Deep Image Sensing and Retrieval Using Suppression, Scale Spacing and Division, Interpolation and Spatial Color Coordinates With Bag of Words for Large and Complex Datasets

KHAWAJA TEHSEEN AHMED [1], **HUMAIRA AFZAL**[1], **MUHAMMAD RAFIQ MUFTI** [2], **ARIF MEHMOOD** [3], **AND GYU SANG CHOI** [4]

[1]Department of Computer Science, Bahauddin Zakariya University, Multan 60800, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Vehari 61100, Pakistan
[3]Department of CS and IT, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[4]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea

Correspondence author: Gyu Sang Choi (castchoi@ynu.ac.kr)

ABSTRACT Intelligent and efficient image retrieval from versatile image datasets is an inevitable requirement of the current era. Primitive image signatures are vital to reflect the visual attributes for content based image retrieval (CBIR). Algorithmically descriptive and well identified visual contents form the image signatures to correctly index and retrieve similar results. Hence feature vectors should contain ample image information with color, shape, objects, spatial information perspectives to distinguish image category as a qualifying candidate. This contribution presents a novel features detector by locating the interest points by applying non-maximum suppression to productive sum of derivative of pixels computed from differential of corner scores. The interest points are described by applying scale space interpolation to scale space division produced from Hessian blob detector resulted after Gaussian smoothing. The computed shape and object information is fused with color features extracted from the spatially arranged L2 normalized coefficients. High variance coefficients are selected for object based feature vectors to reduce the massive data which in fuse form transformed to bag-of-words (BoW) for efficient retrieval and ranking. To check the competitiveness of the presented approach it is experimented on nine well-known image datasets Caltech-101, ImageNet, Corel-10000, 17-Flowers, Columbia object image library (COIL), Corel-1000, Caltech-256, tropical fruits and Amsterdam library of textures (ALOT) belong to shape, color, texture, and spatial & complex objects categories. Extensive experimentation is conducted for seven benchmark descriptors including maximally stable extremal region (MSER), speeded up robust features (SURF), difference of Gaussian (DoG), red green blue local binary pattern (RGBLBP), histogram of oriented gradients (HOG), scale invariant feature transform (SIFT), and local binary pattern (LBP). Remarkable outcomes reported that the presented technique has significant precision rates, recall rates, average retrieval precision & recall, mean average precision & recall rates for many image semantic groups of the challenging datasets. Results comparison is presented with research techniques and reported improved results.

INDEX TERMS Image retrieval, scale spacing, image descriptor, content based image retrieval, bag-of-words, spatial coordinates mapping, features fusion.

I. INTRODUCTION

Digital media is increasing and demanding now a days due to its applications in many parts of life [1]. The advancements in digital image processing are required for the efficient image searching and indexing in the large databases. Generally images are extracted in three different ways namely: content based retrieval, semantic based retrieval and text tagged oriented retrieval. The increasing demand of the digital images need very specific data representation and

The associate editor coordinating the review of this manuscript and approving it for publication was Sunil Karamchandani [ID] .

retrieval; for this reason image indexing and retrieval has promoted as an active research area. For this reason, image fetching and retrieval has are direct and effective role of image searching from huge databases. Content-Based Image Retrieval (CBIR) has been an important procedure to detect the matched image primitive features based on the visual properties [2]. CBIR system extract features to represent an image. The feature extraction process is also known as image preprocessing. The visual features are classified into two categories: global features, also known as overall characteristics and local features, known as visual property of an image [3].

Mostly CBIR system uses global and local features including shapes, edges, spatial coordinates and texture information, color channels while other uses local features such as region, segmented features and interest points to extract similar images. Texture features represent neighborhood relationships as a combination of pixels and categorized into spatial texture and spectral texture. Shape features are also categorized into two types [4] including regioned and contoured where region-based method mostly applied with color features [4] and extract shape keypoints from the whole area of interest. Contour-based methods are sensitized to noise [5] which extract shape based anchors from the corners and edges of the image. Moreover, color histograms representations are rotation and scale invariant. Spatial distribution cannot be represented by color channels only. The major problem with the global features is that they are unable to reduce the semantic gap. Global features cannot represent all the characteristics of an image. For the reason global features are not applicable for the partial matching of images from a retrieval system. Moreover local features reduce the semantic gap. To overcome the drawback of global feature extraction interest point detectors are used that represent the local features of an image. Interest point based algorithms are Hessian [6], Harris [7], affine invariant [8] and scale invariant [6]. For object recognition global and local features of the image are combined to contribute the maximum image contents [9]. The proposed method also uses interest point detector and global feature descriptor.

The contribution presents corner detector to locate interest points by taking derivative of every pixel. Feature extractor algorithm is used as a global feature descriptor which uses Gaussian smoothing. Color image is converted into grey level and L2 normalization is applied on RGB channels. Principal component analysis is performed on redundant features. Bag of visual words architecture is engaged to retrieve relevant images from the visual BoW repository after indexing. The remaining article is consolidated as follow: Section 2 shows the related work on robust corner and feature detector and descriptor. In Section 3 presented methodology is explained. The experimental results with graphical representation are presented in Section 4 and conclusion of the findings are discussed in Section 5.

II. RELATED WORK

The remote sensing research is aimed to deeply learn the potential for the primitive feature synthesis for high resolution images. Several techniques have been proposed to implement the image contents matching. Similar to the work presented in the proposed paper, local features for detection and classification are investigated in different ways. A Harris with Laplace based cornered combined support vector machine based feedback method is presented by [10]. Harris-Laplace corner detector is used to extract image corner at first and then density ratio is used to obtain salient region for all distinguished parts of the image. Furthermore for the initial retrieval shape features along with color information are merged to detect the salient regions. Lastly, Support Vector Machine (SVM) classification is used to compute the relevance feedback for CBIR. Harris corner detector is used with Bi directional Decomposition technique for CBIR demonstration [11]. Harris corner detector detect corners and BEMD technique extract edge information. Extracted features from these two techniques are merged for required retrieve from database. Experiments are performed on COIL-100 database. Fisher vectors are introduced for image classification [12]. In which super-pixels approach with edges and Zernike filters repository are inducted for efficient image retrieval and used classification benchmarks including Harris, Hessian and difference of Gaussian (DoG) detectors. Results of this research shows that the condensed descriptor is remarkable for blob and super-pixel extraction if the patches are considered along the edges. In another approach [13], a visual features attentive technique is proposed by applying salient points detection for CBIR. Corel-10K and GHIM-10K databases are used to test the superiority of the presented algorithms which shows improved performance than the Bag-of-Words and descriptor for micro levels. Local invariant features are evaluated in [14] for geographic image retrieval and reported on the effects of tuned parameters on BoW structure and also performed comparisons on specific typed standardized data for primitive features.

In past, recognition based work is presented using speeded up robust features (SURF) descriptor. SURF works as an interest point descriptor and detector for images as proposed in [15], [16]. SVM and NN (Neural Network) are used in [15] for classification. SURF detector applies to extract required images and matching feature points from the image. Results show better accuracy as compared to existing methods. SURF for feature extraction and Multiple Instance Learning Support Vector Machine proposed in [16] for image classification. In the presented approach image is segmented by quad-tree method and with codebook of Lindae-Buzo-Gray (LBG) technique. Similarity measurement is performed using Histogram Intersection (HI). An attempt of visual words usage of SURF and SIFT is presented by [17]. SURF and SIFT visual words integration adds the robustness to change in rotation, scale and illumination for image extraction. In this method statistical comparisons

are made on image benchmarks including Corel-2000, Corel-1000, Torralba, Corel-1500 are used validate the efficiency of the presented method. A new image probing scheme is presented in [18] to extract image features which use the fusion of Advanced SURF with dominant color description. Evaluation of the proposed approach uses simulation as F-score and average precision. Proposed method results in stability and high accuracy. Invariants image moments are uses in [19] as which are affine and describes the localities for image regions. Proposed method is evaluated using three different setups. The retrieval results are evaluated using UCID and UKBench datasets which gives promising results compared with other extensively used local descriptors. Medical Image Retrieval system which uses SURF features for the medical databases is described by [20]. SURF algorithm is applied as a detector and descriptor to extract referenced the images and the corresponding image feature points. Experiments are performed on medical images using SURF features and produces improved results. An operative deep learning framework is presented in [21] to produce hash codes in binary for efficient the image extraction. This method learns point wised hash codes and image representations. Experiments are performed on CIFAR-10 and MNIST datasets which result high accuracy. Furthermore 1 million clothing images are used to demonstrate scalability and efficacy. Robust Visual Descriptor with Whitening (RVD-W) is proposed in [22] in which local descriptors are used to assign the ranks to clusters. Furthermore a new normalization method also proposed to improves reparability between the matched global descriptors and unmatched values. Moreover the accumulation framework is established using SIFT signatures to perform with Convolutional Neural Network (CNN) features.

Another CBIR scheme was introduced in [23] which uses Multi-scale Geometric Analysis (MGA) of Contourlet Transformation to retrieve images. Relevance Feedback (RF) mechanism was used to improve retrieval performance. Experiments were performed on three datasets and results were tested with state-of-the- art methods when images were corrupted by noise. A new method to retrieve relevant images in three stages is presented by [24]. Color feature similarity measure was used to retrieve a fixed number of images. Then texture and shape features were used to find relevancy of images. Additionally, global and region features were joined to obtain accurate retrievals. Experiments were performed on COREL and CIFAR datasets. Varying illumination is proposed in [25] to compute feature descriptor, which uses HSI to compute its channel intensity and used red green blue channel to eliminate the variations in image intensity. To show the robustness and the uniformity of illuminated changes, the experimental results were tested with benchmarks feature descriptors. A rapid model for CBIR, composed of four phases is presented by [26]. The phases include features abstraction, dimensionality reduction, ANN classifier and matching strategy. Experiments demonstrate improved performance with less computation time. CBIR algorithm based on improved Histograms of Oriented Gradients (HOG) was

introduced by [27]. The method uses a sliding HOG window to adjust the HOG structure and principle component analysis (PCA) technique to reduce feature dimension. The experimental results shows better performance. An accurate and rapid model is presented in [28] for CBIR process, depends on a new matching method. The new model combine four major phases called feature extraction, ANN classifier, dimensionality reduction and matching strategy. The method presents considerable results with less computational time. A novel framework is presented in [29] to retrieve color images by using low level features. To capture texture, color and shape from an image, Angular Radial Transform (ART) and Color Difference Histogram (CDH) are exploited. The presented framework combines the experimental results of the standard descriptors by using some post-classification techniques such as Min–max, Borda Count Method, and Z-score normalization. For scene categorization a new mechanism called mCENTRIST was proposed by [30]. Moreover, Sobel information was embedded into the opponent color space to improve results. Experiments were performed on RGB-near infrared databases which include aerial ortho-imagery, indoor, and outdoor scene category recognition tasks. In grid computing environment [31] presents a multiple support vector machine (SVM) based architecture for CBIR. To extract features in-depth texture analysis was used and for the image representation Gabor filters, wavelet packets and curvelet were used. The proposed work is compared with the research methods to endorse the efficiency of the presented approach.

The method presented in this paper concentrates on: 1) finding effective interest points using the presented technique and describe these anchors to precisely produce the reflective signatures; 2) applying L2 normalization on RGB channel and results of L2 norm are used for the spatial arrangement; and 3) presenting a new technique to index the images by detecting their interest points and primitive and global features effectively. In the proposed method, Principal Component Analysis is applied on redundant features. Bag of features is used to provide significant results.

III. METHODOLOGY

The first step in every CBIR system is to convert a query image to grey level. The proposed method converted the color image into grey level because in greyscale image each pixel contains the intensity information. These images are also called black-and-white or monochrome images which consist of grey shades, varying from black to white ranging from 0 to 255 values. RGB coefficients are converted to grey level by discarding the hue saturation and by maintaining the luminance values.

Harris and Stephens introduced a corner detector in 1988 [22]. Harris detector was proposed by Schmid and Mohr for interest point detection. The method developed is to capture the image regions with texture and other salient image attributes. Harris detector employs the idea of applying the auto-correlation pixels detection. Harris detector localizes

corners, by gradient values that change different directions for a potential corner. The second moment matrix is scale reformed version named as Harris matrix [7]. The equation-1 shows the distribution of gradient values in the neighborhoods [7]:

$$M = \delta_{\mathbb{D}}^2 . k\, (\delta_I) * \begin{bmatrix} \mathbb{D}_q^2\, (q\delta_{\mathbb{D}}) & \mathbb{D}_q\mathbb{D}_r\, (q, \delta_{\mathbb{D}}) \\ \mathbb{D}_q\mathbb{D}_r\, (q, \delta_{\mathbb{D}}) & \mathbb{D}_r^2\, (q, \delta_{\mathbb{D}}) \end{bmatrix} \quad (1)$$

where the integration scale, derivative, differentiation scale is denoted by δ_I, $\mathbb{D}\_z$ and $\delta\_\mathbb{D}$ respectively. q and r shows the direction for the derivate. The derivatives with Gaussian kernels are calculated by using $\delta\_\mathbb{D}$. To smooth and average the point neighborhoods Gaussian window is used. The gradient signal changes are depicted by eigenvalues which is oriented in orthogonal directions. The different levels of eigenvalues represent an edge and equal eigenvalues point to a corner. Consequently the value intensities creates different edges and corners which are later described as interest points [32]. Each detected interest is surrounded by a square block of twenty pixels.

The proposed method utilizes the corner detector to spot points of interest in the images. The proposed method uses the strength of Harris and Stephens method for potential interest point detection. The significance of this method is to perform the texture analysis along with the detection of salient points to segment the image regions for potential shape formation. It is adopted because of its sub-optimization and computational efficiency as contrast to correlation approach. Another advantage of choosing this algorithm is the simplicity of operations using directional corner scores instead of applying the iterative expensive shifted patches. The differential of the corner scores is computed for directions in the introduced method. The advantage of using this mechanism is to find the repeating patterns and disturbance in the series of pixels to ultimately conclude the similarities and dissimilarities. In a simple step it is applied to 2D images and results promptly without losing the generality. It computed x and y derivatives and calculates the derivatives' products for each pixel. Then it calculates the sums of the derivatives product and adds matrix definition for pixels. Furthermore, it calculates the detector output for resulting pixels and initial values of response that is NMS - non-maximal suppression. NMS is applied due to its novelty to eliminate the cascading proposals which otherwise creates ambiguity as successive candidate regions. It overcomes the problem of neighborhood windows which generates hundreds of bounding boxes. Another advantage of NMS is its controls over the recall rates by fixing the repeating proposals.

In order to detect interest point Hessian matrix approximation is used. The reason to choose blob technique is its novelty to find the image regions with constant properties. This similarity leads to the formation of objects. It also performs the comparison of properties like brightness and colors with surrounding regions. Another effectiveness of Hessian blob detection is automatic scale selection for saddles reaction. The proposed method uses integral images which lessens the computation time proposed by Viola and Jones [33]. Simard *et al.* [34] proposed the same type of images to be adjusted in boxlets. Integral images are used for prompt calculation of square-size convolution filters. In the input image I an integral image $I\_\sum$ (k) represents the aggregate of all the pixels at a point k = (x,y)^T within a rectangular region as shown in eq. 2. The computed integral image takes three sums to compute the addition of the intensities [34].

$$I_{\sum} (k) = \sum_{j=0}^{j \leq a} \sum_{k=0}^{k \leq b} I(a, b) \quad (2)$$

It uses the Hessian matrix [34] to obtain better accuracy. Blob is embedded to this scheme to find the higher value of determinant. It results in optimal scale selection in image transformations which is better than the Laplacian operator. For some point k = (a,b) the Hessian matrix H(k, λ) in k for λ scaling is defined as below in equation 3 [34].

$$H(k, \lambda) = \begin{bmatrix} C_{xx}(k, \lambda) & C(k, \lambda) \\ C_{xy}(k, \lambda) & C_{yy}(k, \lambda) \end{bmatrix} \quad (3)$$

where, $C\_xx\,(k,\lambda)$, $C\_xy\,(k,\lambda)$ and $C\_yy\,(k,\lambda)$ are the Gaussian based convolution with second derivative $\partial^2/[\partial p]^2\, g(\lambda)$. These derivatives are also called Laplacian of Gaussians. For scale-space analysis Gaussians are the optimal but in reality they are discretized and cropped [35], [36]. Moreover, Hessian matrix approximation is also used with box filters to evaluate the computational cost and to approximate the second order Gaussian derivatives. The nine square box filters are used for Laplacian of Gaussians and present the lowest scale for calculating the blob response maps.

Interest points are required at different scales where scale spacing is treated as pyramids. To repetitively smoothen the images Gaussian is employed and furthermore sub-sampled to attain corners of higher level. Gaussian smoothing is applied at this step to perform image enhancement. Another advantage of Gaussian smoothing is its scaling at different levels to obtain the maximum image information. The decency of Gaussian that it is two pass process and results in convolving with single pass in fewer calculations. It reduces the computational cost by selecting fewer samples by small kernel size so the resultant feature vectors are small and efficient. Hereafter, the box filters is applied on the image. Box filtering is applied due to its uniqueness of linear filtering over spatial domain where the resultant pixel values are generated by averaging its neighbors and produce the sharp edged information. It's another strength is the convolutional patterns which are time and computation efficient. Box filtering is also adopted due to its equal weight attributes which produce simple accumulation significantly faster than global sliding window fashion algorithm. Moreover, its bonding with Gaussian smoothing is much compatible and faster. Thus the scale spacing is inspected by increasing the size of filter instead of shrinking the image size in steps. The nine squared filter provide the output that is measured approximating Gaussian derivatives with scale λ. The nine squared filter calculates the blob values for the lowest level. The image is filtered with
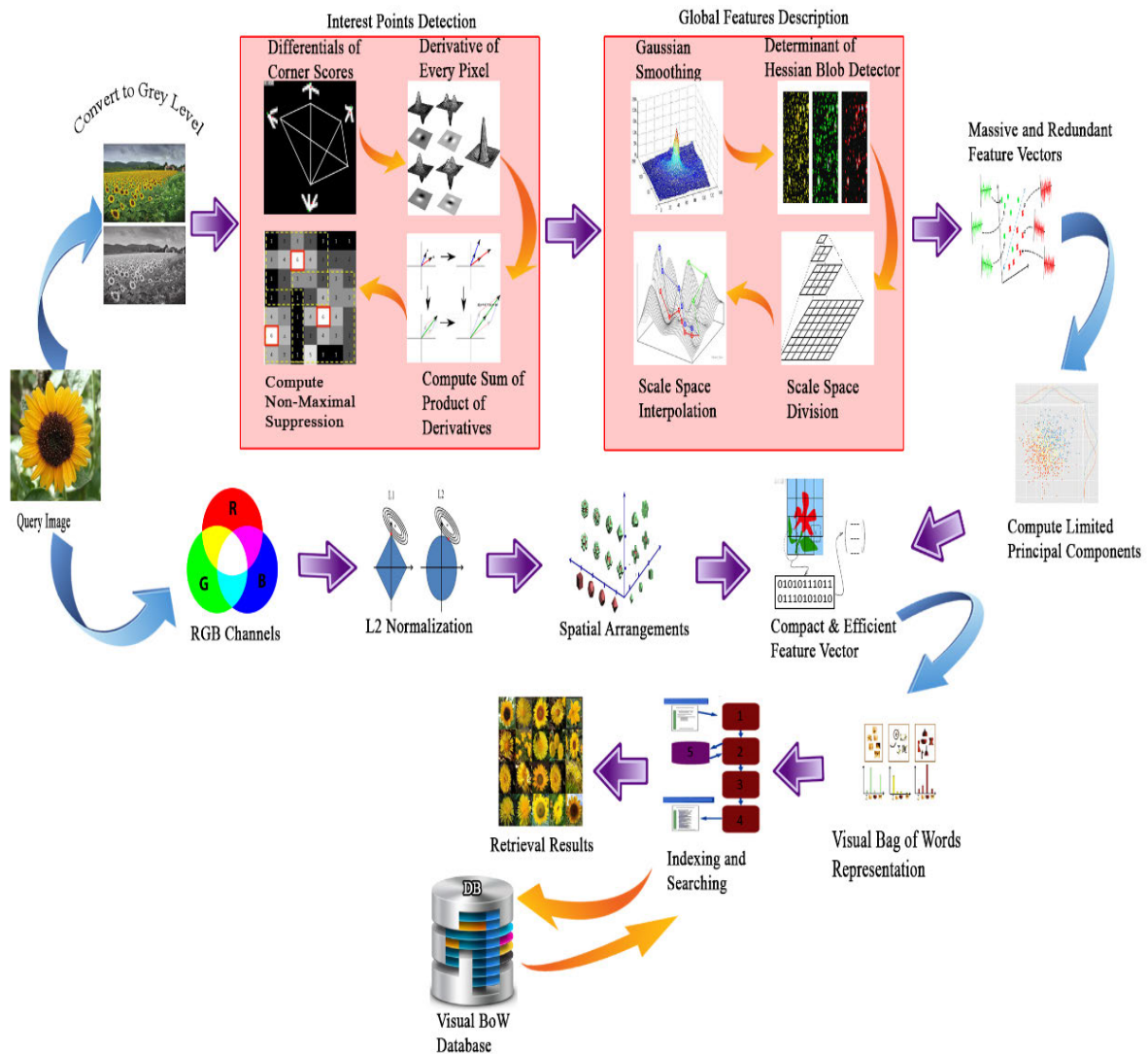
FIGURE 1. Proposed model for efficient image retrieval using signature aggregate with Bow.

gradually bigger masks to obtain the outputted layers, by considering the specific structure of the filters and the specific type of integral images. The step size for the succeeding mask also is scaled as per steps. This step of scale space reveals the benefits of suppressed fine scaling achieved by applying the parametric smoothed kernels with minimum returned parameters. These scale parameters support to generate fine level of scaling with maximum image information. Moreover, it's another advantage is that it is widely applicable and can be derived from limited axioms. The scale space is divided by highlighting a sequence of filter actions correspondence acquired by applying the convolution. Octave incorporates a scaling factor of size 2 and is equally divided into scale levels. Since integral images are of having distinct attributes, the smallest difference of scale is in the direction of derivation of the partial second order derivative between two successive scales of the positive or negative lobes. For the 9×9

filter the length 10 is 3. D non-maximum suppression [37] spatially applied and the neighboring scales in the image to localize the interest points. The beginning and ending Hessian response maps are used for comparison only. Therefore, Hessian matrix determinants interpolation is applied for scaling and image spacing for the maxima response computation so that by applying the interpolation, the lesser obtained scale is λ.

Similar consideration also needed for the related octaves. For the octaves, the increase in the filter size is doubled each time. Moreover, by increasing the sampling intervals by twice minimizing the computational time and increases the accuracy comparably better than the traditional approach. Other octave can be calculated in the same way. In a typical scenario of scale space analysis, the interest points detected on an octave decreases very quickly.

The large level changes are between these octaves and first filters which reduces the scale sampling. Scale spacing is implemented by applying improved sampling rates of scales to calculate the specified image. Filter size of fifteen is optimally used as the first octave. The lowest scale is computed using quadratic interpolation for better accuracy.

The Frobenius norms are already scale normalized as they remain constant for the proposed filters at any size. The Hessian matrix maxima determinants are interpolated in scaling and spacing with [37]. Scale space interpolation is specifically important, since the difference is relatively large between the first layers of every octave. In the scale space division process some pixel values are missed and required to be estimated. To best approximate the missing intensity of the pixels from their neighbors, interpolation is inducted at this step. It calculates the missing points from the known data. This scale space interpolation facilitates the composition of real intensities for correct feature vector generation that leads to better precision. Interpolation at this step is the novelty of the proposed model which otherwise produces less reflective image signatures with discarded samples. The problem occurred during subsampling, reduction, and truncation are approximated using interpolated values. To reduce the feature vector size, principal component analysis (PCA) is performed by applying Eigen coefficients and cyclic steps to calculate the principal components. It is an orthogonal transformation where uncorrelated coefficients are formed from the correlated variables. These computed interrelated coefficients are called principal components. It is a fact that the computed PCs are normally lesser than the original discrete values; however these can be equal to original number of input values. The maximum variance is found in the first component, then in second and it decreases serially. The following variables are orthogonal to their previous serial neighbors and possess less reflections. The results are irrelevant to dependencies if data no less internal relations which is also not convergent to original values. The choice of the PCA over independent component analysis (ICA), linear discriminant analysis (LDA) is due to strong data covariance computation and factor scores. Moreover, ICA searches for separable components instead of successive ones. Separately, an RGB image is considered as RGB components to be treated each as a channel to represent those features. RGB channels are carriers of primary colors to represent the image features. The significance of proposed method is that it equally collects the color channel coefficients along with grey level intensities. The proposed approach perform spatial mapping of these colors to reveals the deep image contents. Color information coupling with grey level values generates maximum image content representation. Color information specifies typical objects and their positioning with spatial coordinates resolves the semantic similarities; which is focused in our approach to obtain better precision and recall rates.

Moreover, a general physical model [38] presents a dense sample of material reflected with a related components can be correctly estimated by a function as in equation 4 [38]:

$$\Re\left(\vartheta,\sigma\right)=\begin{cases}M_S\left(\vartheta\right)C_S\left(\sigma\right) & \dots\text{i}\\ M_S\left(\vartheta\right)C_S\left(\sigma\right)+M_B\left(\vartheta\right)C_B\left(\sigma\right) & \dots\text{ii}\end{cases} \quad (4)$$

where ϑ represents dependency on the angles and σ represents the wavelength [39]. Where S and B indicate the surface reflection and body reflection respectively. Equation-4 part 2 represents the results for non-homogeneous material. Since surface reflection from an inhomogeneous dielectric is focused in a single direction [38], the contribution of MS(ϑ) illuminated from a single direction to measured. To measure the sensor values at each (x,y), it can be shown as equation 5 and 6 [38], [39]:

$$s_i\left(x,y\right)=\int_\sigma f_i\left(\sigma\right)L\left(\sigma\right)M\left(\vartheta\right)C\left(\sigma\right)d\sigma \quad (5)$$

$$=M\left(\vartheta\right)\int_\sigma f_i\left(\sigma\right)L\left(\sigma\right)C\left(\sigma\right)d\sigma \quad (6)$$

where ϑ is a function of image points (x, y). The material surface given by (5.5) highlighted by a spectral distribution $L(\sigma)$.

Let S = $(s_0,\dots\dots,s_{N-1})$ denote the directional line then L2 normalization is represented in equation-7 [38], [39]:

$$\hat{S}=\frac{S}{\sqrt{s_0^2+\dots+s_{N-1}^2}}=\left(\hat{s}_0,\dots,\hat{s}_{N-1}\right) \quad (7)$$

Let,

$$K_i=\int_\sigma f_i\left(\sigma\right)L\left(\sigma\right)C\left(\sigma\right)d\sigma \quad (8)$$

then from equation 6 [38], [39]:

$$\hat{s}_i\left(x,y\right)=\frac{M\left(\vartheta\right)K_i}{\sqrt{M\left(\vartheta\right)^2(K_0^2+\dots+K_{N-1}^2)}} \quad \text{(i)}$$

$$=\frac{K_i}{\sqrt{K_0^2+\dots+K_{N-1}^2}} \quad \text{(ii)} \quad (9)$$

and $\hat{s}_i$ (x,y) depends on the input sensor, the brightness and the reflective output; it is not dependent to ϑ. L1 normalization is used to produce the color space coordinates which are also called chromatic coordinates [40], [41], so that

$$\hat{S}=\frac{S}{s_0+\dots+s_{N-1}} \quad (10)$$

L1 normalized coordinates like the L2 normalized are not fully reflect the scene geometry. The proposed model applied L2 normalization instead of L1 because L2 focuses the optimization of mean cost rather than median. The choice of L2 results in performance gain. Comparatively, overall error rate is lower in L2 regularization by limiting the outliers. L1 has the problem of limited differentiation due to preventive outfitting and sparsity enforcement. L2 has smoothens it and shows invariance with better coverage. Another advantage of applying L2 is its nature of squaring the input that is closed form while L1 pairwise absolute function; therefore

L1 is computationally expensive. In L1 norm, for two different material the distance between the coordinates in depends upon line location. Two texture based materials with $\theta 1$ angle; where it denotes the angle between materials. For L1 normalization and Euclidean distance dl the equation becomes [39]:

$$d_1 = \sqrt{2} \left[\frac{t(\theta + \theta_1)}{1 + t(\theta + \theta_1)} - \frac{t\theta_1}{1 + t\theta_1} \right] \quad (11)$$

For L2 normalization the Euclidean distance d$'$ between is as [39]:

$$d' = k \times s \left(\frac{\theta}{k} \right) \quad (12)$$

where d' is dependent to θ, $k = 2$ and s is *sine* computation. For higher dimensional sensor spaces the situation is same. It is evident that the two colors on the surface will be same even if one point is directly highlighted and the other is in a shadow [42]. It is also seen that the normalized color behave differently as corresponding to dark points. Two undesirable properties using the L1 norm associated with computing normalized color has analyzed. First property is that the color proportions in sensor space has a zero point $s_0 = s_1 = \ldots = s_{N-1} = 0$. The next and the final step is to perform the indexing and retrieval of images using bag-of-words (BoW) architecture.

Bag-of-words or bag-of-visual-words framework is applied in contrast to support vector machine (SVM) because of the multiclass nature of classification and retrieval. BoW uses occurrences as features instead of class-by-class binary matching. Moreover, BoW applies k nearest neighbors (KNN) model that stores the current instances and classifies based on similarity and produces the results efficiently. BoW has the strength to show the image with local patches which thereby treated as numeric vectors by our approach. These vectors are candidate descriptors to gauge and handle the variations with invariances which otherwise difficult to manage in binary classification schemes. BoW is efficient due to its clustering and codewords modeling where the learned patches are mapped to codewords using clustering. Moreover, BoW is also a powerful solution against other models including AdaBoost [43] and pyramid matching [44]. The BoW representation is histogram description with each local descriptor is allocated to visual word. In the offline training staged {s_1,...,s_n } of n clusters trained by K-means. The descriptors from a given image are the vectors quantized into a pre-structured vocabulary. A histogram of local descriptors is constructed to form a fixed length of representation with n bins of an image and based on mapping each descriptor is assigned to the nearest cluster. For efficient comparison of BoW representations the inverse document frequency is applied with inverted valued list. BoW images are indexed and relevant results are searched from visual BoW database and display retrieval results.

This research work has the following contributions:

1. The model comprehensively collects and analyses the entire image contents including texture, color, shape, object and spatial information which actively produces the highest precision and recall rates.
2. Introduced a light-weight feature detection and description model that efficiently retrieves the relevant results from complex and cluttered datasets.
3. A novel image feature fusion method is incorporated by assembling the spatial coordinates with primitive candidates.
4. First time presented a technique performs suppression, scaling, and interpolation together to obtain the deep finer image content details.
5. To enforce the semantic difference, a new method is introduced with spatial color mapping to highlight the objects.
6. An innovative methodology is presented that successfully returns remarkable performance on tiny object, similar textures, complex background objects, overlay ambiguous objects, resized/enlarged images, cluttered patterns, color dominant arrangements, mimicked, occluded and cropped objects.
7. The strength of the presented technique is to reveal only the relevant image contents information from anchor translation rather than complete image iterations.
8. A unique recipe that works over color channels and grey levels simultaneously to act upon the symmetric content representation strategy.
9. A time, computation and storage efficient retrieval system is introduced that retrieved the results in fraction of seconds.
10. A new idea to accumulate the strength of normalized scaled features with bag-of-words architecture to stimulate the indexing and classification.

IV. EXPERIMENTATION
A. DATASETS
For accurate image retrieval, the selection of an appropriate dataset is a crucial task. The results precision proportional to image attributes like size, quality, location, color, cluttering and occlusion. Results of a retrieval system based on the selection of subsets which represents the respective database and include various categories from different areas. Therefore experiments were performed on several standardized datasets, including Caltech-101 [48], ImageNet [45], [46], FTVL [53], ALOT [51], Caltech-256 [47], Corel-10,000 [50], Corel-1000 [49], COIL [52], and 17-Flowers [54].

1) COREL-1,000 DATASET
Corel-1000 dataset is a renowned standard used for image classification and retrieval [49]. The dataset comprises of 1,000 images divided in 10 categories namely Africa, bus, beach, dinosaur, horse, food, building, elephant, flower, mountains. All categories include hundred images of size 384×256 or 256×384 as shown in figure 2(a).
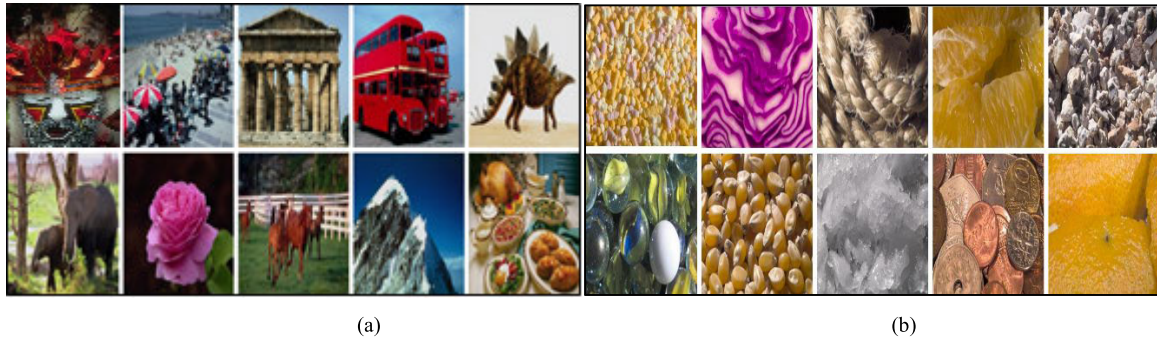
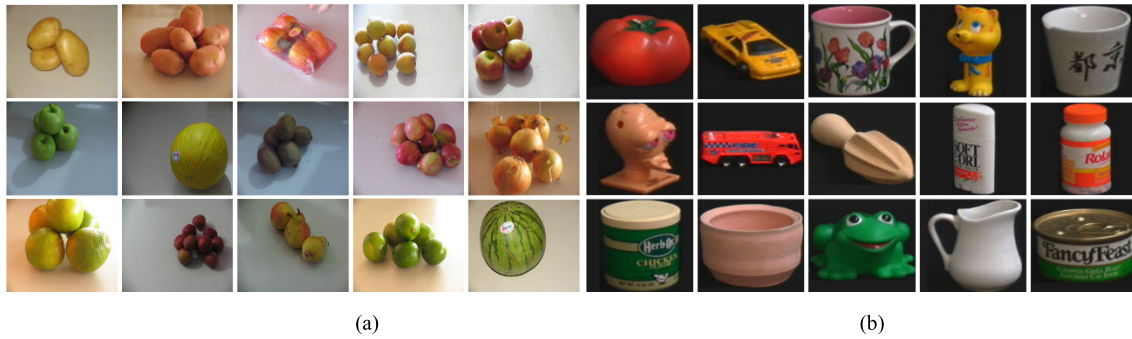**FIGURE 2.** Image samples from all categories (a) Corel-1000 dataset (b) ALOT dataset.

FIGURE 3. Image samples from all categories (a) FTVL dataset (b) COIL-100 dataset.

2) ALOT DATASET

Amsterdam Library of Textures (ALOT) [51] is a color image dataset of 25,000 images with 250 rough textures, used for the scientific purposes and is available to download. In order to capture the sensory variation in the object recognition 10 categories with 100 images of each material were selected. The selected categories include fruit-sprinkles, rope, red-coal, orange-parts, toy-marbles, coins, corn, stones, ice-thick-layer and mandarin-pee as shown in figure 2(b).

3) FTVL DATASET

The supermarket database of fruits and vegetables [53] comprises of 2,612 images of size 1024 × 768. FTVL dataset consists of 15 different classes such as Spanish Pear, Plum, Diamond Peach, Fuji Apple, Taiti Lime, Agata Potato, Watermelon, Onion, Asterix Potato, Kiwi, Granny Smith Apple, Orange, Honneydew Melon, Cashew, Nectarine and Fuji Apple as shown in figure 3(a).

4) COIL-100 DATASET

Columbia Object Image Library (COIL-100) [52] is a standardized database which includes 7200 images from hundred different objects. The dataset corresponds to 72 types of rotations for each object. Sample dataset images of COIL-100 are depicted in figure 3(b). For the experimental purpose 15 objects were chosen. Each objects is turntable at the rotated 360 degree having black noncomplex background. To change the object position images were taken at rotation 5 degrees.

5) COREL-10,000 DATASET

Corel-10K dataset [50] is the most widely used database representing various scenes and subjects to test CBIR performance. Corel-10K database consists of 100 categories and 10,000 images. Every category has 100 images in JPG format of size 85 × 128 or 128 × 85. Images in the dataset are from diverse contents such as sunset, planets, flowers, butterfly, cars, hospital, flags, trees, food, texture, etc. represented below in figure 4(a).

6) IMAGENET SYNSET

ImageNet [45], [46] is a large scale dataset (synset) with over fifteen million high resolution images which belong to hundred thousand categories; used to index and retrieve multimedia data. The repository contains a huge collection of more than 14,197,120 images. For experimentation 15 synsets each containing 100 images were randomly selected which include dust bag, aeria, cherry radio-telephone, nard, tomato, coffee cup, dish, car, gas fixture, flower, golf ball, scootie, wooly bear-caterpillar, flag, and Walnut. These classes were chosen by to their complex nature, textures and art, versatility, and object features. Sample images are shown in fig. 4(b).
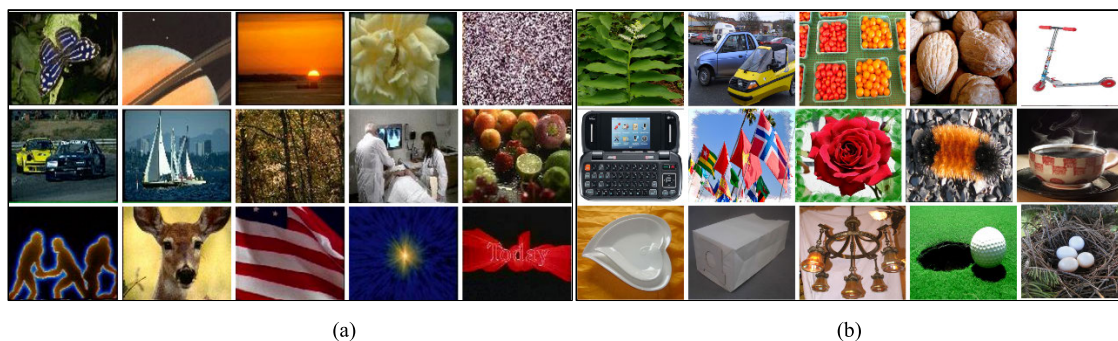
FIGURE 4. Image samples from all categories (a) Corel-10,000 Dataset (b) ImageNet Synset.
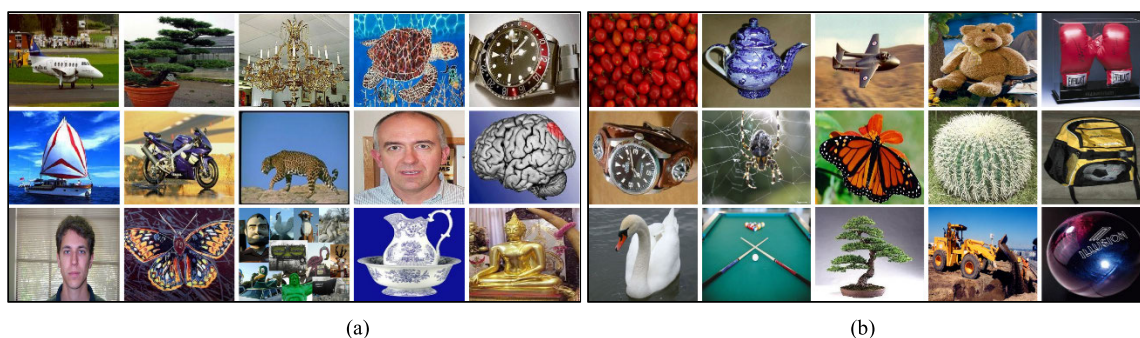

FIGURE 5. Image samples from all categories: (a) Caltech-101 Dataset (b) Caltech-256 Dataset.

7) CALTECH-101 DATASET

Caltech-101 [48] image database is used for the image retrieval, image classification, object matching and recognition tasks as shown in figure 5(a) with sample images. It contains more than nine thousand images belonging to more than hundred distinct image categories. 15 categories with 80 images in each were selected for the retrieval task, including face, airplane, bonsai, face-easy, brain, ketch, chandeliers, things, buddha, tortoise, motorbikes, leopard, wrist watch, butterfly, and ewer. These categories were selected to test the superiority of the presented technique which have ability to share the spatial values, rounded and multi-shaped objects with texture information integrated with color channels.

8) CALTECH-256 DATASET

Caltech-256 dataset [47] contains more than 30 thousand images; which are assigned to 257 categories. Caltech-256 database is a more complex than Caltech-101 dataset with having variation in image categories and contents. Experiments are performed on 15 diverse categories including airplane, swan, back-pack, boxing gloves, bonsai, spider, billiards, tomato, cactus, bulldozer, teapot, butterfly, and teddy-bear. The selected semantic groups represented in figure 5(b) belong to many areas of real life. All categories in dataset are important due to their foreground and background objects and texture patterns. A total of 1500 images with 100 images per category were selected for experimentations.

FIGURE 6. 17-Flowers Dataset: Image samples from all categories.

9) 17-FLOWERS DATASET

The 17-flowers dataset [54] contains 1360 images of flowers with different sizes belonging to 17 numerous classes containing 80 images. These classes show higher level of image changes within the class and possess resemblance with other classes. The dataset images were gathered by web surfing and taking pictures. Images from the dataset are represented in figure 6.

B. EXPERIMENTAL RESULTS

Experiments were performed on 9 standard datasets and results were compared with the research techniques and benchmark descriptors. These descriptors include HOG [57], LBP [61], SURF [56], SIFT [55], DoG [60], RGBLBP [58]

FIGURE 7. Retrieved results for Corel-10,000 Dataset, 17-Flowers Dataset, Caltech-101 Dataset, Corel-1000 Dataset, Caltech-256 Dataset, ALOT Dataset, FTVL Dataset, ImageNet Synset and COIL Dataset.

and MSER [59]. and Experimental results of top 20 outcomes against the query image for the each dataset are shown in figure 7.

The accuracy of the proposed approach is calculated by applying the challenging measures. These are average precision (AP), mean average precision (mAP), average recall (AR), mean average recall (mAP), precision & recall (P&R), average retrieval precision (ARP) and average retrieval recall (ARR). Precision is calculated by dividing relevant results to sum of retrieved results. Average precision is the ratio of precision in respective image category to total number of iterations. Average retrieval precision is the ratio of average precision of image category to total number of categories in which each category precision is summed up to first category. Before plotting, these ARP values are sorted to show the gain or loss gradually.

$$Precision = \frac{Image_{rel(I)}}{Image_{ret(I)}} \quad (13)$$

$$Avergae\ Precision = \frac{For\ each\ C_jP}{\#iterations} \quad (14)$$

$$ARP_k = \frac{For\ each\ C_jAP}{\#categories} \quad (i)$$

$$ARP = Sort(For\ each\ ARP_k) \quad (ii) \quad (15)$$

In equation 13, *rel* and *ret* are the relevant matched and retrieved images. In equation 14, C_jP represents C as category and j is the number of category with P standing the precision in the relevant category. In equation 15 (i), C_jAP represents C as category and AP is the average precision in that category and #categories are the total number of categories upto that iteration. Equation 15 (ii) applies the sort function on all calculated values.

Recall is calculated by dividing relevant plus related results to relevant results. Average recall is the ratio of recall in the respective image category to total number of iterations. Average retrieval recall is the ratio of average recall of image category to total number of categories in which each category recall is summed up to first category. Before plotting, these ARR values are sorted to show the gain or loss gradually.

$$Recall = \frac{Image_{rel(I)} + Image_{ret(I)}}{Image_{rel(I)}} \quad (16)$$
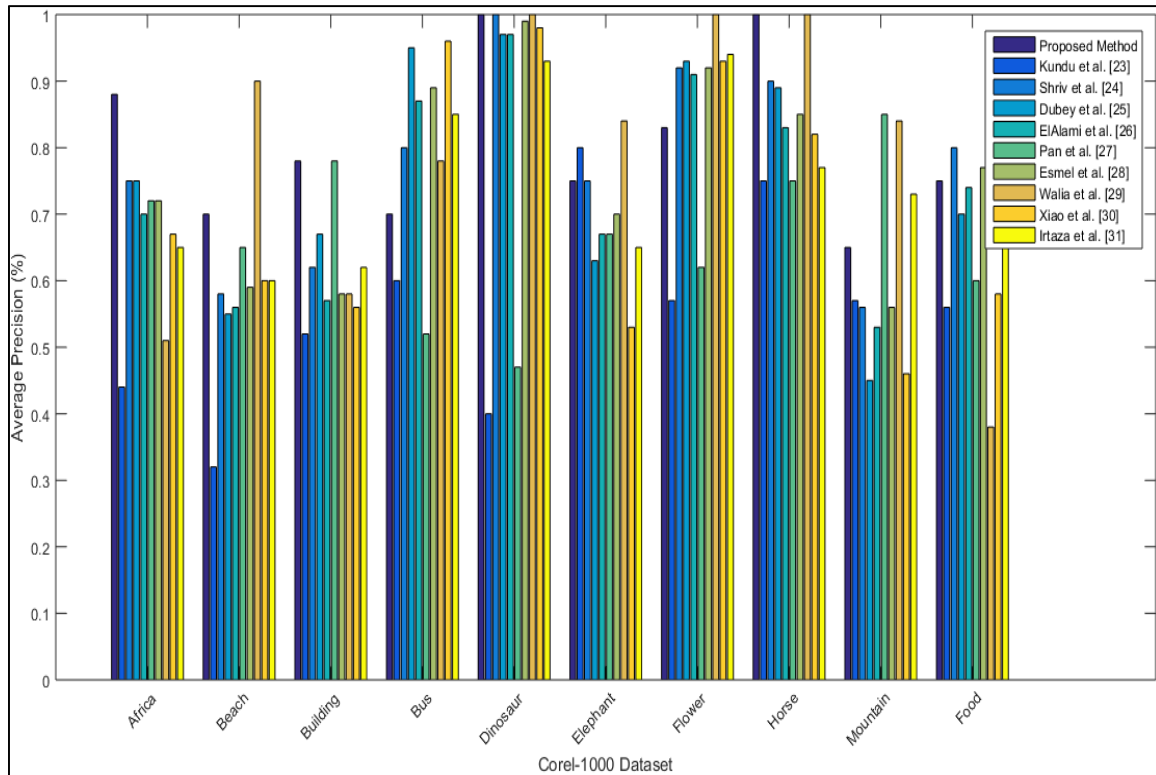
FIGURE 8. Corel-1000 Dataset: Average Precision.

$$Avergae\ Recall = \frac{For\ each\ C_jR}{\#iterations} \quad (17)$$

$$ARR_k = \frac{For\ each\ C_jAR}{\#categories} \quad (i)$$

$$ARR = Sort(For\ each\ ARR_k) \quad (ii) \quad (18)$$

In equation 16, *rel* and *ret* are the relevant matched and retrieved images. In equation 17, C_jR represents C as category and j is the number of category with R standing the recall in the relevant category. In equation 18 (i), C_jAR represents C as category and AR is the average recall in that category and #categories are the total number of categories upto that iteration. Equation 18 (ii) applies the sort function on all calculated values.

Mean average precision (mAP) is computed by summing the average precision in all image group and then divided by the total image semantic group. Similarly the mean average recall is computed by replacing the precision with recall. Precision and recall (P&R) is the ratio of average precision to average recall.

$$mAP = \frac{Sum(AP@C)}{\#categories} \quad (19)$$

$$mAR = \frac{Sum(AR@C)}{\#categories} \quad (20)$$

$$P\&R = \frac{@C(AP)}{@C(AR)} \quad (21)$$

In equation 19-21, $@C$ represents each category and AP, AR are the manipulated average precision and average recall rates.

Retrieved Results for Corel 10,000 dataset and ALOT dataset are shown in figure 7 with up to 95% accuracy. 17-Flowers dataset, Caltech-101 dataset, FTVL dataset and COIL dataset has up to 100% accuracy rates as shown in figure 7. Corel-1,000 dataset has up to 90% accuracy as represented in figure 7. Caltech-256 has accuracy rate of 70% against the query image of category Airplanes as presented in figure 7. It can be shown from Fig 7(h) that ImageNet synset has 65% accuracy rate in complex category of tomatoes. Top 20 images retrieval time taken by the presented method is $\sim 0.3 - 2.19$ sec. The variation of in the computation time is due to the size, and number of images in the datasets. The experiments are conducted on core-i5 @2.5Ghz with 8GB RAM.

1) EXPERIMENTAL RESULTS OF THE COREL-1000 DATASET VS. RESEARCH METHODS

Corel-1000 benchmark is used to measure the accuracy of the presented technique. Results of the presented method are shown in comparison with research methods presented in literature named Kundu *et al.* [23], Shrivastava and Tyagi [24], Dubey *et al.* [25], ElAlami *et al.* [26], Pan *et al.* [27], ElAlami *et al.* [28], Walia and Pal [29], Xiao *et al.* [30], Irtaza *et al.* [31]. Results are graphically represented in figure-8 which shows that the presented approach works

TABLE 1. Corel-1000 Dataset: Average Precision (%) in comparison with standard research techniques

		Corel-1000 Dataset —Mean Average Precision (mAP) (%)								
	Proposed Method	Kundu [23]	Shriv [24]	Dubey [25]	ElAlami [26]	Pan [27]	Esmel [28]	Walia [29]	Xiao [30]	Irtaza [31]
Average	0.80 ± 0.001	0.55 ± 0.002	0.76 ± 0.003	0.74 ± 0.005	0.73 ± 0.003	0.66 ± 0.001	0.75 ± 0.002	0.78 ± 0.001	0.70 ± 0.003	0.75 ± 0.004
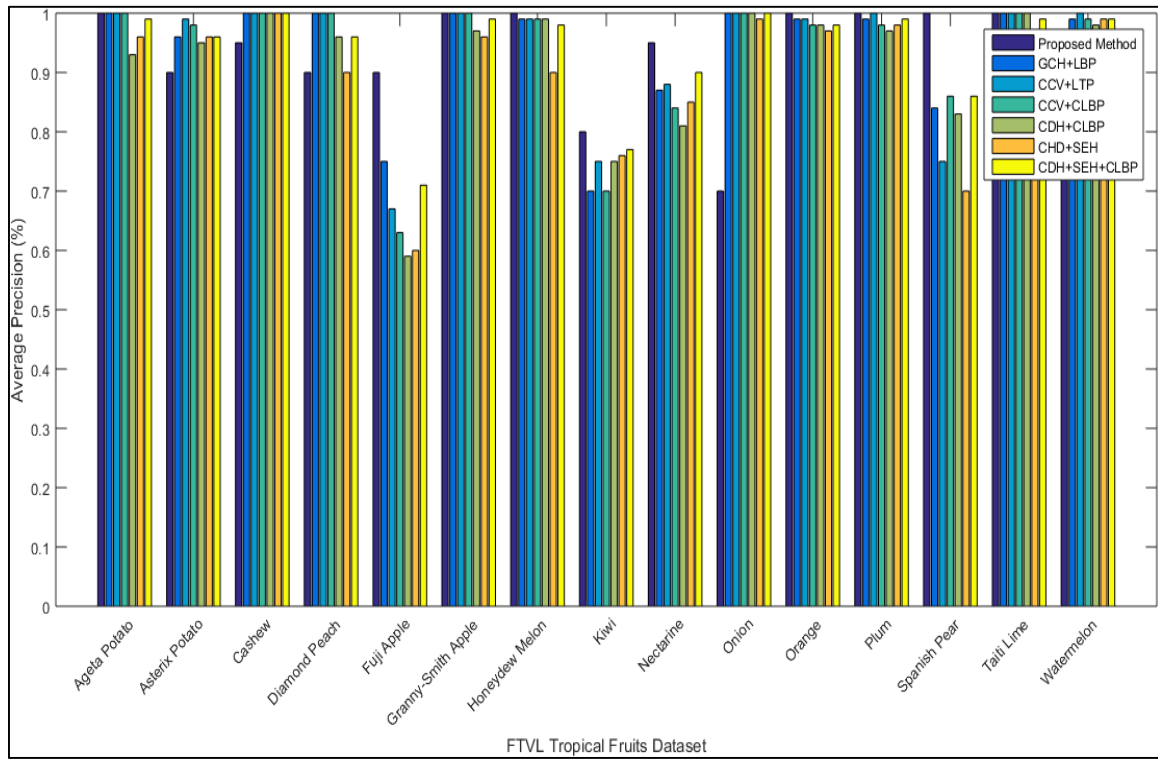
FIGURE 9. FTVL Dataset: Average Precision.

better than many research methods. The presented technique shows better performance in 'dinosaur' and 'horse' for average precision rates for the comparison research techniques. Mean Average Precision (mAP) for the presented approach is also compared with other research techniques is shown in table 13. The presented approach shows improved mAP value of 0.804 than the research techniques. Cumulative results are also remarkable for the presented technique.

Average precision of the presented approach is compared with research techniques is represented in Table 1. The existing methods also show good results in the categories of bus, dinosaur, flower and horse. Shrivastava and Tyagi [24] reported improved rates for dinosaur category because it works well with plain background images but the similar results are not observed in all categories due to specific content analysis while missing the overlay object detection. However, the presented approach shows improved results in dinosaur and other image groups due to its uniqueness of finding the foreground and background objects.

Moreover, Pan *et al.* in [27] reported better precision for building and mountain category because their scheme focuses the scene domain and lacks in object recognition. Walia and Pal [29] provided better precision in dinosaur, flower and horse by showing the strength in color and uncluttered images and resulted low accuracy in texture and complex images. Xiao *et al.* [30] reported the highest precision in bus category and lacks in spatial domain. Irtaza *et al.* shows [31] has the highest precision rates in food and misses to collect the scene domain attributes. Comparatively, the proposed method reports improved precision in scene, texture, color, and spatial features by its entire content analysis methodology.

2) EXPERIMENTAL RESULTS OF THE FTVL DATASET IN COMPARISON WITH EXISTING METHODS

FTVL dataset [53] is used for the experimentations due to its illumination differences, pose variations, partial occlusions and cropped object. Figure 9 graphically depicts the average precision rates of the presented technique with the

TABLE 2. FTVL Dataset [53]: Average Precision (%) in comparison with other standard retrieval methods [62]

		FTVL Database −Mean Average Precision (%)						
	Proposed Method	GCH + LBP	CCV +LTP	CCV +CLBP	CDH +CLBP	CDH +SEH	CDH+ SHE +CLBP
Average	0.937 ± 0.002	0.93 ± 0.001	0.93 ± 0.002	0.93 ± 0.001	0.91 ± 0.002	0.89 ± 0.003	0.93 ± 0.002

TABLE 3. 17-Flowers Dataset [54]: Average Precision (%) in comparison with other standard retrieval methods

		17-Flowers Database −Mean Average Precision (%)					
	Proposed Method	Gao [63]	ScSPM [64]	ScS w/B [64]	LLC w/B [65]	JDL[66]	JDL w/B [66]
Average	0.87 ± 0.005	0.72 ± 0.002	0.52 ± 0.001	0.62 ± 0.004	0.65 ± 0.001	0.68 ± 0.002	0.71 ± 0.005

significant results in comparison with the existing literature research techniques [62]. Table 2 presents the average precision results of the presented approach in assessment with existing research techniques. Some methods CDH + SEH show the low accuracy due to missing nature of their method in cropped objects. Some methods incorporate the texture attribute and reports average precision but mixed the ambiguous objects. Two methods with deep texture patterns reported above average precision by also considering the lacks in other methods. However objects with similar shape and color are still difficult to recognize for them. The proposed method takes into consideration the color coordinates with texture and shape properties to and reports 0.937 mean average precision.

3) EXPERIMENTAL RESULTS OF THE 17-FLOWERS DATASET IN COMPARISON WITH EXISTING METHODS

17-Flowers dataset is used for colored images with texture and shape experimentation. Figure 16 shows comparative results of the presented approach with research techniques for average precision and average recall. The approach in [64] performs spatial matching and computes differences based on this criteria but lacks in color with shape matching therefore results low precision. The other approaches [64]–[66] incorporates linear coding and reports average results. These methods lack shape with texture pattern analysis. The fine-grained [63] approach reports improved results and can be more improved by adding the deeper color and shape details. The average precision rate of the presented approach outperforms in many flower groups by putting spatial color and texture patterns with shape information. Table 3 shows the average precision rates for the presented approach versus the existing research techniques. The proposed method reports 0.876 mAP rates higher than all competitive methods.

4) EXPERIMENTAL RESULTS OF BENCHMARK DESCRIPTORS VERSUS THE PROPOSED METHOD

Experimentation is performed on 7 benchmark descriptors and detectors. These widely used key point descriptors and detectors are MSER [59], LBP [61], DoG [60], SIFT [55], SURF [56], RGBLBP [67], and HOG [57]. Image retrieval systems use feature descriptors and detectors for the texture, object recognition and detection. Image features such as edge, corner or blob are extracted using detectors and descriptors from the interest points. Speeded up robust feature (SURF) is used for the image retrieval [18] and presented at the ECCV. Histogram of oriented gradients is used for classification, object detection and recognition [68], [69] and image retrieval [70]; which was presented at CVPR. To detect colors in an image, RGB model is used in which red, green and blue channels are fused to produce different colors. Image is represented in RGB model as a matrix of X × Y × 3 pixels for the each color component where X & Y are rows and columns of pixels. Maximally Stable Extremal Region (MSER) was described by [59] and is used for blob detection to find correspondences between image features. Difference of Gaussians (DoG) enhances the features by subtracting a less blurred image from the original image. For texture classification local binary patterns (LBP) is used, presented in 1994 [71]. LBP is a type of the texture descriptor. Scale invariant feature transform (SIFT) is used for the object detection [72] and content based image retrieval [73], [74] which was presented in ICCV. These descriptors are compared with the proposed method to measure the efficiency of the presented approach.

a: AVERAGE PRECISION (AP) & AVERAGE RETRIEVAL PRECISION (ARP)

In pattern recognition, image retrieval and object classification Average Precision (AP) is somewhat tricky to interpret. Precision is the average probability among the retrieved instances. AP is the further averaged of all queries and represented as a single score. To measure the versatility of the proposed method, experiments were performed on 9 standardized dataset. Figure 11 graphically represents the average precision rate of top 20 retrievals, compared with benchmarks. Input validation images are collected from all categories to compute the P&R rates for the all semantic groups.
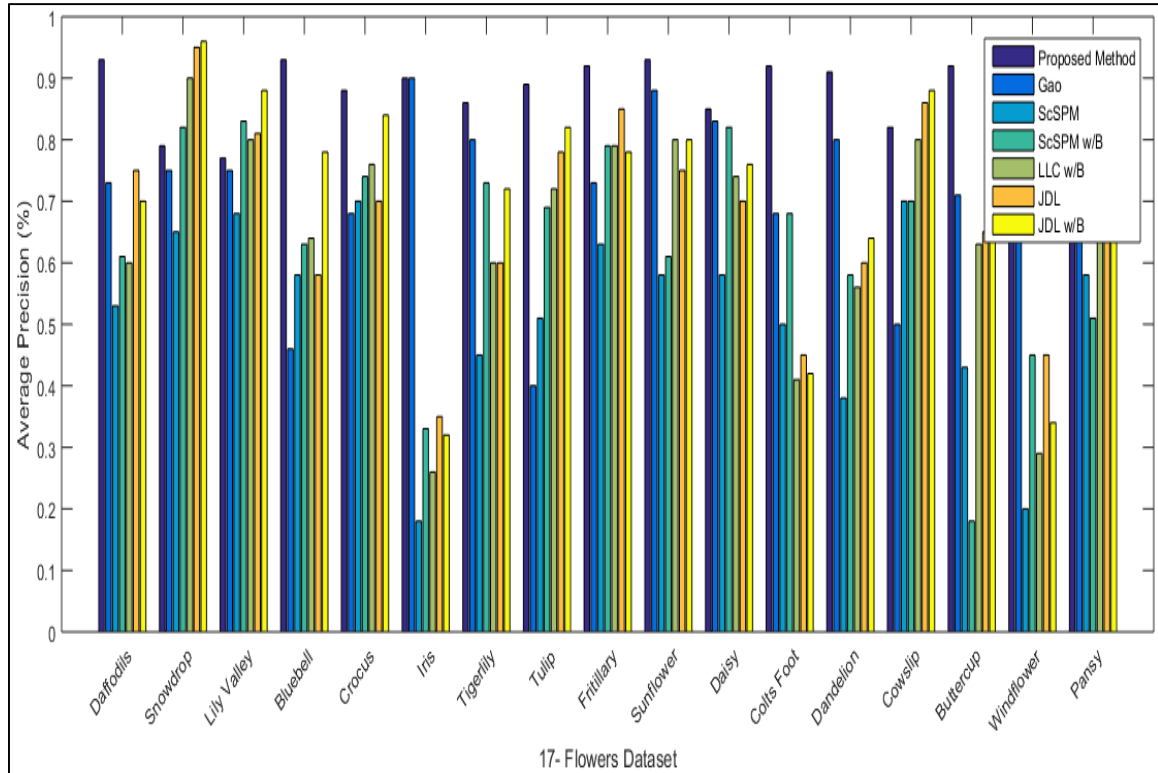
FIGURE 10. 17-Flowers Dataset: Average Precision.

Figure 11(a) represents that the presented approach outperforms in the semantic groups of Horse, dinosaur and Africa from Corel-1,000 due to its superior object recognition capability. Figure 11(b) shows that ALOT database has remarkable average precision rate for texture images. FTVL database shows significant results in most of the image groups as shown in figure 11(c). It is observed from the results represented in figure 11(d) that COIL database has significant results in all the categories. Average Retrieval Precision is shown graphically in Figure 12 and compared with benchmarks. Figure 12(a) represents that Corel-1000 database outperforms in all categories. Figure 12(b) shows that ALOT database has remarkable average precision rate for texture images. FTVL database shows remarkable results in most of the semantic groups as shown in figure 12(c). It is observed from the results represented in figure 12(d) that COIL database has significant results in all the categories.

Average precision for Corel 10,000 database shown in figure 13(a), has improved results in many categories due to its spatial features. Figure 13(b) represents that ImageNet synset has considerable precision rate in complex categories. Caltech-101 dataset contains the images with versatile contents, textures patterns and shapes. The presented approach reflects better precision rate in several image groups as shown in Figure 13(c). Caltech-256 dataset has improved results in many categories due to its robustness as represented in figure 13(d). The proposed method is showed

in figure 15(a) outperforms than the others in many semantic groups. ARP for Corel 10K database shown in figure 14(a) has improved results in many categories. Figure 14(b) represents that ImageNet synset has considerable ARP in complex categories. Caltech-101 dataset includes images with similar objects and complex backgrounds. The presented approach reports significant precision for this dataset as shown in Figure 14(c). Caltech-256 dataset has improved results in many categories due to its robustness as represented in figure 14(d). The proposed method shown in figure 15(b) outperforms than the others in most of the image categories. Results reported that the descriptors suitable for object recognition show better AP and ARP in such categories of image datasets. This fact is reflected from the graphs, as figure-11 (a) shows highest AP for HOG in bus category of Corel-1000, rope and red-coal of ALOT in (b), granny smith apply in (c) of tropical fruit. Similarly average and below average AP is returned for color and texture image classes like crocus and iris figure 15(a). Descriptors suitable for texture and colored textured are RGB and RGBLBP report improved AP and ARP in most of the categories of Corel-10000 and ALOT figure 12 (a,b). Other than the descriptor domain, significance performance is not returned like FTVL figure 12 (c), COIL figure 12 (d). Gaussian differences application on tiny cropped object are better for DoG figure 12 (c) and below average for cross domain categories figure 12 (d) and figure 15 (b). SIFT is used for object
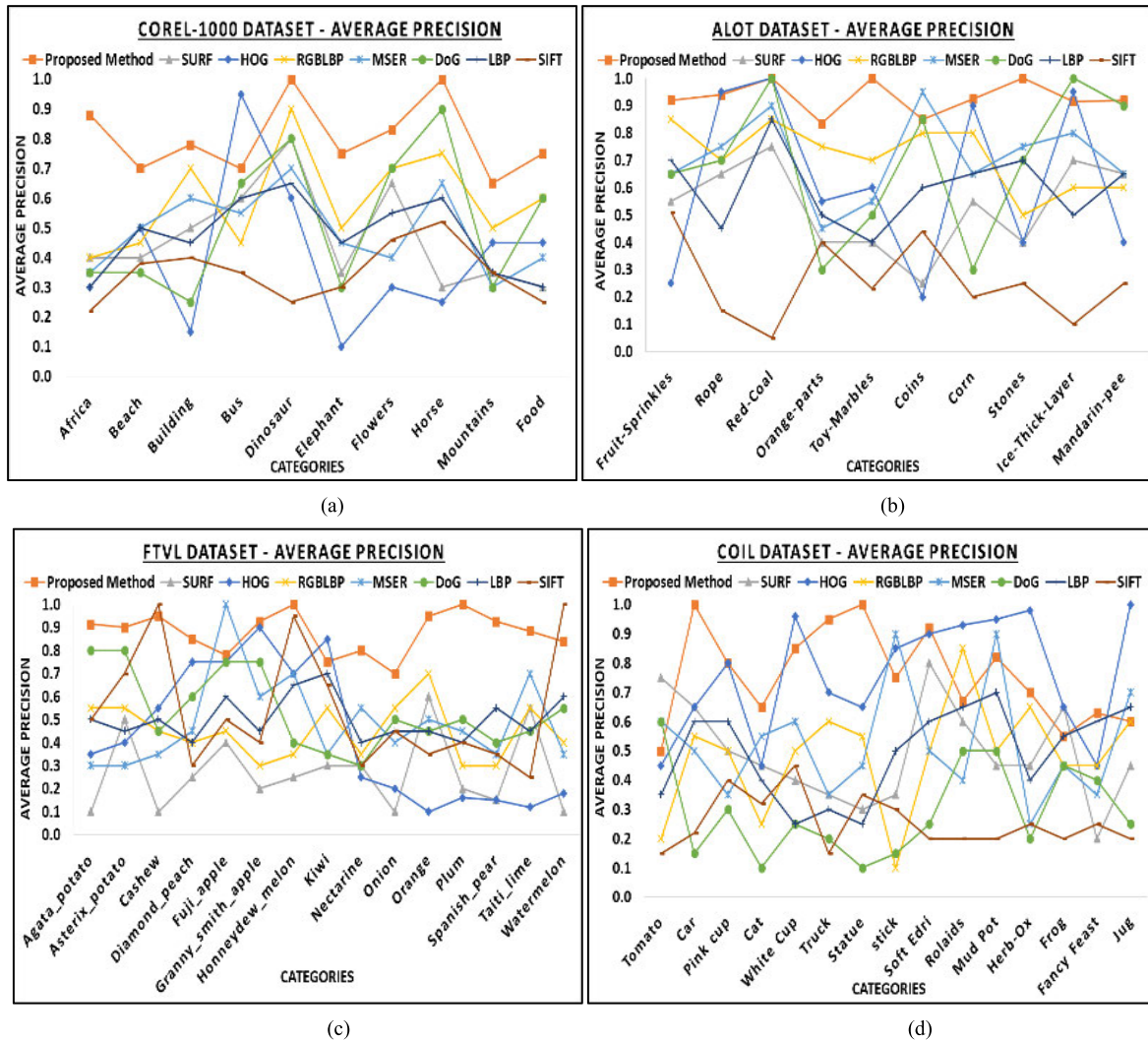
FIGURE 11. Average Precision (AP): (a) Corel-1000 Dataset (b) ALOT Dataset (c) FTVL Dataset (d) COIL Dataset.

detection tasks; hence reports better AP and ARP in most of the categories of object dominant datasets Caltech-101 and Caltech-256 in figure 14 (c, d). However in color texture dataset 17-flowers, below average results are reported in figure 15 (b). COIL with rotational objects in figure 11 (d) has better precision rates by SURF due to its specialty; whilst report average results in out of domain datasets as depicted by figure 14 (a-c). The proposed method reports significant AP and ARP in most of the descriptor domains including object, texture, color and shape as observed by the figure 11-15. The proposed approach is capable to search for cluttered, cropped objects from foreground and background in figure 13 (c, d). It is also able to distinguish to the texture patterns as shown by the figure 12 (b) and figure 15 (a, b). Tiny similar object are successfully distinguished in figure 11, 12 (c). Objects in small and large size images are accurately identified by the proposed method which is endorsed by the figure 11, 12 (a) and figure 13 (a-d).

b: AVERAGE RECALL (AR) &AVERAGE RETRIEVAL RECALL (ARR)

Recall is the fraction of retrieved matched images divided by the total number of matched images. Average Recall (AR) is the average probability of complete retrieval, graphically represented in Figure 16, 18 and 20. In AR graphs, the presented method is compared with the benchmark descriptors. It is evident from the outcomes that the presented method has significant recall rates in all databases.

Figure 16(a) represents significant recall rates in different image groups of Corel-1000 dataset. ALOT Dataset has significant recall rate as shown in Figure 16(b). FTVL Dataset has improved recall repsented in figure 16(c). Figure 16(d) shows considerable racall rate for the COIL Dataset. Figure 17 shows Average Retrieval Recall (ARR) rate against the each category. It is evident from the graphs that the presented approach has significant recall rate in all databases. Figure 17(a) shows significant ARR for
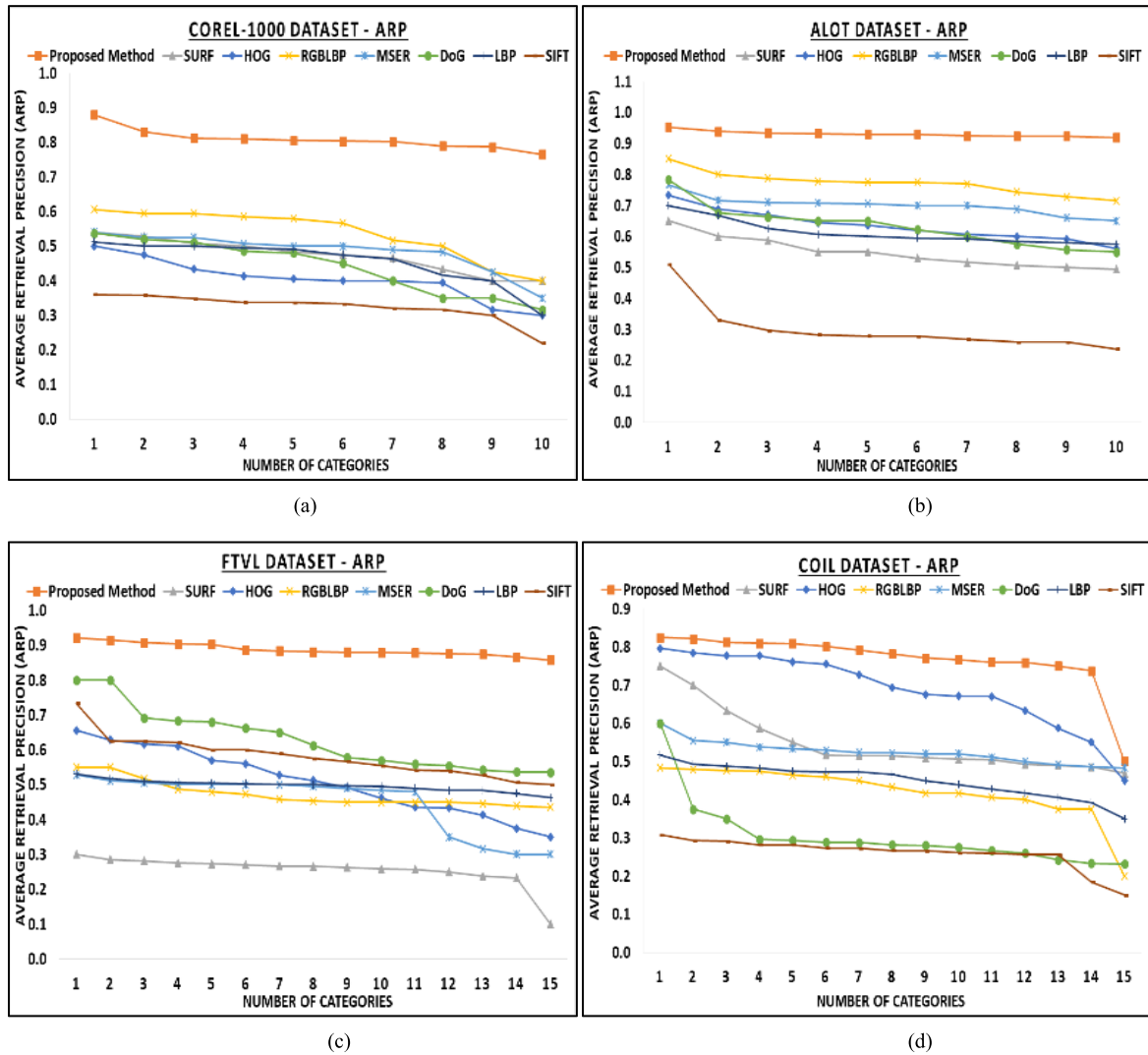
FIGURE 12. Average Retrieval Precision (ARP): (a) Corel-1000 Dataset (b) ALOT Dataset (c) FTVL Dataset (d) COIL Dataset.

Corel-1000 dataset. ALOT dataset also has better ARR as represented in fig. 17(b) and fig. 17(c) presents remarkable ARR for FTVL dataset. COIL dataset has better ARR as shown in fig. 17(d).

Figure 18(a) represents better recall rate in many of the categories of Corel-10,000 Dataset. ImageNet synset has significant recall rate as shown in figure 18(b). Caltech-101 Dataset has improved recall represented in figure 18(c). Figure 18(d) shows considerable recall rate for Caltech-256 dataset.

Results reported in figure 20(a) that 17-Flowers Dataset has remarkable recall rate. ARR for Corel-10,000 dataset is presented in figure 19(a) which shows considerable results. ImageNet synset has better ARR as represented in fig. 19(b). Caltech-101 and Caltech-256 datasets has improved ARR, as shown in figure 19 (c, d). It is observed from fig. 20(b) that 17-Flowers dataset has remarkable ARR.

c: PRECISION AND RECALL RATIO

In an image processing system precision is part of the output images and the relevant content images which are indexed to fetch. While recall is the part of relevant content images with the retrieved images. Precision and recall are inversely proportional set-based procedures, computed using unordered results. A retrieval system needs to achieve a balance between precision and recall. This situation, leads to the need of Precision-Recall (PR) graph based on the measure and understanding of relevance. Figure 21 is the precision and recall depiction of Corel-1000, A LOT, FTVL and COIL datasets.

PR rates for Corel-1000 dataset are between 65% and 100% as shown in fig. 21(a). PR rate for ALOT dataset shown in fig. 21(b) is from 75% to 100%. Figure 21(c) represents PR for FTVL dataset which is from 70% to 100% and COIL dataset has more than 80% PR rate in many image groups as presented in figure 21(d).

FIGURE 13. Average Precision: (a) Corel-10,000 Dataset (b) ImageNet Synset (c) Caltech-101 Dataset (d) Caltech-256 Dataset.

Corel-10,000 dataset has more than 70% PR rate in many categories as shown in figure 22(a). PR rates for ImageNet synset and Caltech-256 dataset are shown in figure 22(b, d) are around 75%. Figure 27(c) represents PR for Caltech-101 dataset which is more than 60%. It is graphically shown in figure 22(d) for 17-Flowers dataset with 60% to 100% PR rates.

d: MEAN AVERAGE PRECISION (mAP) AND MEAN AVERAGE RECALL (mAR) RATES

Mean Average Precision (mAP) is a very popular standard single-number performance measure in information retrieval for comparing search algorithms. mAP is the average of AP across a set of queries. A single bar in the graph represents overall mAP of all categories for the presented approach and the benchmarks.

For ALOT benchmark, the presented apporach shows 0.93 mAP that is greater than in comparison benchmarks

as depicted by figure 23. SIFT shows lowest mAP rates since it shows the lowest results in Rope and Ice-Thick-Layer categories. Table 4 shows mAP of the presented approach and the research based benchmarks. It can be shown that MSER and RGBLBP descriptors report almost equal mAP. Moreover HOG, DoG and LBP report average mAP.

For COIL dataset, the mAP of the presented appraoch is 0.93 as represented in figure 24 that is 15% higher than HOG. HOG shows the second highest mAP that is 0.77. SIFT and DoG shows the lowest mAP that is 0.20 and 0.30 respectively. SIFT shows low results in tomato and truck categories for the reason of their same color. Furthermore DoG reports mAP rates in the categories having same color features. Table 5 shows the mean average precision rates for all seven descriptors experimented along with the proposed method.
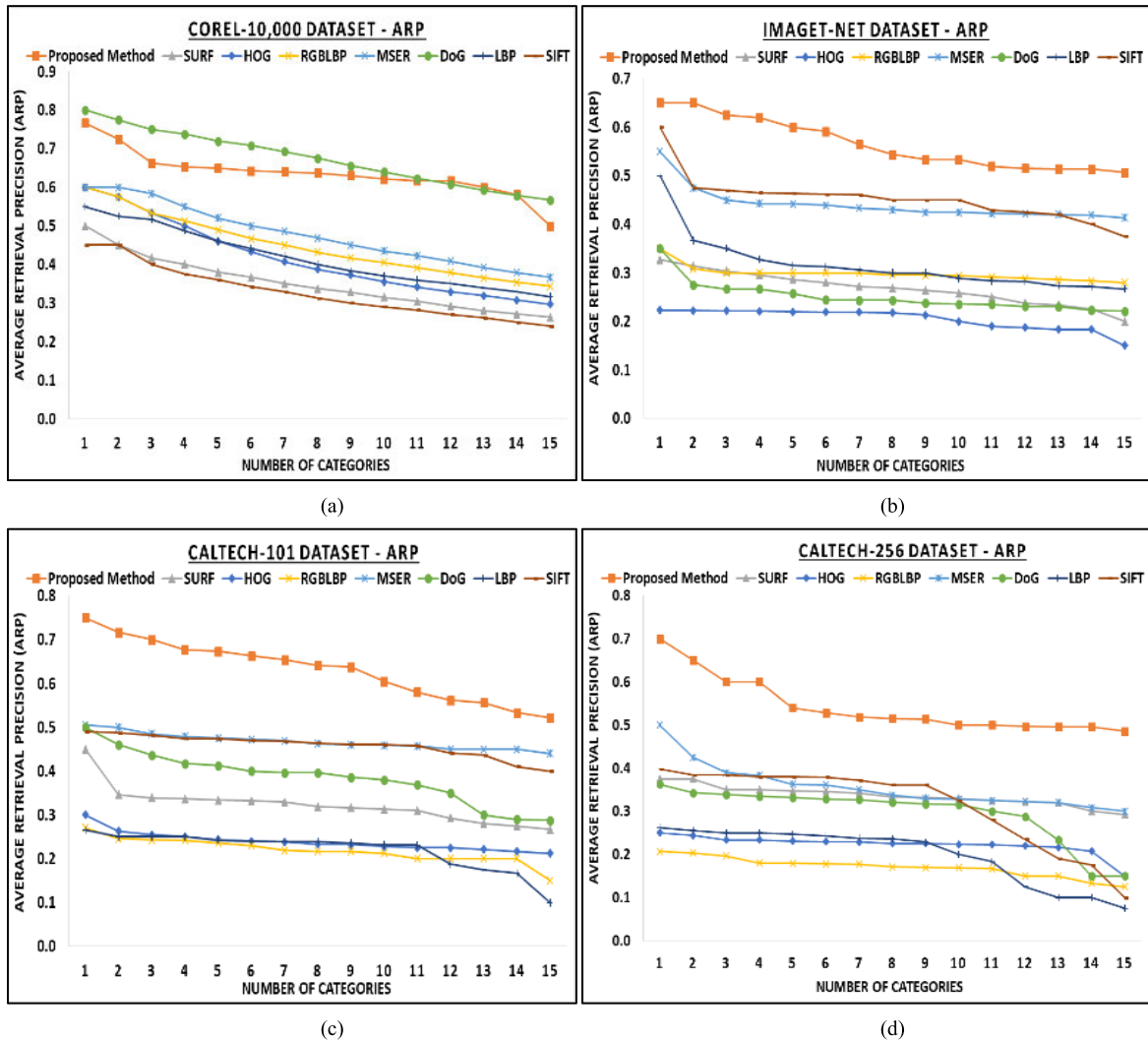
FIGURE 14. Average Retrieval Precision: (a) Corel-10,000 Dataset (b) ImageNet Synset (c) Caltech-101 Dataset (d) Caltech-256 Dataset.
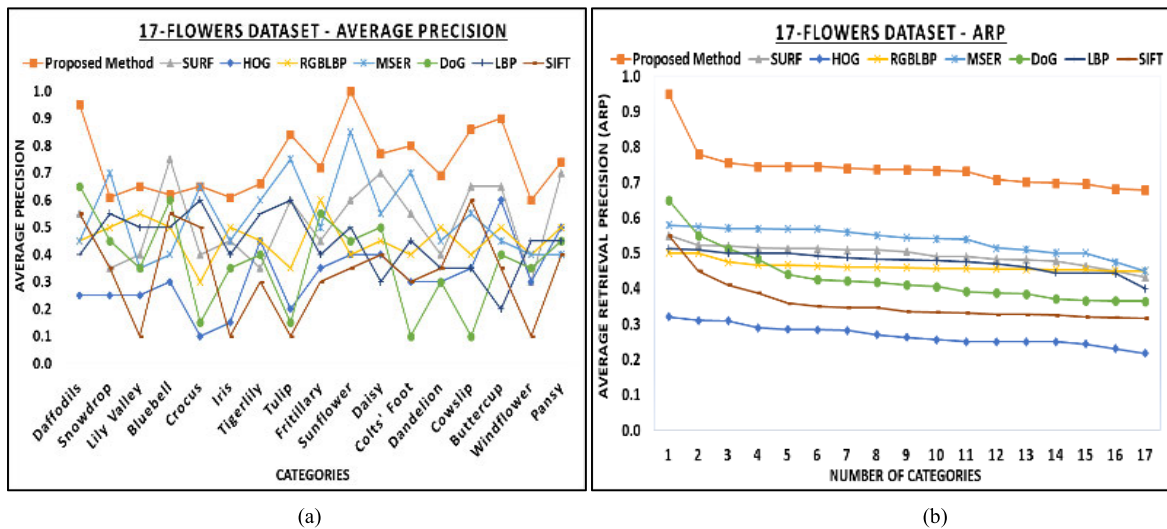

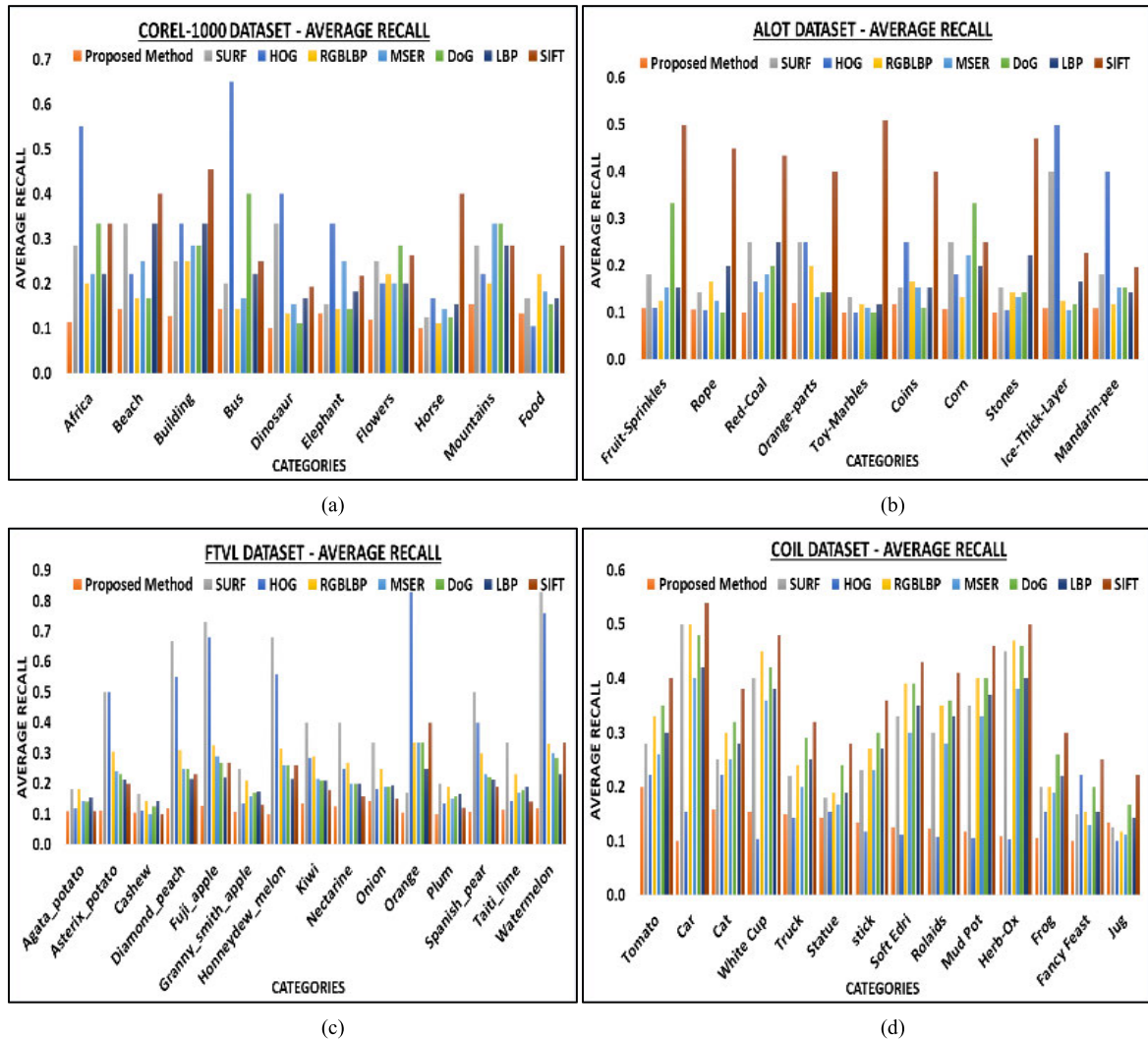**FIGURE 15.** 17-Flowers Dataset: (a) Average Precision: (b) Average Retrieval Precision.

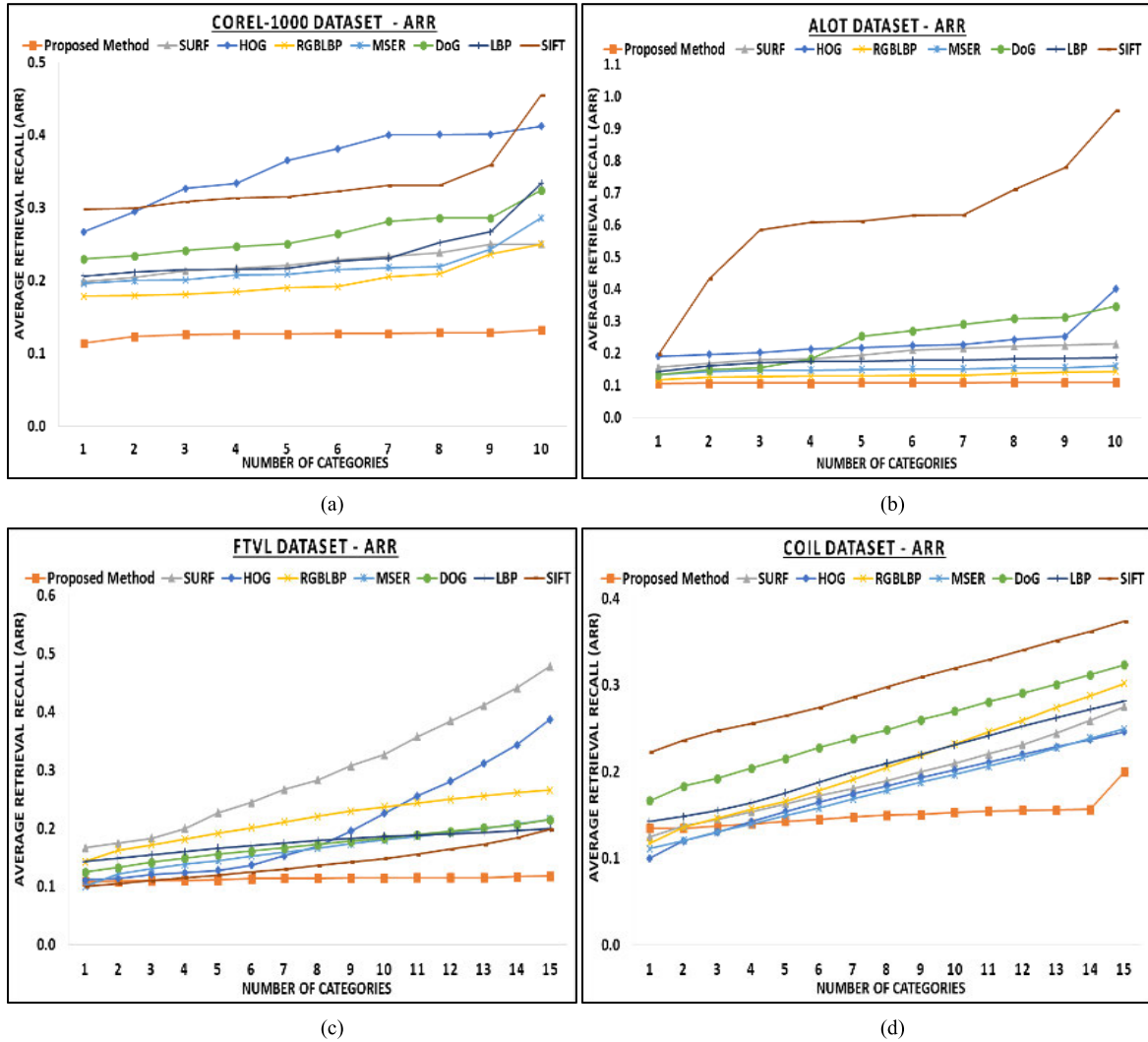**FIGURE 16.** Average Recall (AR): (a) Corel-1000 Dataset (b) ALOT Dataset (c) FTVL Dataset (d) COIL Dataset.

TABLE 4. ALOT Dataset: Mean Average Precision (%)

Proposed Method	SURF	HOG	RGBLBP
0.93 ± 0.001	0.53 ± 0.001	0.62 ± 0.002	0.71 ± 0.004
MSER	DoG	LBP	SIFT
0.71 ± 0.002	0.65 ± 0.004	0.60 ± 0.001	0.30 ± 0.005

TABLE 5. COIL Dataset: Mean Average Precision (%)

Proposed Method	SURF	HOG	RGBLBP
0.93 ± 0.003	0.49 ± 0.001	0.77 ± 0.006	0.48 ± 0.006
MSER	DoG	LBP	SIFT
0.52 ± 0.002	0.30 ± 0.001	0.49 ± 0.002	0.20 ± 0.001

Figure 25 shows that the mAR of the presented approach for Corel-10,000 dataset is 0.16. Other methods report lower mAR as compared to DoG and the proposed method. DoG reports the second highest mAR that is 0.18. SURF, HOG and SIFT reports 0.297 and 0.317 mAP as shown in table 6 while RGBLBP, MSER and LBP report mAR between 0.317 and 0.367.

FIGURE 17. Average Retrieval Recall (ARR): (a) Corel-1000 Dataset (b) ALOT Dataset (c) FTVL Dataset (d) COIL Dataset.

TABLE 6. Corel-10,000 Dataset: Mean Average Precision (%)

Proposed Method	SURF	HOG	RGBLBP
0.65 ± 0.001	0.26 ± 0.002	0.29 ± 0.001	0.34 ± 0.001
MSER	DoG	LBP	SIFT
0.36 ± 0.002	0.56 ± 0.001	0.31 ± 0.002	0.24 ± 0.001

TABLE 7. ImageNet Synset: Mean Average Precision (%)

Proposed Method	SURF	HOG	RGBLBP
0.50 ± 0.005	0.32 ± 0.004	0.21 ± 0.003	0.28 ± 0.001
MSER	**DoG**	**LBP**	**SIFT**
0.43 ± 0.002	0.24 ± 0.005	0.28 ± 0.001	0.47 ± 0.002

For ImageNet synset the mAP for MSER, SIFT is 0.433 and 0.47 that is among the highest results. The proposed method reports the highest mAP that is 0.5 as represented in figure 26. Other descriptors show an average results since ImageNet synset is difficult to categorize. Table 7 show the mean average precision in tabular format for the
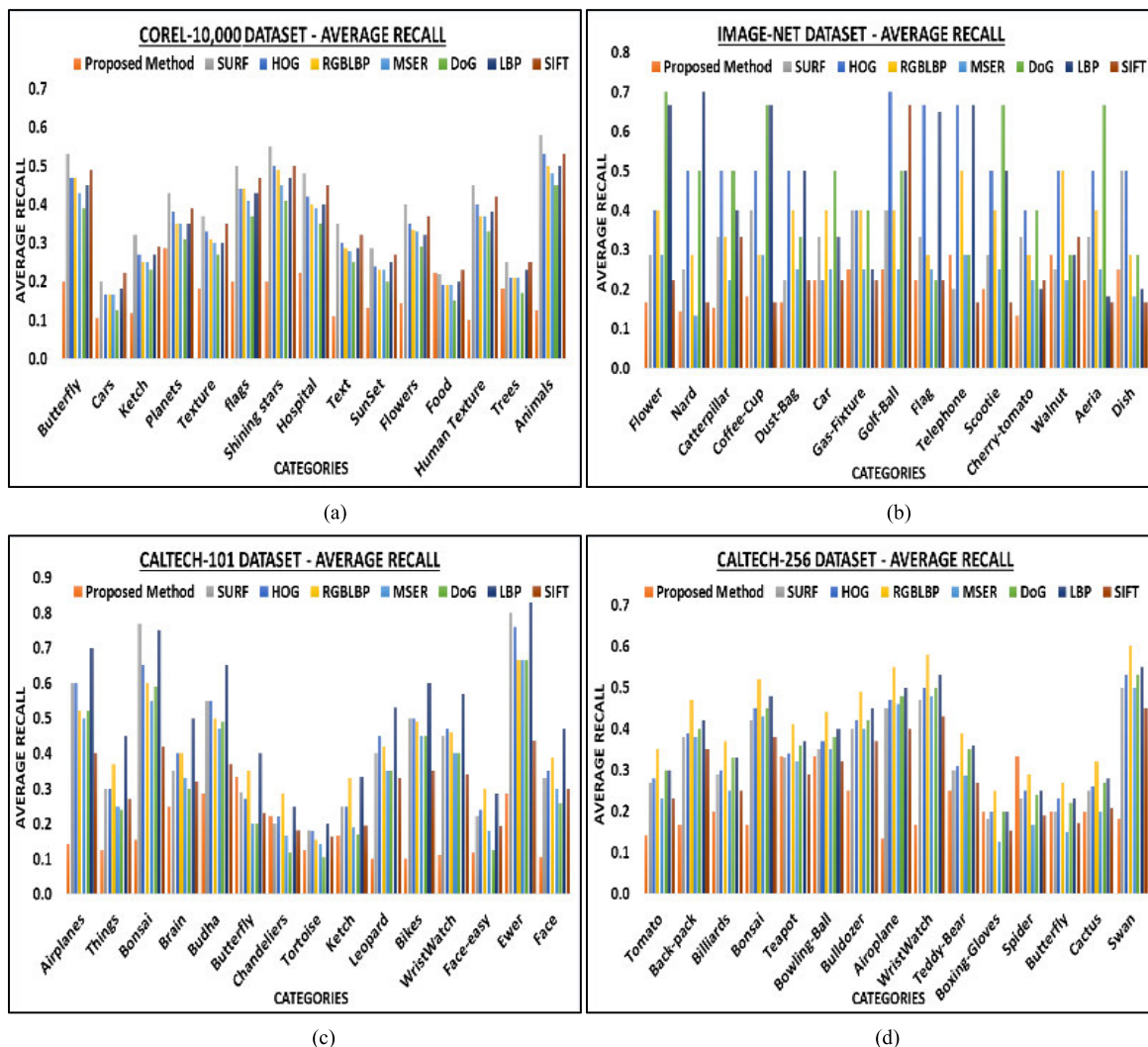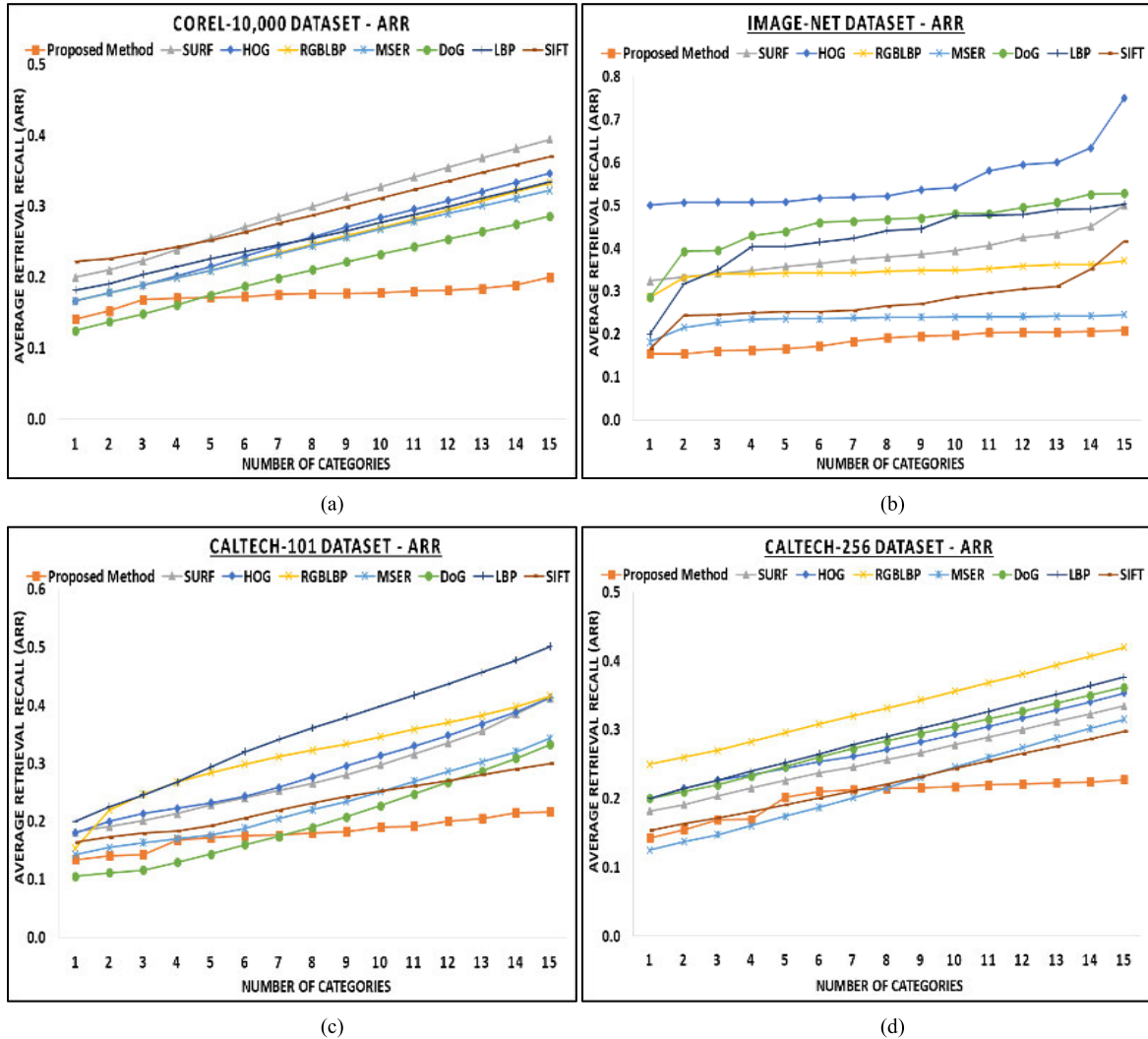
FIGURE 18. Average Recall (AR): (a) Corel-10,000 Dataset (b) ImageNet Synset (c) Caltech-101 Dataset (d) Caltech-256 Dataset.

TABLE 8. Caltech-101 Dataset: Mean Average Precision (%)

Proposed Method	SURF	HOG	RGBLBP
0.67 ± 0.02	0.34 ± 0.001	0.24 ± 0.005	0.27 ± 0.001
MSER	DoG	LBP	SIFT
0.46 ± 0.002	0.38 ± 0.001	0.25 ± 0.001	0.43 ± 0.005

TABLE 9. Caltech-256 Dataset: Mean Average Precision (%)

Proposed Method	SURF	HOG	RGBLBP
0.50 ± 0.001	0.35 ± 0.002	0.22 ± 0.001	0.21 ± 0.002
MSER	DoG	LBP	SIFT
0.38 ± 0.002	0.33 ± 0.001	0.28 ± 0.006	0.36 ± 0.005

presented approach in comparison with benchmark descriptors. ImageNet shows average results due to its complex nature of images which falls into different semantic group at the same time. It is however probable that the image is correctly indexed depending upon the individual contents; however on matching the label from the respective category

FIGURE 19. Average Retrieval Recall (ARR): (a) Corel-10,000 Dataset (b) ImageNet Synset (c) Caltech-101 Dataset (d) Caltech-256 Dataset.
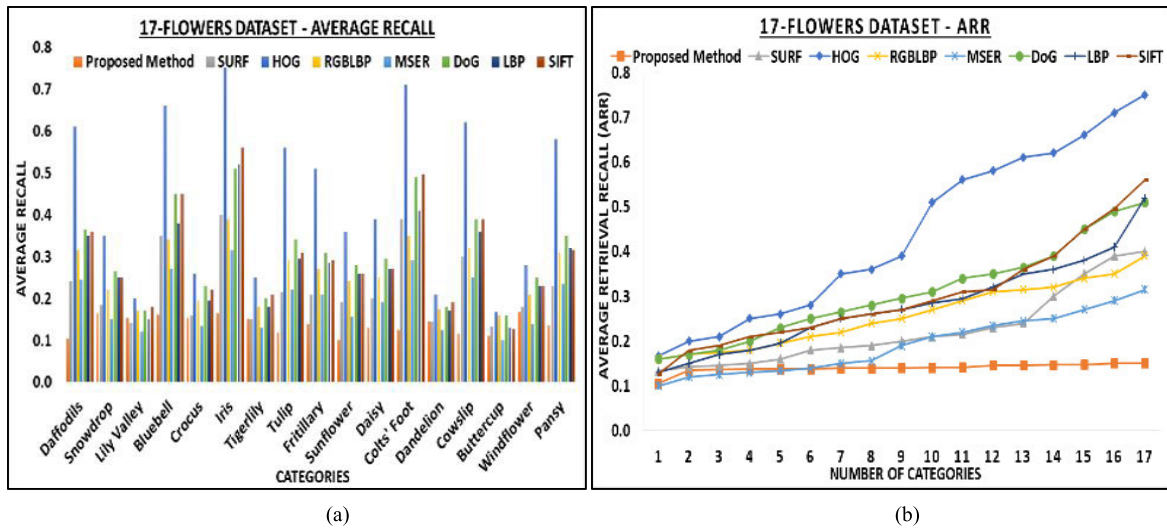
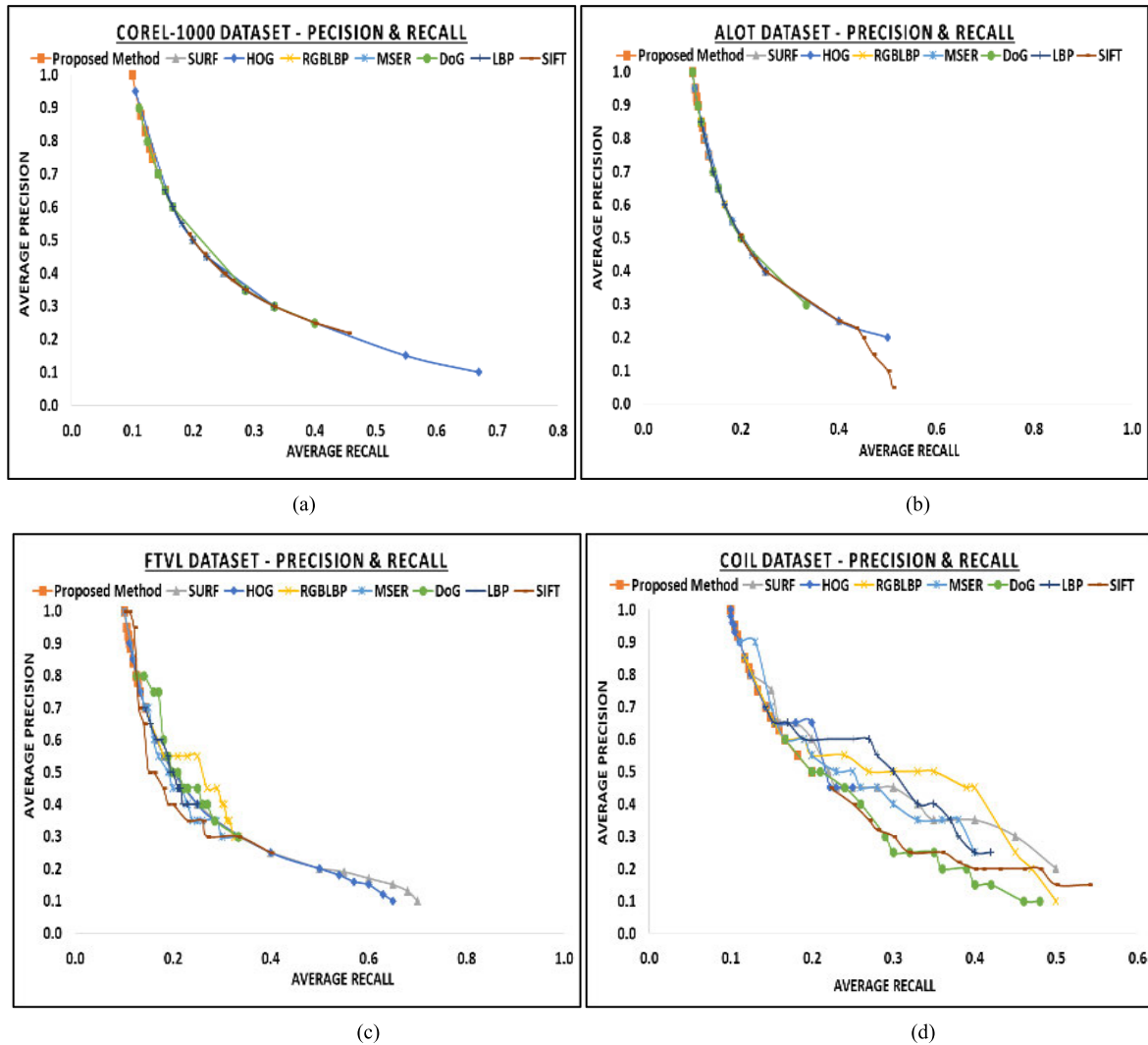**FIGURE 20.** 17-Flowers Dataset: (a) Average Recall: (b) Average Retrieval Recall.

FIGURE 21. Precision & Recall: (a) Corel-1000 Dataset (b) ALOT Dataset (c) FTVL Dataset (d) COIL Dataset.

TABLE 10. ALOT Dataset: Mean Average Recall (%)

Proposed Method	SURF	HOG	RGBLBP
0.10 ± 0.005	0.21 ± 0.001	0.21 ± 0.005	0.14 ± 0.004
MSER	DoG	LBP	SIFT
0.14 ± 0.005	0.22 ± 0.001	0.17 ± 0.005	0.39 ± 0.003

TABLE 11. COIL Dataset: Mean Average Recall (%)

Proposed Method	SURF	HOG	RGBLBP
0.10 ± 0.001	0.23 ± 0.002	0.14 ± 0.001	0.25 ± 0.001
MSER	DoG	LBP	SIFT
0.21 ± 0.001	0.41 ± 0.002	0.22 ± 0.001	0.43 ± 0.001

makes the results false. In our case bag of words approach collects and index the image using KNN approach in which the presented approach returns better results in comparison with challenging descriptors.

The mAP of the proposed method for Caltech-101 dataset is 0.67 as shown in figure 27. HoG reports less than 0.5 mAP rates in most of the categories for this reason HoG reports below average mAP rates. SIFT and MSER report better mAP
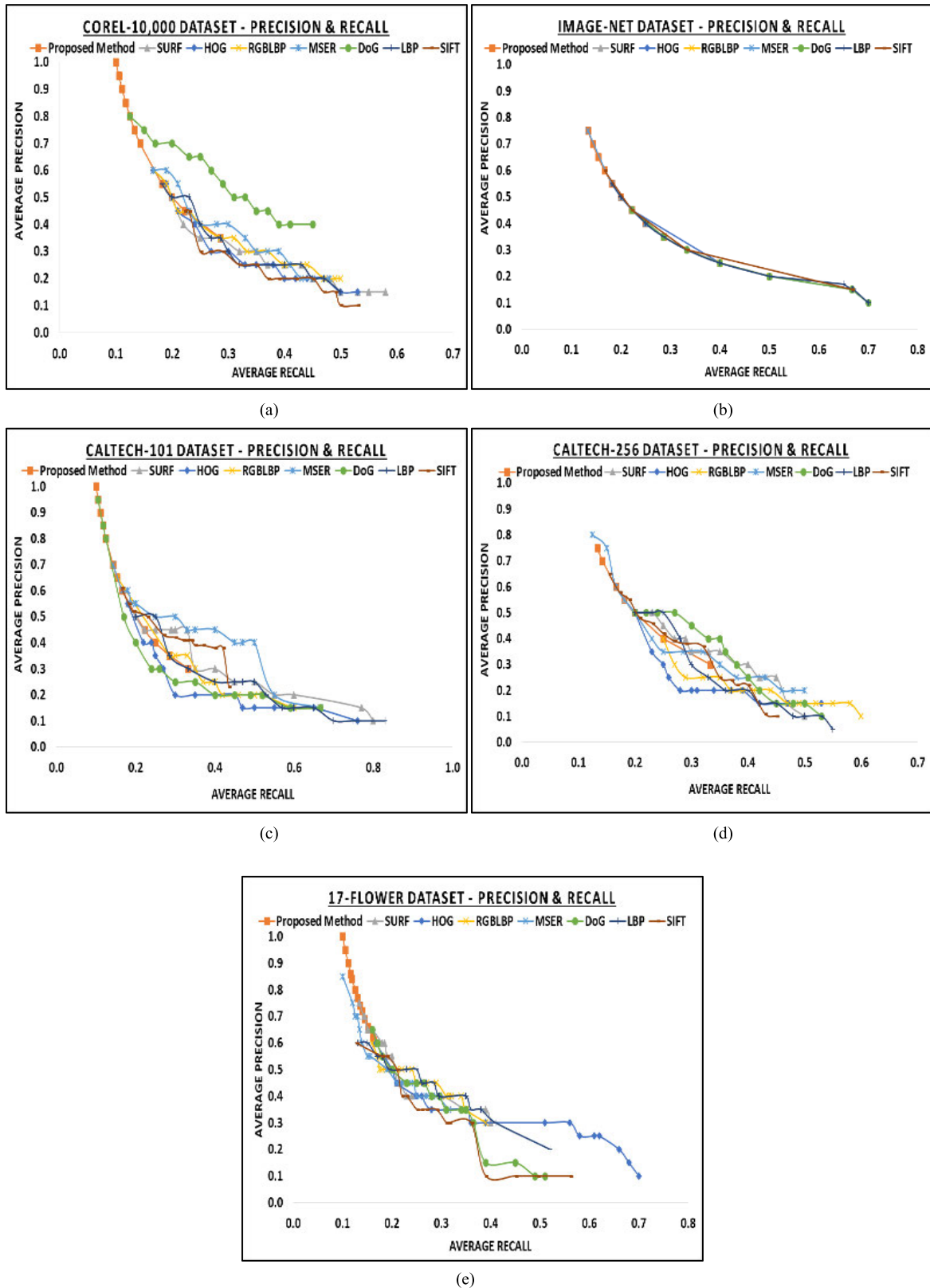
FIGURE 22. (a) Precision & Recall: (a) Corel-10,000 Dataset (b) ImageNet Synset (c) Caltech-101 Dataset (d) Caltech-256 Dataset. (b) Precision & Recall: 17-Flowers Dataset.

TABLE 12. Corel-10000 Dataset: Mean Average Recall (%)

Proposed Method	SURF	HOG	RGBLBP
0.16 ± 0.001	0.43 ± 0.005	0.40 ± 0.001	0.32 ± 0.005
MSER	DoG	LBP	SIFT
0.31 ± 0.001	0.18 ± 0.005	0.36 ± 0.001	0.46 ± 0.001

TABLE 13. ImageNet synset: Mean Average Recall (%)

Proposed Method	SURF	HOG	RGBLBP
0.20 ± 0.001	0.32 ± 0.002	0.49 ± 0.003	0.37 ± 0.001
MSER	DoG	LBP	SIFT
0.23 ± 0.003	0.45 ± 0.001	0.44 ± 0.004	0.24 ± 0.003
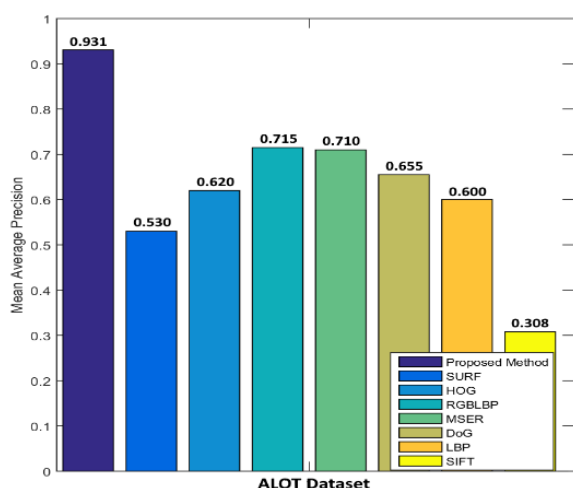
FIGURE 23. mAP: ALOT Dataset.
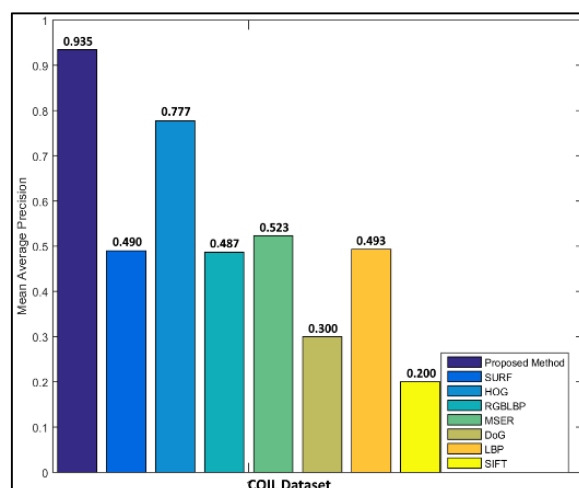
FIGURE 24. mAP: COIL dataset.

rates 0.437 and 0.460 as compared to the other benchmark descriptors as depicted in table 8. HOG, RGBLBP and LBP reports the mAP below than 0.270. Moreover SURF and DoG has mAP in the range of 0.38 that is an average rate as compared to the other methods.

Caltech-256 is results are shown to show the competitive outcomes of the presented technique in contrast with benchmarks. The mAP for the presented approach is 0.50. Other benchmarks show below 50% results. HOG and RGBLBP shows the lowest mAP for Caltech-256 dataset as shown in table 9.

For ALOT dataset, the presented approach returns 0.10 mAR that is better than the competitive benchmarks as shown in figure 29. SIFT shows the lowest mAR rates. Table 10 shows the mAR for the presented approach and the existing benchmarks. It can be concluded that MSER and RGBLBP descriptors report almost equal mAR. Moreover HOG, DoG and LBP report mAR up to 0.17.

For COIL dataset, the mAR of the proposed method is 0.10 as shown in figure 30. HOG shows the second highest mAR that is 0.14. SIFT and DoG report the lowest mAR
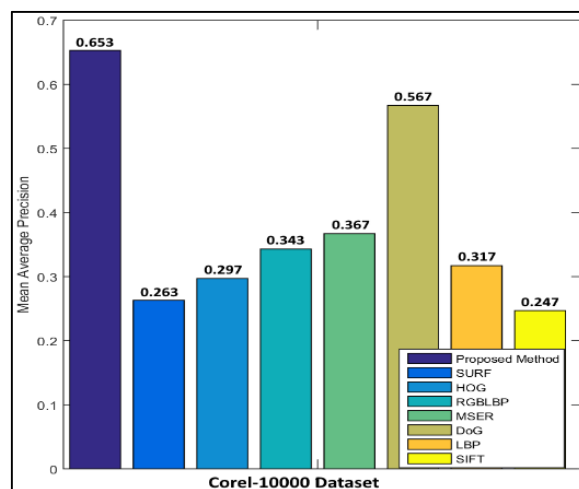

FIGURE 25. mAP: Corel-10,000 dataset.

with 0.43 and 0.41 respectively. Mean average recall of COIL dataset is shown in table 11 for the presented technique and the benchmark descriptors.

TABLE 14. Caltect-101 Dataset: Mean Average Recall (%)

Proposed Method	SURF	HOG	RGBLBP
0.17 ± 0.001	0.35 ± 0.002	0.48 ± 0.001	0.42 ± 0.003
MSER	DoG	LBP	SIFT
0.25 ± 0.001	0.37 ± 0.002	0.47 ± 0.004	0.24 ± 0.001

TABLE 15. Caltect-256 Dataset: Mean Average Recall (%)

Proposed Method	SURF	HOG	RGBLBP
0.21 ± 0.002	0.33 ± 0.001	0.50 ± 0.001	0.51 ± 0.005
MSER	DoG	LBP	SIFT
0.31 ± 0.004	0.38 ± 0.001	0.44 ± 0.001	0.005
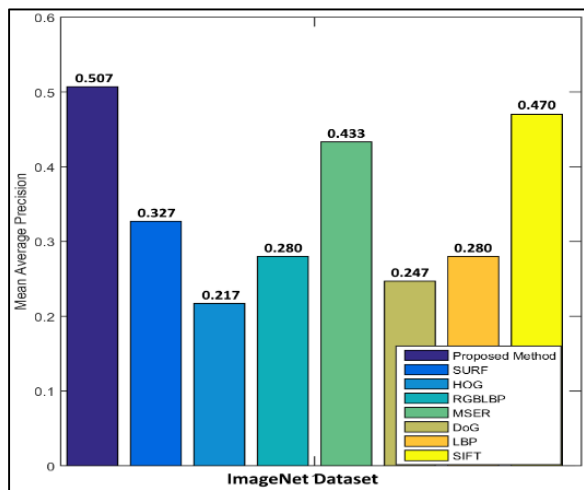
FIGURE 26. mAP: ImageNet synset.
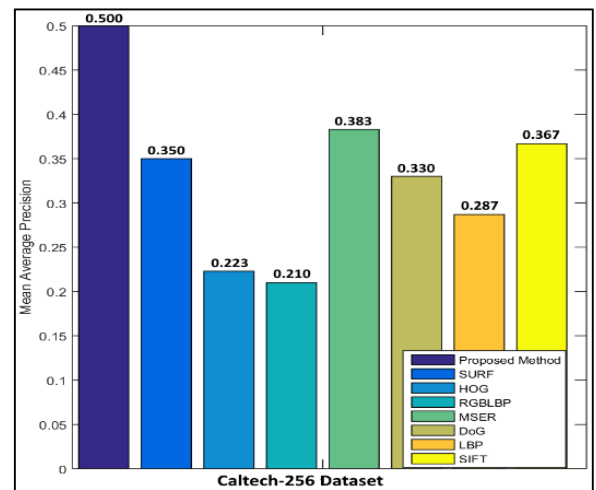
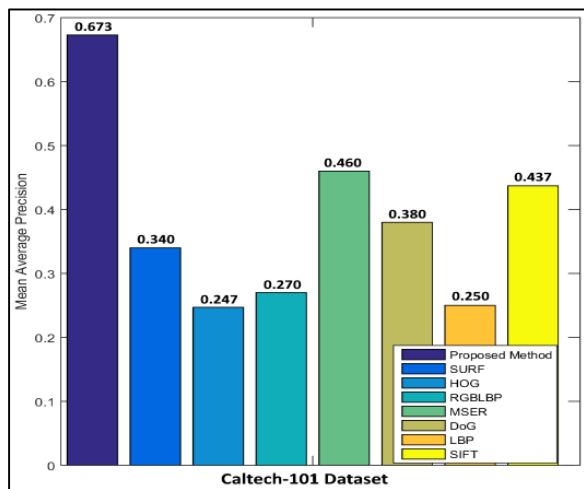**FIGURE 28.** mAP: Caltech-256 dataset.
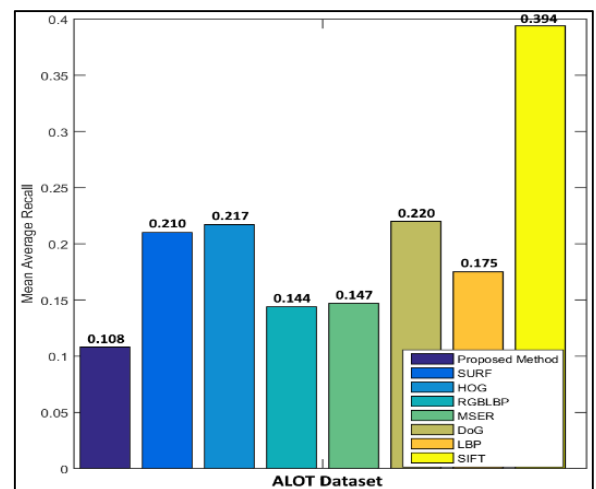

FIGURE 27. mAP: Caltech-101 dataset.

FIGURE 29. mAR: ALOT Dataset.

Figure 31 shows the mAR of the presented approach for Corel-10,000 dataset that is 0.65. Other methods shows much less mAR except DoG. Difference of Gaussian shows the second highest mAR that is 0.56. SURF, HOG and SIFT reports mAR between 0.40 and 0.46 as shown in table 12. Moreover

RGBLBP, MSER and LBP reports 0.32, 0.31 and 0.36 mAR rates.

mAR for ImageNet synset of MSER, SIFT and proposed method is between 0.20 and 0.24. The proposed method shows the highest mAR that is 0.20 as represented
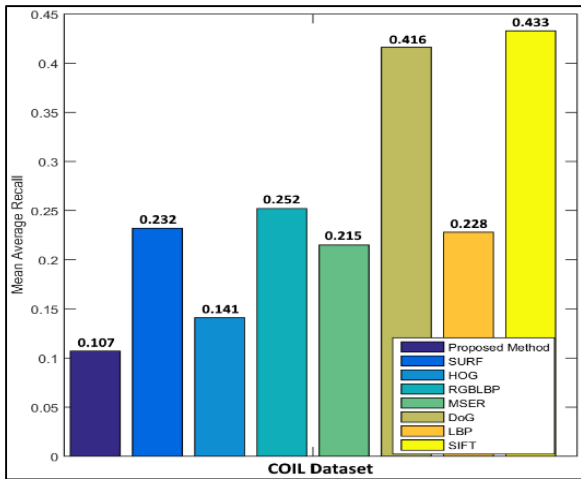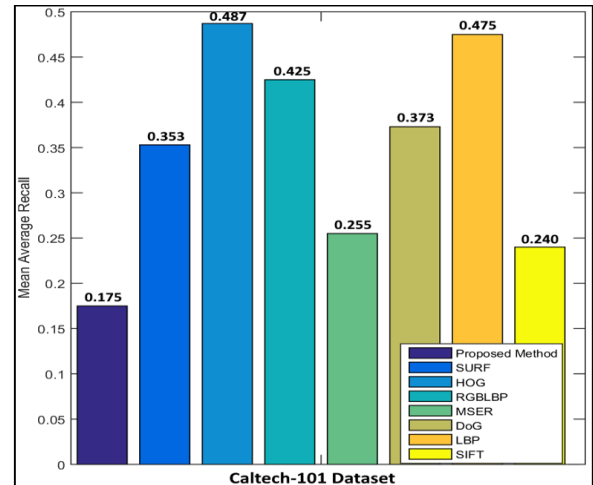
FIGURE 30. mAR: COIL Dataset.
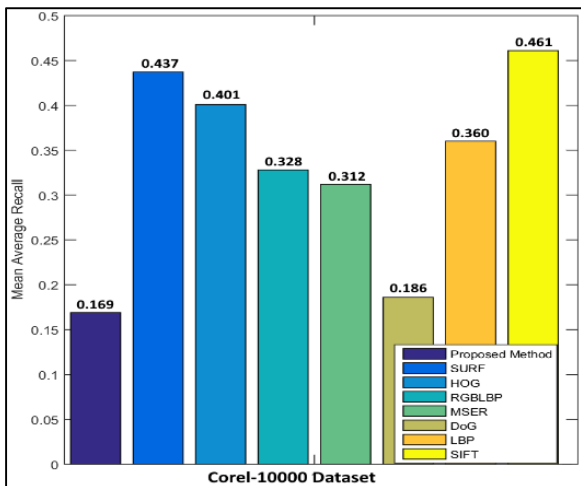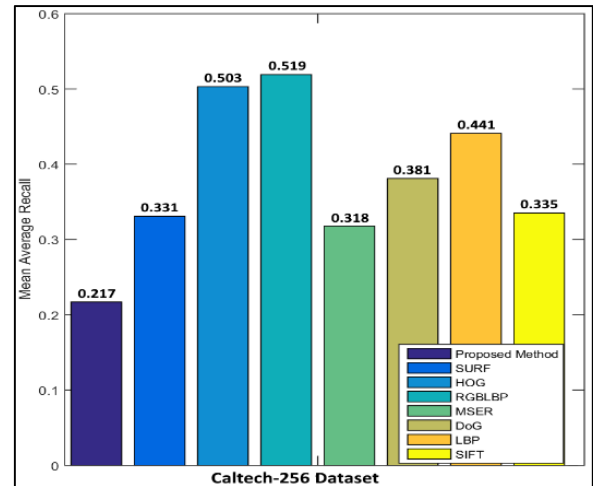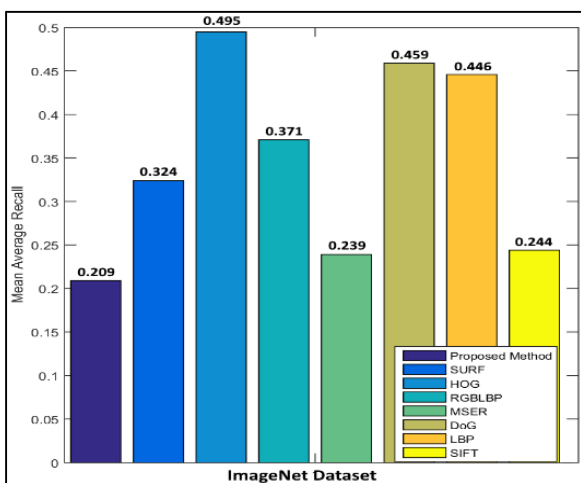

FIGURE 31. mAR: Corel-10,000 Dataset.

FIGURE 32. mAR: ImageNet synset.

in figure 32. Other descriptors show the lowest results since ImageNet synset is comparatively difficult to categorize. Results of all descriptors are shown in table 13.

FIGURE 33. mAR: Caltech-101 dataset.

FIGURE 34. mAR: Caltech-256 dataset.

The mAR of the proposed method obtained for Caltech-101 dataset is 0.17 as shown in figure 33. HoG reports the lowest mAR. SIFT and MSER shows almost the same mAR rates as shown in table 14. HOG, RGBLBP and LBP report up to 0.42 mean average precision. Moreover SURF and DoG report mAR 0.35 and 0.37 respectively.

For Caltech-256 dataset, the results shown in figure 34 are less than 50% mAR for the proposed method. HOG and RGBLBP shows the lowest mAR 0.50 and 0.51 for Caltech-256 dataset as shown in table 15. SURF, SIFT, MSER and DoG shows mAR between 0.31 and 0.38.

V. CONCLUSION

This paper introduced an approach to retrieve the images by discriminatively recognizing the shapes, objects, texture and spatial color information. The described method extracts the spatial image color components, distinctive shape and object features for the cluttered objects and complex background images from diverse image categories. The precision is computed in all respect including average, mean and average retrieval and the same is executed for recall rates against

seven challenging benchmarks. Results are also compared with many others from the latest literate. Experimentation results for nine highly recognized benchmarks showed that the presented approach yielded outstanding performance as compared to the research techniques and benchmark descriptors. Results reported that the fused spatial color and shape features can distinctively retrieve the images from texture, shape, color and object datasets. An extension to the presented work will be the integration of convolution networks to achieve more improved results.

REFERENCES

[1] A. A. Halin, M. Rajeswari, and M. Abbasnejad, "Soccer event detection via collaborative multimodal feature analysis and candidate ranking," *Int. Arab J. Inf. Technol.*, vol. 10, no. 5, pp. 493–502, 2013.

[2] K. T. Ahmed, A. Irtaza, and M. A. Iqbal, "Fusion of local and global features for effective image extraction," *Int. J. Speech Technol.*, vol. 47, no. 2, pp. 526–543, Sep. 2017.

[3] H. Kavitha and M. Sudhamani, "Content based image retrieval-a survey," *Int. J. Adv. Res. Comput. Sci.*, vol. 4, no. 1, pp. 14–20, 2013.

[4] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognit.*, vol. 37, no. 1, pp. 1–19, Jan. 2004.

[5] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognit.*, vol. 45, no. 1, pp. 346–362, Jan. 2012.

[6] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.

[7] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, Manchester, U.K., 1988, pp. 147–151.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[9] K. T. Ahmed and M. A. Iqbal, "Region and texture based effective image extraction," *Cluster Comput.*, vol. 21, no. 1, pp. 493–502, Mar. 2018.

[10] Y. Zhang, "Image retrieval algorithm based on harris-laplace corners and SVM relevance feedback," in *Proc. 7th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Aug. 2016, pp. 337–340.

[11] H. Kavitha and M. V. Sudhamani, "Image retrieval based on object recognition using the harris corner and edge detection technique," in *Proc. Int. Conf. Commun., VLSI Signal Process.*, 2013, pp. 181–184.

[12] A. Iscen, G. Tolias, P.-H. Gosselin, and H. Jegou, "A comparison of dense region detectors for image search and fine-grained classification," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2369–2381, Aug. 2015.

[13] G.-H. Liu, J.-Y. Yang, and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recognit.*, vol. 48, no. 8, pp. 2554–2566, Aug. 2015.

[14] Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 818–832, Feb. 2013.

[15] S. Kaur and M. P. Verma, "Content based image retrieval: Integration of neural networks using speed-up robust feature and SVM," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 1, pp. 243–248, 2015.

[16] B. Shanmugam, "An efficient perceptual of CBIR system using MIL-SVM classification and surf feature extraction," *Int. Arab J. Inf. Technol.*, vol. 14, no. 4, pp. 1–7, 2017.

[17] N. Ali, K. B. Bajwa, R. Sablatnig, S. A. Chatzichristofis, Z. Iqbal, M. Rashid, and H. A. Habib, "A novel image retrieval based on visual words integration of SIFT and SURF," *PLoS ONE*, vol. 11, no. 6, 2016, Art. no. e0157428.

[18] Y.-H. Lee and Y. Kim, "Efficient image retrieval using advanced SURF and DCD on mobile platform," *Multimedia Tools Appl.*, vol. 74, no. 7, pp. 2289–2299, Apr. 2015.

[19] E. G. Karakasis, A. Amanatiadis, A. Gasteratos, and S. A. Chatzichristofis, "Image moment invariants as local features for content based image retrieval using the bag-of-visual-words model," *Pattern Recognit. Lett.*, vol. 55, pp. 22–27, Apr. 2015.

[20] S. Govindaraju and G. P. Ramesh Kumar, "A novel content based medical image retrieval using SURF features," *Indian J. Sci. Technol.*, vol. 9, no. 20, pp. 1–8, 2016.

[21] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 27–35.

[22] S. S. Husain and M. Bober, "Improving large-scale image retrieval through robust aggregation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1783–1796, Sep. 2017.

[23] M. K. Kundu, M. Chowdhury, and S. R. Bulò, "A graph-based relevance feedback mechanism in content-based image retrieval," *Knowl.-Based Syst.*, vol. 73, pp. 254–264, Jan. 2015.

[24] N. Shrivastava and V. Tyagi, "An efficient technique for retrieval of color images in large databases," *Comput. Electr. Eng.*, vol. 46, pp. 314–327, Aug. 2015.

[25] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4018–4032, Sep. 2016.

[26] M. E. ElAlami, "A novel image retrieval model based on the most relevant features," *Knowl.-Based Syst.*, vol. 24, no. 1, pp. 23–32, Feb. 2011.

[27] S. Pan, S. Sun, L. Yang, F. Duan, and A. Guan, "Content retrieval algorithm based on improved HOG," in *Proc. 3rd Int. Conf. Appl. Comput. Inf. Technol./2nd Int. Conf. Comput. Sci. Intell.*, Jul. 2015, pp. 438–441.

[28] M. E. ElAlami, "A new matching strategy for content based image retrieval system," *Appl. Soft Comput.*, vol. 14, pp. 407–418, Jan. 2014.

[29] E. Walia and A. Pal, "Fusion framework for effective color image retrieval," *J. Vis. Commun. Image Represent.*, vol. 25, no. 6, pp. 1335–1348, Aug. 2014.

[30] Y. Xiao, J. Wu, and J. Yuan, "MCENTRIST: A multi-channel feature generation mechanism for scene categorization," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 823–836, Feb. 2014.

[31] A. Irtaza, M. A. Jaffar, E. Aleisa, and T.-S. Choi, "Embedding neural networks for semantic association in content based image retrieval," *Multimedia Tools Appl.*, vol. 72, no. 2, pp. 1911–1931, Sep. 2014.

[32] H. Madzin and R. Zainuddin, "Medical image analysis using local invariant features-interest points," in *Proc. Int. Conf. Softw. Eng. Comput. Syst.*, 2009, pp. 303–307.

[33] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2001, pp. 1–11.

[34] P. Simard, "Boxlets: A fast convolution algorithm for signal processing and neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 571–577.

[35] J. J. Koenderink, "The structure of images," *Biological*, vol. 50, no. 5, pp. 363–370, 1984.

[36] T. Lindeberg, "Scale-space for discrete signals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 3, pp. 234–254, Mar. 1990.

[37] M. Brown and D. Lowe, "Invariant features from interest point groups," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 1–10.

[38] G. Healey, "Using color for geometry-insensitive segmentation," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 6, no. 6, pp. 920–937, Jun. 1989.

[39] G. Healey, "Segmenting images using normalized color," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 1, pp. 64–73, 1st Quart., 1992.

[40] J. R. Kender, *Saturation, Hue, and Normalized Color: Calculation, Digitization Effects, and Use Unknown Binding*. Pittsburgh, PA, USA: Carnegie-Mellon Univ., Dept. of Computer Science, 1976, pp. 1–38.

[41] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Hoboken, NJ, USA: Wiley, 1982.

[42] M. Ali, W. N. Martin, and J. K. Aggarwal, "Color-based computer analysis of aerial photographs," *Comput. Graph. Image Process.*, vol. 9, no. 3, pp. 282–293, Mar. 1979.

[43] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 994–1000.

[44] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Oct. 2005, pp. 1458–1465.

[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[47] G. Griffin and A. P. Holub Perona, "Caltech-256 object category dataset," Tech. Rep., 2007.

[48] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Comput. Vis. Image Understand.*, vol. 106, no. 1, pp. 59–70, Apr. 2007.

[49] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1075–1088, Sep. 2003.

[50] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, 2001.

[51] G. J. Burghouts and J.-M. Geusebroek, "Material-specific adaptation of color invariant features," *Pattern Recognit. Lett.*, vol. 30, no. 3, pp. 306–313, Feb. 2009.

[52] S. Nayar, S. Nene, and H. Murase, "Columbia object image library (coil 100)," Dept. Comput. Sci., Columbia Univ., Tech. Rep. CUCS-006-96, 1996.

[53] M. S. Bhatta and T. P. Pataliab, "Content based image classification in agriculture industry," *Int. J. Control Theory Appl.*, vol. 9, no. 42, pp. 513–520, 2019.

[54] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2006, pp. 1447–1454.

[55] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.

[56] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[57] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[58] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2000, pp. 404–420.

[59] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.

[60] A. Bundy and L. Wallen, "Difference of Gaussians," in *Catalogue of Artificial Intelligence Tools*. Berlin, Germany: Springer, 1984, p. 30.

[61] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Proc. 12th Int. Conf. Pattern Recognit.*, vol. 1, Oct. 1994, pp. 582–585.

[62] S. R. Dubey and A. S. Jalal, "Fruit and vegetable recognition by fusing colour and texture features of the image using machine learning," *Int. J. Appl. Pattern Recognit.*, vol. 2, no. 2, pp. 160–181, 2015.

[63] S. Gao, I. W.-H. Tsang, and Y. Ma, "Learning category-specific dictionary and shared dictionary for fine-grained image categorization," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 623–634, Feb. 2014.

[64] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1794–1801.

[65] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3360–3367.

[66] N. Zhou and J. Fan, "Jointly learning visually correlated dictionaries for large-scale visual recognition applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 715–730, Apr. 2014.

[67] S. Süsstrunk and R. S. Buckley Swen, "Standard RGB color spaces," in *Proc. Color Imag. Conf.* Springfield, VA, USA: Society for Imaging Science and Technology, 1999, pp. 127–134.

[68] Y. Wei, Q. Tian, J. Guo, W. Huang, and J. Cao, "Multi-vehicle detection algorithm through combining Harr and HOG features," *Math. Comput. Simul.*, vol. 155, pp. 130–145, 2019.

[69] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2005, pp. 886–893.

[70] R. Hu and J. Collomosse, "A performance evaluation of gradient field HOG descriptor for sketch based image retrieval," *Comput. Vis. Image Understand.*, vol. 117, no. 7, pp. 790–806, Jul. 2013.

[71] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[72] S. Kim, K.-J. Yoon, and I. S. Kweon, "Object recognition using a generalized robust invariant feature and Gestalt's law of proximity and similarity," *Pattern Recognit.*, vol. 41, no. 2, pp. 726–741, Feb. 2008.

[73] X. Wangming, W. Jin, L. Xinhai, Z. Lei, and S. Gang, "Application of image SIFT features to the context of CBIR," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, Dec. 2008, pp. 552–555.

[74] P. Xu, L. Zhang, K. Yang, and H. Yao, "Nested-SIFT for efficient image matching and retrieval," *IEEE MultimediaMag.*, vol. 20, no. 3, pp. 34–46, Jul. 2013.

KHAWAJA TEHSEEN AHMED received the M.S. degree in computer science from Bahauddin Zakariya Univeristy, Multan, Pakistan, in 2010, and the Ph.D. degree in computer science from the University of Central Punjab (UCP), Lahore, Pakistan, in 2019. He is currently working as an Assistant Processor with the Department of Computer Science, Bahauddin Zakariya University. His research interests include computer vision, deep learning, pattern recognition, image enhancement, and machine learning.

HUMAIRA AFZAL received the M.Sc. degree (Hons.) in computer science from Bahauddin Zakariya University, Multan, Pakistan, in 1997, the M.Sc. degree in computer engineering from CASE, University of Engineering and Technology, Taxila, Pakistan, in 2010, and the Ph.D. degree in computer science from the School of Electrical Engineering and Computer Science, University of Bradford, U.K., in August 2014. She is currently working as an Assistant Professor with the Department of Computer Science, Bahauddin Zakariya University. Her research interests include MAC protocol design for cognitive radio network, performance modeling, queuing theory, network security, software engineering, and sliding mode control.

MUHAMMAD RAFIQ MUFTI received the M.Sc. degree in computer science from Bahauddin Zakariya University, Multan, Pakistan, and the M.Sc. degree in computer engineering from CASE, University of Engineering and Technology, Taxila, Pakistan, in 1994 and 2007, respectively, and the Ph.D. degree in electronic engineering from Mohammad Ali Jinnah University (MAJU), Islamabad, in 2012. He is currently a Faculty Member of COMSATS University Islamabad, Vehari Campus, Pakistan. His research interests include sliding mode control, fractional control, neural network cognitive radio network, and network security.

ARIF MEHMOOD received the Ph.D. degree from the Department of Information and Communication Engineering, Yeungnam University, South Korea, in November 2017. He is currently working as an Assistant Professor with the Department of Computer Science and IT, The Islamia University of Bahawalpur, Pakistan. His recent research interests are related to data mining, mainly working on AI and deep learning-based text mining, and data science management technologies.

GYU SANG CHOI received the Ph.D. degree from the Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA, in 2005. He was a Research Staff Member with the Samsung Advanced Institute of Technology (SAIT) for Samsung Electronics from 2006 to 2009. Since 2009, he has been a Faculty Member with the Department of Information and Communication, Yeungnam University, South Korea. His research areas include non-volatile memory and storage systems.

● ● ●