Fusion of Local and Global Features for Effective Image Extraction

Khawaja Tehseen Ahmed
University of Central Punjab,
Lahore, Pakistan
Email: tehseen@bzu.edu.pk

Aun Irtaza
Department of Computer Science,
University of Engineering and
Technology, Taxila, Pakistan
Email: aun.irtaza@gmail.com

Muhammad Amjad Iqbal,
Faculty of IT,
University of Central Punjab,
Lahore, Pakistan
Email: amjad.iqbal@ucp.edu.pk

Abstract— *Image extraction methods rely on locating interest points and describing feature vectors for these key points. These interest points provide different levels of invariance to the descriptors. The image signature can be described well by the pixel regions that surround the interest points at the local and global levels. This contribution presents a feature descriptor that combines the benefits of local interest point detection with the feature extraction strengths of a fine-tuned sliding window in combination with texture pattern analysis. This process is accomplished with an improved Moravec method using the covariance matrix of the local directional derivatives. These directional derivatives are compared with a scoring factor to identify which features are corners, edges or noise. Located interest point candidates are fetched for the sliding window algorithm to extract robust features. These locally-pointed global features are combined with monotonic invariant uniform local binary patterns that are extracted a priory as part of the proposed method. Extensive experiments and comparisons are conducted on the benchmark ImageNet, Caltech-101, Caltech-256 and Corel-100 datasets and compared with sophisticated methods and state-of-the-art descriptors. The proposed method outperforms the other methods with most of the descriptors and many image categories.*

Keywords— ***Image extraction, interest point detection, image descriptor, principal component coefficients, sliding window, support vector machine.***

I. Introduction

Several studies have contributed to computer vision and rely on object recognition, texture classification, scene understanding, symmetry detection and related domains that are based on detecting interest points, edges and corners for feature description. Images are described by their features to extract useful hidden patterns to produce symbolized signatures at different levels of abstraction and understanding. Different levels of image processing metrics involve different methods of image description and image synthesis. Image description techniques include global, regional and local metrics, and image synthesis uses texture analysis methods.

Texture analysis methods are categorized as statistical, structural, and spectral. *Statistical* methods based on gray level statistical moments describe point pixel area properties, and histograms and scatter plots are used to represent the values. *Structural* techniques use structural primitives, such as parallel lines and regular patterns. *Spectral* methods work in the frequency domain to represent data. Local and global descriptors [1-3] are primitive image descriptors that work in the statistical, structural and spectral domains. Local descriptors describe patches and portions within an image, and global descriptors describe an entire image. Color histograms [7], shape features [8] and textures [9] are used for local feature extraction. However, local features are unable to produce accurate results in different image categories. Global descriptors [1] describe objects for recognition and classification. Local and global features cam be employed together to represent images in a much powerful way. There are several applications of this hybrid scheme for feature extraction, such as the whole-object approach, which uses local interest point detectors, digital correlation, and scale space super-pixels [37]. Another approach is the partial-object method, which is derived from gray level corner detection methods [46], image moments [47], and scale space theory [48]. Interest point descriptors [2,3] are an extension of these approaches that quantify the light intensity, local area gradients, local area statistical features, and the histogram of the local gradient directions. Applications of these extended descriptors have shown better performance in object detection, face recognition, medical image retrieval, and specialized tasks. However, these descriptors typically involve intense computations and require significant memory resources. For image retrieval, these descriptors capture low level image attributes, such as color, texture or spatial information, for optimal performance that is particularly domain specific. Consequently, they result in low performance when the same image descriptor is tested on image categories with complex, overlapping and background objects.

Detectors use *maxima and minima points*, such as gradient peaks and corners; however, edges, ridges, and contours are also considered as key points for better image understanding. For these points, [4] presented an interest points taxonomy that includes intensity-based region methods (IBRs), edge-based region methods (EBRs) and shape-based regions (SBRs). Features were extracted by pixel intensity based on the saturation value of the pixel in [5]. Image retrieval is performed by applying this feature model to image segmentation and histogram generation. Image detectors extract features with a diverse invariance to occlusion, rotation, illumination and scale. These feature detectors are employed in image classification, object recognition, and semantic interpretation based on their specialty. The quality and effectiveness of interest point detection methods were evaluated against standard databases and state-of-the-art methods in [6].

This contribution uses local features along with global feature description by combining texture values to extract images from multiple categories. Useful image patterns are detected by finding edges and corners based on local interest points. These key points are identified using pixel intensities. Pixel intensity-based detectors are more powerful interest point detectors than other methods [6]. A fine-tuned sliding window algorithm is applied to the interest points to extract the image signatures. The texture analysis results are combined with the signatures to comprehensively reflect the image patterns. A novel dimension reduction technique is used to calculate the limited covariant coefficients. The proposed method provides remarkable results on benchmarks, existing methods, renowned databases and state-of-the-art descriptors.

The remainder of this paper is organized as follows. Section 2 presents related work on feature extraction, and Section 3 explains the proposed methodology. The experimental results are provided in Section 4, and we summarize our findings in Section 5.

II. Related Work

A significant amount of research has been performed on Content Based Image Retrieval (CBIR) by analyzing interest points that are composed of corners, edges, contours, maxima shapes, ridges or global features, visual contents and semantic interpretation. These detectors [1], descriptors [2] or extractors [3] can be characterized as invariant or covariant, local or global. Local features are specific and context oriented. Current Content Based Image Retrieval (CBIR) systems require image retrieval from versatile image categories, images with complex overlapping objects, cluttered images, and foreground and background objects. Solutions are normally tested on a specific dataset or selected categories, and the results are uncertain for other benchmarks. A combination of global and local features that uses the Haar discrete wavelet transform (HDWT) and gray level co-occurrence matrix (GLCM) was presented [64], and the results were computed for the Corel-100 dataset [14]. In another image retrieval scheme [17], LBPs are collected and combined from each channel to describe color images. Decoded LBPs are introduced to reduce the highly dimensional patterns that are returned from multiple channels. Experiments were performed on Corel-1k and other benchmarks. The reported precision for the Corel-1k benchmark was 0.749. A computationally practical approach for capturing image properties from two multichannel images was contributed by [18]. Tradeoffs were executed at the feature and channel levels to avoid redundant image information, and the mean average precision for the Corel-1k benchmark was 0.709 [17]. For image retrieval, discrete cubic partitioning of the image was performed in the HSV space [19]. The data were then hierarchical mapped using the hierarchical operator, and a similarity-based ranking scheme was used for the resultant features. A Mean Average Precision (mAP) of 0.797 was reported for the Corel-100 dataset. A three stage method was proposed to identify similar images by first finding the images by their color features [20]. To improve the results, the images are matched by their texture and shape features. This method accumulates global and regional features for better accuracy. The reported precision for the Corel-1k benchmark is 0.766. In [21], images were abstracted based on their statistical features. The Non-subsampled Contourlet Transform (NSCT) was used to compute the features of this Multi-scale Geometric Analysis (MGA). A graph-theoretic approach-based relevance feedback system was also incorporated for retrieval performance. A mAP of 0.553 was reported for this technique for the Corel-1000 benchmark. A method was presented to characterize an image as a generalized histogram quantized by Gaussian Mixture Models (GMMs) [22]. This method learns from training images using the Expectation-Maximization (EM) algorithm, and the number of quantized color bins is determined by the Bayesian Information Criterion (BIC). The method gave a mAP of 0.801 for the Corel image dataset. Color, texture and shape information was incorporated using the Color Difference Histogram (CDH) and Angular Radial Transform (ART) features [23]. The mAP using min-max normalization on the Corel-1k benchmark was 0.783. Histograms of triangles were used to add spatial information to the inverted index of a bag-of-features by [24]. An image was divided into two and four triangles that were evaluated separately. Experiments were performed on the Corel-1000 dataset with an average precision of 0.82. The color co-occurrence matrix (CCM) and the difference between pixels of scan pattern (DBPSP) were used to extract color and texture features [25]. To eliminate redundant features, selective features were chosen by finding their high dependency on the target class. This approach reported a mean average precision of 0.757 for the Corel-1000 dataset. A content-based image retrieval approach was presented in [65] for biometric security based on color histogram, texture and moment invariants. Color histograms were used for color features, a Gabor filter was used for the texture features, and the moment invariants were used for shape information. This approach reported improved results for biometric security. A method for CBIR using Local Binary Pattern (LBP), Hu-moments and radial Chebyshev moments by focusing shapes and textures was presented in [66]. Ten categories from the COIL dataset [67] were used for experiments, and the method reported a 3% higher accuracy than previous results. A method to retrieve images using color features by dividing images into non-overlapping blocks and to determine the dominant color of each block using the k-means algorithm was presented in [68]. A gray-level co-occurrence matrix was used for texture feature extraction, and Fourier descriptors were extracted from the segmented images for the shape representation. The final feature vector was composed of these extracted features. The results of experiments performed on the Corel-1000 dataset [14] were compared with the results of histogram-based methods. An 8% improvement in precision was achieved with a 4.5 second retrieval time. A descriptor that adds spatial distribution information of the gray-level variation between pixels in LBP for image retrieval was presented in [26]. This spatial texture descriptor constructs statistic histograms of pattern pairs between the reference pixel and its neighboring pixels. Spatial information combined with texture features produced relatively effective results. A descriptor based on shape and texture features was presented in [27] by employing the Discrete Wavelet Transform (DWT) and Edge Histogram Descriptor (EHD) features of MPEG-7. The wavelet coefficients were calculated for the input image, and the Edge Histogram Descriptor was then used on the coefficients to determine the dominant edge orientations. This combination of DWT and EHD was tested on the Corel-1000 dataset.

HOG [1], SIFT [2] and SURF [3] are interest point detectors and image descriptors that are used in combination with local and global descriptors for content-based image retrieval. The time and the computational costs are barriers to using these famous descriptors in CBIR systems with complex and cluttered images of different sizes. However, dimension reduction techniques are employed to overcome the computation time constraint. A multilayer feed forward neural network-based CBIR system that incorporates the strength of SIFT for object detection was introduced by [28]. SIFT object detection was used for CBIR by reducing the large number of key points generated by SIFT to improve the retrieval performance [29]. Salient image parts were extracted by a saliency-based region detection system, and the final results were tested on VOC2005. A CBIR system that integrates the yc_bc_r color histogram, edge histogram and shape descriptor as a global descriptor with surf salient points using SURF as a local descriptor to enhance the retrieval results was proposed by [30]. Experiments were performed on the Corel-1000 and the Uncompressed Color Image Database (UCID) databases. Velmurugan et al. [31] combined SURF with color moments by calculating the first and second order color moments for SURF key points, and experiments were performed on the COIL-100 dataset. The Histograms of Oriented Gradients feature has been used in pedestrian detection [32], face recognition [33] and object detection [34]. For CBIR, Shujing et al. [35] used HOG by transforming the sizes of images and calculated a feature vector of 3780 dimensions. Orthogonal lower dimension data were achieved by applying PCA, and experiments were performed on the Corel-1000 image dataset. A CBIR called Local Tetra Patterns (LTrPs) was proposed by Murala et al. [36] by calculating the first order derivatives in the vertical and horizontal directions on reference pixels and their neighbors. The performance on benchmark databases was compared by combining this method with the Gabor transform.

The technique presented in this paper focuses on: 1) finding suitable key points to produce useful feature sets to effectively classify images from multiple categories with remarkable precision; 2) identifying foreground and background objects in complex images for better accuracy; and 3) introducing a new mechanism to search images by their local and global features with low computational cost and by storing and comparing compact image signatures for efficient retrieval. In the proposed method, corners are detected by pixel intensities to avoid unwanted key points. A fine-tuned sliding window technique returns identifiable features for robust classification, and a useful texture patterns analysis supports the proposed method to provide more accurate results.

III. Methodology

A. *Intensity-based local interest point detection*

The first corner detector was introduced by Moravec [49]. It returns points with local maxima of the directional variance measure and determines the average change in intensity by moving a local preset detection window in different directions. This idea was also employed in [50] to investigate the local statistics of the variations in the directional image intensity using first order derivatives. This method results in better subpixel precision and provides better localization and corner detection. Our method uses the approach of Moravec [49] by expanding the average intensity variance and computing the Sobel derivatives and Gaussian window.

First, intensity-based local interest points are detected. Local features provide identifiable and localized interest points. An anchor point can be a point on a curve, the end of a line and a corner. It can also be an identified point of local intensity that has the maximum curvature of the points on the curve. An auto-correlation matrix best describes the local image features and structure. The following matrix describes the gradient distribution in the local neighborhood of an interest point:

$$M=\sigma^{2}{}^{D}g(\sigma)*\begin{bmatrix} I_x^2(X,\sigma D) & I_xI_y(X,\sigma D) \\ I_xI_y(X,\sigma D) & I_x^2(X,\sigma D) \end{bmatrix} \qquad (1)$$

with,

$$I_x(X,\sigma D)=\frac{\partial}{\partial x}g(\sigma D)*I(x)$$

$$g(\sigma)=\frac{1}{2\Pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (3)$$

Local image derivatives are computed with Gaussian kernels of scale σ D [37]. In the neighborhood of a point, the derivatives are averaged using a Gaussian window. The eigenvalues determine the principal signal changes in both orthogonal directions in the neighborhood of the point σ I. Therefore, corners are found when the image signals vary or the eigenvalues are large. Harris [37] proposed a less computationally expensive metric that uses two eigenvalues.

B. *Global feature detection using an optimized sliding window*

For image classification, the entire image is of interest for global features. Global feature computations for an entire image have large time and computational costs. Local features describe image patches around interest points, while global features describe an image as a single vector. With an increasing number of local features, large numbers of feature vectors are generated, which are difficult to match and store. To overcome these problems, we used local and global features in an intuitive way to compute the global features only for the detected local features of interest.

A sliding window slides a fixed size frame across an image. The object's size, location, positioning and scaling are directly impacted by the block size, cell size within the block, orientation angle and block overlap. The optimized values for these parameters correctly classify an image. Our technique tunes the sliding window technique against these parameters for the datasets. This optimized sliding window extracts feature vectors for the detected intensity-based interest points. For the global feature detection using the local intensity points, discrete values of quantized visual features are represented in histograms. Pixel edges of 8 bin histograms are used as cells.

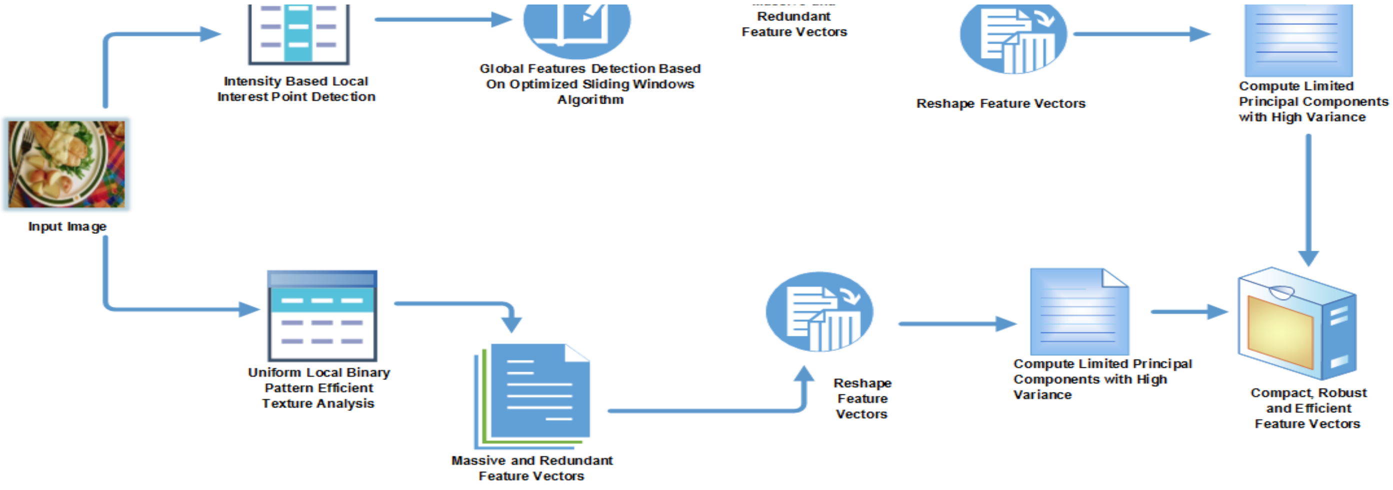

Intensity Based Local Interest Point Detection
Global Features Detection Based On Optimized Sliding Windows Algorithm
Redundant Feature Vectors
Reshape Feature Vectors
Compute Limited Principal Components with High Variance
Input Image
Uniform Local Binary Pattern Efficient Texture Analysis
Massive and Redundant Feature Vectors
Reshape Feature Vectors
Compute Limited Principal Components with High Variance
Compact, Robust and Efficient Feature Vectors

Figure 1: Proposed method showing the step-by-step feature extraction process for an input image.

The edge magnitude and orientation are computed using a first order Sobel kernel filter. Histograms are constructed using the following equation, where y denotes bins, and z denotes the cells:

$$h(y,z)=\sum_{x\in z}\begin{vmatrix} \|\nabla I(x)\| \ if \ \frac{\Theta(x)}{T}=y \\ 0 \ else \end{vmatrix}$$

(4)

where $\Theta(x)$ is the orientation of the edge, and $\|\nabla I(x)\|$ is the magnitude. A histogram of the gradient orientation is computed for each cell. Histogram normalization is then performed by accumulating the local histograms for each block and applying them over all of the cells in the block. A cell size of 4×4 is used to capture the small-scale spatial information. Significant local pixel information is captured by using a block size of 2×2. This helps to express local illumination changes that are not lost when the small block size is averaged. Parts of the overlapping adjacent blocks produce better contrast normalization. To produce non-massive feature vectors, an overlap of half of the block size is used. An increased number of orientation bins results in large feature vectors, so our optimized technique uses 7-9 bins, which produce a relatively small number of feature vectors with respect to the number of blocks and the cell size. Better orientation results are not observed when values are evenly spaced between 180 degrees, so in our technique, values are placed between 0 and 180 degrees by placing minus values in the positive bins.

C. Efficient uniform local binary pattern texture analysis

An operator that is invariant to monotonic transformations of the gray scale is used for the texture analysis. It works on a circularly symmetric neighboring 3×3 set of eight pixels. The strength of the technique is that it uses a limited subset of pixels to achieve computation simplicity. The selected subset of pixels is 'uniform' in patterns, which results in rotational invariance. The fewer spatial transitions employed in uniform patterns are more tolerable to rotation changes [40]. In this process, a local 3×3 neighborhood with a gray level distribution of nine pixels is selected first. In the circularly symmetric neighborhood of eight pixels, the gray values of the diagonal pixels are computed by interpolation. To achieve gray scale invariance, the gray values of these eight pixels are subtracted from the center pixel. This can be calculated by:

$$T= p(v_0, v_1\text{-}v_0, v_2\text{-}v_0, v_3\text{-}v_0, v_4\text{-}v_0, v_5\text{-}v_0, v_6\text{-}v_0, v_7\text{-}v_0, v_8\text{-}v_0) \quad (5)$$

where *T* is the texture, *p* is the pixel, and *v* is the gray scale value.

The center pixel v_0 contains the overall luminance of the image, which is not required for the local texture analysis; therefore, it is discarded for the gray level distribution calculation [38]. Invariance is easily achieved if the sign of the gray scale difference is only noted for the pixel pairs to be differentiated. The minor change to the LBP [39] can be illustrated as [40]:

$$LBP_8=\sum_{i=1}^{8} sign(v_i-v_0)2^{i-1}$$

(6)

LBP$_8$ and the traditional LBP [41] have different indexed neighbors and interpolated diagonal values. Both differences form a base of the rotational invariance of LBP$_8$. The circular symmetric neighborhood set of eight pixels produces 2^8 outputs. The gray pixel value is moved along the perimeter. A right rotation operation is performed on the pixel values so the bits values have a maximum of zero starting from the eighth bit. This process is formulated as [40]:

$$LBP_8^{ri36} = \min\{ROR(LBP_8, i) \mid i = 0,1,\ldots,7\} \quad (7)$$

As mentioned in [42], these 36 unique rotation invariant local binary patterns carry micro-scale features; for example, the least significant bit shows a bright spot, the most significant bit shows a dark spot, and diagonals show edges. Moreover, in some cases, suboptimal results are observed due to rotated values at 45°. A large number of spatial transactions occurs when the uniformity value is large, so considering this dependency, a value of 2 is used for uniformity. Thus, LBP$_8^{ri36}$ is reformed as:

$$LBP_8^{riu\,2} = \sum_{i=1}^{8} sign(v_i - v_0) \;\; if \;\; U(LBP_8) \le 2 \quad (8)$$

Similarly LBP$_{16}$ is calculated as:

$$LBP_{16} = \sum_{i=1}^{16} sign(v_i - v_0)\, 2^{i-1} \quad (9)$$

A distribution of 16 bits along the perimeter produces 65,536 output values, which contain 243 different patterns. Similarly, defining a uniform rotation produces invariant patterns of the 16 bit version:

$$LBP_{16}^{riu\,2} = \sum_{i=1}^{16} sign(v_i - v_0) \;\; if \;\; U(LBP_{16}) \le 2$$

(10)

Mapping of LBP16 to LBP16 riu2 is performed using a lookup table.

D. Generation of feature vectors

a) The sliding window extracts feature vectors for the intensity-based detected local interest points. The interest points are edges and corners, which are a limited set of the image pixels. Feature extraction is performed on these interest points. The global extraction depends on the number of bins, the number of overlapping blocks, and the cell and block sizes. These generated values are very small compared to the signatures generated by state-of-the-art descriptors [2,3], which impacts the computational time for large databases. These points are extracted once for each image, but this process takes a long time if they need to be used for classification and image retrieval purposes. It can also take a long time to access these feature vectors if they are stored and retrieved.

b) Similar feature vector strengths are produced using texture analysis with a uniform local binary pattern, as was described in C-III.

Steps *a)* and *b)* represent the image with distinct features. Their concatenation results in better classification, but it is computationally and time intensive. Therefore, these feature vectors must be reduced with minimum information loss before concatenation for efficient processing.

E. Feature vector optimization algorithm and calculating the principal components

The concatenated feature vectors contain useful information about the images. The coefficients returned by principal component analysis are often large enough to utilize high computational power. This is even more crucial in large database scenarios. The returned coefficients are directly proportional to the number of observations. The descriptors [1,2,3] return voluminous rows and columns, which provide a base for the coefficient calculations. A smaller set of principal components can be obtained by limiting the number of observations. By row elimination, the signature subset or similar approaches result in fewer observations but significant information loss of the original data, which cause worse prediction and classification results.

The algorithm in our technique is trained for different image datasets, image dimensions, image resolutions, pixel intensities and image formats. The algorithm primarily checks all of the required and related information for the dataset on which the experiments are performed. After this examination, the program has an understanding of the images. The algorithm takes the feature vectors as inputs, and they are reshaped based on the optimal number of observations as inputs for the coefficient calculations before computing the principal components. The number of observations is optimized based on the image attributes that were described previously (e.g., dimension, resolution). The reshaped feature vectors are then input to the principal component calculation and result in fewer coefficients with large variances. These limited principal components identify the feature vectors comprehensively with very little information loss. The novelty of the technique for limited coefficients is that it provides nearly the same precision as using the complete set of feature vectors without subset optimization. The feature vectors produced by the texture analysis are optimized using the feature vector optimization algorithm on which the principal components are computed. The texture information additionally carries monotonic and rotational invariance characteristics, which can perform better prediction along with the global features. To achieve this, the texture features are concatenated with intensity-based interest points extracted with the sliding window. The coefficient-based reduced, optimized and concatenated feature vectors contain texture and object recognition capabilities for simple, overlapping and complex images. The image retrieval time is also reduced due to the slim size of the feature vectors. In cases of image retrieval from thousands to millions of images, these compact signatures are efficient as well as storage friendly.

The number of coefficients generated by PCA is proportional to the dimensions of the input feature. The feature vectors returned by different local and global methods vary in size. These hybrid feature vectors produce a large number of principal components. The number of PCs varies from hundreds to thousands per image. Even after the reduction process, classification of such a large number of feature vectors is time intensive. For the datasets, our algorithm reorders the feature vectors and generates between 80 and 120 observations depending on several characteristics, such as the size and type of image attributes and the pixel intensities. Based on the number of observations, each image is represented by 80-120 coefficients. This image signature size is very small compared to the output produced by descriptors [1,2] or the standard PCA.

F. Image classification using Support Vector Machine (SVM)

A Support Vector Machine is a discriminative classifier that separates two classes of points by a hyperplane. It is a supervised learning model that analyzes data that are used for classification and regression analysis tasks.

Let the input belong to one of two classes as [51]:

$$\{(x_i, \mathrm{y}_i)\}_{i=1}^{N} yi = \{+1, -1\} \qquad (11)$$

where x_i is the input set, and y_i are the corresponding labels. Hyperplanes are assigned the values of the weight vectors '*w*' and bias '*b*' as follows:

$$w^T .\ x + b = 0 \qquad (12)$$

and a maximum margin of $2/\|w\|$ hyperplanes are found such that the two classes can be separated from each other; i.e.,

$$w^T .\ x_i + b \geq +1 \qquad (13)$$

$$w^T .\ x_i + b \leq -1 \qquad (14)$$

or

$$y_i(w^T .\ x + b) \geq +1 \qquad (15)$$

The kernel version solution of the Wolfe dual problem is then found with the Lagrangian multiplied by α_i:

$$Q(\alpha) = \sum_{i=1}^{m} \alpha_i - \sum_{i\ j=1}^{m} \alpha_i \alpha_j\ y_i y_j\ K(x_i . x_j)/2 \qquad (16)$$

where $\propto_i \geq 0$, and $\sum_{i=1}^{m} \propto_i y_i = 0$.

Based on the kernel function, the SVM classifier is given by:

$$F(x) = Sgn(f(x)) \qquad (17)$$

where $f(x) = \sum_{i=1}^{l} \propto_i y_i K(x_i, x) + b$ is the output hyperplane decision function of the SVM. High values of *f(x)* represent high prediction confidence, and low values of *f(x)* represent low prediction confidence.

IV. Experimentation

A. Datasets

Most of the datasets are tailored for custom tasks depending on the nature of the project. Many contributions use domain-based image types. Experiments performed on a dataset are difficult to compare with those performed on another dataset. The accuracy of the results is directly affected by the image attributes, such as color, object location, quality, size, overlapping, occlusion, and cluttering [43]. In our case, widely-used datasets and their respective categories are selected by considering the following characteristics:

- Diverse image categories
- General content-based image retrieval usage
- Categories contain many types of textures, foreground and background objects, colors and spatial features
- Images from different areas of life to test the descriptor's effectiveness

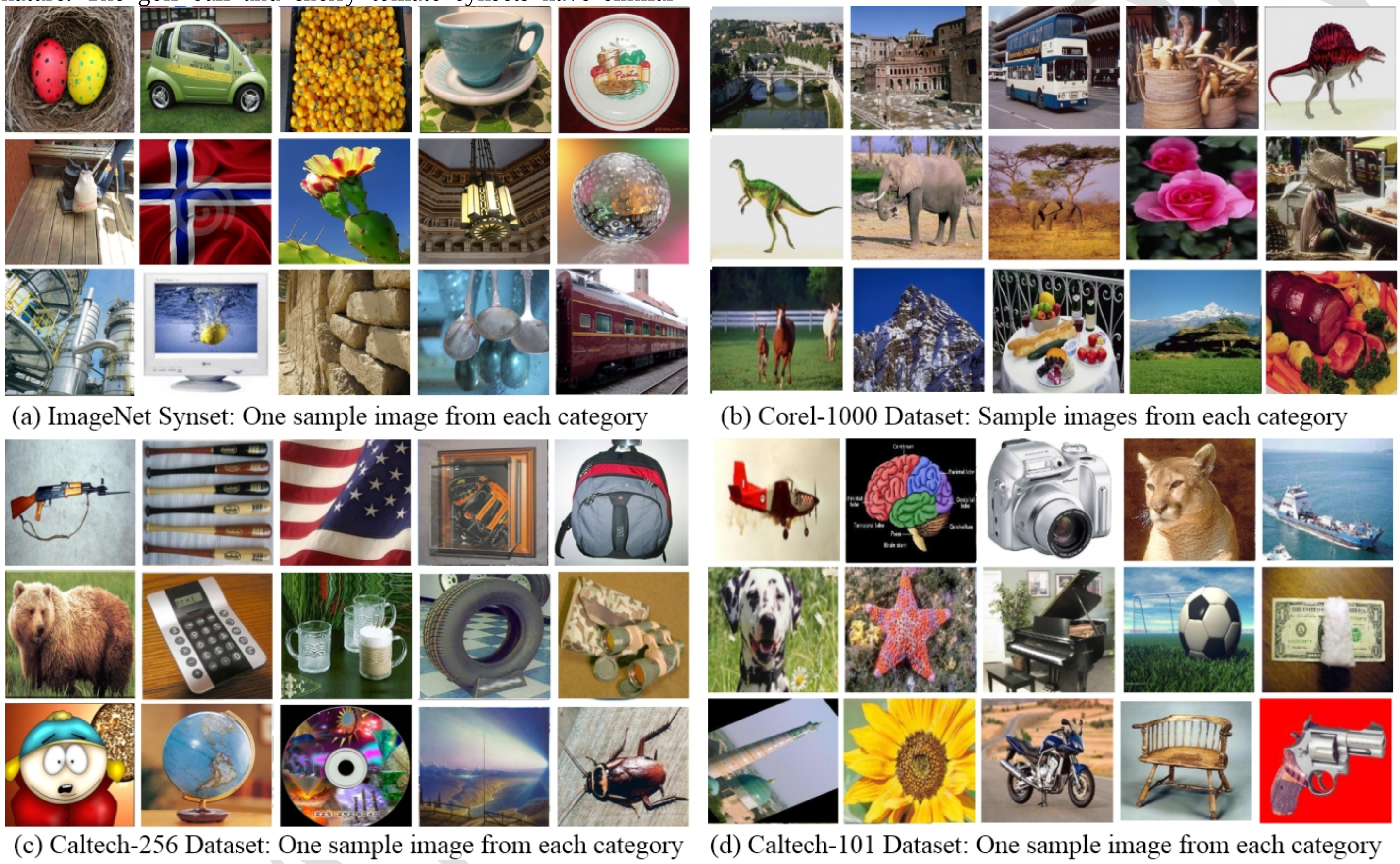
The selected subsets are representative of the respective datasets and include diverse categories from different areas, object orientations, shapes and textures, and global and spatial information. Therefore, the results that are based on the selected categories are representative of the entire dataset. Experiments are performed on a variety of standardized datasets, including ImageNet [13], Caltech-256 [16], Caltech-101 [15] and Corel-1000 [14]. The sampling, object categories, and image characteristics of each category are described below.

1) Corel-1000 dataset

The Corel-1000 dataset is a benchmark that is widely used in the literature for classification tasks [17-20,44]. The Corel database includes many semantic groups, including scene, nature, people, flowers, animal, and food. It consists of 1000 images in 10 categories. Each semantic category consists of 100 images with a resolution of 256 × 384 pixels or 384 × 256 pixels. Our algorithm randomly selects 70% of the images from each category for training and 30% for testing. A total of 660 images from all of the categories is used for training, and 330 images are used for testing.

2) ImageNet Synset

The ImageNet synset [13] is a large-scale image database that is used to index, retrieve, organize and annotate multimedia data. It is organized based on the WordNet hierarchy. Each meaningful concept in WordNet that can be described by multiple words or word phrases is called a synset. WordNet contains more than 100,000 synsets, which are dominated by nouns (80,000+). The repository contains an enormous collection of more than 14,197,120 images. Experiments were performed on 15 synsets downloaded from the ImageNet repository [13], including aerie, car, cherry tomato, coffee cup, dish, dust bag, flag, flower, gas fixture, golf ball, heat exchanger, monitor, mud ceramics, spoon and train. These synsets were selected from the semantic groups of plants, natural objects, artifacts, devices, containers, ceramics, arms, and equipment. These synsets were selected due to their versatility, textures, complexity and object orientation features. The cherry tomato contains small and medium foreground and background objects as well as overlapping objects. The flag synset contains specific color-oriented objects; in other words, color and texture are both used to classify this category. The gas fixture and aerie synsets are complex and cluttered object categories. Both contain spatial information due to their hanging nature. The golf ball and cherry tomato synsets have similar round object orientations with color differences. It is challenging to distinguish these synsets. The artifacts group contains structural complexities with semantics associations; therefore, classification of this synset requires careful analysis of local and global features. The equipment and devices groups are sometimes semantically the same. However, the object- and texture-based training leads to better classification. These 15 synsets contain 13,554 images, from which 100 images were randomly selected from each synset for the experiments; i.e., 1500 (100×15) images were used for training and testing.



(a) ImageNet Synset: One sample image from each category

(b) Corel-1000 Dataset: Sample images from each category

(c) Caltech-256 Dataset: One sample image from each category

(d) Caltech-101 Dataset: One sample image from each category

Figure 2: (a) ImageNet Synsets with 15 image samples (one image from each category). (b) Corel-1000 dataset showing15 sample images from 10 categories. (c) Caltech-256 dataset showing 15 sample images from 15 categories (one image per category). (d) Caltech-101 dataset showing 15 sample images from 15 categories (one image per category).

A total of 1050 images were used for training, and 450 were used for testing, by randomly selecting them from each category. The positive training samples included two-thirds of the candidates, which were randomly selected from each category. The negative training samples included one-third of the total.

1) *Caltech-256 dataset*

This dataset is a challenging set of 256 object categories that contain 30,607 images [16]. It is a successor to the Caltech-101 dataset. Image classification in Caltech-256 is more difficult than in Caltech-101 [15] because it has more variations. We performed experiments on 15 diverse categories, including AK47, American flag, backpack, baseball bat, baseball glove, bear, mug, binocular, calculator, car tire, Cartman, CD, cockroach, desk globe and comet. The semantic groups were selected carefully to represent many areas of real life. These categories contain animals, flags, guns, accessories, tires, insects, computer accessories, daily used entities and images with complex and overlapping objects. Some of the categories are important because of their texture patterns, whereas others are important because of their foreground and background objects. The desk globe, car tire and CD are round objects. Their classifications are based on their orientations and textures. The cockroach was selected from the insect category. Recognizing an insect in an image requires the technique to have object recognition capabilities. The American flag contains specific color and texture information that can be used to classify it. Cartman and the binocular are normally in complex backgrounds and contain overlapping objects. A total of 1050 images were used for the experiments by selecting 70 images per category. Our algorithm randomly selects 70% of the images from each category for training and 30% of the images for testing. A total of 735 images from all of the categories were used for training, and 315 were used for testing. In the training phase, positive samples are chosen randomly from the respective category. Of each category, 70% is used for positive training, and the remaining 30% are negative training samples. The negative training samples are gathered randomly from the rest of the categories by selecting an equal proportion from each semantic group.

3) *Caltech-101dataset*

Caltech-101 [15] is a benchmark that is widely used for image categorization, recognition and classification. It contains a total of 9146 images in 101 distinct categories. Fifteen categories were selected the classification, including airplane, ferry, camera, brain, cougar face, grand piano, Dalmatian, dollar bill, starfish, soccer ball, minaret, motorbikes, revolver, sunflower and Windsor chair. These categories were chosen due to their ability to contribute spatial information, rounded objects, and objects with different shape, texture and color information to test the effectiveness of the proposed method. The brain and sunflower groups were considered because of their textures. The dollar bill and cougar face are categories with complex object structures and orientations. The camera, revolver and Windsor chair categories require specific object recognition capability. The minaret and airplane share spatial and texture information for classification. A total of 1050 images were used for the experiments by selecting 70 images per category. Our algorithm randomly selects 70% of the images from each category for training and 30% of the images for testing. A total of 735 images from all of the categories was used for training, and 315 images were used for testing.

B. *Results*

1) *Input process*

In the first step, the color space is converted to gray scale for efficient computation. The gray scale image is then processed to detect the intensity-based local interest points. Global features are extracted for these interest points using the optimized sliding window. The extracted features are concatenated with the texture features that are invariant to monotonic and rotation changes. The feature vector concatenation is followed by applying the proposed feature reshaping technique. Coefficients are generated for the restructured observations. These data are passed to the support vector machine for classification. The support vector machine is involved in two phases: training and testing. During the training phase, the fused and reduced extracted feature vectors are input to the support vector machine. The positive training samples are randomly selected from the respective categories, and the negative training samples are collected from the other categories. Two times more positive training samples are used than negative training samples. Each training sample is labeled as belonging to one or the other sample type. The supervised learning model of the support vector machine learns new examples of one or the other category, which makes it a non-probabilistic binary linear classifier.

2) *Precision and recall evaluation*

Precision is the specificity measure or positive predicted value, and recall is the sensitivity measure or true positive rate evaluation. Precision and recall are calculated on each image category and also for small and large databases. The precision and recall results are tested on different sets of training and testing data.

$$precision=\frac{N_{A(q)}}{N_{R(q)}} \tag{11}$$

$$recall=\frac{N_{A(q)}}{N_t} \tag{12}$$

where $N_{A(q)}$ represents the relevant images that match the query image, $N_{R(q)}$ represents the images retrieved against the query image, and N_t is the total number of relevant images available in the database.

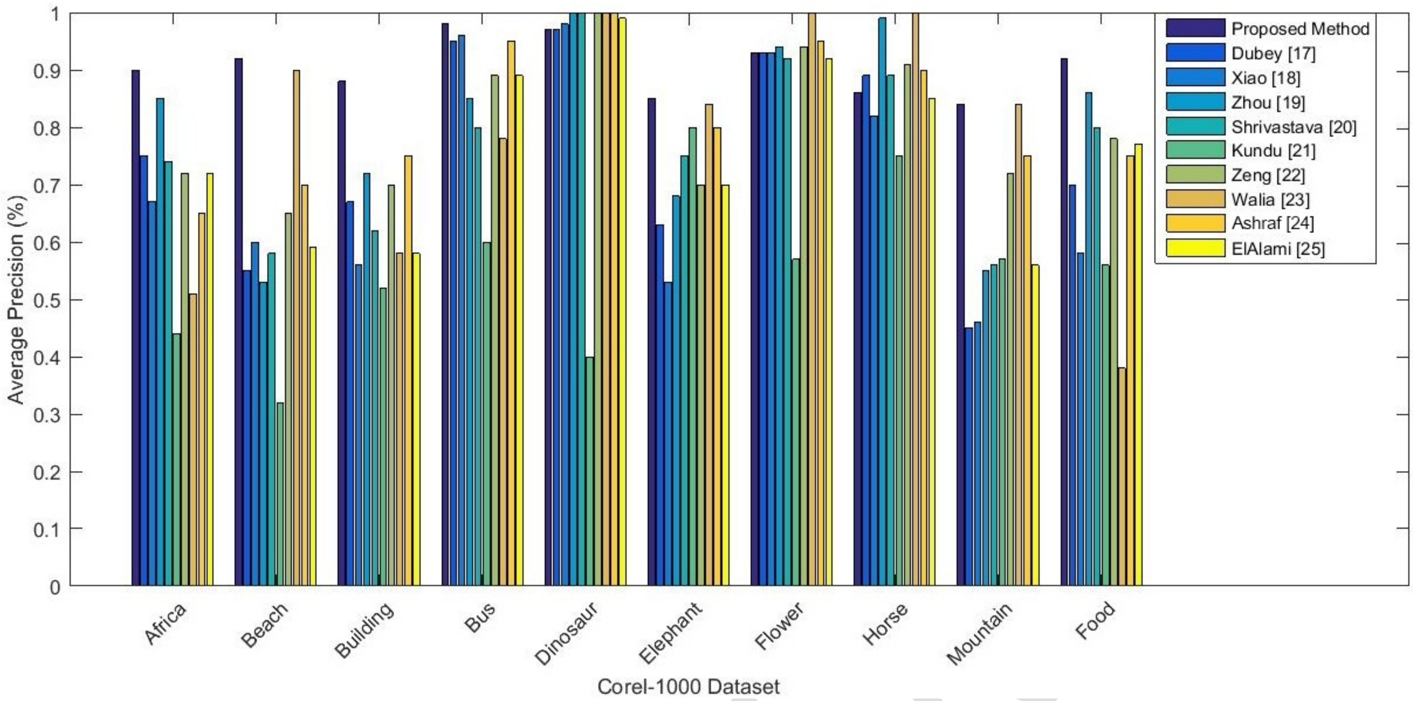


Figure 3 (a): Comparison of the average precisions obtained by the proposed method and other standard retrieval systems using the Corel-1000 dataset.

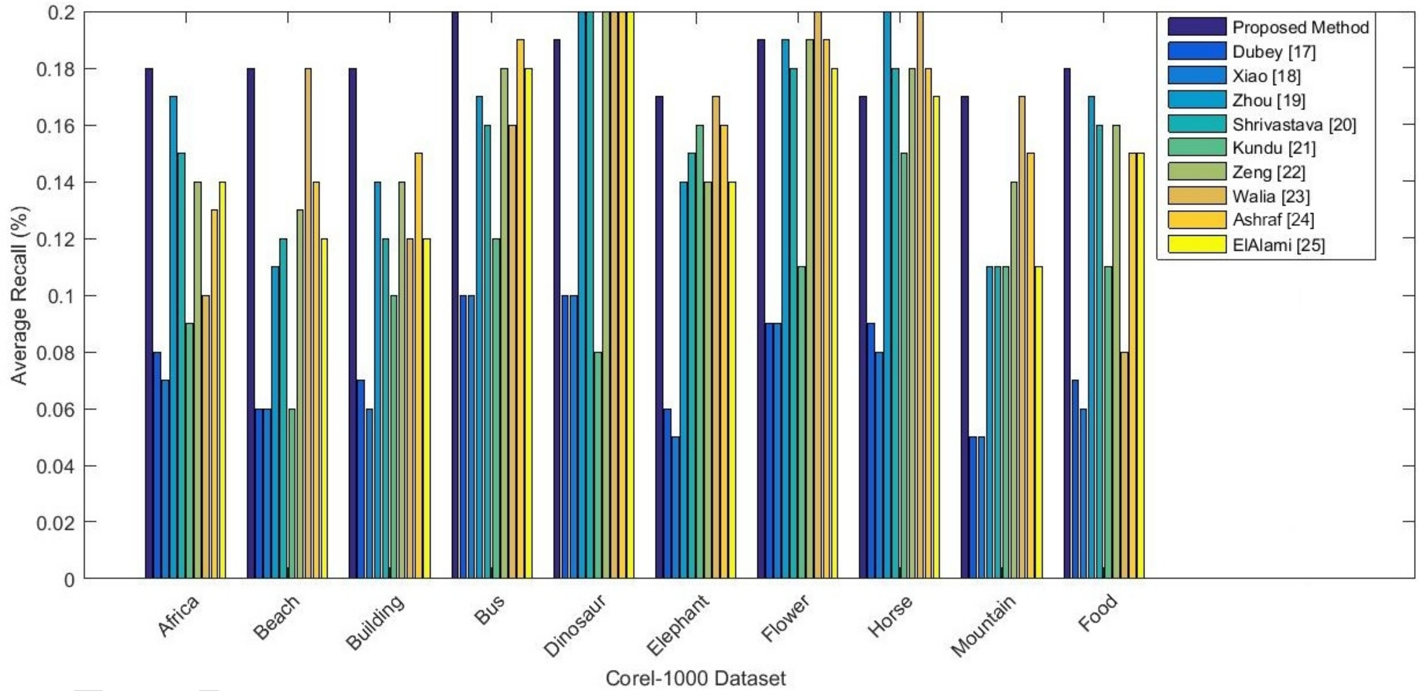


Figure 3 (b): Comparison of the average recalls obtained by the proposed method and other standard retrieval systems using the Corel-1000 dataset.

C. Experimental Results

1)Results of the Corel-1000 dataset with existing methods

To determine the accuracy of the proposed technique, we performed experiments on widely-used benchmarks. The experimental results are compared with those from existing methods as well as with the state-of-the-art descriptors SIFT, SURF, and HOG. The results are also compared with those of Dubey et al. [17], Xiao et al. [18], Zhou et al. [19], Shrivastava et al. [20], Kundu et al. [21], Zeng et al. [22], Walia et al. [23], Ashraf et al. [24] and ElAlami et al. [25] whose methods achieved remarkable performance. Their standardized work has also been cited by current researchers [59-62]. Figure 3 shows a graphical representation of the results of the proposed method compared to those from existing state-of-the-art methods. The results show that the proposed method outperforms most of the other methods. Figure 3(a) shows the average precision rates in comparison with those of existing methods. The proposed method shows remarkable performance in most of the image categories. The average recall rates are shown in Figure 3(b). The results show that the proposed method has better recall rates in most of the categories and that the mean average recall is higher than those of other methods.

Table 1: *Comparison of the average precision obtained by the proposed method and other standard retrieval systems on the top 20 results.*

Class	**Proposed Method**	**Dubey [17]**	**Xiao [18]**	**Zhou [19]**	**Shriv [20]**	**Kundu [21]**	**Zeng [22]**	**Walia [23]**	**Ashraf [24]**	**ElAlami [25]**
Africa	**0.90**	0.75	0.67	0.85	0.74	0.44	0.72	0.51	0.65	0.72
Beach	**0.92**	0.55	0.60	0.53	0.58	0.32	0.65	0.90	0.70	0.59
Building	**0.88**	0.67	0.56	0.72	0.62	0.52	0.70	0.58	0.75	0.58
Bus	**0.98**	0.95	0.96	0.85	0.80	0.60	0.89	0.78	0.95	0.89
Dinosaur	0.97	0.97	0.98	**1.00**	**1.00**	0.40	**1.00**	**1.00**	1.00	0.99
Elephant	**0.85**	0.63	0.53	0.68	0.75	0.80	0.70	0.84	0.80	0.70
Flower	0.93	0.93	0.93	0.94	0.92	0.57	0.94	**1.00**	**0.95**	0.92
Horse	0.86	0.89	0.82	0.99	0.89	0.75	0.91	**1.00**	0.90	0.85
Mountain	**0.84**	0.45	0.46	0.55	0.56	0.57	0.72	0.84	0.75	0.56
Food	**0.92**	0.70	0.58	0.86	0.80	0.56	0.78	0.38	0.75	0.77
Average	**0.904**	0.749	0.709	0.797	0.766	0.553	0.801	0.783	0.820	0.757

Table 1 shows a comparison of the average precision of the proposed method with those of the standard retrieval systems. The proposed system provides better precision in most of the semantic groups; it outperforms in the semantic groups of Africa, beach, building, bus, elephant, mountain and food. The proposed method extracts local texture and global features, which provide better results. The existing methods provide better results in some categories; for example, [20] gives better results for dinosaur and flower. However, the proposed method provides better results in these and other categories. Similarly, [17] provided a good precision rate in horse classification. The proposed method also has good accuracy in this category. Overall, the proposed method provides an increase in the mean average precision of 0.084%.

Table 2: *Comparison of the average recalls obtained by the proposed method and other standard retrieval systems on the top 20 results.*

Class	**Proposed Method**	**Dubey [17]**	**Xiao [18]**	**Zhou [19]**	**Shriv [20]**	**Kundu [21]**	**Zeng [22]**	**Walia [23]**	**Ashraf [24]**	**ElAlami [25]**
Africa	**0.18**	0.08	0.07	0.17	0.15	0.09	0.14	0.10	0.13	0.14
Beach	**0.18**	0.06	0.06	0.11	0.12	0.06	0.13	**0.18**	0.1	0.12

									4	
Building	**0.18**	0.07	0.06	0.14	0.12	0.10	0.14	0.12	0.15	0.12
Bus	**0.20**	0.10	0.10	0.17	0.16	0.12	0.18	0.16	0.19	0.18
Dinosaur	0.19	0.10	0.10	**0.20**	**0.20**	0.08	**0.20**	**0.20**	**0.20**	**0.20**
Elephant	**0.17**	0.06	0.05	0.14	0.15	0.16	0.14	**0.17**	0.16	0.14
Flower	0.19	0.09	0.09	0.19	0.18	0.11	0.19	**0.20**	0.19	0.18
Horse	0.17	0.09	0.08	**0.20**	0.18	0.15	0.18	**0.20**	0.18	0.17
Mountain	**0.17**	0.05	0.05	0.11	0.11	0.11	0.14	**0.17**	0.15	0.11
Food	**0.18**	0.07	0.06	0.17	0.16	0.11	0.16	0.08	0.15	0.15
Average	**0.181**	0.075	0.071	0.159	0.153	0.111	0.160	0.157	0.164	0.151

Table 2 shows the average recall rates obtained by the proposed methods and standard retrieval systems. The proposed method has remarkable recall rates in seven of the ten categories. Better classification leads to improved recall rates even in the complex semantic groups, such as Africa, mountain and food. The dinosaur and elephant categories are relatively easy to classify, and most of the existing methods provide better results in these categories. The proposed method provides high recall rates in the dinosaur and bus categories as well as in complex groups, such as flower and beach.

Figure 4 shows the mean average precision and recall rates for the proposed method and the existing methods. Figure 4 (a) shows that the proposed method has a higher mean average precision rate than the existing methods, and Figure 4 (b) shows that it has significantly better mean average recall rates. The recall rate is improved by 0.017% over those from the existing methods [20].

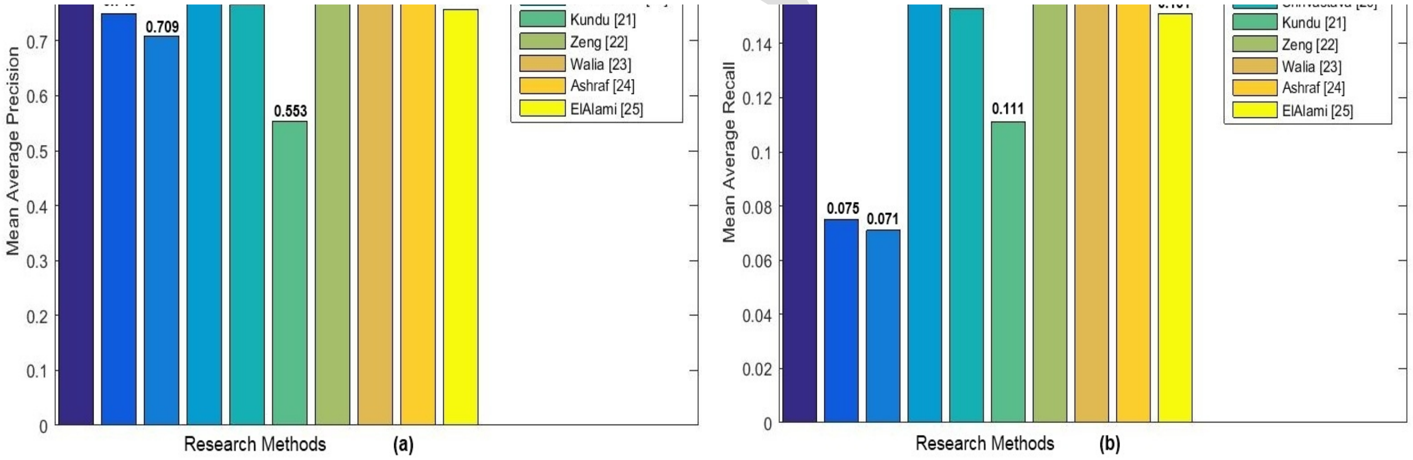


Figure 4: ***(a)*** *Graphical representation of the mean average precisions on the Corel dataset.* ***(b)*** *Graphical representation of the mean average recalls on the Corel dataset.*

2) ImageNet Synset results

Experiments were performed on ImageNet synsets to check the robustness and versatility of the proposed method. The results are shown for the top 20 images. In the testing phase, feature vectors of an input image are extracted using the proposed method. The support vector machine classifies the input image based on the training data. Input images are selected from each category to check the precision and recall rates for each category, and the results are computed for 20 images. The classified images for each category yield the precision and recall rates for that category. For this benchmark, the mean average precision is 0.735%, and the mean average recall is 0.147%.

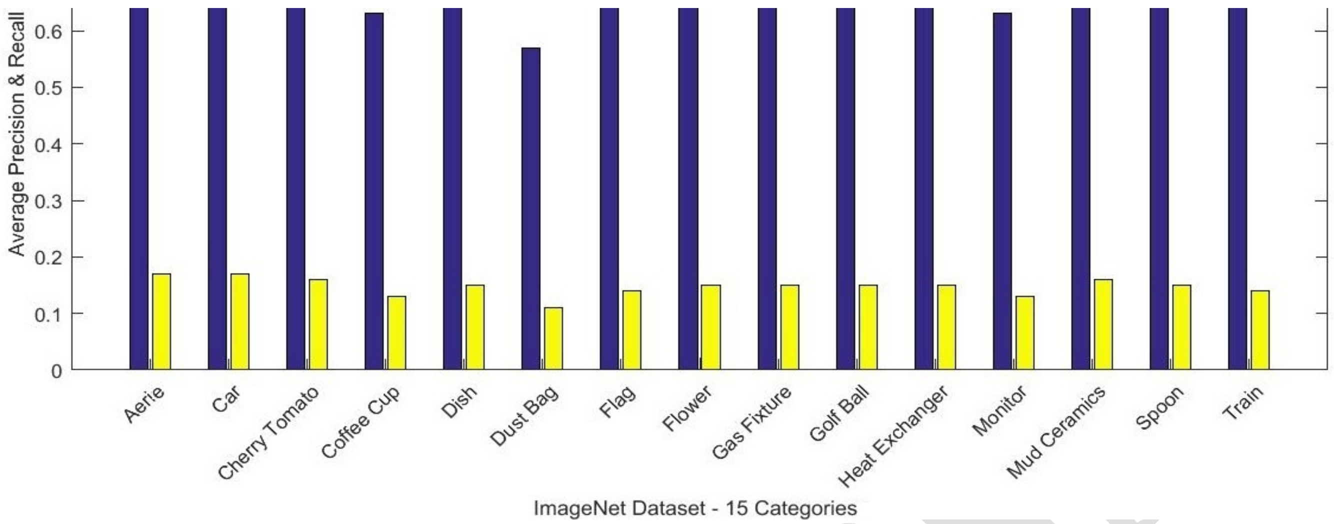


Figure 5: *Average precisions and recall rates for the ImageNet synset. The results are computed for the proposed method with 15 synsets.*

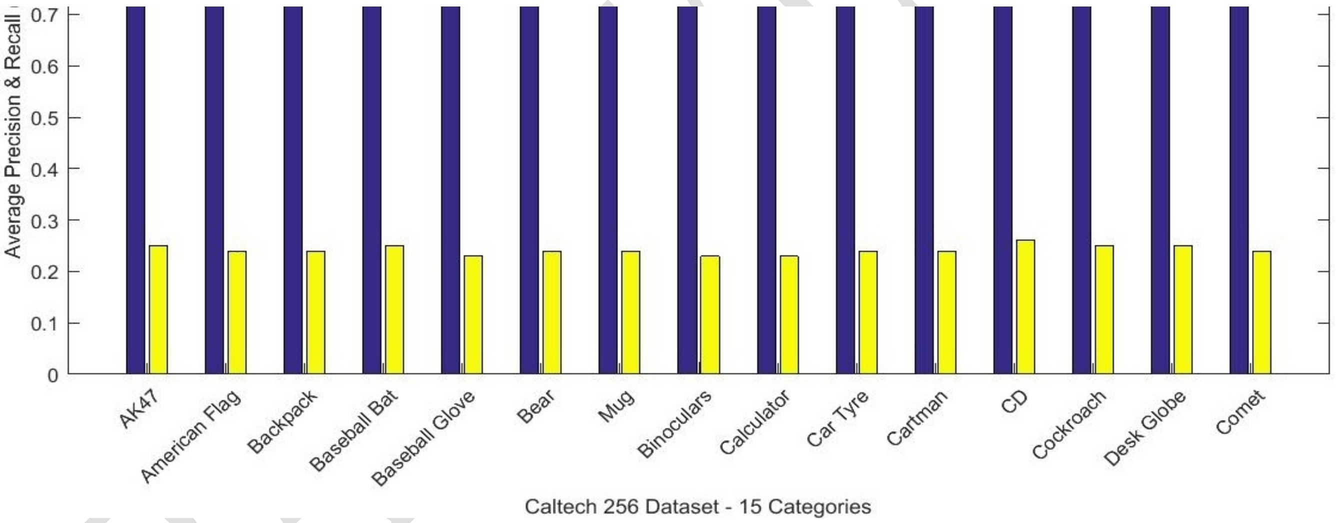


Figure 6: *Average precision and recall rates of the proposed method on 15 categories of the Caltech-256 dataset.*

1) Caltech-256 dataset results

To check the effectiveness of the proposed method, the results are compared with those from state-of-the-art methods. A total of 1050 images are randomly selected from 15 preselected image categories for training and testing. A batch of 14 images is used to test each category. A total of 15 such batches are used to obtain the precision and recall rates for each category. The results are shown for the top 14 images from the batch of 50 relevant images. The proposed method outperforms the others in most of the image categories. The results show a mean average precision of 0.865% and a mean average recall rate of 0.242%. Caltech-256 is considered a challenging dataset that contains complex images. The proposed method provides exceptional results for the AK47, baseball bat, desk globe, car tire and CD image categories, which contain uncrowded backgrounds and objects with clear boundaries. Sample images from these categories are shown in Figure 7(a). However the results of the proposed method are equally good for the other categories, which include cluttered objects, overlapping objects, and complex backgrounds as shown in Figure 7(b).

(a) **(b)**

Figure 7: (a) *Sample images from the categories with exceptional results from Caltech-256.* ***(b)*** *Sample images with overlapping objects and complex backgrounds in Caltech-256.*

3)Caltech-101 dataset results

The average precision and recall rates for 15 categories of the Caltech-101 dataset are shown in Figure 8. Images with different foregrounds and backgrounds, object shapes, and textures are selected for classification. The proposed technique provides better precision in all of the categories by processing the local features with global values. The recall rates for Caltech-101 are also promising. Most of the categories have high recall rates, while a few have average rates. The Windsor chair and camera have average rates due to the complex backgrounds and cluttered objects. The mean average precision obtained for this benchmark is 0.884%, and the mean average recall is 0.248%.

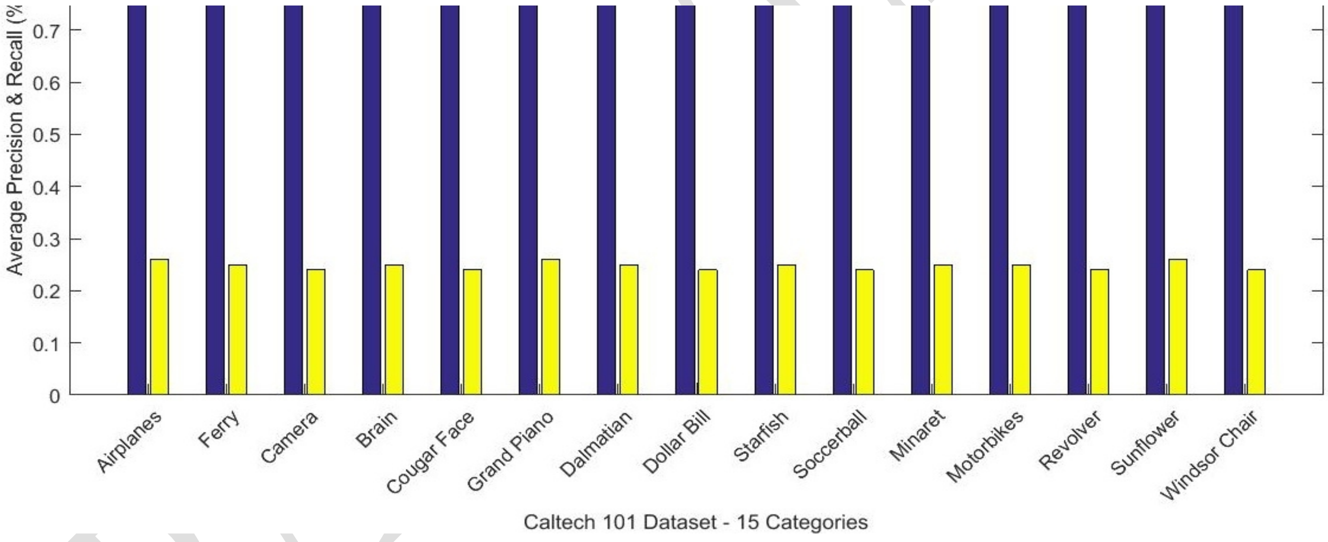


Figure 8: *Average precision rates of the proposed method on 15 categories of the Caltech-101 dataset.*

D. Comparative analysis against key point detectors and descriptors

Feature detectors and descriptors are used in object detection and recognition. Detectors refer to the tool that extracts the features from the image, such as corner, blob or edge detectors. Extractors are used to read the features from the interest points. HOG [1], SIFT [2], and SURF [3] are well-known object detectors and descriptors that are widely used in many applications. HOG was presented at the Conference on Computer Vision and Pattern Recognition (CVPR) and is used for object detection [52], image classification [53] and image retrieval [54] tasks. SIFT was presented in the proceedings of the International Conference on Computer Vision (ICCV) and is used for content-based image retrieval [55, 56] and object detection tasks [57]. SURF was presented at the European Conference on Computer

Vision (ECCV) and is used for image retrieval [58] and related tasks. These descriptors are compared to test the effectiveness of the proposed method. For the experiments, 1050 images are randomly selected from 15 categories, and each category contains 70 images. Our algorithm randomly selects 2/3 of the images from each category for training and 1/3 of the images for testing. A total of 735 images from all of the categories was used for training, and 315 were used for testing. In the training phase, positive samples are taken randomly from the respective category. Positive samples make up 70% of each category, and the negative training samples (30%) are selected from the rest of the categories.

1) *Computational Load*

Experiments are performed with HOG, SIFT, and SURF, and the results are compared to those of the proposed method. These descriptors, particularly SIFT, produced results with very high computational times. Moreover, redundant and massive feature vectors are produced, which require large amounts of processing time and system resources for computation and classification. The proposed method performed the classification with very low time and computation costs. The computational efficiency achieved by processing a limited set of feature vectors from the proposed reordering algorithm generated a compact input that was used to obtain compact coefficients. The computational load is an aggregate of the gray level conversion of the input image, the feature extraction using the image descriptor, feature reduction and comparison with the dataset for classification. The proposed method consumed a total computation time of 0.70083 sec/image, which is 35.5%, 71.22% and 59.7% less than HOG, SIFT, and SURF, respectively.

2)*Precision Rates*

Descriptors are unable to perform equally well in all image categories due to their limits of effectiveness. Descriptor [7] is suitable for local features, but it is unable to provide accurate results for global features. Similarly, the detector with the best ability to predict texture patterns is unable to accurately recognize objects. In addition, the descriptors that are suitable for finding edges and corners are not good candidates for texture analysis. Therefore, none of the state-of-the-art descriptors are ideal candidates for feature extraction in versatile image categories. However, the proposed descriptor is able to find the textures, edges, corners, and pixel intensities and recognize complex and overlapping objects.

Figure 9(a) shows the results of the proposed method in comparison with those of the state-of-the-art descriptors for 15 categories of the Caltech-101 dataset. Some of the detectors show better results in some image categories because they were designed for those categories. The descriptors perform well in their areas of specialty.

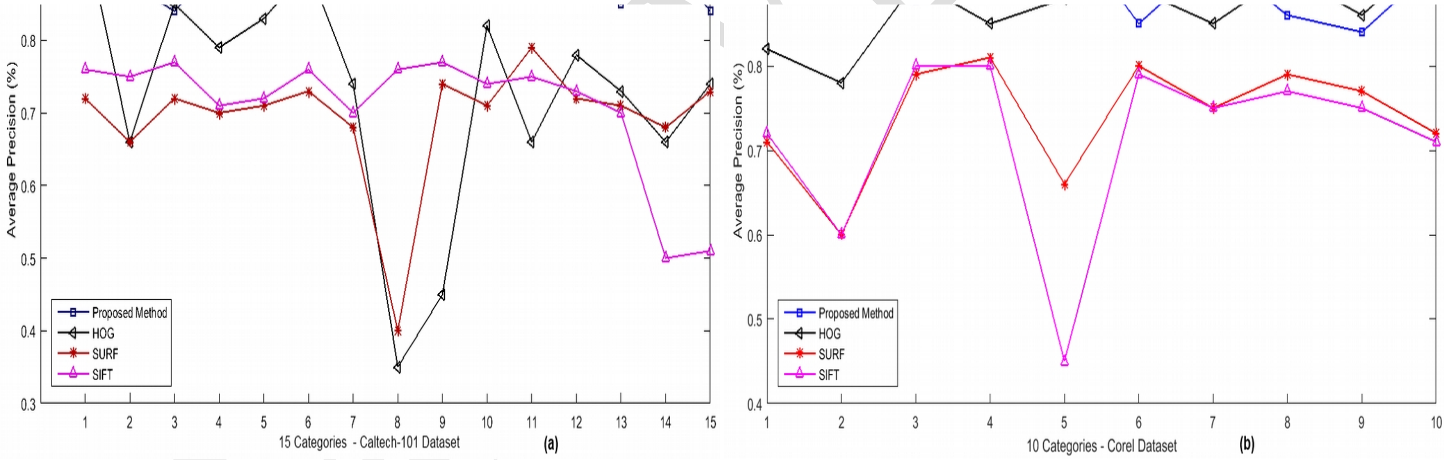


Figure 9: (a) *Comparison of the average precisions obtained by the proposed method compared with those of the state-of-the-art descriptors on 15 categories of the Caltech-101 dataset.* ***(b)*** *Comparison of the average precisions obtained by the proposed method compared with those of the state-of-the-art descriptors on 10 categories of the Corel-1000 dataset.*

Table 3: *Comparison of the average precisions obtained by the proposed method compared with those from the state-of-the-art descriptors on 15 categories of the Caltech-101 dataset.*

Class	Proposed Method	HOG [1]	SURF [3]	SIFT [2]
Airplanes	0.94	**0.95**	0.72	0.76
Ferry	**0.88**	0.66	0.66	0.75
Camera	0.84	**0.85**	0.72	0.77
Brain	**0.91**	0.79	0.70	0.71
Cougar Face	**0.87**	0.83	0.71	0.72

Grand Piano	**0.92**	0.90	0.73	0.76
Dalmatian	**0.90**	0.73	0.68	0.70
Dollar Bill	**0.86**	0.35	0.40	0.76
Starfish	**0.88**	0.45	0.74	0.77
Soccer Ball	**0.87**	0.82	0.71	0.74
Minaret	**0.88**	0.66	0.79	0.74
Motorbikes	**0.90**	0.78	0.72	0.73
Revolver	**0.85**	0.73	0.71	0.70
Sunflower	**0.92**	0.66	0.68	0.50
Windsor Chair	**0.84**	0.74	0.73	0.51
Average	**0.887**	0.728	0.695	0.710

Table 3 shows a comparison of the average precisions of the proposed descriptor with those of the state-of-the-art descriptors HOG, SIFT and SURF. Experiments are performed with all of the descriptors to check the strength of the proposed descriptor. The proposed method shows remarkable performance in the sunflower, motorbike, starfish, ferry and brain categories. The mean average precision obtained by the proposed descriptor is 0.158% higher than that of the HOG descriptor.

***Table 4:** Comparison of the average precisions obtained by the proposed method compared with those from the state-of-the-art descriptors on the Corel-1000 dataset.*

Class	**Proposed Method**	**HOG [1]**	**SURF [3]**	**SIFT [2]**
Africa	**0.90**	0.82	0.71	0.72
Beach	**0.91**	0.78	0.60	0.60
Building	0.87	**0.90**	0.79	0.80
Bus	**0.95**	0.85	0.81	0.80
Kangaroo	**0.97**	0.88	0.66	0.45
Elephant	0.85	**0.89**	0.80	0.79
Flower	**0.93**	0.85	0.75	0.75
Horse	0.85	**0.91**	0.76	0.77
Mountain	0.83	**0.86**	0.77	0.75
Food	**0.98**	0.94	0.72	0.71
Average	**0.907**	0.871	0.738	0.716

Table 4 compares the experimental results of the proposed method with those of the state-of-the-art descriptors using the Caltech-256 dataset. The proposed method has better precision than the existing methods in 13 of the 15 categories. For the other two categories, the precision is almost the same as that reported by SURF. The results of the Corel-1000 collection are shown to check the effectiveness of the proposed method compared to those of the state-of-the-art descriptors. The proposed method provides better results for most of the image categories. The proposed descriptor has a 0.036% better mean average precision for the proposed method.

3) Recall Rates

Figure 10 shows the recall rates for Caltech-101. The results show that the state-of-the-art descriptors provide better performance in some image categories and below average performance in others. However, the proposed method provides better recall rates for most of the categories in both datasets. Hence, the proposed method provides better classification results for all of the image categories. Low recall rates are observed for the dollar bill and sunflower categories using the HOG and SURF descriptors.

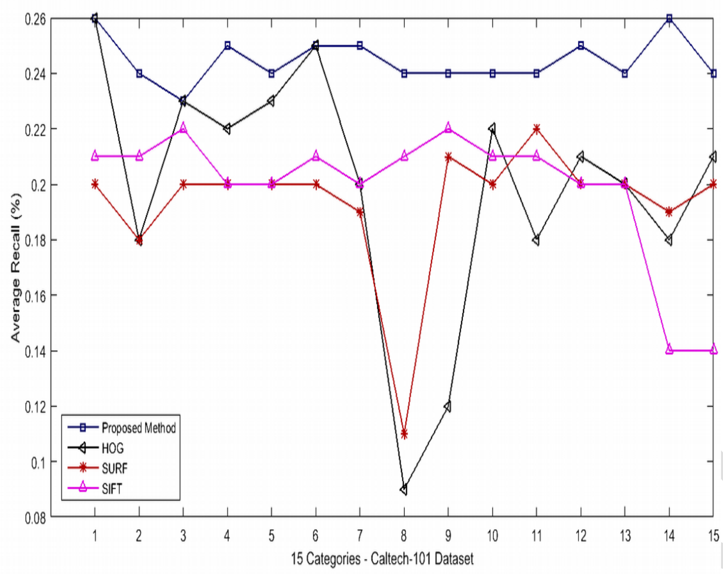


Figure 10: *Comparison of the average recall rates obtained by the proposed method compared with those of the state-of-the-art descriptors on 15 categories of the Caltech-101 dataset.*

In these categories, complex image backgrounds with overlapping objects are difficult to classify. However the proposed method provided better recall rates in these categories. Hence, the proposed method intuitively combines local and global features by selecting local features based on the pixel intensity level and texture values and selecting global features using the sliding window. The local features help in the texture and shape analysis, whereas the global features are robust to object recognition. Assembling local values with global depiction detects the hidden patterns of an image as well as the distinctive objects. The concatenation of the local and global features is performed after computing the high variance coefficients. The proposed reshaping algorithm limits the inputs for the component analysis. Thus, the combined feature vectors are compact and represent an image efficiently.

V. Conclusion

In this paper, we proposed a novel method for effective and accurate feature vector extraction and image classification. The descriptor is able to perform classifications with significant precision in diverse categories of the benchmark datasets ImageNet, Caltech-256, Caltech-101 and Corel-1000. The descriptor accurately distinguishes corners, edges, and lines and performs texture analysis and object recognition for complex and overlapping images. The proposed method was compared with other sophisticated methods and provided remarkable precision in most of the image categories due to its superior nature. The proposed descriptor was also compared with the state-of-the-art descriptors SIFT, SURF and HOG and outperformed them in all of the datasets. The experimental results showed that the state-of-the-art descriptors perform well in some image categories due to their specialization in those areas but are unable to provide good results in other categories due to their limitations with those image attributes. The proposed method provides reliable and remarkable precision and recall rates in most of the image categories of the benchmark datasets.

References

1. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886-893. IEEE, 2005.
2. Lowe, David G. "Object recognition from local scale-invariant features." In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150-1157. Ieee, 1999.
3. Bay, Herbert, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. "Speeded-up robust features (SURF)." *Computer vision and image understanding* 110, no. 3 (2008): 346-359.
4. Tuytelaars, Tinne, and Luc Van Gool. "Matching widely separated views based on affine invariant regions." *International journal of computer vision* 59, no. 1 (2004): 61-85.
5. Sural, Shamik, Gang Qian, and Sakti Pramanik. "Segmentation and histogram generation using the HSV color space for image retrieval." In *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 2, pp. II-589. IEEE, 2002.
6. Mikolajczyk, Krystian, and Cordelia Schmid. "A performance evaluation of local descriptors." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, no. 10 (2005): 1615-1630.
7. Liu, Guang-Hai, and Jing-Yu Yang. "Content-based image retrieval using color difference histogram." *Pattern Recognition* 46, no. 1 (2013): 188-198.
8. Chaudhary, Manoj D., and Abhay B. Upadhyay. "Integrating shape and Edge Histogram Descriptor with

Stationary Wavelet Transform for Effective Content Based Image Retrieval." In *Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on*, pp. 1522-1527. IEEE, 2014.

9. Agrawal, Deepak, Anand Singh Jalal, and Rajeev Tripathi. "Trademark image retrieval by integrating shape with texture feature." In *Information Systems and Computer Networks (ISCON), 2013 International Conference on*, pp. 30-33. IEEE, 2013.
10. Zheng, Liang, and Shengjin Wang. "Visual phraselet: Refining spatial constraints for large scale image search." *Signal Processing Letters, IEEE* 20, no. 4 (2013): 391-394.
11. Irtaza, Aun, M. Arfan Jaffar, and Muhammad Tariq Mahmood. "Semantic image retrieval in a grid computing environment using support vector machines." *The Computer Journal* (2013): bxt087.
12. Shen, Xiaohui, Zhe Lin, Jonathan Brandt, Shai Avidan, and Ying Wu. "Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking." In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 3013-3020. IEEE, 2012.
13. Stanford Vision Lab, http://image-net.org/ last accessed on October 2016.
14. Li, Jia, and James Z. Wang. "Automatic linguistic indexing of pictures by a statistical modeling approach." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25, no. 9 (2003): 1075-1088.
15. 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. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004.
16. Griffin, Gregory, Alex Holub, and Pietro Perona. "Caltech-256 object category dataset." (2007).
17. Dubey, Shiv Ram, Satish Kumar Singh, and Rajat Kumar Singh. "Multichannel Decoded Local Binary Patterns for Content-Based Image Retrieval." IEEE Transactions on Image Processing 25, no. 9 (2016): 4018-4032.
18. Xiao, Yang, Jianxin Wu, and Junsong Yuan. "mCENTRIST: A multi-channel feature generation mechanism for scene categorization." IEEE Transactions on Image Processing 23, no. 2 (2014): 823-836.
19. Zhou, Yan, Fan-Zhi Zeng, Hui-min Zhao, Paul Murray, and Jinchang Ren. "Hierarchical visual perception and two-dimensional compressive sensing for effective content-based color image retrieval." Cognitive Computation 8, no. 5 (2016): 877-889.
20. Shrivastava, Nishant, and Vipin Tyagi. "An efficient technique for retrieval of color images in large databases." Computers & Electrical Engineering 46 (2015): 314-327.
21. Kundu, Malay Kumar, Manish Chowdhury, and Samuel Rota Bulò. "A graph-based relevance feedback mechanism in content-based image retrieval." Knowledge-Based Systems 73 (2015): 254-264.
22. Zeng, Shan, Rui Huang, Haibing Wang, and Zhen Kang. "Image retrieval using spatiograms of colors quantized by Gaussian Mixture Models." Neurocomputing 171 (2016): 673-684.
23. Walia, Ekta, and Aman Pal. "Fusion framework for effective color image retrieval." Journal of Visual Communication and Image Representation 25, no. 6 (2014): 1335-1348.
24. Ashraf, Rehan, Khalid Bashir, Aun Irtaza, and Muhammad Tariq Mahmood. "Content based image retrieval using embedded neural networks with bandletized regions." *Entropy* 17, no. 6 (2015): 3552-3580.
25. ElAlami, M. Esmel. "A new matching strategy for content based image retrieval system." Applied Soft Computing 14 (2014): 407-418.
26. Xia, Yu, Shouhong Wan, and Lihua Yue. "A New Texture Direction Feature Descriptor and Its Application in Content-Based Image Retrieval." In *Proceedings of the 3rd International Conference on Multimedia Technology (ICMT 2013)*, pp. 143-151. Springer Berlin Heidelberg, 2014.
27. Agarwal, Sankalp, Anil Kumar Verma, and Prashant Singh. "Content based image retrieval using discrete wavelet transform and edge histogram descriptor." In *Information Systems and Computer Networks (ISCON), 2013 International Conference on*, pp. 19-23. IEEE, 2013.
28. Jadhav, Pratima, and Rashmi Phalnikar. "SIFT based Efficient Content based Image Retrieval System using Neural Network." *Artificial Intelligent Systems and Machine Learning* 7, no. 8 (2015): 234-238.
29. Awad, Dounia, Vincent Courboulay, and Arnaud Revel. "Saliency filtering of sift detectors: Application to cbir." In *Advanced Concepts for Intelligent Vision Systems*, pp. 290-300. Springer Berlin Heidelberg, 2012.
30. Saad, M. H., H. I. Saleh, H. Konber, and M. Ashour. "CBIR SYSTEM BASED on INTEGRATION BETWEEN SURF AND GLOBAL FEATURES." (2013).
31. Velmurugan, K., and Lt Dr S. Santhosh Baboo. "Content-based image retrieval using SURF and colour moments." *Global Journal of Computer Science and Technology* 11, no. 10 (2011).
32. Tudor Barbu. Pedestrian detection and tracking using temporal differencing and HOG features. Computers and Electrical Engineering, 2013.
33. Albiol, Alberto, David Monzo, Antoine Martin, Jorge Sastre, and Antonio Albiol. "Face recognition using HOG–EBGM." *Pattern Recognition Letters* 29, no. 10 (2008): 1537-1543.
34. Felzenszwalb, Pedro F., Ross B. Girshick, David McAllester, and Deva Ramanan. "Object detection with discriminatively trained part-based models." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32, no. 9 (2010): 1627-1645.
35. Pan, Shujing, Shutao Sun, Lifang Yang, Fengfeng Duan, and Anqi Guan. "Content Retrieval Algorithm Based on Improved HOG." In Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (ACIT-CSI), 2015 3rd International Conference on, pp. 438-441. IEEE, 2015.

36. Murala, Subrahmanyam, R. P. Maheshwari, and R. Balasubramanian. "Local tetra patterns: a new feature descriptor for content-based image retrieval." *Image Processing, IEEE Transactions on* 21, no. 5 (2012): 2874-2886.
37. Harris, Chris, and Mike Stephens. "A combined corner and edge detector." In *Alvey vision conference*, vol. 15, p. 50. 1988.
38. Ojala, Timo, Kimmo Valkealahti, Erkki Oja, and Matti Pietikäinen. "Texture discrimination with multidimensional distributions of signed gray-level differences." *Pattern Recognition* 34, no. 3 (2001): 727-739.
39. Ojala, Timo, Kimmo Valkealahti, Erkki Oja, and Matti Pietikäinen. "Texture discrimination with multidimensional distributions of signed gray-level differences." *Pattern Recognition* 34, no. 3 (2001): 727-739.
40. Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Gray scale and rotation invariant texture classification with local binary patterns." In *Computer Vision-ECCV 2000*, pp. 404-420. Springer Berlin Heidelberg, 2000.
41. Ojala, Timo, Matti Pietikäinen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions." *Pattern recognition* 29, no. 1 (1996): 51-59.
42. Pietikäinen, M., Ojala, T., Xu. Z.: Rotation-Invariant Texture Classification Using Feature Distributions. Pattern Recognition 33 (2000) 43-52.
43. Oertel, C., Colder, B., Colombe, J., High, J., Ingram, M., Sallee, P., Current Challenges in Automating Visual Perception. Proceedings of IEEE Advanced Imagery Pattern Recognition Workshop 2008.
44. Lai C-C, Chen Y-C (2011) A user-oriented image retrieval system based on interactive genetic algorithm. IEEE Trans Instrum Meas 60:3318–3325.
45. Krig, Scott. "Interest Point Detector and Feature Descriptor Survey." In *Computer Vision Metrics*, pp. 217-282. Apress, 2014.
46. Wang, Han, and Michael Brady. "Real-time corner detection algorithm for motion estimation." *Image and Vision Computing* 13, no. 9 (1995): 695-703.
47. Khotanzad, Alireza, and Yaw Hua Hong. "Invariant image recognition by Zernike moments." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12, no. 5 (1990): 489-497.
48. Rosten, Edward, and Tom Drummond. "Machine learning for high-speed corner detection." In *Computer Vision–ECCV 2006*, pp. 430-443. Springer Berlin Heidelberg, 2006.
49. Moravec, Hans P. "Visual mapping by a robot rover." In *Proceedings of the 6th international joint conference on Artificial intelligence-Volume 1*, pp. 598-600. Morgan Kaufmann Publishers Inc., 1979.
50. Förstner, Wolfgang, and Eberhard Gülch. "A fast operator for detection and precise location of distinct points, corners and centres of circular features." In*Proc. ISPRS intercommission conference on fast processing of photogrammetric data*, pp. 281-305. 1987.
51. Stejić, Zoran, Yasufumi Takama, and Kaoru Hirota. "Genetic algorithm-based relevance feedback for image retrieval using local similarity patterns." *Information processing & management* 39, no. 1 (2003): 1-23.
52. Dalal, Navteen, and Bill Triggs. "Object detection using histograms of oriented gradients." In *Pascal VOC Workshop, ECCV*. 2006.
53. Dalal, Navteen, and Bill Triggs. "Object detection using histograms of oriented gradients." In *Pascal VOC Workshop, ECCV*. 2006.
54. Hu, Rui, and John Collomosse. "A performance evaluation of gradient field hog descriptor for sketch based image retrieval." *Computer Vision and Image Understanding* 117, no. 7 (2013): 790-806.
55. Wangming, Xu, Wu Jin, Liu Xinhai, Zhu Lei, and Shi Gang. "Application of Image SIFT Features to the Context of CBIR." In *Computer Science and Software Engineering, 2008 International Conference on*, vol. 4, pp. 552-555. IEEE, 2008.
56. Xu, Pengfei, Lei Zhang, Kuiyuan Yang, and Hongxun Yao. "Nested-SIFT for efficient image matching and retrieval." *IEEE MultiMedia* 20, no. 3 (2013): 34-46.
57. Kim, Sungho, Kuk-Jin Yoon, and In So Kweon. "Object recognition using a generalized robust invariant feature and Gestalt's law of proximity and similarity." *Pattern Recognition* 41, no. 2 (2008): 726-741.
58. Lee, Yong-Hwan, and Youngseop Kim. "Efficient image retrieval using advanced SURF and DCD on mobile platform." *Multimedia Tools and Applications* 74, no. 7 (2015): 2289-2299.
59. Ali, Nouman, Khalid Bashir Bajwa, Robert Sablatnig, and Zahid Mehmood. "Image retrieval by addition of spatial information based on histograms of triangular regions." *Computers & Electrical Engineering* (2016).
60. Walia, Ekta, and Aman Pal. "Fusion framework for effective color image retrieval." *Journal of Visual Communication and Image Representation* 25, no. 6 (2014): 1335-1348.
61. Dubey, Shiv Ram, Satish Kumar Singh, and Rajat Kumar Singh. "A multi-channel based illumination compensation mechanism for brightness invariant image retrieval." *Multimedia Tools and Applications* 74, no. 24 (2015): 11223-11253.
62. Thepade, Sudeep, Rik Das, and Saurav Ghosh. "Novel technique in block truncation coding based feature extraction for content based image identification." In *Transactions on Computational Science XXV*, pp. 55-76. Springer Berlin Heidelberg, 2015.
63. Datta, Ritendra, Dhiraj Joshi, Jia Li, and James Z. Wang. "Image retrieval: Ideas, influences, and trends of the new age." *ACM Computing Surveys (CSUR)* 40, no. 2 (2008): 5.
64. Gupta, Ekta, and Rajendra Singh Kushwah. "Combination of global and local features using DWT with SVM for CBIR." In *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on*, pp. 1-6. IEEE, 2015.

65. Iqbal, Kashif, Michael O. Odetayo, and Anne James. "Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics." Journal of Computer and System Sciences 78, no. 4 (2012): 1258-1277.
66. Neelima, N., and E. Sreenivasa Reddy. "An improved image retrieval system using optimized FCM & multiple shape, texture features." In *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1-7. IEEE, 2015.
67. Youssef, Sherin M. "ICTEDCT-CBIR: Integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval." *Computers & Electrical Engineering* 38, no. 5 (2012): 1358-1376.
68. Lande, Milind V., Praveen Bhanodiya, and Pritesh Jain. "An effective content-based image retrieval using color, texture and shape feature." In *Intelligent Computing, Networking, and Informatics*, pp. 1163-1170. Springer India, 2014.