Archives of Computational Methods in Engineering (2021) 28:1147–1161
https://doi.org/10.1007/s11831-020-09409-1

ORIGINAL PAPER

2D Object Recognition Techniques: State-of-the-Art Work

Monika Bansal[1] · Munish Kumar[2] · Manish Kumar[3]

Received: 26 July 2019 / Accepted: 10 February 2020 / Published online: 18 February 2020

Abstract

Object recognition is one of the research area in the field of computer vision and image processing because of its varied applications in surveillance and security systems, biometrics, intelligent vehicle system, content based image retrieval, etc. Many researchers have already done a lot of work in this area, but still there are many issues like scale, rotation, illumination, viewpoint, occlusion, background clutter among many more that draw the attention of the researchers. Object recognition is the task of recognizing the object and labeling the object in an image. The main goal of this survey is to present a comprehensive study in the field of 2D object recognition. An object is recognized by extracting the features of object like color of the object, texture of the object or shape or some other features. Then based on these features, objects are classified into various classes and each class is assigned a name. In this paper, various feature extraction techniques and classification algorithms are discussed which are required for object recognition. As the deep learning has made a tremendous improvement in object recognition process, so the paper also presents the recognition results achieved with various deep learning methods. The survey also includes the applications of object recognition system and various challenges faced while recognizing the object. Pros and cons of feature extraction and classification algorithms are also discussed which may help other researchers during their initial period of study. In this survey, the authors have also reported an analysis of various researches that describes the techniques used for object recognition with the accuracy achieved on particular image dataset. Finally, this paper ends with concluding notes and future directions. The aim of this study is to introduce the researchers about various techniques used for object recognition system.

1 Introduction

Object recognition is a process of recognition of all objects in a given image and labeling them with their associated classes. In other words, from a given image, the system will recognize and label all the objects with their name. A human can easily recognize the real world things, but it is very difficult for a machine to recognize those things accurately. So, for this purpose the machine has to be trained by using a number of feature extracted by feature extraction algorithms. Objects having similar features are grouped under same class and that class is assigned a name. This method is called classification under supervised machine learning. But, if a new object having unknown features is inputted, then a classification under unsupervised learning is to be done for object recognition. As a human can easily recognize the image by seeing its color, shape, texture or some other feature, the same way machine first extracts the features of the object and then it applies the classification algorithm to label a particular class of the recognized object according to the extracted features. When a dataset size is small, it takes less time to train the data. But, when the size of the dataset is large, training process will take more time and it increases the computational complexity. To make computations easier, object recognition adopts several feature selection and reduction algorithms to reduce the size of feature vector. The object recognition system should also consider all the issues that occurred while recognizing the object such as occlusion, cluttering, scale variation, rotation, illumination etc. The object recognition system identifies the existence of all objects in an image. So, this recognition process is a

✉ Munish Kumar
munishcse@gmail.com

Monika Bansal
write2monikabansal@gmail.com

1 Department of Computer Science, Punjabi University, Patiala, India

2 Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

3 Department of Computer Science, Baba Farid College, Bathinda, Punjab, India

combination of image segmentation and object classification. Image Segmentation is used to separate all objects from the background image. Then these objects are matched with trained images using some classifier and/or some similarity measure. The object is classified based on the minimum distance of the object from a trained image. In this paper, various feature extraction and classification methods for object recognition are discussed with their various pros and cons. This paper is organized as follows. Section 2 considers diverse applications of object recognition. In Sect. 3, several issues and challenges are discussed in the object recognition process. In Sect. 4, a literature review is presented for object recognition. Section 5 presents numerous methods used for feature extraction. In Sect. 6, varied classification techniques for object recognition have been discussed. Section 7 depicts the accuracy achieved on particular dataset using different methods. Finally, in Sect. 8, we have presented concluding notes of this paper.

2 Applications of Object Recognition System

Object recognition system helps in various areas such as content based retrieval system, security and surveillance system, auto-driving system, biometrics, etc. In object recognition system, an object may be a person, animal, chair, table, tree, vehicle or any real world thing. In this section, we present various applications of object recognition system. Object Recognition System helps to count all occurrences of a particular object in an image such as counting of persons, animals, buildings, etc. In Visual Search Engine, an image query is used to search similar images from the image dataset. This system helps to identify and count some specific objects from an aerial image such as trees, buildings, etc. In Intelligent Vehicle System (IVS), this recognition system helps to drive a vehicle automatically and even without the use of specialized sensors. The system helps to recognize pedestrian in a crowded scene. Object recognition can be used in biometric recognition. Biometrics are used to identify an individual based on some biological feature such as eye, ear, fingerprint, etc. The system helps in surveillance systems as it is used to recognize the suspected person or vehicle from a still image or a video. Nowadays object recognition becomes a vital tool to improve security in various places such as airports, government offices, etc. In airports, automated systems could be used to check luggage. Same in banks, the presence of guns and other weapons can easily be identified with the help of an object recognition system. Object Recognition System makes it easy to recognize the parts of machinery and even they can be monitored for any kind of defect or damage in its functioning. It also helps in biomedical signal processing. It includes the activities regarding detection of tumor in MRI images, skin cancer detection and many other diseases and displaying images more clearly captured through X-rays, MRI or from some other machine. Optical Character Recognition (OCR) is also a type of object recognition system wherein characters are identified from scanned documents. The characters may be written in any language. Object Recognition System also helps in assistive devices. In a real-time environment, this system makes it possible for computers to interact with human by storing various human gestures in the system. It also helps the blinds or the persons who have low vision or deaf and dumb persons in their routine life. Object Recognition System also helps in Content Based Image Retrieval (CBIR).

3 Issues and Challenges

A number of issues and challenges are identified during working on the object recognition process. These challenges provide more interest to researchers to solve these problems. A lot of work has already been done to solve these issues, but still it is not completely solved. It is a challenging task to acquire the high accuracy rate and low error rate in an object recognition task. In this section, we will discuss about these issues and challenges. Results of object recognition depend on the feature extraction technique. In real life, an object may not be recognize based on only one feature extraction method. Some objects can be recognized based on color while some can be based on texture or some on shape and so on. So, a combination of various features should be used for gaining more accuracy. Image segmentation is also a major challenge in object recognition due to under-segmentation and over-segmentation of the objects in the image. In under-segmentation case, a region of interest being segmented, may comprise more than one object as a reason of background clutter or occlusion. In over-segmentation, the cluttered objects or objects segmented from the background are segmented into subcomponents. So an efficient image segmentation method should be adopted for image segmentation. In a multi-class object recognition, it is very difficult to recognize all kinds of objects. Intra-class variation is a big challenge in object recognition. Two objects belong to the same class, but the system identifies them as a different class. As the objects are identified by matching the object segmented or test image with the trained image, it requires a lot of computation and memory. The purpose of object recognition system is also to reduce the computations and memory requirement. Scale variation affects the detection process, i.e. the objects of any size should be identified. Occlusion of many objects in an image is a major problem of object recognition. The objects occluded may be of the same kind or it may be of a different kind. Same object has different views from different viewpoints so the recognition system should consider all viewpoints. Deformation of an object means the shape of the object is changed due to stretching, elasticity or some any reason. The recognition system should take into account the articulated object as belong to the right class. Inter-class variation means two different objects seem to belong to the same class, but actually they are not in the same class. It is easy for machine to recognize the object from labeled dataset but if a new object that is not already known is found then it will be difficult for machine to recognize that object [9]. There is one more challenge occurred when number of classes and images in classes are increased. In this case, the complexity may increase and recognition rate may decrease. Convolutional Neural Network based object recognition systems work on machines having CPU with GPU to process the data and they take more processing time and more memory. So, these systems cannot be designed for small devices with limited memory like cell phones [45].

4 Related Work

[58] defined a function on combining a global matching shape measure and local dissimilarity measure for the object recognition task. This two stage experiment shows more accurate results than one stage method. A comparative study is done on 9 hand tools with five functions on these measures. This approach is based on string matching algorithm. Fergus [20] used the constellation model and a learning algorithm to recognize the objects in parts. Helmer and Lowe [23] presented a part based modeling approach for object recognition. It is a probabilistic model that decomposes the object class into various parts and then these parts are added incrementally. This model overcomes the problem of occlusion, scale and background cluttering. Thureson and Carlsson [56] used a histogram of qualitative shape descriptors for object recognition. This descriptor is a combination of triplets of locations and gradient directions in an image. Then the results are computed using modified nearest-neighbor classifier. This proposed method gives better results on two classes (car and motorcycle) as compared to other methods. Afterwards cropping is applied to the object in images, this shows more accuracy in all four classes of object (car, motorcycle, faces and airplanes). Jurie and Triggs [27] created an efficient codebook that helps in visual recognition of objects. SVM classifier is used to find the labels for the objects according to created 600 code words. Sivic et al. [50] used probabilistic Latent Semantic Analysis (pLSA) for object recognition in an unsupervised manner. This model used a bag of words for identifying the object in an image. This is experimented on two datasets- Caltech 101 dataset that has one object per image data collection and other is MIT dataset that has multiple objects per image data collection.

Chum and Zisserman [12] proposed an exemplar model that uses visual words to find object classes in test images. This model detects the region of interest (ROI) to recognize multi class objects in an image. ROI is generated using appearance patches based on Hessian-Laplace operator and SIFT and edge detection based on the canny edge detector. A hierarchical spatial pyramid histogram measures the visual similarity between images. They used the vocabulary of 3000 visual words trained on Pascal VOC 2006 training sets. This model is experimented on Pascal VOC 2006 detection challenges and Caltech-4. Arnow and Bovik [1] developed a gray scale object recognition system. This system is an object recognition system that uses corner feature and a modified Lowe's SIFT algorithm is used to accomplish this task. They suggest adding a global feature that improves to accuracy. Yakhnenko and Honavar [63] proposed multimodal hierarchical Dirichlet Process model (MoM-HDP). This model recognizes the objects from an image and label them appropriately using the codebook collected from the trained images. Kim et al. [29] presented a comparative analysis between k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM). The authors experimented both classifiers on Caltech-4 cropped dataset and proved SVM works more accurately than k-NN. Roy and Mukherjee [47] proposed a new scheme of feature extraction that is a combination of color histogram, color coherence vector and Sobel edge detector for Content Based Image Retrieval. These three feature vectors help to identify the similar images. The Manhattan Distance method is used to compute the similarity measure. In their next research, they will work on identifying the similar kind of objects. Uijlings et al. [57] proposed a selective search algorithm for object recognition. This strategy is experimented on PASCAL VOC 2010 dataset. They used the bottom up segmentation technique. A comparative view is taken among various state-of-the-art recognition techniques where the authors' proposed technique outperforms in case of recognition of an object and computation time. Ijjina and Mohan [26] presented 3D color histogram as a feature detector and CNN for object classification. This work is experimented on ALOI dataset and outperforms the results with average 97% accuracy. Muralidharan [40] constructed feature vector by using a canny edge detection algorithm after converting the colored image into a greyscale image and then applying Hu's Seven Moment Invariants, Center of the object and the dimension of the object in the extracted edge information. Eigenvalues are computed from the computed feature vector using Principle Component Analysis. K-nearest neighbor classifier is used to recognize the object by comparing the computed Eigenvalue of the query instance with the stored database of Eigenvalue. A comparative results of k-NN are shown against Fuzzy k-NN and Back Propagation Network.

Diplaros et al. [16] proposed a combination of color and shape feature extraction scheme for object recognition. This proposed scheme also recognizes the object in the presence of object occlusion and cluttering. This is experimented on various datasets and proved more accurate with fast recognition. Prajapati et al. [42] introduces shape feature for recognizing the object in an image. Shape descriptor is extracted by combining three features – Edge Histogram, Geometric moments and Sobel edge detector. The computed feature vector of the query image is compared with the stored feature vectors of image database and obtained the best match images from the database. Euclidean distance metric is used to compute the similarity. In future, this method will be experimented on large dataset. Rastegari et al. [45] developed two approximations to Convolutional Neural Networks – Binary-Weight-Networks and XNOR-Networks. Binary-Weight-Networks a version of AlexNet performs same like AlexNet but the authors compared their proposed binarization method with other binarization methods- BinaryConnect and BinaryNets. And the proposed methods outperform on ImageNet by more than 16% in top-1accurcay. The proposed method reduced the size of network by ~ 32 × and it made possible to run the recognition system on small machines. Chao et al. [9] proposed conventional zero-shot learning(CZSL) that overcome the problem of generalized zero-shot learning (GZSL). In the paper, authors discussed to identify the object from seen images as well as unseen images. The experiment was done on AwA, CUB and ImageNet datasets in the wild. Authors used stacking classifiers for the classification of seen and unseen objects. Wei et al. [61] proposed a novel framework for object detection and framework. This framework is based on contour shape descriptor and this also works accurately in cluttered images. This is experimented on several datasets–ETHZ shape classes, INRIA horses, Weizmann horses, two classes of caltech101.

5 Feature Extraction Techniques

Each object in our real life is associated with one or more features which help to identify its class or name. For example, a human face has two eyes, one nose, one mouth, skin of brown color, round shape, etc. These extracted features are the information that help in the task of object recognition and classification. In image processing, features are defined as a quantitative description of an image which is represented by feature vector. Similarity between features extracted from the input image and computed features of stored image help to identify the particular class of the image. There are various types of feature extraction techniques with respect to 2-D image which help in computation of the features of the image. The features extracted from an image are categorized into two types, namely, general features and domain specific features. General features are not dependent on some specific application like color, texture and shape. General features are further categorized into Local Features and Global Features. If the whole image is considered, then the system uses global features of the image. Global features help to recognize the object from the background, but these features work well only when there is no problem of occlusion, illumination. Area, perimeter, compactness and Euler number are considered here. If a selected region of the image is considered, then the system extracts local features of the object. Edge and shape are considered here. Domain specific features are specific to any particular application such as human faces, automobile detection, fingerprint etc. These features are often a combination of low-level features of a specific domain. The performance of these applications also depends on the type of classifier used. Classifier is selected based on the type of feature for individual application. In this paper, a few state-of-the-art feature extraction algorithms are explained in detail that are grouped into three categories i.e. feature based, textured based and shape based.

5.1 Feature Based Algorithms

5.1.1 Color

Color is one of the important features through which one can easily identify the object. For example – a bird having a green color is a parrot, a fruit having the red color is apple [43]. It is easy to extract, analyze and represent an object with color feature. The color descriptor of an image can be local or global. Color descriptors are represented using color moments, color coherent vector, color correlogram [24], fuzzy color moments, color histogram etc.

- Color moments have been successfully used in many retrieval systems. The first order (mean), the second (variance) and the third order (skewness) color moments have been proved to be efficient and effective in representing color distribution of images. This feature is proposed by [17, 51, 64] for image retrieval. All the three orders can be calculated either for each color band of an image separately or for gray band. If it is used separately for each of color bands, then there will be $3 \times p$ color moments making a feature vector. Where p is equal to the number of color bands in an image. In case of gray band, there will be only 3 color moments.

  Pros:

- It offers computational simplicity, speedy retrieval, and minimum storage.
- Since only 9 (three moments for each of the three color components) numbers are used to represent the color content of each image, color moments are a very compact representation compared to other color features.
- They are very robust to the complex background.
- They are independent of image size and orientation.

Cons:

- Two images with similar color histograms can possess different contents.
- It is more sensitive to the color and intensity of the light source.

- Color Histogram describes the distribution of pixel color within a whole image or within a region of interest. For the given color image, the color histogram is obtained by discretizing the image colors into a number of bins (or image array) and counting the number of times each discrete color occurs in the image array. Swain and Ballard [53] introduce the histogram based recognition methods.
  Pros:

  - The histogram is invariant to rotation, translation and scaling of an object.
  - It takes less computational time as compared to classical color.

  Cons:

  - Two images with similar color histograms can belong to different contents.
  - It is more sensitive to the color and intensity of the light source.
  - Huge memory is required as the number of bins is increased.
  - It shows the results without considering spatial and shape information.
  - It increases the computational cost.

5.1.2 Corner

Corner descriptor is used as interest points in the image through which the similarity of one image is matched with the other image. This feature is a global descriptor. By using corner descriptor, all issues regarding scale, rotation and deformation can be eliminated.

- Harris Corner detects the little patch of the image that generates a large variation when moved around. The Harris corner detector has a corner selection criterion. It was introduced by Chris Harris and Mike Stephens in 1988. In Harris Corner detector, a score is determined for each pixel, and if the score is above a certain value, the pixel is considered as a corner. Two Eigen values are used to calculate the score.
  The score for Harris corner detector was calculated like this (R is the score):

$$R = detM - k(traceM)^2$$
$$detM = \lambda_1 \lambda_2$$
$$traceM = \lambda_1 + \lambda_2$$

  Kitti et al. [30] proposed Harris Corner detector for object identification and shape recognition.
- *Shi-Tomasi corner detector* The Shi-Tomasi corner detector is based entirely on the Harris corner detector. A slight change has been made in selection criteria that gives much better results than the Harris corner detector. This detector was developed by Shi and Tomasi. For Shi-Tomasi, R, the score is calculated as:

$$R = min\left(\lambda_1, \lambda_2\right)$$

  Shi and Tomasi demonstrated the performance of their work experimentally in their paper. If R is greater than a certain predefined value, it can be marked as a corner.
  The FAST Harris algorithm was only detecting corners, it does not determine the connectivity of feature-points through which major level descriptors such as surfaces and objects can be obtained. A new corner detector algorithm called FAST (Features from Accelerated Segment Test) was presented by Rosten and Drummond [46] to solve the issues of Harris algorithm. FAST is a computationally efficient algorithm as compared to other algorithms. Machine learning algorithm uses FAST for efficient execution of the algorithm.
  Pros:

  - It is faster than other corner detection algorithms.
  - High levels of repeatability under large aspect changes and for different kinds of features.

  Cons:

  - It does not give accurate results in the presence of noise.
  - It is dependent on a threshold.

5.1.3 Edge

The edge detection feature helps to locate the edges of the object. Edge descriptor captures the information of five types of edges in an image. The detected edges are used for recognition of the object. This feature reduces the extracted data as useless edges are not preserved for matching the test image with the trained image. There are many edge detection techniques like Robert, Sobel, Prewitt, Kirsch, Robinson, MarrHildreth, LoG and Canny Edge Detection. Edge feature works for scale invariant object recognition (Fig. 1).

- Canny edge detection is a multistage algorithm that is used to detect the edges of the image. It was developed by John F. Canny in 1986. This algorithm reduces the amount of data to be processed for edge detection compared to other algorithms. It is computationally expensive than other edge detection algorithms. It performs 5 operations one after another, as following:-

  1. The Gaussian filter is applied on the image to smoothen the image.
  2. Intensity gradients of the image are found
  3. Non-maximum suppression (NMS) is applied.
  4. The threshold is applied to find potential edges.
  5. Then finally, the edges are tracked by hysteresis.

  Performance of canny edge detector is presented by showing the comparative study among various edge detectors in a research paper of [31,44].
- Sobel operator computes 2-D spatial gradients of an image and so emphasizes regions of high spatial frequency that corresponds to the edges. It was introduced by Irwin Sobel and Gary Feldman in 1968. The operator used 3×3 matrix which is convolved with the original image to produce separate measurements of the approximation of the gradient in both changes – horizontal (say Gx) and vertical changes (say Gy). Then these calculated gradients are combined together to find the absolute magnitude of the gradient at each point and orientation of the gradient. The gradient magnitude is given by

$$|G| = \sqrt{Gx^2 + Gy^2}$$

  Using this information, gradient's direction is computed as

$$\theta = \operatorname{atan}\left(\frac{Gy}{Gx}\right)$$

  Pros:

  - It is invariance to affine geometric and photometric transformation.
  - It performs well by rejecting false matches.

  Cons:

  - It does not perform well in case of background cluttering.
  - It is not scale invariant.

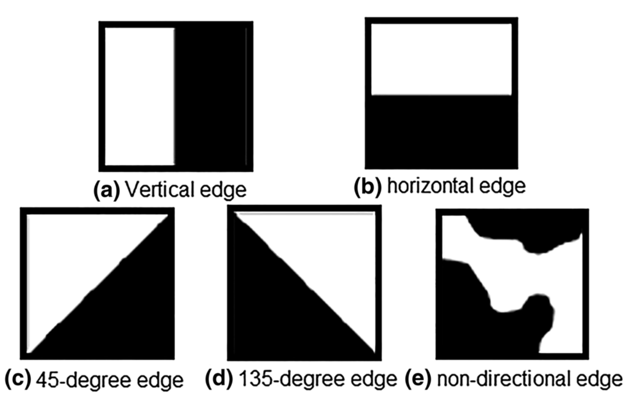

Fig. 1 Five type of edges

5.1.4 Texture Based Algorithms

When various intensities of interactions within a region are required, then textual descriptor is used. It is a regional descriptor that is used in the image retrieval process. Texture represents the properties of an object in terms of its smoothness, coarseness, and regularity. It alone cannot be used to find similar objects rather, it is combined with another feature descriptor for the identification process. Texture representation can be divided into two categories-structural and statistical. Statistical methods, including Fourier power spectra, gray level co-occurrence matrices(GLCM), Haar wavelet, SIFT-invariant principal component analysis (SPCA), Tamura features, World decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor filters [52], Gabor and wavelet transform [62], describe texture by the statistical distribution of the image intensity. Textural feature improves the classification accuracy and it reduces the computational time.

- Gray level co-occurrence matrix (GLCM) computes the texture information from an image in two dimensional arrays in the form of intensity of the pixel and describes the spatial relationship between these. It is a statistical method that extracts second order texture information from images. This method shows the relationship between two neighboring pixels. Then various features can be selected from GLCM extracted features like energy, entropy, contrast, homogeneity, correlation, shade, prominence etc. Saraswat et al. [48] proposed a methodology to combine GLCM with Haar wavelet. This method will improve classification accuracy and reduce the computation cost. Syahputra et al. [54] proposed GLCM for recognizing plant based on leaf shape.

- Haar Wavelets is a statistical method that uses a windowing technique to select the different regions in an image. It is a discrete wavelet transform that uses a series of low pass filters and high pass filters to perform wavelet analysis and divide the image into various frequency bands. This method describes the local features of the image.

  Pros:

  - It has high computational speed.
  - It works in multi scale analysis of the image.
  - There is no loss of information in the processing of data

  Cons:

  - It does not provide continuous data.
  - It produces a large number of data for all objects in the image.

5.1.5 Shape Based Algorithms

An object is more accurately identified using the shape of the object. But, the object shape is difficult to be determined in a quantifiable manner. Therefore, it works in two steps – first the features are extracted and then a measure of similarity between these features is done. This descriptor combines two or more features to extract the shape information like color and edges, texture and edges etc. Shape descriptor has two categories-region based shape descriptor and contour based shape descriptor. In region based method, complete area of the object is determined by considering all pixel information within the shape region while contour based use local feature as they need an extraction of boundary information. A region based model is used for the segmentation while contour based model needs edge detection algorithm. Belongie et al. [4] introduced shape context descriptor for object recognition.

5.2 Some Feature Descriptors

5.2.1 HOG

Histogram of Oriented Gradients (HOG) is used to determine the object appearance and the shape of the image. The shape of the image is described by the distribution of intensity gradients. The image is split into sub regions and intensity value for each pixel in that sub region is computed in all directions and put in a bin. Then the values or directions in the bin are used to find the edges in the image. This process is repeated for all sub regions and at last, all these HOG of individual sub regions are summed up to determine the final HOG descriptor. Dalal and Triggs [13] proposed HOG for human detection. Dalal and Triggs [14] presented the results of HOG on various image objects. There are other variations of HOG as RIHOG proposed by Vashaee et al. [59] that overcomes the issues of HOG.

Pros:

- It is invariant to geometric and photometric transformation.
- It shows better results in invariance to changes in illumination or shadowing.

Cons:

- It takes more time to extract features and to train using the classifier.
- It is variant to object rotation.
- It has more computation cost.
- It makes the false rate recognition of object in an image ([60].

5.2.2 SIFT

Scale Invariant Feature Transform (SIFT) receives image as $N \times N$ data and produces a set of features from that inputted image. The extracted features are local features. SIFT output is invariant to scale, rotation, lighting and camera viewpoint. So it detects an object in case of background clutter and occlusion. This algorithm is developed by Lowe [34]. An array of oriented histogram is created of size 4×4 with 8 bins. 8 bins are the 8 angles of orientation. So SIFT produces $4 \times 4 \times 8 = 128$ features. There are other variations of SIFT that provide more accurate results than SIFT like PCA-SIFT, RIFT, Gauss-SIFT etc. Shivakanth and Mane [49] presented the results of SIFT descriptor for identifying the objects.

Pros:

- SIFT is invariant to rotation, scale, lighting and viewpoint.
- It is more accurate than other descriptors.

Cons:

- It is mathematically complicated.
- It takes more time to compute the feature vector.
- Sometimes it does not give accurate results on the blurred image.

5.2.3 SURF

Speeded Up Robust Feature (SURF) is a speed up version of SIFT. SURF follows the same steps of SIFT but, will do little change in methods in each step. SURF was first introduced by [3] for object recognition. SURF uses a LOG with Box filter to detect scale invariant characteristic points. SURF used the Hessian matrix to evaluate points of interest. The features extracted are also local features same as SIFT but SURF do all calculations faster than SIFT. For orientation assignments, SURF works in following steps as follows:-

- It uses a Haar wavelet algorithm to detect responses in both horizontal and vertical directions in a neighborhood of size 6 s where s is a scale at which the point of interest was detected.
- Gaussian weights are applied to it.
- All responses within a sliding orientation window of angle 60 degrees are summed to estimate the dominant orientation.
- These summed responses, then yield a local orientation vector. The size of the descriptor is 64 bins so it takes less time for computations.

Pros:

- SURF has shown good results in rotation, blur and illumination as compared to SIFT.
- It is fast compared to SIFT.

Cons:

- It has less results in the scale area as compared to SIFT.

6 Classification Techniques

The features extracted using various algorithms as discussed in Sect. 5 are further classified for recognition task. Various classification techniques are used to classify the different objects in an image that is performed using some machine learning or statistical learning technique. Classification comes under supervised or unsupervised learning. In supervised learning, we have trained data that is matched with input data to find the correct label for the object. Classification under supervised learning involves Naïve Bayes classifier, Bayesian network, SVM, KNN, Decision Tree. In unsupervised learning, there are no trained data available. Classification under unsupervised learning involves k-means, Fuzzy clustering, and hierarchical clustering. In this section, some classifiers are discussed.

6.1 Supervised Classifiers

These classifiers need a set of training data to identify the label for the object in the query image.

6.1.1 BayesNet

Bayesian Network is a probabilistic graphical model. It represents a set of variables (features) and their relationship via directed acyclic graph (DAG). It is assumed that all features are statistically dependent on each other. Bayesian creates a full network of all extracted features so it improves the accuracy by a small amount, but it also increases the computational cost for classification tasks. Here, the classification is performed using Bayes theorem:-

$$p(\theta_i|F) = \left(p(F|\theta_i)p\left(\theta_i\right)\right)/p(F)$$

Elazary and Itti [19], Fergus [20], Miller et al. [39], Tao and Hung [55] proposed Bayesian network for image classification.

Pros:

- It is easy to implement and fast.
- It requires less training data.
- It can make probabilistic predictions.
- It is more accurate than naïve Bayes method.
- It can be used for both binary and multi-class classifiers.

Cons:

- It requires more memory than naïve Bayes classifier.
- It increases the computational cost.

6.1.2 Decision Tree

Decision Tree is a learning method that builds a classification model in a tree structure. It breaks the dataset into smaller subparts and these subparts are further divided into subparts. This division of nodes goes till no further decision obtained. In this tree, the nodes that have two or more branches is called a decision node and the node that has no branch is called a leaf node. Cheong et al. [11], Madzarov et al. [35] and Madzarov and Gjorgjevikj [36], presented the use of SVM based on Binary Decision Tree (SVM-BDT) for multi-class classification problem.

Pros:

- It is easy to interpret the tree.

- It can handle both numerical and categorical dependent and independent features.
- It is easy to access the features of a particular object.

Cons:

- It may favor variables that have more categories.
- It is generally less preferable than other statistical classifiers when higher precision and recall are required.

6.1.3 CNN

Convolutional Neural Networks is a variant of MLP that requires minimum preprocessing. It is composed of three or more layers—one input layer, one output layer and in between these layers are one or more than one hidden layers. Hidden layers of CNN typically consist of convolution layer, pooling layer, fully connected layers, and normalization layers. Convolutional layers take the use of the local spatial coherence of the input to cut down on the number of parameters by sharing weight. Pooling layer combines the output of nodes in one layer to the nodes in its next higher level layer. The last layer of CNN determines the presence of objects in an image. There are other variations of CNN – R-CNN, Fast R-CNN. Huang and LeCun [25] combines two supervised learning methods—SVM and CNN for recognition of object. Atabay [2] proposed CNN for binary shape classification problems.

Pros:

- CNN improves the performance of various complex problems in computer vision fields.
- Its three layers makes the task of feature extraction and classification easier.
- When there is large dataset, even then it performs well.

Cons:

- The computation complexity is more comparative to other classifiers.
- It requires more memory to store the data used for object recognition.

6.1.4 *k*-NN

k-NN (*k*-Nearest Neighbor) classifier is a supervised learning method for classifying the objects by assigning the label to the object. The label chosen is more frequent among the *k* training samples nearest to that feature vector of unlabeled object. Here the data is trained by extracting various features of the set of training images and then applying PCA to reduce the dimensionality of the data. Finally, it produces the Eigen values of the image. *K* is a positive integer that is used to classify the object by using a majority vote of its neighbors. Euclidian distance is used to compute the distance between the query instance and the k training data. All categories of trained data are taken out which falls under *k*. On last majority of nearest neighbor is used for predicted value as a label for the object. Kim et al. [29] used SIFT descriptors and SVM and k-NN classifiers for object recognition.

Pros:

- It performs well in multi-model classes.
- It shows good accuracy as compared to other classifiers as it considers all features of a model.
- This classifier is easy to understand and implement.
- It requires less memory to store the feature vector for training data because of the use of PCA.

Cons:

- It can lead to classification error when there is only a small subset of features.

6.1.5 MLP

Multilayer Perceptron is a supervised learning technique and a class of artificial neural network. It is a feed-forward net with three or more layers of nodes (an input and an output layer with one or more hidden layers) as shown in Fig. 2.

Each layer in MLP consists of a number of nodes in parallel. An input layer has N weighted nodes and each node in the layer connects to each node in its next higher level layer. The output of multilayer perceptron is computed by summing the weights of all inputs at each layer level and passed the results to the next layer. The results also include threshold. In the output layer, all nodes be set to 0 except for the node that is marked to correspond to the class the input is from. That desired output is 1. Khotanzad and Lu [28] proposed MLP for classification of images. Frias-Martinez

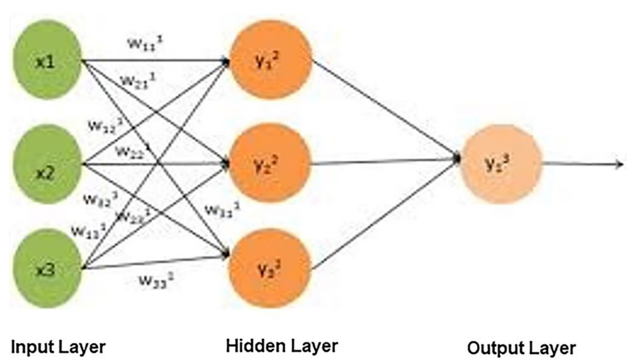

Fig. 2 A multilayer perceptron (MLP) with one layer

et al. [21] introduced SVM and MLP and also compared SVM over MLP.

6.1.6 Naïve Bayes

Naïve Bayes classifier is a probability classifier that learns the conditional probability of the object against the multiple features of the object extracted. It is called naïve as it assumes that the features extracted are independent from each other. Before applying classifier, some data is trained and the test images are experimented on these training images. This classifier is based on Bayes' probability theorem. e.g. if we are identifying the fruit objects in an image, then the specific object, whether it is apple, orange, etc. can be easily identified based on the color, texture, or shape, such as a fruit is apple if red color is found or it may be an orange if orange color is found. This is a technique to label the object from an available set of labels. McCann and Lowe [37] presented local Naïve Bayes nearest neighbor algorithm for image classification.

Naïve Bayes classifier is expressed by the following equation:-

$$P\left(X_1, X_2, \dots X_{n,} C_i\right) = P\left(X_1 | X_2, \dots X_{n,} C_i\right) \cdot P\left(X_2 | X_3, \dots X_{n,} C_i\right) \cdot P\left(X_{n-1} | X_n, C_i\right) \cdot P\left(X_{n,} C_i\right) \cdot P\left(C_i\right)$$

where X1, X2….Xn are the features extracted.

Pros:

- Image dataset can be trained easily.
- It takes less time for computations.
- It is easy to understand and develop

Cons:

- It is very confident for independent features that it makes.
- It computes less classification accuracy.

6.1.7 Random Forest

Random forest is a machine learning algorithm that combines a number of decision trees. The height of the tree is grown by using some randomization method. Here prediction about the class of an object depends on the majority of decisions made by a k-decision trees. The random forest algorithm works in two stages. First, Image dataset is trained by randomly dividing the dataset into k-decision tree. The leaf nodes of each tree are labeled by estimates of the posterior distribution over the image classes. Second, test image is passed down each random decision tree until it reaches the leaf node. All the posterior probabilities are then combined and a final prediction is derived by considering the prediction that appears the most time in decision trees. Lepetit and Fua [33] proposed random forest classifier for 3-D object detection and pose estimation. Bosch et al. [6] compared the performance of random forest/ferns classifier over SVM classifier.

Pros:

- It provides accuracy even in case of availability of inconsistent data.
- It runs efficiently on large database.
- It is able to handle both numerical and categorical data without scaling and transformation.

Cons:

- A large number of decision trees can slow the task of predicting the class.
- It has poor performance on imbalance data.

6.1.8 SVM

Support Vector Machine (SVM) is the most used algorithm for object classification. It is a supervised machine learning algorithm. SVM builds hyperplane of some dimensional space that generates the partitions of the object. Distance between training data points and hyperplane is considered to classify the object that will depend on choice of right hyperplane. And the decision of right hyperplane will depend on the maximum distance between the nearest data points and hyperplane. The goal of SVM is to classify the objects with minimum error rate.

Considering the above example where features of the two classes that are represented in two different types of shape (like a star and circle) are to be separated by taking three hyperplanes (A, B, C). In Fig. 3a, hyper plane B appears to separate the two classes more clearly as comparative to hyperplane A and C. Same in Fig. 3b, hyper plane C is chosen as best classifier as it is at a maximum distance from two classes as compared to other hyper planes. It is a binary classifier and can also be used for multi class problems. [38] proposed hierarchical SVM for object recognition.

Pros:

- It gives the more accurate results than other methods.

Cons:

- It takes time to train data.
- It requires extra memory to store the features of trained images.

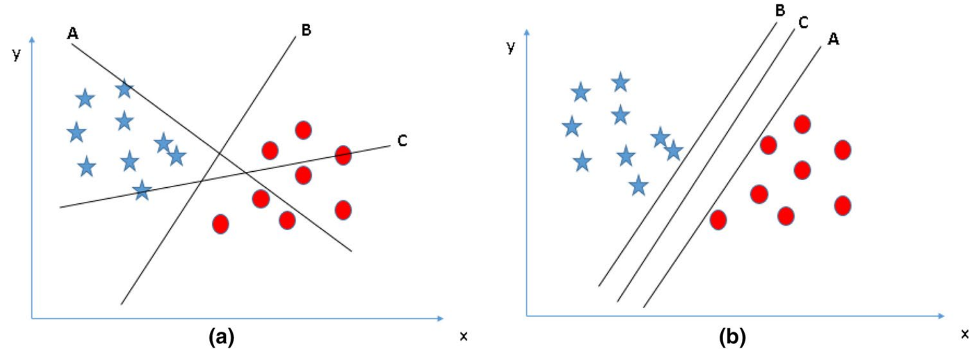

Fig. 3 Depicting SVM classification through three hyper planes A, B, C

6.2 Un-Supervised Classifiers

In unsupervised learning, there are no training data available with which query data can be matched for classification. Rather in unsupervised learning, we aim to identify regularities in data. The identification task is done by using a similarity measure. It is a task of describing unknown features from unlabeled data. The best method of unsupervised learning is done with clustering. Clustering is a statistical technique where an entire dataset is divided into a number of clusters based on common feature or class. The items belonging to the same cluster are similar. Rather the items in different clusters are considered as dissimilar.

6.2.1 K-Means

K-means is a clustering algorithm that is used when we have unlabeled data. It is one of the famous hard clustering algorithm. This method divides the whole dataset into k groups based on the extracted feature. K-means clustering algorithm outputs centroid of k clusters that is used to train the data and creates labels for both train and test data. It is applicable in those areas where one data belongs only to one cluster or class. Leibe et al. [32] proposed K-means based clustering algorithm and a fast NN search method for object recognition. Chandhok et al. [8] presented K-means clustering algorithm for image segmentation. Charaniya and Rathod [10] proposed SIFT for feature extraction and various segmentation methods such as GraphCut, K-means, and Linde–Buzo–Gray (LBG) algorithm.

Pros:

- This is a simple and popular clustering algorithm.
- It is computationally fast.
- It works to segment large datasets.

Cons:

- Number of clusters, i.e. K should be already known.
- It is a linearly separating algorithm.

6.2.2 Fuzzy Clustering

Fuzzy clustering is applicable in those areas where one data item may belong to two or more clusters. So this is called soft clustering. When the data is more complex, it becomes difficult to obtain the correct classification. In the complex dataset, one data item may belong to many clusters. In this case, the degree of belongings will not be 0 or 1, rather than it will be between 0 and 1. Dell'Agnello et al. [15] proposed the bag-of-visual-words model for image classification. They presented the comparative view of k-means, Fuzzy C-means (FCM) and the kernalized version of FCM (i.e. KFCM). Hatori and Sato-Ilic [22] proposed the use of a fuzzy clustering algorithm for object classification when a complicated similarity data is given. Bora and Gupta [5] presented a comparative study between Fuzzy clustering and k-means clustering and proved the results of k-means over Fuzzy clustering.

Pros:

- Gives best result for overlapped data set and comparatively better then k-means algorithm.
- Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more then one cluster center.

Cons:

- Apriori specification of the number of clusters.
- Euclidean distance measures can unequally weight underlying factors.

6.2.3 Hierarchical Clustering

Hierarchical clustering is a statistical method that groups similar objects into clusters and then groups into similar clusters. This process continues until all objects are not in the same cluster. It works in two approaches- agglomerative (top-down approach) and Divisive (bottom-up approach). In

Table 1 Results of object recognition using various feature extraction techniques

Authors	Data set	No. of images	Recognition technique	Feature extraction technique	Classifier	Accuracy
[20]	Corel Dataset	2235	Part based	Entropy based	Bayesian	90–96.4%
[23]	http://www.robots.ox.ac.uk/~vgg/data/	2900	Part based approach	Entropy based		More accuracy in car object class
[56]	Caltech	2476	Appearance based	combination of triplets of sampled locations and gradient directions	Edited nearest neighbor	89–96%
[27]	Side views of cars Xerox 7 ETH80	1000 1776	Bag of words	SIFT	SVM	86.6–98.6%
[50]	Caltech 101 MIT	4090 2873	Bag of words	SIFT		Above 96%
[12]	Pascal VOC 2006 Caltech 4		Appearance based	ROI (SIFT + canny edge detector)	SVM	Above 99%
[1]	hand tools from Sclaroff and the Rutgers tool database, and images of fighter and passenger jets	78	Shape based	Corner + SIFT		60–100%
[63]	VOC 2007 LabelMe	2501 (training) 4952 (testing) 7373 (training) 1513 (testing)	Feature based	SIFT	multi-modal hierarchical Dirichlet Process model (MoM-HDP)	41% 28.45%
[29]	Caltech-4 cropped dataset	2800 (training) 700 (testing)	Bag of words model	SIFT	SVM, k-NN	91.9%
[47]	No standard dataset	50	Feature based	Shape feature (Color Histogram, color coherence vector, sobel edge detector)		More than 83%
[57]	PASCAL VOC 2010		Appearance based	Bag of words	SVM	99%
[40]	COIL-100	2000 (training) 560 (testing)	Shape based	Canny edge detection	kNN	97%
[26]	ALOI		Appearance based	3D Color histogram	CNN	95.7–99.6%
[18]	RGB-D	35000 (training) 7000 (testing)	Deep Learning		CNN	91.0%
[16]	Amsterdam COIL-100 COREL	500 7200 2600	Appearance based	Color and shape		99% (avg ranking percentile in case of occlusion and cluttering)
[42]	Wang	1000	Feature based	Shape feature (Edge histogram, geometric moments, sobel edge detector)		60%

Table 1 (continued)

Authors	Data set	No. of images	Recognition technique	Feature extraction technique	Classifier	Accuracy
[61]	ETHZ	255	Shape based (template based)	Edge fragment, shape descriptor (contour based)	SVM to Pyramid histogram of oriented gradient	97%
	INRIA horses	340				
	Weizmann horses	456				
	caltech101 (anchors and cups)	99				
	MPEG-7	1400				
	Kimia's 99	99				
	Kimia's 216	216				

hierarchical agglomerative clustering (HAC), similar clusters are merged into a new cluster. [32] suggest HAC with a partitioned clustering strategy to improve the matching speed in multi-class object recognition. In Hierarchical Divisive Clustering (HDC), the process starts from one cluster and splits into a number of clusters recursively. Zaboli et al. [65] proposed clustering and matching method for object recognition. This is done by extracting shape descriptor. Namratha and Prajwala [41] make a comparison between k-means and hierarchical clustering algorithm by mentioning its advantages and disadvantages. Bouza and Cernuschi [7] propose hierarchical clustering and asymmetric dissimilarity function to detect all common visual descriptors in two images using a directed graph.

Pros:

- Prior knowledge of the number of clusters is not needed.
- It requires only a similarity measure.

Cons:

- It has slow run time.

7 Recognition Results

A complete survey on various research methodologies for object recognition is depicted through Table 1. Table 1 shows the feature extraction methods and similarity measurement algorithm with accuracy presented by many researchers.

8 Conclusion and Discussion

In this paper, we have presented a survey on various feature extraction and classification techniques used for object recognition. We have also surveyed various classifiers under supervised and unsupervised machine learning that accurately labeled the object with proper class. A lot of work is done on single object recognition but multi-object recognition problem is not accurately solved till now. During surveying, we also determine that one method does not always recognize all objects with high accuracy. So, in future work, a combination of feature extraction and classification techniques may be propose to solve the problem of multi-object recognition and to report the improved recognition results for benchmark datasets. Papers surveyed are experimented on standard datasets like Caltech-101, COIL-100, Pascal VOC 2007, etc.

Compliance with Ethical Standards

We have surveyed on various object recognition techniques. Papers surveyed are experimented on standard datasets like Caltech-101, COIL-100, Pascal VOC 2007, etc.

Conflict of interest Authors have no conflict of interest.

References

1. Arnow T, Bovik AC (2008) Foveated object recognition using corners. In: Proceedings of the IEEE southwest symposium on image analysis and interpretation, 53–56
2. Atabay HA (2016) Binary shape classification using convolutional neural networks. IIOAB J 7(5):332–336
3. Bay H, Ess A, Tuytelaars T, Van-Gool L (2008) Speeded-up robust features (SURF). Comput Vis Image Understand 110(3):346–359
4. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(4):509–522
5. Bora DJ, Gupta AK (2014) A comparative study between fuzzy clustering algorithm and hard clustering algorithm. Int J Comput Trends Technol (IJCTT) 10(2):108–113
6. Bosch A, Zisserman A, Munoz X (2007) Image classification using random forests and ferns. In: Proceedings of the IEEE 11th international conference on computer vision (ICCV), 1–8
7. Bouza M, Cernuschi FB (2015) Common visual pattern recognition using hierarchical clustering methods with asymmetric networks. In: Argentine symposium on artificial intelligence (ASAI), vol 44, pp 192–199
8. Chandhok C, Chaturvedi S, Khurshid AA (2012) An approach to image segmentation using K-means clustering algorithm. Int J Inf Technol (IJIT) 1(1):11–17
9. Chao WL, Changpinyo S, Gong B, Sha F (2016) An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In: European conference on computer vision, 52–68
10. Charaniya NA, Rathod MY (2016) Object recognition using image segmentation. Int J Electr Electr Eng (IJEEE) 8(1):12–18
11. Cheong S, Oh SH, Lee SY (2004) Support vector machines with binary tree architecture for multi-class classification. Neural Inf Processing-Lett Rev 2(3):47–51
12. Chum O, Zisserman A (2007) An exemplar model for learning object classes. In: Proceedings of IEEE conference on computer vision and pattern recognition, 1–8
13. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the international conference on computer vision and pattern recognition (CVPR), vol 1, pp 886–893
14. Dalal N, Triggs B (2006) Object detection using histograms of oriented gradients. In: Proceedings of the Pascal VOC Workshop, ECCV, 1–5
15. Dell'Agnello D, Carneiro G, Chin TJ, Castellano G, Fanelli AM (2013) Fuzzy clustering based encoding for visual object classification. In: Proceedings of the In Joint IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS), 1439–1444
16. Diplaros A, Gevers T, Patras I (2016) Combining color and shape information for illumination-viewpoint invariant object recognition. IEEE Trans Image Process 15(1):1–11
17. Duanmu X (2010) Image retrieval using color moment invariant. In: Proceedings of the seventh international conference on information technology, 200–203
18. Eitel A, Springenberg JT, Spinello L, Riedmiller M, Burgard W (2015) Multimodal deep learning for robust RGB-D object recognition. In: International conference on intelligent robots and systems (IROS), 681–687
19. Elazary L, Itti L (2010) A Bayesian model for efficient visual search and recognition. Vis Res 50(14):1338–1352
20. Fergus R, Perona P, Zisserman A (2003) Object class recognition by unsupervised scale-invariant learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 1–8
21. Frias-Martinez E, Sanchez A, Velez J (2006) Support vector machines versus multi-layer perceptrons for efficient off-line signature recognition. Eng Appl Artif Intell 19(6):693–704
22. Hatori T, Sato-Ilic M (2014) A Fuzzy clustering method using the relative structure of the belongingness of objects to clusters. Proc Comput Sci 35:994–1002
23. Helmer S, Lowe DG (2004) Object class recognition with many local features. In: Proceedings of the international conference on computer vision and pattern recognition, 187–187
24. Huang J, Kumar SK, Mitra M, Zhu W, Zabih R (1997) Image indexing using color correlograms. In: Proceedings of the international conference on computer vision and pattern recognition, 762–768
25. Huang FJ, LeCun Y (2006) Large-scale learning with SVM and convolutional for generic object categorization. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 1:284–291
26. Ijjina EP, Mohan CK (2014) View and illumination invariant object classification based on 3D color histogram using convolutional neural networks. In: Proceedings of the Asian conference on computer vision, 316–327
27. Jurie F, Triggs B (2005) Creating efficient codebooks for visual recognition. Proc Tenth IEEE Int Conf Comput Vis (ICCV) 1:604–610
28. Khotanzad A, Lu JH (1990) Classification of invariant image representations using a neural network. IEEE Trans Acoust Speech Signal Process 38(6):1028–1038
29. Kim J, Kim BS, Savarese S (2012) Comparing image classification methods: K-nearest-neighbor and support-vector-machines
30. Kitti T, Jaruwan T, Chaiyapon T (2012) An object recognition and identification system using the harris corner detection method. Int J Mach Learn Comput 2(4):462–465
31. Kumar I, Rawat J, Bhadauria HS (2014) A conventional study of edge detection technique in digital image processing. Int J Comput Sci Mob Comput 3(4):328–334
32. Leibe B, Mikolajczyk K, Schiele B (2006) Efficient clustering and matching for object class recognition. In: Proceedings of the British machine vision conference, 789–798
33. Lepetit V, Fua P (2006) Keypoint recognition using randomized trees. IEEE Trans Pattern Anal Mach Intell 28(9):1465–1479
34. Lowe D (1999) Object recognition from local scale invariant features. In: Proceedings of the international conference on computer vision, 1–8
35. Madzarov G, Gjorgjevikj D, Chorbev I (2009) A multi-class SVM classifier utilizing binary decision tree. Informatica 33(2):233–241
36. Madzarov G, Gjorgjevikj D (2010) Evaluation of distance measures for multi-class classification in binary SVM decision tree. In: Proceedings of the international conference on artificial intelligence and soft computing (ICAISC), 437–444
37. McCann S, Lowe DG (2012) Local naive Bayes nearest neighbor for image classification. In: Proceedings of the international conference on computer vision and pattern recognition (CVPR), 3650–3656
38. Mele K, Maver J (2003) Object recognition using hierarchical SVMs. In: Proceedings of the workshop on computer vision, 109–114
39. Miller D, Nicholson L, Dayoub F, Sünderhauf N (2017) Dropout sampling for robust object detection in open-set conditions. In:

Proceedings of the IEEE international conference on robotics and automation. Arxiv preprint arXiv:1710.06677
40. Muralidharan R (2014) Object recognition using K-nearest neighbor supported by eigen value generated from the features of an image. Int J Innov Res Comput Commun Eng (IJIRCCE) 2(8):5521–5528
41. Namratha M, Prajwala TR (2012) A comprehensive overview of clustering algorithms in pattern recognition. IOSR J Comput Eng (IOSRJCE) 4(6):23–30
42. Prajapati N, Nandanwar AK, Prajapati GS (2016) Edge histogram descriptor, geometric moment and sobel edge detector combined features based object recognition and retrieval system. Int J Comput Sci Inf Technol (IJCSIT) 7(1):407–412
43. Qiu G, Sudirman S (2002) A binary color vision framework for content-based image indexing. In: Proceedings of the international conference on advances in visual information systems, 50–60
44. Ramadevi Y, Sridevi T, Poornima B, Kalyani B (2010) Segmentation and object recognition using edge detection techniques. Int J Comput Sci Inf Technol (IJCSIT) 2(6):153–161
45. Rastegari M, Ordonez V, Redmon J, Farhadi A (2016) Xnor-net: imagenet classification using binary convolutional neural networks. In: European conference on computer vision, 525–542
46. Rosten E, Drummond T (2006) Machine learning for high-speed corner detection. In: Proceedings of the European conference on computer vision, 430–443
47. Roy K, Mukherjee J (2013) Image similarity measure using color histogram, color coherence vector, and sobel method. Int J Sci Res (IJSR) 2(1):538–543
48. Saraswat M, Goswami AK, Tiwari A (2013) Object recognition using texture based analysis. Int J Comput Sci Inf Technol (IJCSIT) 4(6):775–782
49. Shivakanth AM (2014) Object recognition using SIFT. Int J Innov Sci Eng Technol (IJISET) 1(4):378–381
50. Sivic J, Russell BC, Efros AA, Zisserman A, Freeman WT (2005) Discovering objects and their location in images. In: Proceedings of the 10th international conference on computer vision, vol 1, pp 370–377
51. Stricker M, Orengo M (1995) Similarity of color images. In: Proceedings of the SPIE storage and retrieval for still image and video databases III, 381–392
52. Sun S, Chen Z, Chia T (2002) Invariant feature extraction and object shape matching using gabor filtering. Recent Advances in Visual Information Systems
53. Swain MJ, Ballard DH (1991) Color Indexing. Int J Comput Vis 7(1):11–32
54. Syahputra H, Harjoko A, Wardoyo R, Pulungan R (2014) Plant recognition using stereo leaf image using gray-level co-occurrence matrix. J Comput Sci 10(4):697–704
55. Tao J, Hung Y (2002) A Bayesian method for content-based image retrieval by use of relevance feedback. In: Proceedings of the international conference on advances in visual information systems, 76–87
56. Thureson J, Carlsson S (2004) Appearance based qualitative image description for object class recognition. In: Proceedings of the European conference on computer vision, 518–529
57. Uijlings JRR, Van de Sande KEA, Gevers T (2013) Selective search for object recognition. Int J Comput Vis (IJCA) 104:1–14
58. Wu W, Wang MJ (1999) Two-dimensional object recognition through two-stage string matching. IEEE Trans Image Process 8(7):978–981
59. Vashaee A, Jafari R, Ziou D, Rashidi MM (2016) Rotation invariant HOG for object localization in web images. Signal Process 125:304–314
60. Vondrick C, Khosla A, Pirsiavash H, Malisiewicz T, Torralba A (2016) Visualizing object detection features. Int J Comput Vis 119(2):145–158
61. Wei H, Chengzhuan Y, Yu Q (2017) Contour segment grouping for object detection. J Vis Commun Image Represent 48:292–309
62. Wu X, Bhanu B (1997) Gabor wavelet representation for 3-D object recognition. IEEE Trans Image Process 6(1):47–64
63. Yakhnenko O, Honavar V (2009) Multi-modal hierarchical Dirichlet process model for predicting image annotation and image-object label correspondence. In: Proceedings of the SIAM international conference on data mining, 281–294
64. Yu H, Li M, Zhang H, Feng J (2002) Color texture moments for content-based image retrieval. In: Proceedings of the IEEE international conference on image processing, 929–932
65. Zaboli H, Rahmati M, Mirzaei A (2008) Shape recognition by clustering and matching of skeletons. J Comput 3(5):24–33

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.