Improved object recognition results using SIFT and ORB feature detector
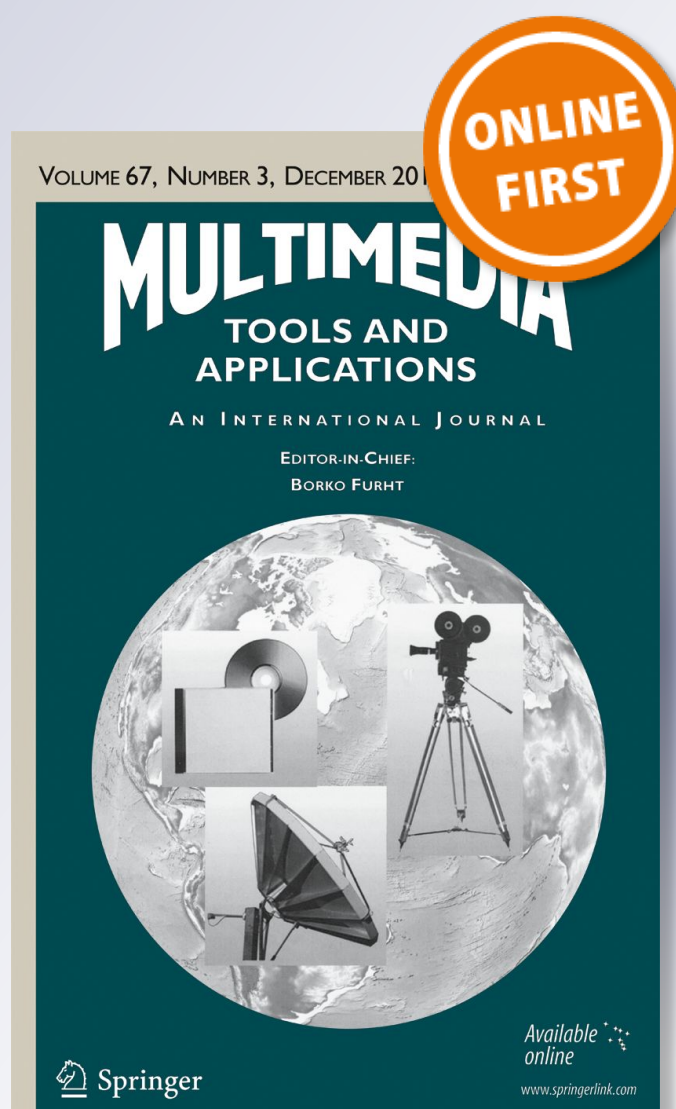
Surbhi Gupta, Munish Kumar & Anupam Garg

VOLUME 67, NUMBER 3, DECEMBER 20

MULTIMEDIA
TOOLS AND
APPLICATIONS

AN INTERNATIONAL JOURNAL

EDITOR-IN-CHIEF:
BORKO FURHT

Available online
www.springerlink.com

ONLINE FIRST

Springer

Springer

Improved object recognition results using SIFT and ORB feature detector

Surbhi Gupta[1] · Munish Kumar[2] · Anupam Garg[3]

Abstract

Object recognition has a wide domain of applications such as content-based image classification, video data mining, video surveillance and more. Object recognition accuracy has been a significant concern. Although deep learning had automated the feature extraction but hand crafted features continue to deliver consistent performance. This paper aims at efficient object recognition using hand crafted features based on Oriented Fast & Rotated BRIEF (Binary Robust Independent Elementary Features) and Scale Invariant Feature Transform features. Scale Invariant Feature Transform (SIFT) are particularly useful for analysis of images in light of different orientation and scale. Locality Preserving Projection (LPP) dimensionality reduction algorithm is explored to reduce the dimensions of obtained image feature vector. The execution of the proposed work is tested by using k-NN, decision tree and random forest classifiers. A dataset of 8000 samples of 100-class objects has been considered for experimental work. A precision rate of 69.8% and 76.9% has been achieved using ORB and SIFT feature descriptors, respectively. A combination of ORB and SIFT feature descriptors is also considered for experimental work. The integrated technique achieved an improved precision rate of 85.6% for the same.

Keywords Object Recognition · ORB · SIFT · K-Means · LPP · k-NN · Decision Tree · Random Forest

✉ Munish Kumar
   munishcse@gmail.com

1   Department of Computer Science & Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

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

3   Department of Computer Science & Engineering, Thapar Institute of Engineering & Technology, Patiala, Punjab, India

1 Introduction

Object Recognition in images is an important concern while learning visual classifications and distinguishing different occurrences of those classifications. Invariantly, all computer vision assignments largely depend on the ability to recognize objects, items, faces and scenes from the image. Visual recognition has many potential applications that involve artificial intelligence and information retrieval. Content-based image search, video data mining or object identification for mobile robots are some of the examples. Visual analysts categorize recognition in two classes as specific and generic as shown in Fig. 1. The presented work focuses on recognition of generic object categories. Humans have the great ability to generalize the objects, even if objects differ and are not exactly the same. Feature extraction in the key while doing object recognition using computer vision. First the useful features are obtained from the images and then matched with the known reference features. Classifiers are used to do the matching and classification job. The extracted features could be hand crafted or machine learned. Hand crafted features are obtained based on the subject knowledge of the algorithm designer ed. as sliding window based, contour based, graph based, fuzzy based or content based [26].

- Forest for same feature set. To compare the results with CN based feature classifiers.

This paper propose a novel strategy for generic object recognition using Scale Invariant Feature Transform features (SIFT) [21] and Fast-Rotated and BRIEF (ORB) [31]. The integrated technique extracts the best robust features from the image. The SIFT descriptor utilizes a 128 element feature dimension in one key point and ORB descriptor uses 32 elements in one key point, which requires a high memory space for storing features and high complexity. Further, K-Means clustering [15] and Locality Preserving Projection (LPP) [13] are used to reduce the complexity of the integrated technique.

This paper consists of seven sections. Section 1 has the introduction of the present work. Related work is exhibited in Section 2. Feature extraction techniques are discussed in

Wild card   **Carl Friedrich Gauss**   **Eiffel Tower**   **My neighbor's car**

Specific objects

car   car   building   building

Generic object categories

Fig. 1 Types of object recognition considered by vision analysts

Section 3. Section 4, presented features dimensionality reduction techniques. A novel framework for object recognition using ORB and SIFT feature extraction techniques has been proposed in Section 5. Section 6 presents experimental results in view of the proposed framework. Finally, Section 7 has the concluding remarks.

2 Related work

There is the vast volume of literary work on feature extraction and classification techniques; a few approaches related to the present work have been discussed in this section. Initially, Ballard [2] has described a technique for circle detection and recognition from images by using Hough transforms. Ullman [30] used basic visual operators like color, shape, texture, etc., which are integrated in different ways, for detection of complicated objects. Peterson and Gibson [25] highlighted the role of acquaintance in object recognition. Minut and Mahadevan [20] recommended a visual attention prototype which was based on reinforcement learning of an image. Belongie *et al*. [3] proposed the shape context, which conquers the spatial distribution of all other points relative to each point on the shape. This semi-local illustration allowed establishment of point-to-point communication between different features of an object even under yielding deformations. Elnagara and Alhajj [9] described an accurate partition path and further the separation of numerals and restoration was proposed. Yu and Shi [32] proposed a technique based on segmentation using partial detection instead of global shape descriptors. Lowe [19] described SIFT (Scale Invariant Feature Transform) for extracting a large collection of feature vectors which were constant to image translation, scaling and rotation and were robust across a substantial range of affine deformations, noise addition and enlightenment changes. The key location of an object was defined as maxima and minima of the result of DOG (Difference of Gaussians). Mori *et al*. [22] discussed about the connection between the shapes of an object and recognized these shapes by finding an aligning alteration for correspondence between points of the two shapes. Leordeanu *et al*. [17] presented an unusual way, by programming relations between all pairs of edges. Nadernejad *et al*. [23] reviewed and compared different edge detection techniques, i.e. the Marr-hildreth Edge Detector, the Canny Edge detector, Threshold and Boolean function based edge detection, color and vector angle based edge detection for object recognition. Ferrari *et al*. [11] presented an object recognition approach which entirely integrates the paired strength presented by shape matchers. Toshev *et al*. [29] proposed a ground segmentation technique for drawing out of image regions that can be similar to the global properties of model border construction and chordiogram. Soltanshahi *et al*. [27] used Scale Invariant Feature Transform (SIFT) for Content Based Image Retrieval (CBIR). They considered SIFT feature extraction technique and used two approaches for SIFT descriptor. In the principal approach, they clustered a SIFT descriptor into 16 clusters and in the second approach, they focused on the registration of descriptors in view of directions. Alhassan and Alfaki [1] introduced a color and texture fusion based technique for CBIR. Peizhong *et al*. [24] extracted texture and color based feature elements from images. Texture features like LBP (Local Binary Pattern) were mostly used and in shade based features like CIF (Color Information Feature) were used. Recently CNN based solution was suggested by

[18]. The machine learned features were used for object detection and classification, Comparable results were obtained for hand crafted and machine learned features.

3 Feature extraction techniques

Feature extraction is the most important phase of the proposed object recognition system. In this phase, the significant features are extracted for object category classification. In the present paper, two feature extraction techniques, namely, ORB and SIFT are considered for object recognition. These techniques are briefly discussed in following sub-sections.

3.1 ORB (Oriented FAST Rotated and BRIEF)

ORB is developed in "OpenCV" which uses FAST (Features from Accelerated Segment Test) key point detector and binary BRIEF descriptor. ORB is used to extract the fewer but the best features from an image. The cost of computation is also less when contrasted with SIFT and SURF, but the magnitude is faster than SURF. ORB first apply the FAST key point detector, which detects the large number of key points and then the ORB uses a Harris corner detector to find good features from those key points. Extracted features generate better results and are less sensitive to noise. The centroid of the image can be calculated using the patch moment in ORB, using eq. (1):

$$m_{pq} = \sum_{x,y} x^p y^q (x, y) \tag{1}$$

The orientation of the corners is calculated using the intensity centroid of image patches using eq. (2):

$$C\left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \tag{2}$$

From the center of the patch to centroid, the angle is given by eq. (3)

$$atan2\left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) = atan2(m_{01}, m_{10}) \tag{3}$$

The ORB uses *OFAST* for the quick key focuses on the direction and *RBRIEF* for the BRIEF descriptor with orientation (rotation) angle. The sample used in BRIEF descriptor is transformed using the orientation.

3.2 SIFT (Scale Invariant Feature Transform)

The SIFT detector extracts a number of attributes from an image in such a way which is reliable with changes in the lighting impacts and perspectives alongside with other imagining viewpoints. The SIFT descriptor will distinguish nearby elements of an image.

The major stages in SIFT are: -

- Detector
  Find Scale Space Extreme Detection.
  Key point localization and filtering

- Descriptor
    Orientation Assignment.
    Feature Description
- Similarity
- Feature Matching

4 Feature dimensions reduction techniques

The following techniques are used to reduce the dimensions of the feature vector, as a large number of features has been extracted using the integrated technique.

4.1 K-means clustering

In K-means clustering algorithm, the Euclidean distance method or the max-min measurement is used to calculate the distance between the centroid of the cluster and an object. A set of n descriptor array and K-means clustering algorithm is used to cluster n descriptor array into a K number of clusters and to minimize the intra cluster variance as described in the following eq. (4).

$$arg_s = min\sum_{i=1}^{k}\sum_{x_j \varepsilon s_i}\left(X_j - \mu_i\right) \qquad (4)$$

Where $k = 1$, clusters $S_i = 1, 2 ... K(64)$ and μ_i is the center of all points X_j in S_i.
    The entire process can be described as follows:

Step 1: Randomly select 64 instances from a set of n descriptor array.
Step 2: Assign every instance to its closest center point.
Step 3: Update a cluster center.
Step 4: Finally obtain 64 clusters by assigning instances according to closest center points.
Step 5: Calculate a mean of each cluster.

4.2 Locality preserving projection

LPP (Locality Preserving Projection) is utilized to reduce the higher dimensional space into lower dimensional space for information storage. LPP takes a shot at neighborhood data about the informational LPP that is keeping the desired information and discarding the undesired data. PCA (Principal Component Analysis) keeps less information about the data than LPP. LPP for dimensionality diminishing works in three stages as follows:

- First an Adjacency graph for data is generated using space $x_i \in R^d$ and nodes for undirected graph $i = 1 .. . N$.
- Weights are selected for graphs which are represented using similarity matrix $[p_{ij}]$.

$$p_{ij} = \begin{cases} e^{-\left(x_i - x_j\right)^2/t}, & if \ x_i \in kNN\left(x_j\right) or \ x_j \in kNN\left(x_i\right) \\ 0, & otherwise \end{cases}$$

- Eigenvector and Eigen value equation are computed:

$$XLX^T \sigma = \lambda XDX^T \sigma.$$

Where $X = \langle x_1 | \ldots\ldots | x_n \rangle$, D is a matrix with sum of P. Using row or column diagonal terms, $d_{ii} = \sum_j p_{ij}$, L is Laplacian matrix $L = D - P$, σ projection matrix.

5 Proposed integrated system for object recognition

The proposed system of object recognition consists of various phases, namely, image acquisition, pre-processing, feature extraction, classifier prediction and lastly object recognition. Block diagram of the proposed framework for object recognition is depicted in Fig. 2. Image acquisition is the process of acquiring an image from the environment from any source (whether using a camera or from already available database) and to convert an acquired image into digital form. The digital image then undergoes the preprocessing stage. Preprocessing is a preliminary phase of an object recognition system. The most crucial stage is feature extraction, which is used to extracts different features (may be hand crafted or machine learned) of the preprocessed image that results in some quantitative information of interest. This phase analyzes a set of features that can be used for distinctively grading the shape present in the image. Extracted features are global and local in nature and aims to detect the objects present in the image. Global features include the geometrical shapes, texture, size, color etc. after feature extraction last stage is to train the classifier using these features and then test the images for accurate classification. In the proposed work, image acquisition and preprocessing is not required as a public dataset Caltech 101 is used for experimentation. Preliminary features from the input image are extracted to recognize an object by using geometrical shape features (stage
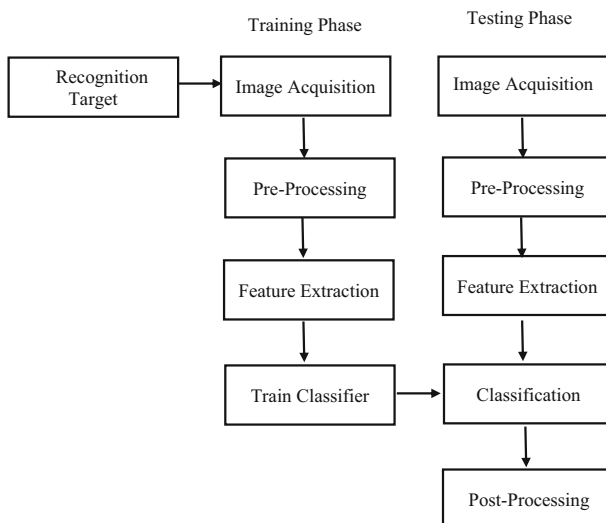
Training Phase   Testing Phase

```
┌──────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Recognition  │─────▶│ Image Acquisition│      │ Image Acquisition│
│   Target     │      └──────────────────┘      └──────────────────┘
└──────────────┘              │                          │
                              ▼                          ▼
                      ┌──────────────────┐      ┌──────────────────┐
                      │  Pre-Processing  │      │  Pre-Processing  │
                      └──────────────────┘      └──────────────────┘
                              │                          │
                              ▼                          ▼
                      ┌──────────────────┐      ┌──────────────────┐
                      │ Feature Extraction│      │ Feature Extraction│
                      └──────────────────┘      └──────────────────┘
                              │                          │
                              ▼                          ▼
                      ┌──────────────────┐      ┌──────────────────┐
                      │ Train Classifier │─────▶│  Classification  │
                      └──────────────────┘      └──────────────────┘
                                                         │
                                                         ▼
                                                ┌──────────────────┐
                                                │ Post-Processing  │
                                                └──────────────────┘
```

Fig. 2 Block diagram of Object Recognition System

3). SIFT and ORB features extraction methodologies are used for extracting image feature descriptors. Further, K-means clustering algorithm is used to generate 'K' number of clusters using the descriptor array. Lastly, LPP is used to reduce the dimensions of the feature vector.

5.1 Proposed algorithm

Input: Query Image.

Step 1: Extract a feature descriptor from images of training dataset using ORB and SIFT feature extracting methods.

Step 2: Use K-means clustering algorithm to generate 64 clusters, for every descriptor array. Then, compute the mean of every cluster to obtain a 64 dimensional feature vector for every descriptor.

Step 3: Use LPP dimensionality reduction algorithm to reduce the feature vectors of 64 units into 8, 16 and 32 components.

Step 4: Integrate the both feature vectors i.e. ORB and SIFT and store the combined feature vector in database.

Step 5: Train the proposed system using a combination of ORB and SIFT feature vectors to obtain the classifier model.

Step 6: Test the trained classifier model by inputting the query object image and extract the ORB and SIFT features of a questioned image.

Step 7: Predict the similarity between query object data features and trained dataset using the model classifier.

6 Dataset and experimental results

This section, presents the experimental results obtained using the proposed object recognition system. A dataset namely Caltech 101 consisting of 9197 samples, has been considered for experimental work [10]. We have taken 8000 samples (100-classes) from this entire dataset. This dataset consists of 100 generic object categories. A few samples of this dataset are shown in Fig. 3. Entire dataset is divided into two categories, namely, training dataset and testing dataset. In training dataset 70% data of the entire database has been used and remaining 30% data is considered as testing dataset.

All object images of the dataset are tested using ORB and SIFT features. Each image produces a different array length due to different sizes of images, so K-Means (K = 64) clustering algorithm is applied to build the uniformity in the dimensions of the feature set. Further, the LPP dimensionality reduction method is used to reduce the dimensions into 8, 16 and 32 dimensional feature vectors. The ORB and SIFT features are extracted independently and finally, the integration of ORB and SIFT features is used for the improvement of the performance of the proposed system. In the classification phase, three different classifiers are considered, namely, k-NN, Decision Tree, and Random Forest. All these classifier algorithms have different advantages and disadvantages. High bias and low variance classifiers e.g., Naive Bayes (Cestnik et al. [6]) are more suitable for smaller training set. But low bias/high variance classifiers e.g. KNN [7] are beneficial when the training set is large. Decision trees [28] are easy to interpret and explain. They can handle feature interactions which are non-parametric. Random forests [5] are most popularly used for problems in classification as they are fast, scalable and no tuning of parameters is required as in SVMs.

Object_Category	Object_Image	Object_Category	Object_Image
Airplane		Cup	
Anchor		Dolphin	
Barrel		Elephant	
Brain		Grand_Piano	
Camera		Helicopter	
Car		Motor Cycle	

Fig. 3 A few samples of dataset

The performance of the proposed system is measured based on four parameters: TPR (True Positive Rate), FPR (False Positive Rate), Precision Rate and AUC (Area under Curve). TPR states the rate at which the positive instances are correctly classified. High recall indicates the class is correctly recognized. FPR states the rate at which the negative class labels, which has been classified correctly. Precision is the ratio of correctly classified positive instances to that of the total number of instances that have been classified as positive by the classifier [8]. All the experimental results are tabulated in Table 1.

Table 1 Experimental Results Based on SIFT and ORB on Caltech 101

Experimental Results Based on SIFT and ORB on Caltech 101

Evaluation Measures	k-NN	Decision Tree (DT)	Random Forest (RF)	k-NN	Decision Tree (DT)	Random Forest (RF)	k-NN	Decision Tree (DT)	Random Forest (RF)
SIFT features		8 features			16 features			32 features	
TPR	58.40%	60.90%	76.60%	59.60%	65.00%	74.80%	67.30%	69.30%	67.30%
FPR	1.60%	1.10%	0.90%	1.70%	1.10%	1.00%	1.20%	1.10%	1.20%
Precision Rate	59.60%	61.30%	76.90%	60.50%	64.30%	75.00%	67.00%	69.40%	67.00%
ROC	85.20%	90.00%	98.90%	79.00%	89.40%	98.80%	98.00%	88.30%	98.00%
ORB features		8 features			16 features			32 features	
TPR	53.40%	58.60%	69.60%	56.20%	63.10%	68.60%	43.80%	65.20%	62.00%
FPR	1.80%	1.40%	1.40%	2.00%	1.40%	1.40%	2.40%	1.40%	1.50%
Precision Rate	54.30%	55.70%	69.80%	57.70%	61.00%	69.10%	46.20%	63.20%	62.00%
ROC	82.30%	91.30%	98.50%	81.80%	91.00%	98.40%	74.10%	89.60%	97.70%
SIFT+ ORB features		8 + 8 features			16 + 16 features			32 + 32 features	
TPR	59.80%	78.80%	85.40%	53.60%	81.10%	80.20%	37.30%	80.50%	69.80%
FPR	1.70%	0.80%	0.80%	2.00%	0.90%	0.90%	2.90%	1.00%	1.10%
Precision Rate	61.10%	79.20%	85.60%	53.80%	81.30%	80.40%	41.40%	80.80%	68.60%
ROC	79.10%	94.80%	99.30%	75.40%	94.20%	99.10%	67.20%	92.20%	98.50%

The precision comparison for all the classifier are illustrated in Fig. 4. The best accuracy is obtained for SIFT+ORB (8 + 8) feature set with random forest, i.e. 85.42% with a low dimensionality of 16 features.

A precision rate of 76.9% is obtained using SIFT feature descriptors and Random forest classifier as shown in Fig. 5. A precision rate of 69.8% has been achieved using ORB feature descriptors and Random forest classifier, as shown in Fig. 6. A combination of these two feature descriptors, namely, SIFT and ORB has been also tested.

Using this combination, the best precision rate of 85.6% has been achieved using 16 feature descriptors (8 feature descriptors of SIFT and 8 feature descriptors of ORB) as depicted in Fig.
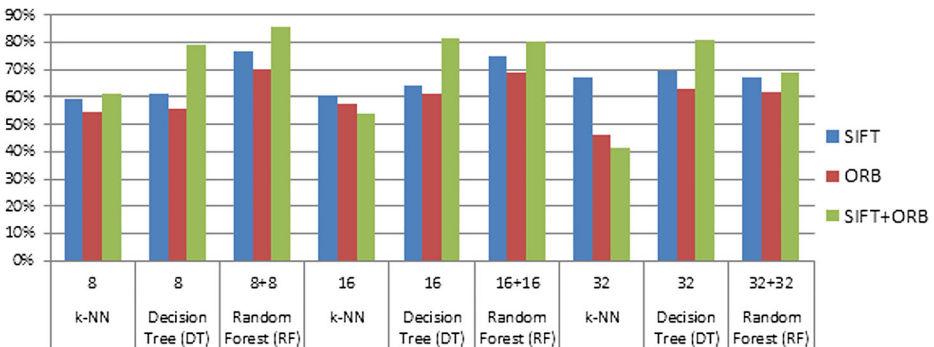

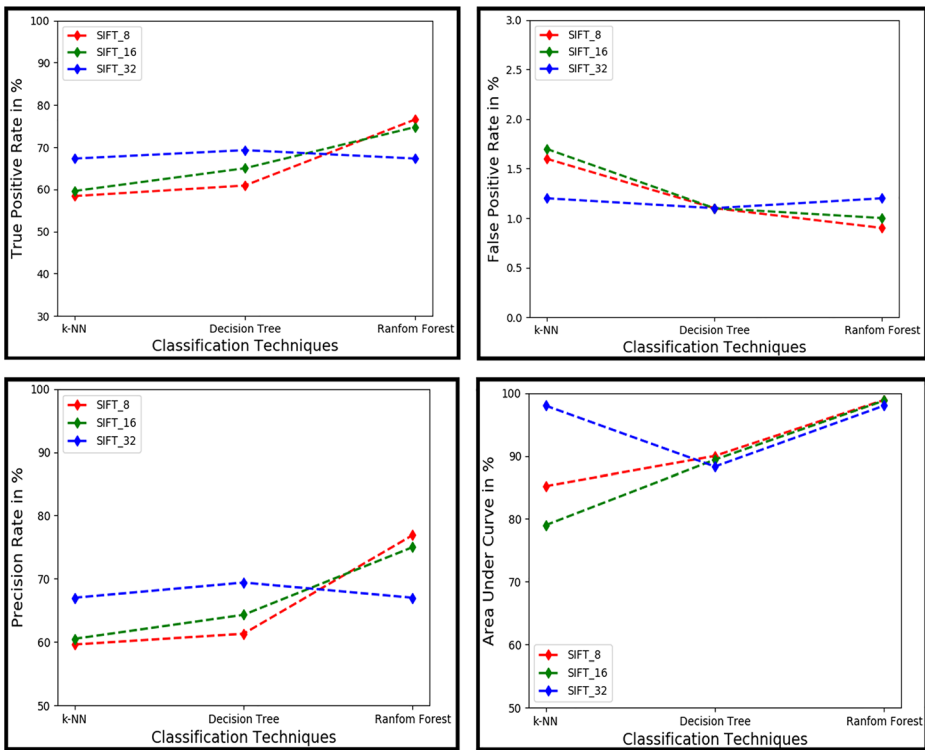**Fig. 4** Precision comparison of proposed features

Fig. 5 Experimental Results Based on SIFT Features

7. Figure 7 shows a decrease of AUC compared with Fig. 6 or Fig. 5, because the AUC is depends on the quality of feature and number of components in feature vector for recognition.

6.1 Comparison with other techniques based on hand crafted features

Table 2. Shows the results for multi-class classification on Cal-Tech101 dataset. Our approach outperforms the existing state-of-the-art techniques for the same setup. Different algorithms utilizing shape matching, texture descriptors, visual similarity and pyramid matching are compared with the proposed technique. The best results on Caltech101 dataset (75.67%) are currently reported by Huang *et al*. [14] based on visual similarity. The experimental setup of Grauman and Darrell [12] and Zhang *et al*. [33] is imitated i.e. training is done on 30 images per class and the number of test images is limited to 50 per class. These results demonstrated that the SIFT8 features have outperformed all the existing state of art hand crafted algorithms. ORB8 performed better than all but visual descriptors. The results for SIFT8 and ORB8 are much encouraging as they further improved the classification accuracy remarkably.

6.2 Comparison with other techniques based on CNN features

Now a days the hand crafted features are getting replaced with CNN based machine learned features. These features are extracted automatically in contrast to features extracted using human knowledge. Although these proved to be robust but still their consistency and
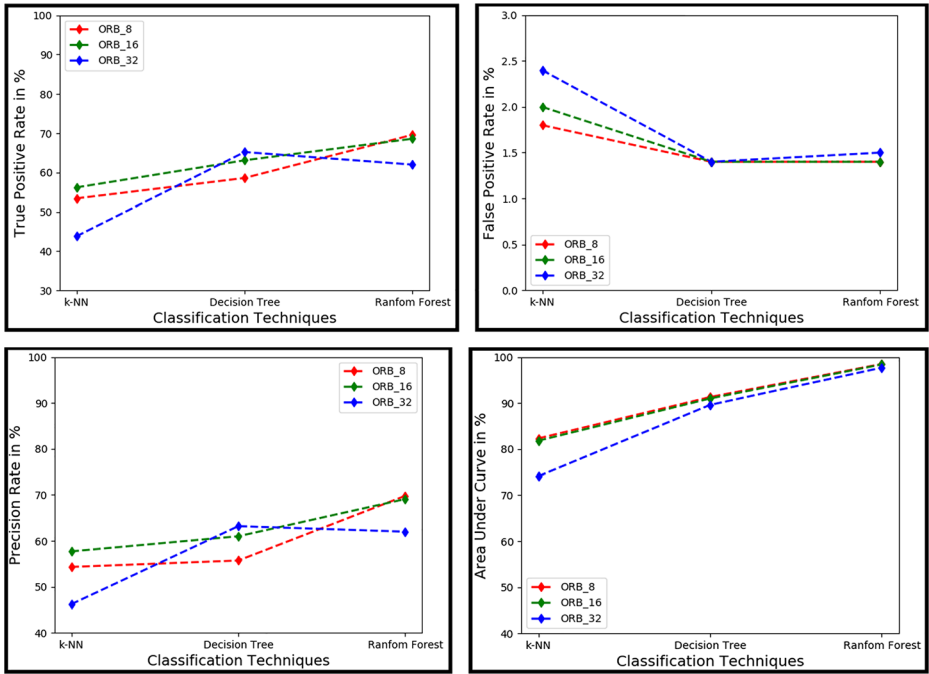
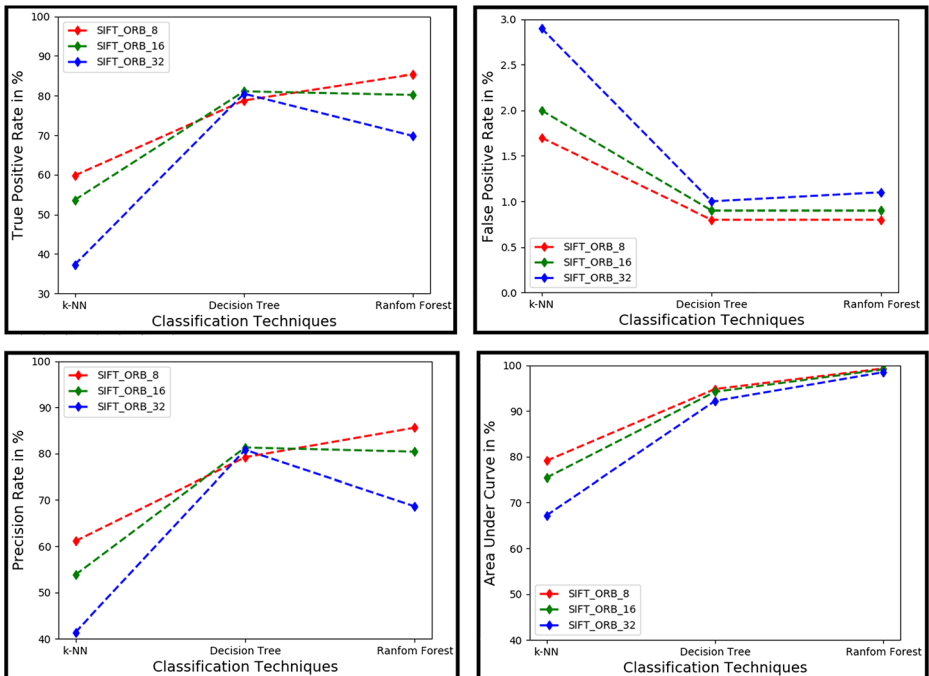**Fig. 6** Experimental Results Based on ORB Features

Fig. 7 Experimental Results Based on SIFT+ORB Features

Table 2 Comparison with existing hand crafted feature based state-of-the-art work

Technique	Proposed SIFT8	Proposed ORB8	Proposed SIFT8 + ORB8 features	Local visual similarity descriptor by Huang et al. [14]	Texture descriptors by Zhang et al. [33]	Spatial pyramid matching Lazebnik et al. [16]	Shape matching by Berg et al. [4]	Pyramid match kernel by Grauman and Darrell [12]
Accuracy	76.94%	69.83%	85.42%	75.67%	53.9%	64.6%	48.5%	43.0%

Table 4 Comparison with existing CNN feature based state-of-the-art work

Technique	Proposed features + DT	Proposed features + RF	AlexNet +DT	Vgg19 + DT	ResNet+DT
Accuracy	81.13%	85.42%	85.20%	75.80%	75.80%

effectiveness needs to be tested thoroughly. Liu et al. has experimented CNN algorithm like ResNet, Vgg19 and AlexNet with different classifiers and dataset to prove their effectiveness for object recognition. We have compared the results of proposed features with those CNN based features using decision Tree classifier. The Accuracy obtained for decision tree classifier is tabulated in Table 3.

The best accuracy obtained with decision tree classifier is using SIFT16 + ORB16 features i.e. 81.13. If we compare this accuracy with CNN based classifier results, these are better than ResNet and Vgg19 results as illustrated in Table 4.

Only AlexNet achieved 85.2% accuracy. But if use random forest instead the accuracy achieved id comparable with AlexNet i.e. 85.42%. Hence, the efficiency of hand crafted features is well proved. Further, it would be interesting to experiment that whether the integration of hand crafted and CNN based features could improve the results further. But, it is out of the scope of present work and can be experimented in future.

7 Conclusion

Computer vision and image processing is a domain where a lot of advancement is possible for feature extraction and classification. One of them is to improve the recognition accuracy. In this paper, a novel integrated technique using ORB and SIFT features are proposed for generic object recognition. Further, feature dimension is reduced by K-means clustering and LPP approach. It is demonstrated that the proposed technique performs better than already existing techniques for object recognition. Each feature component is firstly evaluated individually as it is of the interest to assess. Then, their integration is analyzed. This evaluation is performed and analyzed thorough experimental work. The performance of proposed methodology is assessed by considering various parameters such as True Positive Rate (TPR), False Positive Rate (FPR), Precision Rate, and Area Under Curve (AUC). Authors have achieved an efficient precision rate and accuracy of 85.6% and 85.42%, respectively, for 8000 samples of 100-class samples of generic objects. The presented work is compared with many state of art algorithms. Encouraging findings are obtained for the proposed classifier when compared with CNN feature based classifiers.

Table 3 Accuracy obtained for decision tree classifier

Features	Accuracy obtained for Decision Tree Classifier			
	8	16	32	Best
SIFT	60.87%	64.96%	69.29%	69.29%
ORB	58.60%	63.07%	65.16%	65.16%
SIFT+ORB	78.78%	81.13%	80.47%	81.13%

References

1. Alhassan AK, Alfaki AA (2017) Color and Texture Fusion-Based Method for Content-Based Image Retrieval. *Proceedings of the International Conference on Communication, Control, Computing and Electronics Engineering*, 1–6
2. Ballard DH (1981) Generalizing the Hough transform to detect arbitrary shapes. Pattern Recogn 13(2):111–122
3. 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
4. Berg AC, Berg TL, Malik J (2005) Shape matching and object recognition using low distortion correspondences. *In Proceedings of Computer Vision and Pattern Recognition,* 1:26–33
5. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
6. Cestnik B, Kononenko I, Bratko I (1987) Assistant 86: A knowledge elicitation tool for sophisticated users. *Proceedings of the 2nd European Working Session on Learning, Bled, Yugoslavia*, 31–45
7. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27
8. Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine learning, Pennsylvania, USA*, 233–240
9. Elnagara A, Alhajj R (2003) Segmentation of connected handwritten numeral strings. Pattern Recogn 36(3):625–634
10. Fei-Fei L, Fergus R, Perona P (2004) Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *In proceedings of the Workshop on Generative-Model Based Vision*. Washington, DC, June 2004
11. Ferrari V, Jurie F, Schmid C (2010) From Images to Shape Models for Object Detection. Int J Comput Vis 87(3):284–303
12. Grauman K, Darrell T (2005) The pyramid match kernel: Discriminative classification with sets of image features. *Proceedings of the 10th IEEE International Conference on Computer Vision,* 2:1458–1465
13. He X, Niyogi P (2004) Locality preserving projections. *Advances in Neural Information Processing systems*, 153–160
14. Huang X, Xu Y, Yang L (2017) Local visual similarity descriptor for describing local region. *Proceedings of theNinth International Conference on Machine Vision* (ICMV 2016), 10341: 103410S
15. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY (2002) An efficient k-means clustering algorithm: Analysis and implementation. IEEE Trans Pattern Anal Mach Intell 24(7):881–892
16. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:2169–2178
17. Leordeanu M, Hebert M, Sukthankar R (2007) Beyond local appearance: Category recognition from pairwise interactions of simple features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–8
18. Liu X, Zhang R, Meng Z, Hong R, Liu G (2019) On fusing the latent deep CNN feature for image classification. World Wide Web 22(2):423–436
19. Lowe DG (2004) Distinctive image features from scale invariant key points. Int J Comput Vis 60:91–110
20. Minut S, Mahadevan S (2001) A reinforcement learning model of selective visual attention. *Proceedings of the fifth international conference on Autonomous Agents*, 457–464
21. Montazer GA, Giveki D (2015) Content based image retrieval system using clustered scale invariant feature transforms. Optik-International Journal for Light and Electron Optics 126(18):1695–1699
22. Mori G, Belongie S, Malik J (2005) Efficient shape matching using shape contexts. IEEE Trans Pattern Anal Mach Intell 27(11):1832–1837
23. Nadernejad E, Sharifzadeh S, Hassanpour H (2008) Edge Detection Techniques: Evaluation and Comparisons. Appl Math Sci 2(31):1507–1520
24. Peizhong L, Guo J, Chamnongthai K, Prasetyo H (2017) Fusion of Color Histogram and LBP-based Features for Texture Image Retrieval and Classification. Inf Sci 390:95–111
25. Peterson MA, Gibson BS (1994) Must Fig.-Ground Organization Precede Object Recognition? An Assumption in Peril. Psychol Sci 5(5):253–259
26. Sharma KU, Thakur NV (2017) A review and an approach for object detection in images. International Journal of Computational Vision and Robotics 7(1/2):196–237
27. Soltanshahi MA, Montazer GA, Giveki D (2015) Content Based Image Retrieval System Using Clustered Scale Invariant Feature Transforms. Optik - International Journal for Light and Electron Optics 126(18):1695–1699
28. Swain PH, Hauska H (1977) The decision tree classifier: Design and potential. IEEE Trans Geosci Electron 15(3):142–147

29. Toshev A, Taskar B, Daniilidis K (2010) Object detection via boundary structure segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 950–957

30. Ullman S (1984) Visual Routines. Cognition 18:97–157

31. Vinay A, Kumar CA, Shenoy GR, Murthy KB, Natarajan S (2015) ORB-PCA based feature extractiontechnique for face recognition. Proc Comput Sci 58:614–621

32. Yu SU, Shi J (2003) Object-specific Fig.-ground segregation. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR), 2:39–45

33. Zhang J, Marszałek M, Lazebnik S, Schmid C (2007) Local features and kernels for classification of texture and object categories: A comprehensive study. Int J Comput Vis 73(2):213–238