostly, the corners of the images are missed out because of the corner pixels show least difference with respect to the neighbouring pixels in terms of scale intensity. The so stated drawbacks are overcome by DKB detector by the following way. The overall algorithm of the proposed object recognition system is presented as follows. The sample images for points of interest detection are presented in Fig. 2.

Proposed object recognition algorithm
Input: Image database
Output: Recognized objects
// Training
Begin
For all train images
Select points of interest by DKB detector pi_n*;*
  For all pi
    Extract 16×16 *neighbourhood pixels* ai_n*;*
      For each ai
        Extract gabor and curvelet features;
      Build knowledge base;
      Train SVM with the built feature set;
    End For;
  End For;
End For;
// Testing
For an image
Select points of interest by DKB detector pi_n*;*
  For all pi
    Extract 16×16 *neighbourhood pixels* ai_n*;*
      For each ai
        Extract gabor and curvelet features;
        Employ SVM to recognize objects;
      End for;
End for;
End;

The DKB detector works by considering the linking the derivative function of the image to the other pixel. This makes sense that the high derivative points in the low derivative neighbourhood are important and considered as points of interest. This kind of operation eliminates the repetitive processing of the image scenes. The processing area of the KB and DKB are the same, except for the operation. Let $\mathcal{F}$ is the histogram mapping derivative function of the second moment matrix and the intensity I of pixel x is denoted by $I(x)$. The intensity derivatives of directions a and $b$are represented by $I(x)_a$ and $I(x)_b$. The second moment matrix is computed for a specific scale μ for a pixel x, which is done by the following equation.

$$S(x) = \left(\sum_{j \in P_\mu(x)} wt(j,x)\right)^{-1} \times \sum_{j \in P_\mu(x)} wt(j,x) \begin{pmatrix} I(j)_a^2 & I(j)_a I(j)_b \\ I(j)_a I(j)_b & I(j)_b^2 \end{pmatrix} \quad (1)$$

In the above equation, $wt(j,x)$ is the Gaussian function to compute weight by considering the points nearer to x, which can further be described by

$$wt(j,x) = e^{\frac{-\|j-x\|^2}{2\mu^2}} \quad (2)$$

While computing the second moment matrix, the derivatives are considered for the 50 pixels such that the image is perceived better. The eigen values $\beta_1$and β_2 of $S(x)$ are to be computed to obtain the image derivative. It is understood that when the values of $\beta_1$and β_2 are greater, then the pixel x is known as a corner. A histogram mapping $\mathcal{F}$ is built by computing the eigen values of all the pixels, which are normalized and discretised initially. The sample images along with the points of interest selection is presented in figure.

The reason for normalization and discretisation is to make the pixels to find place in the$reg_Y \times reg_Y$ histogram. $\mathcal{F}$ connects the eigen values of $S(x)$ to the histogram bins of $reg_Y \times reg_Y$. The histogram of the eigen values adapts to the scale, which makes it possible to detect areas based on derivatives and thereby repetitive areas of images are eliminated. Hence, the points of interest are extracted from the images and the areas surrounding the points of extraction are focussed. The area surrounding the points of interest are varied with different window sizes such as 8×8, 16×16 and 32×32. The features of the pixel window are extracted to recognize the objects and the feature extraction process is explained as follows.

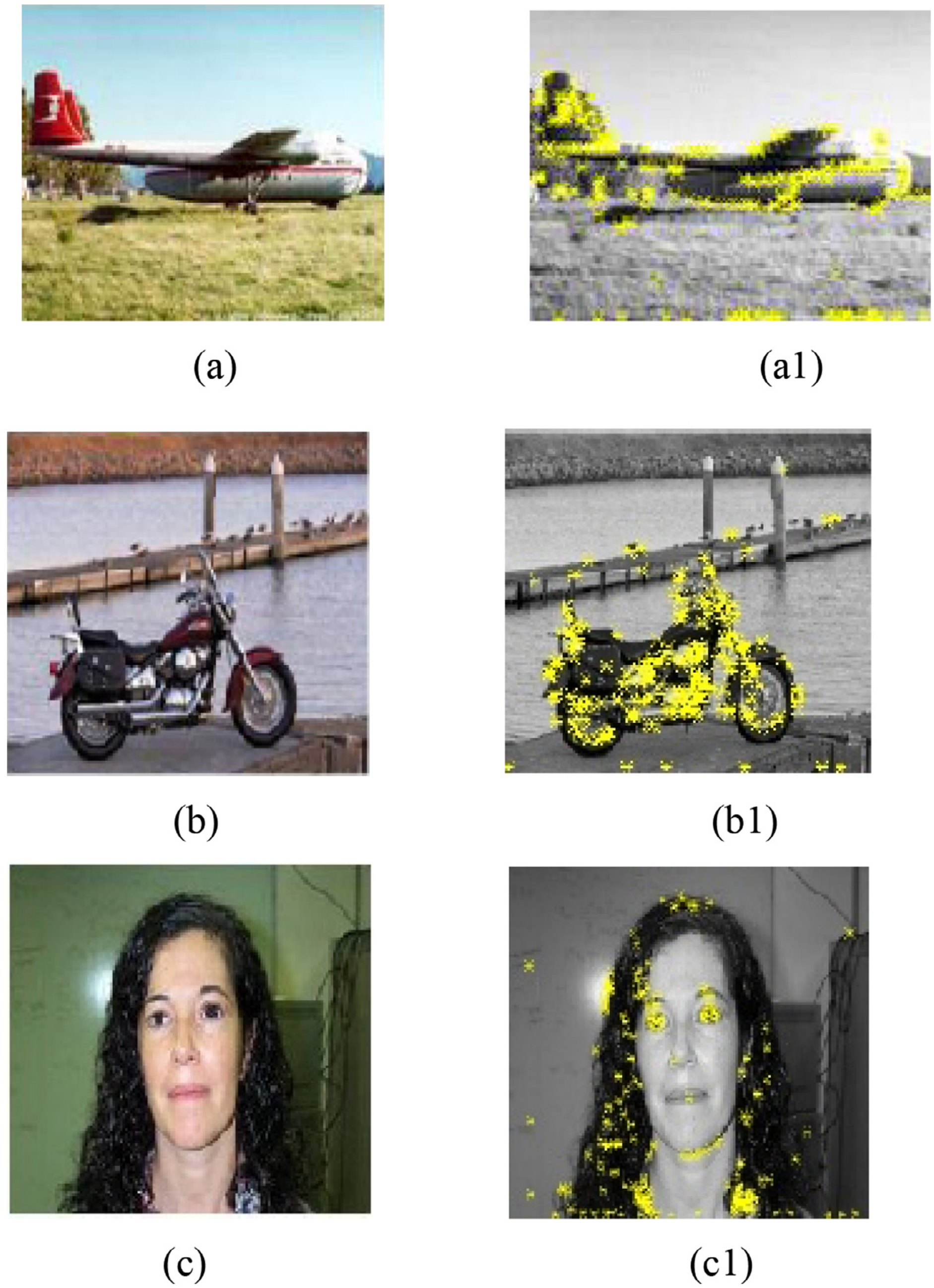

Fig. 2. Sample points of interest extracted images.

3.3. Feature extraction phase

3.3.1. Gabor features

Gabor filter is a potential texture feature extraction technique, which minimizes the joint ambiguity issues in both spatial and frequency domain. Additionally, the joint uncertainty is minimized in both spatial and frequency domain. The 2D gabor function $g(x, y)$ is created as follows.

$$g(a,b) = \frac{1}{2\pi\sigma_a\sigma_b}\exp\left[-\frac{1}{2}\left[\frac{a^2}{\sigma_a^2}+\frac{b^2}{\sigma_b^2}\right]+2\pi jM_f a\right] \tag{3}$$

M_f is the modulation frequency and the 2D fourier transform is defined by

$$G(u,v) = exp\left[-\frac{1}{2}\left[\frac{(u-w)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right];$$
$$\sigma_a = (2\pi\sigma_a)^{-1}, \sigma_b = (2\pi\sigma_b)^{-1} \tag{4}$$

Let $g(a, b)$ is the mother wavelet and the filters are constructed with various rotational degrees as presented in the following equation.

$$g_{xy}(a,b) = k^{-x}g(a', b') \tag{5}$$

In the above equation, a' and b' denotes

$$a' = k^{-x}(acos\theta + bsin\theta) \quad (6)$$

$$b' = k^{-x}(-asin\theta + bcos\theta); k > 1 \quad (7)$$

$$\theta = \frac{x\pi}{X}; x = 0, 1, 2, .., X - 1; y = 0, 1, 2, \cdots, Y - 1 \quad (8)$$

where x and $y$denote the gabor filter's scale and orientation. The size of the employed gabor filter is $3 \times 3, 5 \times 5, 7 \times 7$. The theta values are varied as 15, 75, 180 with 60 as frequency. In order to extract features, forty gabor filters are produced with five and eight scales and orientation respectively.

3.3.2. Curvelet features

Curvelet is the improvised version of Ridgelet transform. The drawback of ridgelet transform is that it works with lines rather than points and is represented by

$$\mathcal{R}_t(p, q, \theta) = \psi_{p,q,\theta}(x, y)I(x, y)dxdy \quad (9)$$

In the above equation, $I(x, y)$ is the image and the ridgelet function is represented by ψ, which is denoted as follows.

$$\psi_{p,q,\theta}(x, y) = p^{-1/2}\psi\left(\frac{xcos\theta + ysin\theta - q}{p}\right) \quad (10)$$

The curvelet bands of the image are obtained by tailoring the ridgelet in various scales and orientations. The curvelet sub-bands are obtained by

$$E(p, \theta) = \sum_x \sum_y |Sb_{p,\theta}(x, y)| \quad (11)$$

where $E(p, \theta)$ is the energy of the pixel in a particular orientation. Gabor filters are powerful in detecting shapes and the curvelets work for an entire frequency spectrum. Hence, the combination of curvelet and gabor features results in a better set of features. These features are considered for recognizing the objects effectively. The collected features act as the base for the classification process, which is presented as follows.

3.4. Object classification by SVM

SVM is one of the promising classifiers, which is based on statistical learning principle. SVM distinguishes between the objects of different classes by means of a hyperplane. The complete classification phase involves two stages, which are training and testing. The training phase intends to impart knowledge to the classifier about different objects by means of the extracted features from the images. In the testing stage, the classifier differentiates between the objects in the image by applying the knowledge it has gained in the training stage. This work employs SVM, as it enjoys numerous merits. Some of the remarkable points about SVM are as follows. SVM is robust against noise, overthrows computational complexity and does not demand perfect training data.

This work employs non-linear SVM classifier to accomplish object recognition. The classification of objects is achieved by means of hyperplane, which separates between different classes and is represented as follows.

$$w.m_i + b \geq +1; n_i = +1 \quad (12)$$

$$w.m_i + b \leq -1; n_i = -1 \quad (13)$$

In Eqs. (12) and (13), m_i is a data point, b is the bias and w is the weight, which may observed normal, as far as the hyperplane is considered. Though, the points cannot be divided linearly and this issue is overcome by the introduction of slack variables $\{\omega_i\}_{i=1}^r$.

In this classification problem, there are $c$possible classes for which the single optimization problem is framed as

$$\min_{w,b,\omega} \frac{1}{2}\sum_{n=1}^{c} w_n^p w_n + C\sum_{i=1}^{l}\sum_{n \neq s_i} \omega_{i,n} \quad (14)$$

The constraints of the above given equation is presented in the following equation.

$$w_{s_i}^p \rho(m_i) + b_{s_i} \geq w_n^p \rho(m_i) + b_n + 2 - \omega_{i,n}; \omega_{i,n} \geq 0 \quad (15)$$

Here, i represents training samples ranging from $1tol$ and $n \in \{1, 2, \cdots, c\}$. The classifier differentiates the data points by applying the following equation.

$$clfn = \max_{n=1,2,..,c} w_n^p \rho(m_i) + b_n \quad (16)$$

This work employs $c(c-1)/2$ classifiers, where c is the total number of classes involved in the classification problem. In case, if a data point is suitable for class X, then the vote of that data point to be in class X is incremented by 1. The decision score of all the classifiers are added and the class which attained maximum votes to accommodate the data point dp is chosen as the residing class of the data point. The multiclass classification issue is converted to a single optimization problem and this work requires lesser support vectors, when compared to one-classifier-one-class strategy of classification. The objects being present in the image are recognized by this way of classification and the following section evaluates the performance of the proposed approach.

4. Results and discussion

The efficiency of the proposed approach is evaluated over the Caltech5 dataset (http://www.vision.caltech.edu/html-files/archive.html). The proposed approach is implemented in Matlab 8.1 and executed on a standalone system with 8 GB RAM. The performance of the proposed approach is tested in terms of precision, recall, F-measure and time consumption by varying the point of interest selection scheme, area of interest and feature extraction techniques. The performance of the proposed approach is compared with the existing works such as shape-growth pattern (Cheddad et al., 2017), CNN (Srinivas et al., 2017), bag-of-words (Hannat et al., 2016).

Table 1
Performance comparison of KB and DKB detector.

Image category	Precision (%)				Recall (%)				F-measure (%)			
	SIFT	SURF	KB	DKB	SIFT	SURF	KB	DKB	SIFT	SURF	KB	DKB
Aeroplane	86.2	88.6	89.3	**98.3**	83.4	85.4	86.9	**96.2**	84.77	86.97	88.08	**97.23**
Motorbikes	84.8	86.2	90.2	**96.8**	80.7	83.2	84.3	**95.9**	82.69	84.67	87.15	**96.34**
Human Faces	83.6	85.4	89.7	**97.3**	73.9	78.1	79.8	**94.9**	78.45	81.58	84.46	**96.08**
Car	81.4	83.2	84.9	**97.9**	78.7	79.3	80.3	**95.3**	80.02	81.2	82.53	**96.58**
Leaves	80.9	85.6	86.3	**96.9**	74.8	76.9	78.3	**94.6**	77.73	81.01	82.10	**95.73**
Average	83.3	85.8	88	**97.4**	78.3	80.58	81.9	**95.3**	80.72	83.08	84.86	**96.39**

The Caltech5 dataset contains images such as airplanes, motorbike, faces, cars and leaves. The dataset contains 1074 aeroplane images, 826 motorbike images, 450 images of human faces, 1155 images of car and 186 images of leaves. The objects of these images vary in terms of orientation, size and location. Some of the sample images of Caltech5 dataset are presented below. This work utilizes fifty percent of the images for training and the remaining fifty percent are used for the testing purpose. For each category of the image, the performance measures are computed to prove the efficiency of the proposed approach.

The performance measures such as precision, recall, F-measure rely on the four fundamental measures True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP rate is measured by the correctly recognised images to the total number of images available in the dataset. TN rates are the images that are correctly recognised, such that the image does not belong to a particular class. FP rate is the ratio of the images that are wrongly classified to be a part of the wrong class. Similarly, FN rate is measured by the images that are wrongly classified such that the image of a particular class is declared as a non-member of the class. Based on these measures, the precision, recall and F-measure are computed.

Table 2
Accuracy and error rate analysis.

Accuracy (%)				Error rate (%)			
SIFT	SURF	KB	DKB	SIFT	SURF	KB	DKB
86.9	89.3	89.5	**95.4**	13.1	10.7	10.46	**4.6**
80.4	81.4	83.3	**96.2**	19.6	18.6	16.7	**3.8**
74.8	78.9	80.1	**94.4**	25.2	21.1	19.9	**5.6**
73.9	77.2	78.4	**93.2**	26.1	22.8	21.6	**6.8**
74.9	77.9	81.1	**94.8**	25.1	22.1	18.9	**5.2**
78.18	80.94	82.4	**94.82**	21.82	19.06	17.51	**5.2**

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 ($\mathcal{R}$), F-measure ($\mathcal{F}$), accuracy (A) and error rates (E)are presented below.

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

$$\mathcal{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{18}$$

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

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

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

An object recognition system should prove greater precision, recall, F-measure and accuracy rates. On the other hand, the error rates and the time consumption must be as minimal as possible. In order to prove the efficiency, the proposed object recognition is evaluated and compared against several existing techniques.

4.1. Performance analysis w.r.t points of interest selection

In this section, the performance of the points of interest selection is varied and the results are analysed. The proposed approach employs DKB detector and the results obtained are compared against KB detector, Scale Invariant Feature Transform (SIFT) and SURF (Speeded Up Robust Features) in terms of precision, recall, F-measure

Table 3
Performance comparison of feature extraction techniques.

Image category	Precision (%)			Recall (%)			F-measure (%)			Accuracy (%)			Error rate (%)		
	G	C	G + C	G	C	G + C	G	C	G + C	G	C	G + C	G	C	G + C
Aeroplane	66.3	72.3	**98.3**	59.3	68.3	**96.2**	62.6	70.2	**97.23**	70.3	76.8	**95.4**	29.7	23.2	**4.6**
Motorbike	64.2	70.2	**96.8**	55.6	63.5	**95.9**	59.5	66.6	**96.34**	68.8	79.2	**96.2**	31.2	20.8	**3.8**
Human Face	69.4	71.6	**97.3**	62.7	69.7	**94.9**	65.8	70.6	**96.08**	69.2	74.6	**94.4**	30.8	25.4	**5.6**
Car	65.3	71.9	**97.9**	60.9	70.2	**95.3**	63	71.1	**96.58**	64.6	79.8	**93.2**	35.4	20.2	**6.8**
Leaf	66.2	74.3	**96.9**	59.7	68.1	**94.6**	62.7	71.0	**95.73**	69.3	81.3	**94.8**	30.7	18.7	**5.2**
Average	66.2	72.0	**97.4**	59.6	67.9	**95.3**	62.7	69.9	**96.3**	68.4	78.3	**94.8**	31.5	21.6	**5.2**

and accuracy, by taking the image category into account (see Tables 1 and 2).

The above presented table illustrates the efficacy of SIFT, SURF, DKB detector over KB detector in various classes of images. The performance of DKB is better than the other different keypoint detectors for all the image categories. The performance of SIFT is not commendable, as SIFT cannot prove itself under poor lighting conditions. The performance of SURF is quite comparable with SIFT and the difference is the time consumption. The SURF works faster than SIFT descriptor.

The main reason for the better performance of DKB is that it takes the information content of the image pixel, which is computed by performing the entropy histogram. On the other hand, the KB detector focuses on complex regions that increase the probability of missing the geometrical points. Mostly, the corners of the images are missed out because of the corner pixels show least difference with respect to the neighbouring pixels in terms of scale intensity. This is the reason for why the DKB is better than other keypoint detectors.

4.2. Performance analysis w.r.t feature extraction techniques

The effectiveness of the proposed feature extraction technique is justified by employing the gabor and curvelet features separately and proven that the combination of gabor and curvelet performs better. As the proposed approach employs the combination of gabor and curvelet features, rich set of features are obtained. The so obtained features are sufficient for the classifier to distinguish between the objects. The experimental results of the proposed approach are presented in Table 3.

From the experimental results, the potential of the combination of gabor and curvelet features is justified. Gabor filters can detect the shape effectively and produce texture rich features. Subsequently, curvelets can deal with the whole frequency spectrum and is not restricted to any geometrical shape. However, the results of these techniques are not convincing when employed independently. The combination of gabor and curvelet features prove a drastic improvement in the performance metrics. The reason for betterment is that the curvelet and gabor filter work hand-in-hand to meet the objective. For instance, Gabor can detect the shape effectively but cannot handle the complete frequency spectrum, whereas the curvelet can do that. The rich set of significant features results in the improvement of the performance rates and is evident from the above presented table.

4.3. Performance analysis w.r.t the size of area of interest

As soon as the points of interest are detected, the areas surrounding the points of interest are taken into account for the process of feature extraction. In order to select the optimal size for the area of interest, the experiments are carried out by choosing different sizes of area of interest such as 8×8, 16×16, 32×32.

Table 4
Performance comparison w.r.t size of area of interest.

Image category	Precision (%)			Recall (%)			F-measure (%)			Accuracy (%)			Error rate (%)		
	8×8	16×16	32×32	8×8	16×16	32×32	8×8	16×16	32×32	8×8	16×16	32×32	8×8	16×16	32×32
Aeroplane	86.3	**98.3**	82.4	89.8	**96.2**	86.3	88.01	**97.23**	84.3	87.7	**95.4**	88.4	12.3	**4.6**	11.6
Motorbike	84.2	**96.8**	80.9	90.6	**95.9**	84.7	87.28	**96.34**	82.75	86.8	**96.2**	82.9	13.2	**3.8**	17.1
Human Face	83.7	**97.3**	78.4	88.4	**94.9**	87.9	85.98	**96.08**	82.87	88.6	**94.4**	86.9	11.4	**5.6**	13.1
Car	87.2	**97.9**	82.4	88.9	**95.3**	87.4	88.04	**96.58**	84.82	87.3	**93.2**	85.8	12.7	**6.8**	14.2
Leaf	85.9	**96.9**	81.9	87.3	**94.6**	85.9	86.59	**95.73**	83.85	89.6	**94.8**	88.9	10.4	**5.2**	11.1
Average	85.46	**97.44**	81.2	89	**95.3**	86.4	87.18	**96.39**	83.71	88	**94.8**	86.5	12	**5.2**	13.4

As shown in Table 4, the precision, recall, accuracy and F-measure rates of the proposed approach is greater, when the size of the area of interest is set as 16×16. The average precision rate of the proposed approach with 16×16 as the size of area of interest is 97.4. Conversely, the precision rate decreases drastically with the size 8×8 and 32×32. Similarly, the huge difference is observed for all the performance measures. The error rates is greater when the size of the area of interest is set as 32×32, which is 13.4. After experimental analysis, this approach finds that the size 16×16 is suitable for the Caltech dataset.

4.4. Comparison with the state-of-the-art object recognition techniques

In the previous sections, the performance of the proposed approach is analysed by varying the techniques being incorporated and to justify the choice of the techniques.

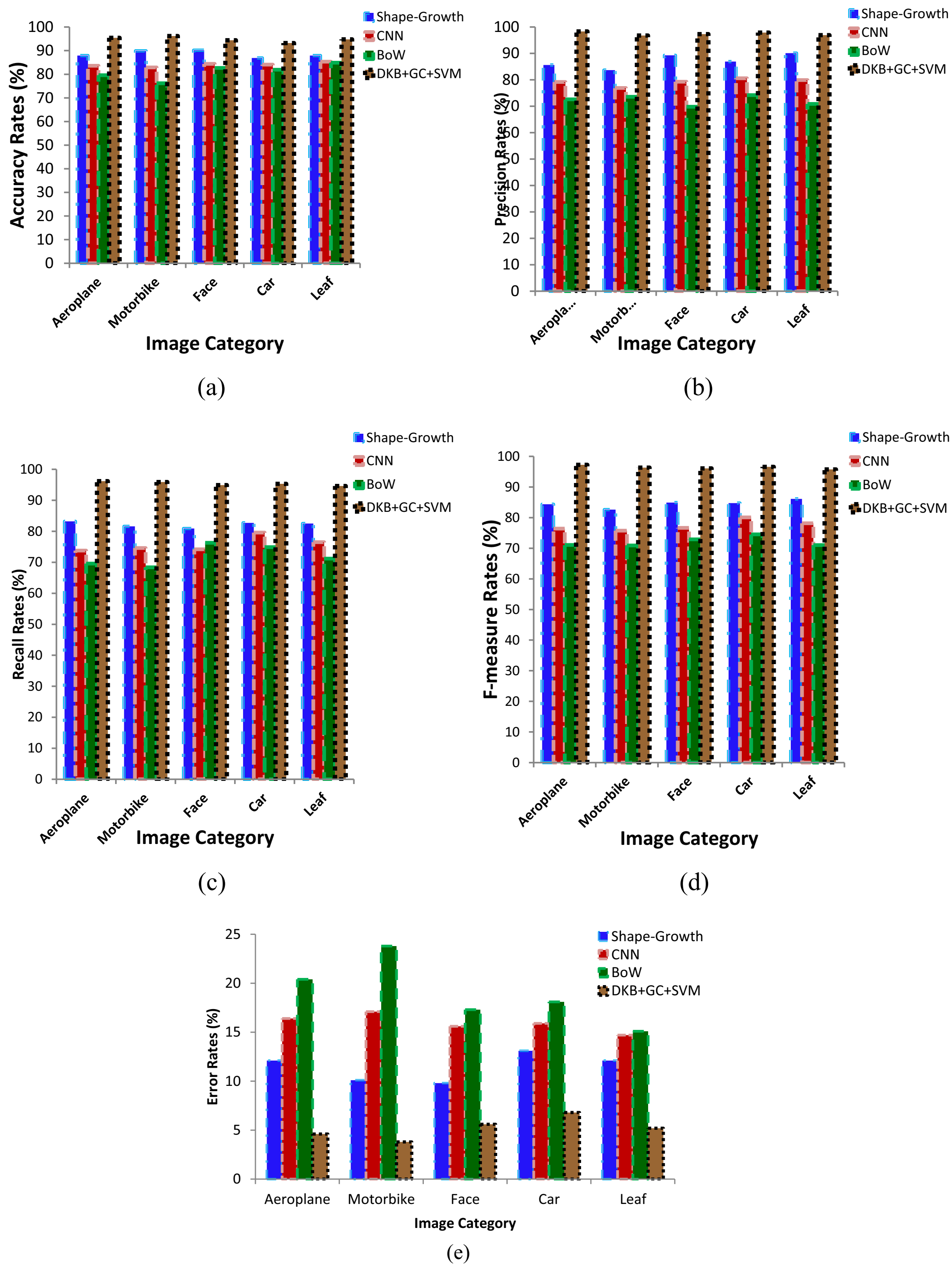

Fig. 3. (a) Performance analysis on accuracy rates (b) Performance analysis on precision rates (c) Performance analysis on recall rates (d) Performance analysis on F-measure rates (e) Performance analysis on error rates.

This section attempts to compare the performance of the proposed object recognition techniques with the existing state-of-the-art object recognition techniques such as shape-growth pattern (Cheddad et al., 2017), CNN (Srinivas et al., 2017), bag-of-words (Hannat et al., 2016). The comparative analysis is carried out in terms of accuracy, precision, recall, F-measure and error rates. The following figure presents the results in graphical representation by plotting the average of the performance measures, which takes all the five categories of images into account (see Fig. 3).

From the experimental results, it is evident that the proposed object recognition system works better in terms of accuracy, precision, recall and F-measure rates. The shape-growth pattern technique works better next to the proposed approach by showing reasonable performance. Though the results of shape-growth pattern are convincing, the computational complexity being involved is greater. The reason is that it relies on geometrical processing of the image pixels. Next to the shape-growth, CNN shows reasonable results. However, CNN expects good labelling process which is a time consuming process. Though this work utilizes dictionary learning and CNN, computational overhead is observed to achieve reasonable results. The BoW approach utilizes SURF features along with SVM classifier. However, this work cannot achieve good results because the features are not good enough to ensure better classification.

On comparison with all the existing approaches, the proposed object recognition technique proves better results. The reasons for the performance of the proposed object recognition technique are no comparative works select interest points to extract features. This work pays more attention in extracting points of interest and the specific pixel window size for feature extraction. The gabor and curvelet features are extracted from the areas of interest and the SVM is trained. The rich and crispy feature set makes it possible for SVM to distinguish between the objects.

5. Conclusion

This article proposes an object recognition system, which is based on points of interest detection, feature extraction and classification. The main substance of this technique is the points of interest detection and the feature extraction from the specific area of interest. The points of interest are selected by means of DKB detector, followed by which the surrounding neighbourhood pixels are processed for feature extraction. The gabor and curvelet features are extracted from the areas of interest and the SVM is trained. With the gained knowledge, the SVM distinguishes between the objects and recognizes them. The experimental evaluation of the proposed approach is done with the help of Caltech5 dataset and tested in terms of accuracy, precision, recall and F-measure. On analysis, it is found that the proposed technique outperforms all the analogous techniques. In future, this work plans to incorporate 3D images for detecting and recognizing the objects.

References

Alom, M. Z., Alam, M., Taha, T. M., & Iftekharuddin, K. M. (2017). Object recognition using cellular simultaneous recurrent networks and convolutional neural network. *International Joint Conference on Neural Networks, 14-19 May, Anchorage, USA*.

Brown, M., Windridge, D., & Guillemaut, J.-Y. (2017). A generalised framework for saliency-based point feature detection. *Computer Vision and Image Understanding, 157*, 117–137.

Bui, H. M., Lech, M., Cheng, E., Neville, K., & Burnett, I. S. (2016). Object recognition using deep convolutional features transformed by a recursive network structure. *IEEE Access, 4*, 10059–10066.

Cheddad, A., Kusetogullari, H., & Grahn, H. (2017). Object recognition using shape growth pattern. *International Symposium on Image and Signal Processing and Analysis, 18-20 Sep Slovenia*.

Hannat, M., Zrira, N., Raoui, Y., & Bouyakhf, E. H. (2016). A fast object recognition and categorization technique for robot grasping using the visual bag of words. *International Conference on Multimedia Computing and Systems, 29 Sep-1 Oct, Morocco*.

Happy, S. L., & Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing, 6*(1), 1–12.

Hsu, Y.-W. (2017). Object recognition system design in regions of interest based on AdaBoost algorithm. *International Conference on Information Fusion, 10-13 July, China*.

http://www.vision.caltech.edu/html-files/archive.html.

Kadir, T., & Brady, M. (2001). Saliency, scale and image description. *International Journal of Computer Vision, 45*(2), 83–105.

Kim, E. K., Lee, H., Kim, S., Cho, H., 2017. Automatic brightness adjustment system by fuzzy inference system for object recognition, In Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS), 27-30 June, Otsu, Japan.

Mikolajczyk, K., & Schmid, C. (2005). A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(10), 1615–1630.

Nikhila, M., & Rawat, S. (2016). Recognition of salient object. *IEEE International Conference on Advances in Electronics, Communication and Computer Technology, 2-3 Dec, Pune, India*.

Srinivas, M., Lin, Y. Y., & Liao, H.-Y. M. (2017). Learning deep and sparse feature representation for fine-grained object recognition. *IEEE International Conference on Multimedia and Expo, 10-14 July, Hong Kong, China*.

Sun, M., Han, T. X., Liu, M.-C., & Khodayari-Rostamabad, A. (2016). Multiple Instance Learning Convolutional Neural Networks for object recognition. *International Conference on Pattern Recognition, 4-8 Dec, Cancun, Mexico*.

Tang, K., Song, P., & Chen, X. (2017). 3D object recognition in cluttered scenes with robust shape description and correspondence selection. *IEEE Access, 5*, 1833–1845.

Wong, S. C., Stamatescu, V., Gatt, A., Kearney, D., Lee, I., & McDonnell, M. D. (2017). Track everything: limiting prior knowledge in online multi-object recognition. *IEEE Transactions on Image Processing, 26*(10), 4669–4683.

Yokono, J. J., & Poggio, T. (2006). A multiview face identification model with no geometric constraints. In, 2006. FGR 2006. In *7th International Conference on Automatic Face and Gesture Recognition* (pp. 493–498). IEEE.

Yörük, E., Öner, K. T., & Akgül, C. B. (2016). An efficient Hough transform for multi-instance object recognition and pose estimation. *International Conference on Pattern Recognition, 4-8 Dec, Cancun, Mexico*.