The Nucleus ***57****, No. 4 (2020) 118-128*

www.thenucleuspak.org.pk

The Nucleus

ISSN 0029-5698 (Print)
ISSN 2306-6539 (Online)

Highly Efficient Multimedia Image Retrieval using Slim Descriptor

K.T. Ahmed [1], H. Afzal [1], S. Iqbal [1], M.G. Hussain [1], M.R. Mufti [2*], A. Karim [3]

[1] *Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan.*

[2] *Department of Computer Science, COMSATS University Islamabad, Vehari Campus Vehari, Pakistan.*

[3] *Department of Information Technology Bahauddin Zakariya University Multan, Pakistan.*

ARTICLE INFO

Article history:
Received: *11 February, 2021*
Accepted: *18 May, 2021*
Published: *02 June, 2021*

Keywords:
Multimedia images retrieval,
Histogram of oriented gradients,
Principal component analysis,
Support vector machine,
Dimension reduction,
Computer vision

ABSTRACT

Efficient multimedia image extraction with high precision compatible with diverse image datasets is an implicit requirement of current image retrieval systems. In this paper, a multimedia image descriptor is introduced to achieve high performance along with high accuracy. For this, Histograms of Oriented Gradients (HOG) are extracted from a dense grid partitioned image by taking edge intensity based orientation histograms as primitive feature vectors. We depleted these massive redundant candidates to linearly uncorrelated variables by applying orthogonal transformation to achieve Principal Components (PC) where succeeding component's constraint dependent orthogonal variance based local descriptors are compact and robust to deformation. A distinctness of our proposed approach is the selection of a single coefficient having largest variance as image descriptor out of returned dimensionally reduced vectors which results in higher performance and less space and time consumption. Supervised learning using Support Vector Machine (SVM) is then applied on non-probabilistic binary linear classification of images. The experimental results show higher precision, low memory consumption and sufficient performance gain.

1. Introduction

Efficient image extraction systems for small and large datasets with high precision, prompt response, less space and time consumption and high throughput is an inevitable requirement of the day. Traditional concept based image indexing is tag based which provides non-semantic results due to limited context orientation of metadata. Medical content based image querying systems extract the images for color, shapes, texture, objects etc. and return semantically relevant results more accurately. Depending upon visual features and domain knowledge for searched contents, current Content Based Image Retrieval (CBIR) systems fetch relevant results to satiate the searcher. Visual content formation is phased for feature processing and feature description by modeling minimum invariant conditions based on user feedback. Probability distribution estimation of quantitative variables is formed in histograms for data density estimations of color, shape and texture features. Color histogram intersection on multicolored images for efficient indexing in a large database [1], correlogram based enhanced histograms [2], fuzzy color histograms [3], color and shape based retrieval [4] and other variations are used for efficient and accurate retrieval. Local and global features [5] possess important information for image retrieval. Moreover, color distribution has some limitations like global color distribution in images is suitable when object and region position is not a concern and it fails when object detection inside image is required. Shape invariant histograms [6], edge histograms for indexing and segmentation involve shape contexts for rapid image matching but segmentation has weighted graph partitioning problem [7]. For region-based feature extraction, image segmentation is employed with variations like weighted graph partitioning, k-mean clustering and cues of texture and contour differences which are computationally complex and crucial for image classification within images. To achieve high performance and precision, local features detection and interest point description; Scale-Invariant Feature Transform (SIFT) [8] descriptor was introduced. It is invariant to uniform scaling and robust to noise and illumination. It also suffers high dimensionality problem. Speeded Up Robust Features (SURF) [9] outperforms and are reliant on integral images for image convolutions that results in low computational time with 64 dimensions to reduce feature description calculation time. It has comparatively low descriptor variance. Gradient Location-Orientation Histogram (GLOH) [10] is a variant to SIFT whose descriptor size is reduced by PCA. It is limited to fifty degrees of affine transformation after which results become unreliable. Intensity gradients distribution description of shapes in the localized portions of images similar to edge orientation histogram is performed by Histogram of Oriented Gradients (HOG) [11], which is invariant to photometric and geometric transformations. All these local descriptors find interest points which are then transformed into feature vectors based on extracted gradient, color, texture, etc. The number of interest points reveals better object classification. Processing time is proportionate to data dimensions generated by these descriptors which generate hundreds to thousands data dimensions for an image. To process it, high computational time, processing power and memory space is required. Medical imaging, videos, DNA and MRI databases contain highly dimensional data. Therefore, generation of up to the mark precision level for these bulk feature vectors becomes inaccessible. For experimentation, testbed databases usually contain few hundreds to thousand images and their results,

*Corresponding author: rafiq_mufti@ciitvehari.edu.pk

and computational power vary for large datasets. A tradeoff is observed between performance and accuracy for high dimensional large datasets.

Performance enhancement is being studied by many researchers and filters, distance measures, dimension reduction, segmentation, semantic gap and relevance feedback are used to achieve it. In many cases, testbed is normally a selective image set [12] or a subset of Corel dataset [13-15].

Histograms had been used in different ways to achieve better performance, efficiency and improve precision. Large databases and rapid performance was focused on by Swain and Ballard [1] using Histogram intersection on multicolored images. They performed quick indexing in a large database and Histogram Back-projection to solve location problem efficiently. Color histograms with some variations named as Color correlogram [2] were used for effective and inexpensive retrieval of changed shape images. Histogram calculations based on fuzzy sets were employed for inexpensive large computations which were robust to noisy interfaces and quantization errors [3].

Data descriptors SIFT [8], SURF [9], GLOH [10] and HOG [11] return high dimensions which result in more space and computation time, low performance and scalability issues. To compact the feature vectors size, different dimension reduction techniques were applied to these signatures [16, 17]. High dimensionality of data limits the speed and scalability of feature matching that is targeted by Kernel Projection Based SIFT (KPB-SIFT) [16] technique which encodes the salient aspects of image gradients in the feature point's neighborhood and kernel projections to orientation gradient patches. Ke and Sukthankar [17] applied PCA to normalize gradient patch in PCA-SIFT. On selective small dataset, Lu and Little [18] used HOG descriptor for athlete images representation and the results were projected to linear subspace using PCA for effective and speedy retrieval. HOG with PCA was pedestrian dataset [19] for which subset is selected for feature reduction with forward or backward sampling. HOG with PCA was tested on 3D human posture [20] where PCA was applied on every HOG block and experimentation results showed that shape context has highest RMS (Root-mean-square) error. Human action detections from a far field of view [21] are detected using HOG for accurate object detection. Supervised Principle Component Analysis (SPCA) is applied to obtain informative principle components. For this, Support Vector Machine (SVM) classifier is trained and tested on selective normal and low resolution image datasets. Manifold learning algorithm based Locality Preserving Projection (LPP) was used with HOG to attain short response time in real time applications [22]. Linear LPP algorithm transformation vector results in reduced computational complexity and less input descriptor redundancies. HOG extraction was performed on local 16×16 pixel regions and then LPP was used to preserve data similarity after projection. MIT CBCL pedestrian database of 1000 images was used for experimentation and the results were better than PCA-SIFT [17]. In a hand gesture recognition system, hand images are first transformed into grid of HOG and then incremental PCA is applied to obtain compact candidates on which particle filter method is applied. Results of this scheme are robust to changes and show better accuracy [23]. For image extraction in large database located at remote server [24], pyramid HOG descriptor is first computed on images and then images are encrypted using RSA double key algorithm and stored at remote server. It causes an overhead for encryption, decryption and authentications.

HOG and LBP are individually used for feature vectors production but sometimes their accumulated outcome is better. For example, in pedestrian detection, PCA is applied to HOG and LBP based extracted feature vectors and then reduced HOG and LBP results are combined to retrieve the images. Thus PCA-HOG-LBP [25] descriptor improves the accuracy but involves in extensive computation. For eye detection, PCA is applied, followed by whitening transformation on HOG feature sets and then discriminative analysis is performed to reduce feature space [26]. Experiments on Face Recognition Grand Challenge (FRGC) are encouraging. Center-Symmetric Local Binary Pattern (CS-LBP) as a feature set is used in combination with Histogram of Oriented Gradients for the detection of crow birds. Linear Support Vector Machine is used for training, testing and better results classification [27]. Based on background elimination using statistical features [28], object area is filtered using local histogram distribution and then trained classifier detects human faces and cars using HOG and PCA. Finally, SVM classifies using ROC-AUC parameters. This approach has better precision and recall rates. ROI (Region of Interest) based method intends to select and interact with the relevant area in an image. ROI based shape and region extraction is easily performed using edge based histogram of oriented gradients. HOG was not applied to the whole image [29]. At first, region of interest is determined and then HOG is applied to that portion only. Optical flow based tracking limits the HOG computation in redundant frames. PCA is then applied to reduce feature vectors and proposed dynamic ROI selection method minimizes detection time and increases accuracy. Fisher criterion and multi-scale method [30] was employed to reduce feature space obtained by HOG where histogram of oriented gradients were applied to selective ROI and support vector machine was adopted to classify the results. Experiments showed improved accuracy due to selective ROI applicability.

This paper presents an agile medical image retrieval system that initially extracts feature vectors of images using local image descriptor by HOG. Secondly, these massive signatures are compacted by applying PCA. Thirdly, the highest variance based data is projected from first coordinate as a sole descriptor. Experimentation is performed to achieve high dimension reduction using PCA, performance gain, accuracy with low space and time consumption.

2. Methodology

Various interest point descriptors SIFT [8], SURF [9],

GLOH [10], KPB-SIFT [16], PCA-SIFT [17], LPP-HOG [22] and HOG-LBP [25] are used in image processing and computer vision for the purposes of object recognition, motion estimation and event detection. Local gradients orientation density counting is well suited for shape and object detection that is employed by HOG [11]. This methodology is based on SIFT [8] descriptor with a variation that it is performed on a dense grid with well-normalized histograms. Practically, an image is sliced into small spatial cells which accumulating a local edge orientation over the pixels of the cell. Large spatial blocks are used for contrast and normalization for better classification. HOG captures edge or gradient orientation that is an attribute of local shape and is performed at local representation with an easily controllable degree of invariance to local geometric and photometric transformations as shown in Fig. 1 [11].

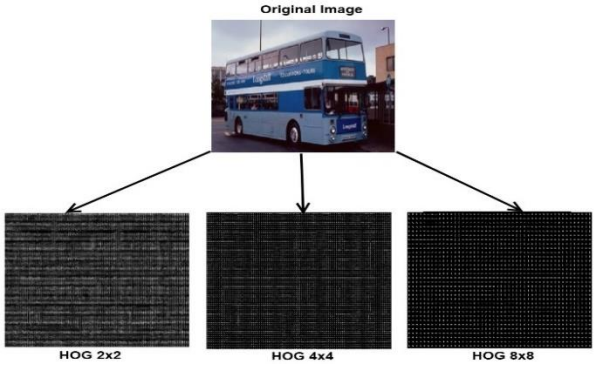

Fig. 1: HOG feature representation in 2×2, 4×4 and 8×8 block sizes.

$$\begin{cases} f_x(x,y) = I(x+1,y) - I(x-1,y) & \forall x,y \\ f_y(x,y) = I(x,y+1) - I(x,y-1) & \forall x,y \end{cases} \quad (1)$$

The strength of gradient *m(x, y)* is calculated as [11]:

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \quad (2)$$

Edge orientation is calculated using Sobel filter at points *x*, *y* [11]:

$$\theta(x,y) = tan^{-1}\left[\frac{f_y(x,y)}{f_x(x,y)}\right] \quad (3)$$

Then, for each image gradient cell, unsigned orientation $\tilde{\theta}(x,y)$ is quantized into orientation bins [11]:

$$\tilde{\theta}(x,y) = \begin{cases} \theta(x,y) + \Pi & if\ \theta(x,y) < 0 \\ \theta(x,y) & \text{otherwise} \end{cases} \quad (4)$$

For each 4×4 pixels size, cell histogram of edge gradient is computed with 8 orientations that results a feature vector (*v*). The feature vectors are computed for each cell and block. Normalized feature vector b'_{xy} for each block is computed as [11]:

$$b'_{xy} = \frac{b_{xy}}{\sqrt{\|v\vec{\ }\|^2 + \varepsilon}} \qquad \varepsilon = 1 \quad (5)$$

where $\varepsilon = 1$ and b_{xy} is unnormalized block histogram of cell at position *x*, *y*.

For each 4×4 pixels cell with 8 orientations, 128 (i.e., 8×4×4) HOG features are extracted as shown in Fig. 2 [11, 31, 32].

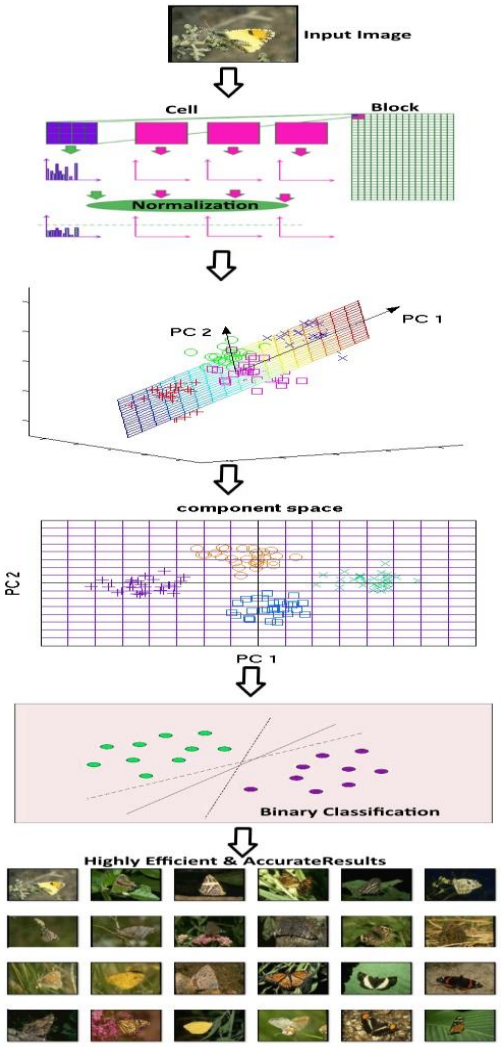

Fig. 2: The proposed descriptor scenario in CBIR system.

Hence, the feature vectors extraction for all grid locations grow from thousands to millions, based on image dimensions. These bulk feature vectors are computed for overlapped blocks so their produced vectors are redundant and unnecessary. These also cause computational overhead, performance degradation and greater time and space consumption. Therefore, correlated variables need a transformation to linearly uncorrelated variables that is achieved in principal components. PCA is a useful, simple, non-parametric statistical analytical tool using in face recognition, image compression and finding patterns in data of high dimensions [33]. It is helpful to explore, sort, group data and to reveal underlying hidden structures. It also quantifies the importance of each dimension and shows total variance with a few components without loss of information. A smaller database is used because only the trainee images are stored in the form of their projections on a reduced basis [34].

Principal components for high dimensional feature vectors are computed as [31]:

$$x = p^T(y_i - \bar{y}) \quad (6)$$

where covariance matrix is used for the projection matrix p. For random vectors X and Y with m-dimensions, the covariance matrix is equal to [35]:

$$\sigma(X,Y) = E[(X - E[X])(Y - E[Y])^T] \quad (7)$$

where T is transpose and $E[X]$ is mean of vector X.

Histograms of Oriented Gradient describe the shape and appearance of local objects by edge orientations. Histogram of gradient directions for all pixels and contrast-normalization for each cell within the block generate enormous and redundant values having highly dimensional data. For example, in Corel data set [36, 37], each image dimension is 384×256, on which gradient orientation counting in localized patterns produce almost 30 thousand values for each image. For this dataset, only 100 images computation takes remarkable calculation time and space consumption. HOG evaluation for thousand images [38-41] in small datasets outcome in performance degradation and resource consumption. Hence, CBIR using large datasets with SIFT and SURF are not workable due to huge, redundant and high dimensional data values [42] and same is the case with HOG. Therefore, it is an essential requirement to compact HOG colossal dimensions without loss of information. PCA [43] comes to front as a potential candidate with its key advantages like low noise sensitivity, less memory requirements, lack of redundancy, reduced complexity and smaller database representation [44]. Principal components returned by PCA are linearly uncorrelated and less than or equal to input observations. We have tested that HOG feature vectors are reduced up to more than 400 times by applying PCA. In other words, for a single image [36, 37], 30000 HOG feature vector values are converted to 60 - 70 principal components. So, for a database of 10 - 50000 images with reduced PCA-HOG components, considerable saving is achieved in term of time and memory consumption.

It is a challenging requirement for large CBIR systems that there should be some slim descriptor to represent an image with a few dimensions. It should also be capable to retrieve results with enhanced performance and better accuracy. To attain this goal, principal components should be selective based on some criteria. Stepwise Forward Selection (SFS) algorithm or Stepwise Backward Selection (SBS) algorithm [19] and other techniques incorporated for generalization performance with the aim of principal component subset selection. These techniques employ 20 to 100 principal components as descriptor candidate. We introduce a novel way to represent an image with only one principal component value.

In our proposed method, HOG is first computed for all grid locations to extract feature vectors. Signatures extracted from all uniformed spaced cells are bulk and performance concerned. For large datasets, these huge volume signatures and highly dimensional data are then projected to linear subspace in compacted principal component form without loss of accuracy by applying PCA. These principal components are descriptor candidates from which selective principal components are chosen. In the present case, one dimension is selected to exclusively represent an image uniquely.

The descriptor is used as an input to linear SVM for classification. The first PC is selected as a local descriptor because it has the largest possible variance [45, 46] that is accounted for 71% of variance of almost all data [47]. Only first principal component is assumed as a measure of economic status [48]. According to Porter et al (46), the first principal components can be used to obtain the highest variance with providing a compact and more efficient description. They experimentally proved on monochromatic aberrations of the human eye in a large population data that first principal component accounts for over 90% of the cumulative variance. Wall et al. [49] observed 90% of variance in first two principal components with the result that the first principal component contains a strong steady-state signal. By Hubert et al. [50], 85% of total variance is accounted for first two principal components.

Principal component factor structure is computed as [31]:

$$S = VL^{1/2} \quad (8)$$

where S is a matrix whose elements are the correlations between the principal components and the variables [43]. V is premultiplier used in the calculation of transformed variables. L is a diagonal matrix with λj in the j^{th} position on the diagonal. So the full eigenstructure of correlation matrix R is given by Veen [51]:

$$RV = V \quad (9)$$

$$V'V = VV' = 1 \quad (10)$$

$$V'RV = L = D_y \quad (11)$$

where D_y is diagonal matrix whose all off-diagonal elements are zero. Thus orthogonal transformation is achieved.

The working methodology involves the following steps:

2.1 Selection of Database

Our algorithm selects some database for the testing and training purpose depending upon the connection string path. In that database the semantic groups of images are in some predefined range. In our case, each image category contains hundred images which are next to each other and the next image category starts just after the last image of a category.

2.2 HOG Computation

HOG is computed on each image and the results are temporarily saved in a file before the computation of PCA. These results are not saved permanently because each image HOG computed value is massive, multidimensional and redundant.

2.3 PCA Computation

PCA performs simplification, un-mixing, prediction, classification, variable selection, outlier detection, modeling

and data reduction. HOG [11] is computed for a semantic batch of images and PCA is calculated for those images and the values are saved in a file. For the next batch of images, the same process is repeated until all image categories are computed. Numbers of principal components are changed to check their impact on computational time, precision and recall, memory space consumption and permanent storage file size for large databases.

2.4 SVM Training and Testing

Support Vector Machine (SVM) performs supervised learning by analyzing data and recognizing the hidden patterns for classification. Set of *n* items using linear SVM are shown as [32]:

$$T = \{(a_x, b_x) |\, a_x \in R_i, b_x \in \{-1, 1\}\}, x = 1 \ldots n \quad (12)$$

where a_x is *i* dimensional real vector and b_x is either -1 or 1 for the training data *T*. a_x belongs to R_i where *i* represents the exponent for which floating values are extracted. These values are iterated from 1 to n; where *n* are the number of items used in SVM data set.

We input training data to SVM with Boolean data labels. Data labels show that the respective image belongs to this category or not, so that, based on this training data, discriminative classification is carried out on test data.

2.5 Compression and Retrieval Mechanism

The localized pattern produces massive and redundant signatures, for example, a JPEG image with dimension 384×256 and 25 to 35 KB in size generates 0.55-1.25 million values. The proposed method intelligently chooses the uncorrelated variables by applying the succeeding component's constraint dependent orthogonal variance. The high value of initial few components are the best candidates to show the variance. The compact values based large datasets are represented by their primary components to retrieve the images. Therefore, less comparisons on binary classification produces very fast and accurate results based on the potential representative coefficients. The large datasets with variety of semantic groups save each image with the single coefficient with highest variance to represent it uniquely. It is, therefore, resulted in quick processing and less space consumption. The unique coefficients or a subset of coefficients are classified by supervised learning for which the top 20 or any number of sorted values based candidates are displayed.

3. Experimentation

3.1 Dataset

Dataset selection, training and testing methodologies impact the results precision. The results tested only at a few number of images or at small databases outcome differently at large databases with diverse image types. For this reason, we used two databases. First, a medium size database is for training, testing and results precision. The objective is to verify the results accuracy along with performance gain, time and space consumption. The results are also validated on a large database for the space and time consumption.

There is no standard testbed for CBIR systems, however, a Corel database is usually used by researchers due to different semantic groups and free availability. Corel datasets contain a variety of semantic groups like people, animal, flower, texture, food and others. The first dataset is Corel [36, 37] containing 1000 images in 10 categories as shown in Fig. 3. All are 384×246 pixel images.

Fig. 3: Corel thousand images dataset

The second dataset is Corel comprising of 10,000 images in 100 categories with 128×85 image size as shown in Fig. 4. The categories include butterflies, scenes, buildings, nature, sunset, flowers, trees, textures, cars, boats etc.

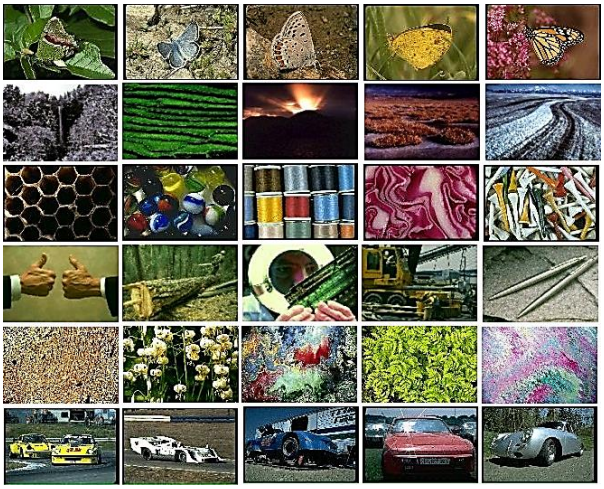

Fig. 4: Corel ten thousand images dataset.

3.2 Results and discussions

3.2.1 Precision and recall evaluation

Precision is the specificity measure or positive predicted values, whereas, recall is the sensitivity measure or true positive rate evaluation. Precision and recall are calculated on each category of images for small or large databases. These results are tested on different sets of training and testing data.

$$Precision = \frac{number\ of\ relevant\ images\ retrieved}{total\ number\ of\ retrieved\ images} \quad (13)$$

$$recall = \frac{number\ of\ relevant\ images\ retrieved}{total\ number\ of\ relevant\ images} \quad (14)$$

3.3 Experimental Results

Experimentation is performed on Corel dataset of 1000 images in 10 semantic groups. For this, Intel 1.5 GHz machine with 4GB RAM was employed. For training, 38 positive images were used for a category and 3 negative samples from each of the rest of categories. Precision and recall are evaluated for each category by applying positive and negative images for training.

Precision and recall rate are also computed for different principal components based descriptors. One principal component per image is our slim descriptor as compared to 3, 5, 10, 50 or 100 principal components to describe an image. Principal components from 1 to 100 are tested and the recall rates for each number of components are shown in the Figs. 5 - 9. It is noted that in all these categories, the slim descriptor has the highest recall rate.

3.3.1 Recall rates

Recall rates are tested for different number of PCs on different image categories and specified sampling rates. Fig. 5 shows the proposed slim descriptor representing an image with highest recall rate. In this experiment, various number of PCs are tested for recall rate on all image categories of Corel image dataset.

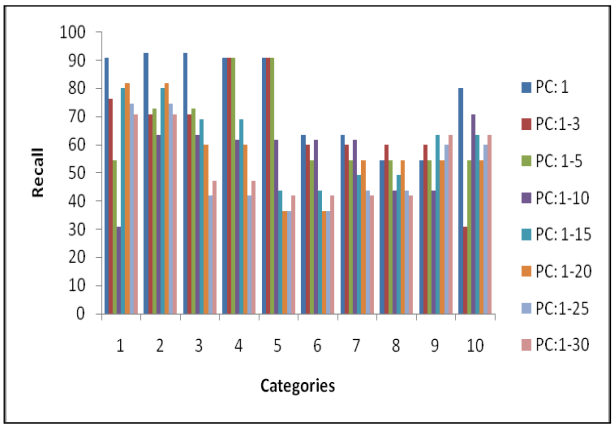

Fig. 5: Recall rates for Corel dataset with 10 categories for the proposed descriptor vs. 1-30 PCs.

Fig. 6 shows the recall rates for the proposed slim descriptor and the rest of principal components descriptors. It is evident from the figure that the proposed slim descriptor has the highest recall rates among all the principal components descriptors in all the image categories. This is due to the highest possible variance of first principal component among all the data [45-48].

Fig. 7 shows the recall rates for the proposed slim descriptor for 10 categories of Corel image dataset. The results tested on 1000 images show that the proposed slim descriptor has outstanding performance and significant recall rates.

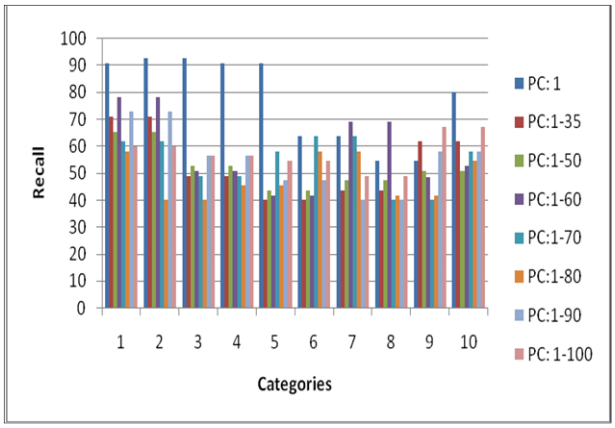

Fig. 6: Recall rates for Corel dataset for 50-100 PCs vs. the proposed descriptor.

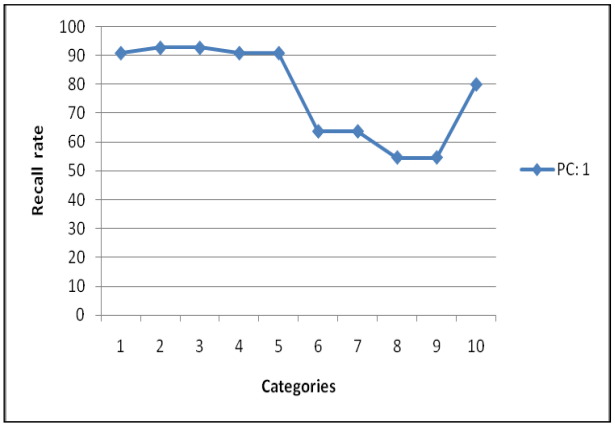

Fig. 7: Recall rates for the proposed slim descriptor tested for 10 semantic groups of thousand images.

For these categories, recall rates of other descriptors are also evaluated. Fig. 8 shows the recall rates for all categories for all descriptors. These results show that our proposed descriptor has still the better recall rates as compared to other descriptors. In categories 6-9, low recall rates using other descriptors can also be observed as shown in Figs. 5 and 6.

Fig. 8 shows the aggregate recall rate for different principal components tested on 10 categories of Corel images dataset. The results indicate that the recall rate for slim descriptor is the highest as compared to different number of principal components describing an image.

Fig. 9 shows the accumulated recall rates in all 10 semantic groups for the proposed slim descriptor and different other descriptors. It is noted that the proposed slim descriptor outperforms in accumulated recall rates. However, significant difference in recall rates can also be observed. These results are produced by performing the tests on different image combinations, sampling rates and image groups.

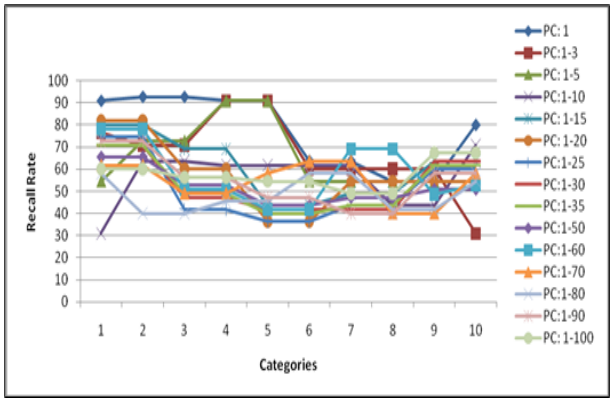

Fig. 8: Recall rate for different PCs using Corel dataset with the proposed descriptor.

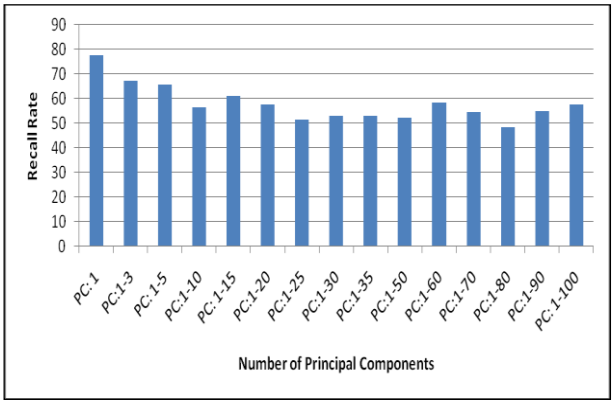

Fig. 9: Accumulated recall rates for the proposed descriptor vs. PCs.

3.3.2 *Precision*

Precision is evaluated for different image categories using different number of principal components. The proposed slim descriptor shows better accuracy and retrieval in the dataset as is shown in Fig. 10. Precision level of the proposed descriptor is remarkable in many categories. It shows a tradeoff between improved precision in large number of PCs and minor precision gap due to slim descriptor in some cases.

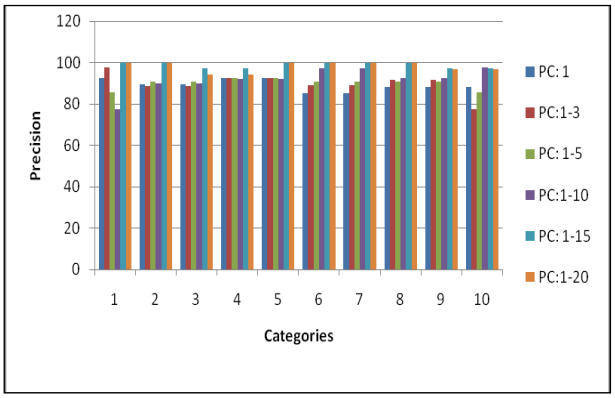

Fig. 10: Precision results using Corel dataset using the proposed descriptor vs. different PCs.

Fig. 11 shows short descriptor precision results. It reveals that the proposed slim descriptor shows an overall better and accurate retrieval. Similarly, precision rates for slim descriptor and other descriptors are shown in Fig. 12, which indicates that slim descriptor has remarkable performance among all possible PCs based combinations. This achievement is due to the single coefficient descriptor with largest variance.

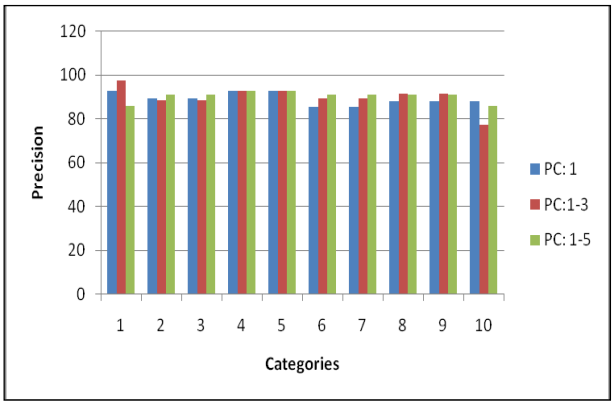

Fig. 11: Precision results of Corel dataset using the proposed descriptor.

Precision of slim descriptor for all 10 categories in 1000 images database is shown in Fig. 13. Our proposed descriptor depicts good performance in all categories. It is also noticed that the precision results are reliable for simple object images to complex natural scenes and textures.

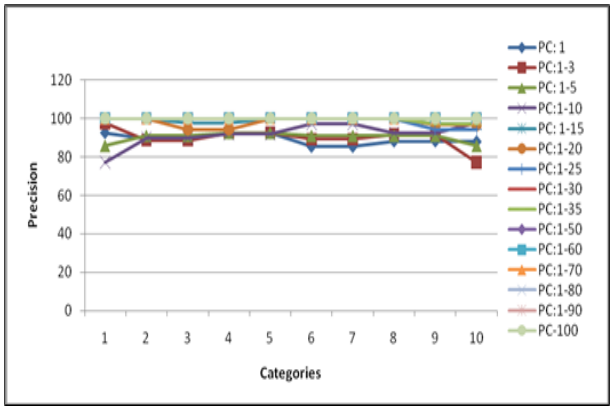

Fig. 12: Precision rates for all categories of Corel dataset with the proposed slim descriptor vs. different PCs.

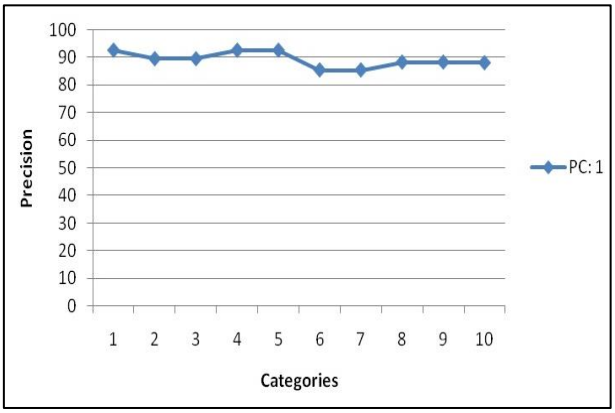

Fig. 13: Precision rates for all categories of Corel dataset with the proposed descriptor.

3.3.3 *Computation time*

Computation time is a main concern for small and especially for large databases. It has an important impact on images listing and retrieval from database, HOG computation, PCA reduction, SVM training and testing for the number of principal components per image and for the total number of images. HOG computation returns a massive and highly dimensional data for each image. Fig. 14 shows the calculation time for HOG values plus the aggregate computation time for the various number of principal components. From these results we note that computational time increases proportionally as the numbers of principal component increase. From the results, it is very much clear that the proposed descriptor not only exhibits reliable precision and recall rates but also minimizes the computation time.

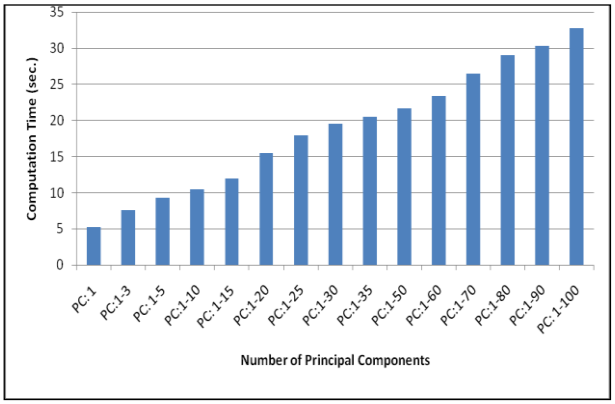

Fig. 14: HOG and PCA computation time for the number of principal components.

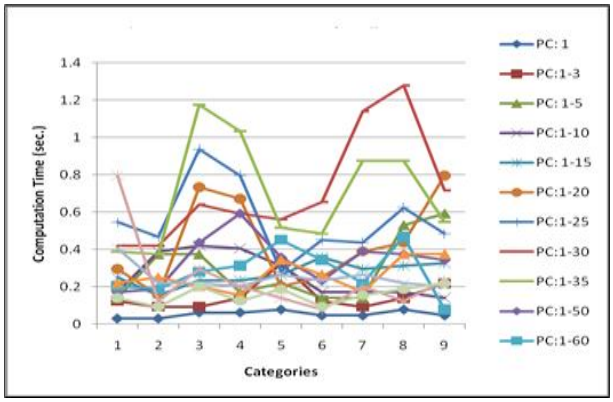

Fig. 15: HOG and PCA computation time for the number of principal components.

HOG and PCA computation time for all categories with all descriptor sizes are shown in Fig. 15. The proposed descriptor shows the lowest computation time which confirms the best precision and recall rates. From these results, it is clear that the proposed descriptor outperforms for small, medium and large databases.

3.3.4 *Data formation time*

Prior to compute HOG, images data is formulated with labels for the purpose of training. These steps take considerable computation time. Fig. 16 shows the calculation and evaluation time. The proposed descriptor performs well in all image categories. It means that the intermediate steps time consumption for proposed descriptor is less as compared to other tested descriptors.

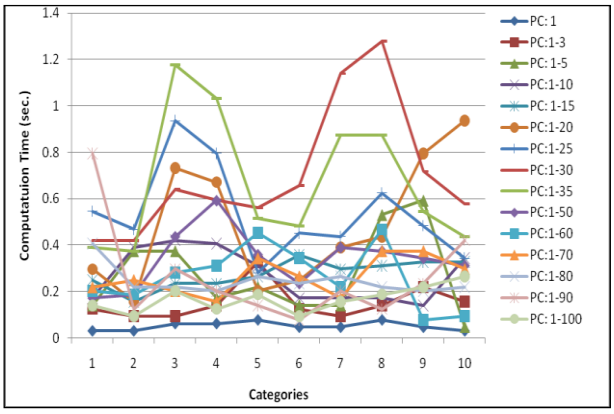

Fig. 16: Data formation, listing and labeling etc. time consumed by descriptors

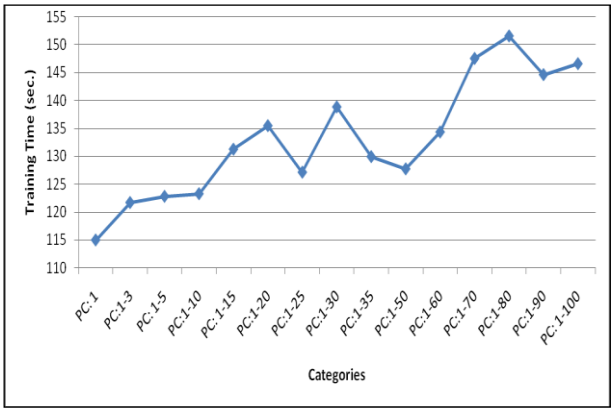

Fig. 17: Accumulated training time for proposed descriptor and other descriptors.

3.3.5 *Training time*

Training time for all image categories for the proposed and other descriptors is shown in Fig. 17, which illustrates that the proposed descriptor has the least training time due to its slim size. A single PC value represents an image thus training is performed for a very few values.

3.3.6 *Memory consumption*

Memory space consumption plays an important role especially in large databases. The large data associated with an image is required to be processed and stored permanently. Intermediate processing steps need paging and data transfer tasks. Bulk data image representation cause massive memory resources in large databases.

Fig. 18 shows memory space consumption to store image data on permanent storage for small, medium and large databases. It is observed that, the proposed descriptor needs very small storage for small, medium and large databases as compared to other descriptor sizes.

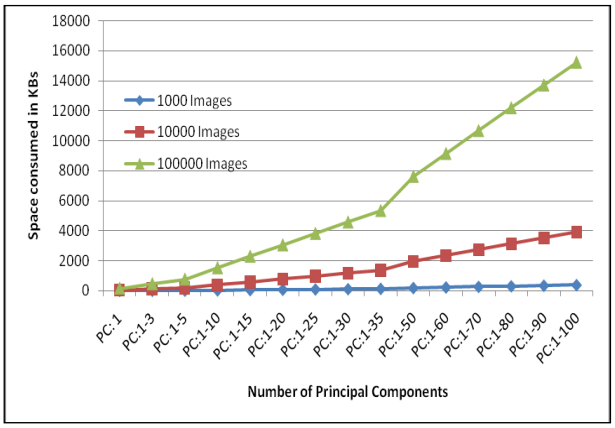

Fig. 18: Storage used by different sizes of databases using different descriptor sizes.

The space consumption in small database with respect to different sizes of descriptors is also shown in Fig. 19. It is clear that the proposed descriptor consumes very small space as compared to big descriptors. For large database, the impact of large descriptor size is given in Fig. 20, which illustrates a difference in database sizes for large content based image retrieval system.

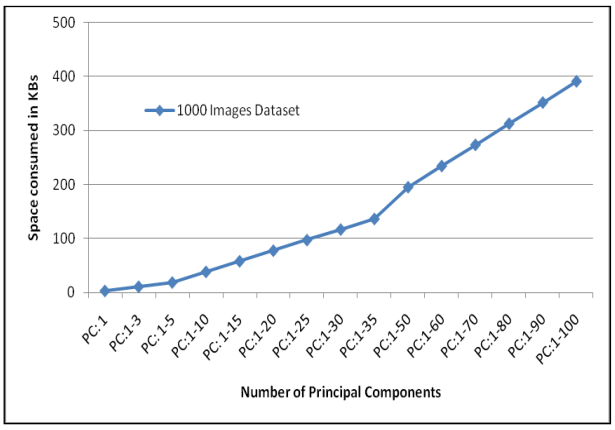

Fig. 19: Memory space consumed by proposed descriptor and other descriptors.

3.3.7 *Comparison with benchmark descriptors*

Proposed descriptor is evaluated experimentally with existing standardized descriptors. The performance parameters like precision, recall rates, performance gain, storage and time consumption are studied for PCA-SIFT [17], PCA-HOG and kernel PCA HOG [52] versus the proposed descriptor. The results are extracted by conducting the experimentation on Corel 1000 images dataset. PCA-SIFT algorithm [13], Kernel PCA and the proposed descriptor are implemented at current testing framework. Fig. 21 depicts the recall rates of the proposed descriptor with respect to PCA-HOG, PCA-SIFT and kernel PCA-HOG.

It can be observed that the recall rates for the proposed descriptor are too high in all benchmarks in all image categories.

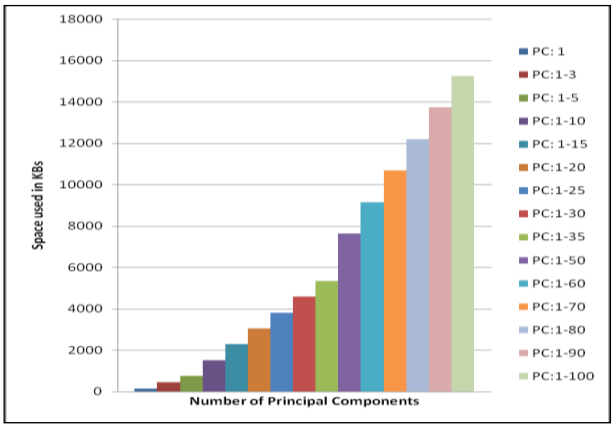

Fig. 20: Space consumption with large database for different descriptors.

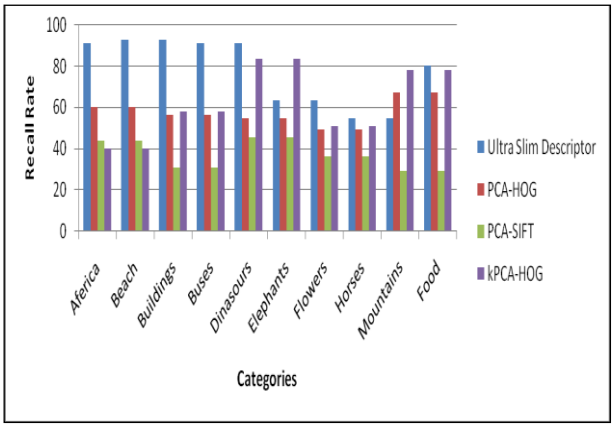

Fig. 21: Recall rates for the proposed descriptor versus benchmark descriptors.

Fig. 22 shows the precision rates for the descriptors. All the descriptors show good precision results. The proposed descriptor has a small precision tradeoff due to its ultra slim size that uniquely represent an image while PCA-HOG, PCA-SIFT and kPCA-HOG contain 250, 109 and 144 values for each image representation, respectively.

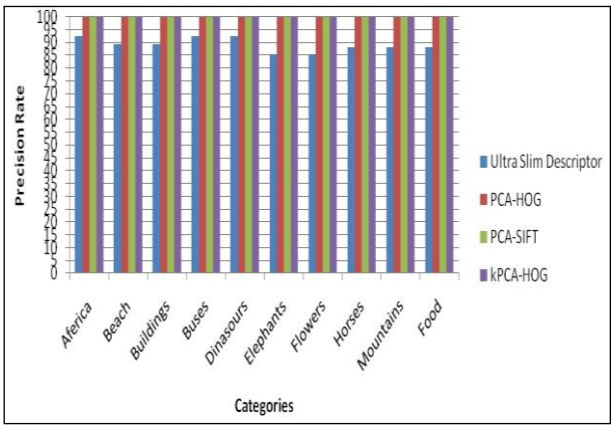

Fig. 22: Recall rates for the proposed descriptor versus benchmarks descriptors.

Feature vectors calculation, training and testing times are shown in Fig. 23. Due to its size, the proposed descriptor has very low signature extraction, training and testing time. For the same dataset, other benchmarks take comparatively more time due to descriptor length. It has a direct impact on large databases with millions of images where the proposed descriptor shows outstanding performance due to its less feature extraction, training, testing and searching times.

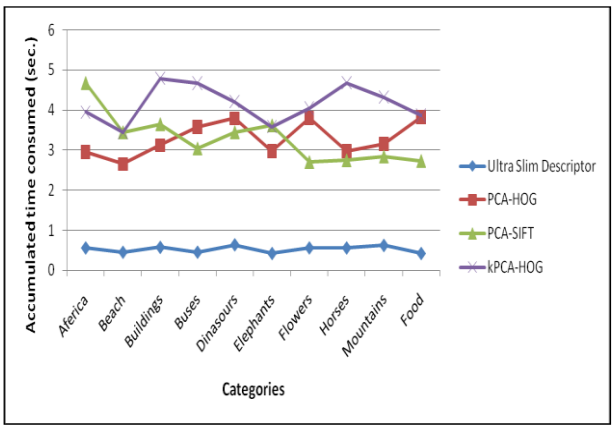

Fig. 23: Feature extraction, training, classification and searching time aggregated for proposed descriptor, PCA-HOG, PCA-SIFT and kPCA-HOG using Corel image dataset.

Descriptor size is directly proportional to storage consumption. Fig. 24 shows the storage consumed by different descriptors for Corel dataset. Image descriptor storage consumption with proposed descriptor is hundred to thousand times less than the existing benchmarks. It can be observed that the proposed slim descriptor has efficient and accurate image retrieval results with low space and time consumption.

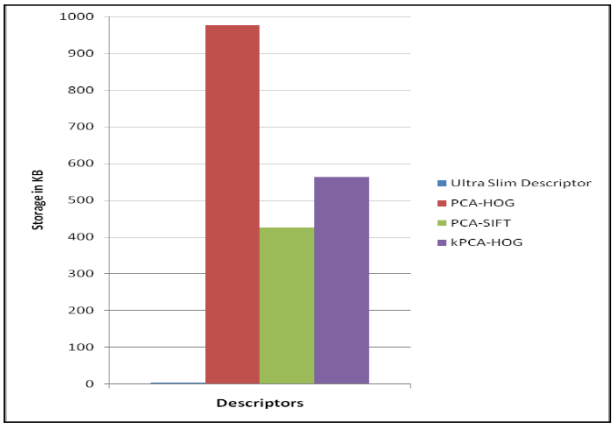

Fig. 24: Corel 1000 images dataset storage used by different descriptors versus proposed descriptor.

4. Conclusions

In this paper, a novel slim descriptor for fast and reliable image retrieval has been presented. This descriptor is capable of producing reliable results at low cost of time and space. The performance of the proposed descriptor has been evaluated with different datasets. As a result, the proposed descriptor shows better precision and recall rates as compared to other descriptors. The significant performance has been gained and efficient image retrieval with good accuracy has also been achieved.

Acknowledgement

This research was supported by the Ministry of Trade, Industry & Energy (MOTIE, Korea) under the Industrial Technology Innovation Program (No.10063130) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1A2C1006159).

Conflicts of Interest

"The authors declare no conflict of interest."

References

[1] M.J. Swain and D.H. Ballard, "Color indexing", Int. J. Comput. Vis., vol. 7, no. 1, pp. 11-32, 1991.

[2] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu and R. Zabih, "Spatial color indexing and applications", Int. J. Comput. Vis., vol. 35, no. 3, pp. 245-268, 1999.

[3] J. Han and K.K. Ma, "Fuzzy color histogram and its use in color image retrieval", IEEE Trans. on image Process., vol. 11, no. 8, pp.944-952, 2002.

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

[5] K.T. Ahmed, A. Irtaza and M.A. Iqbal, "Fusion of local and global features for effective image extraction", Appl. Intell., vol. 47, no. 2, pp. 526-543, 2017.

[6] T. Gevers and A.W. Smeulders, "Pictoseek: Combining color and shape invariant features for image retrieval", IEEE Trans. on image Process., vol. 9, no. 1, pp. 102-119, 2000.

[7] J. Shi and J. Malik, "Normalized cuts and image segmentation", IEEE Trans. on Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 888-905, 2000.

[8] D.G. Lowe, "Object Recognition from Local Scale-Invariant Features", IEEE Int. Conf. Comput. Vis., vol. 2, pp. 1150-1157, 1999.

[9] H. Bay, T. Tuytelaars and L. Van Gool, "Surf: Speeded up robust features", Eur. Conf. Comput. Vis., pp. 404-417, Springer, Berlin, Heidelberg, 2006.

[10] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors", IEEE Trans. on Pattern Anal. Mach. Intell., vol. 27, no. 10, pp. 1615-1630, 2005.

[11] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection", IEEE Comput. Society Conf. Comput. Vis. Pattern Recog., vol. 1, pp. 886-893, 2005.

[12] K.S. Goh, E. Chang and K.T. Cheng, "Support vector machine pairwise classifiers with error reduction for image classification", Proc. of the 2001 ACM workshops on Multimed.: multimed. Inf. Retr., pp. 32-37, 2001.

[13] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Color- and texture-based image segmentation using EM and its application to content-based image retrieval", 6th Int. Conf. Comput. Vis., pp. 675-682, 1998.

[14] F. Jing, B. Zhang, F. Lin, W.Y. Ma and H.J. Zhang, "A novel region-based image retrieval method using relevance feedback", Proc. of the 2001 ACM workshops on Multimed.: Multimed. Inf. Retr., pp. 28-31, 2001.

[15] M. Ortega, Y. Rui, K. Chakrabarti, K. Porkaew, S. Mehrotra, and T.S. Huang, "Supporting ranked boolean similarity queries in MARS", IEEE Trans. Knowl. Data Eng., vol. 10, no. 6, pp. 905-925, 2002.

[16] G. Zhao, L. Chen, G. Chen and J. Yuan, "KPB-SIFT: a compact local feature descriptor", Proc. of the 18th ACM Int. Conf. on Multimed., pp. 1175-1178, 2010.

[17] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors", IEEE Comput. Society Conf. on Comput. Vis. Pattern Recognit., vol. 2, pp. 1063-6919, 2004.

[18] W.L. Lu and J.J. Little, “Simultaneous tracking and action recognition using the pca-hog descriptor”, Proc. of 3rd IEEE Can. Conf. on Comput. Robot Vis. (CRV'06), pp. 6-6, 2006.

[19] T. Kobayashi, A. Hidaka and T. Kurita, “Selection of histograms of oriented gradients features for pedestrian detection”, Int. Conf. on neural Inf. Process., pp. 598-607, 2007.

[20] K. Onishi, T. Takiguchi and Y. Ariki, “3D human posture estimation using the HOG features from monocular image”, 19th IEEE Int. Conf. on Pattern Recognit., pp.1-4, 2008.

[21] C.C. Chen and J.K. Aggarwal, "Recognizing human action from a far field of view”, IEEE Workshop on Motion and Video Comput., WMVC'09, pp. 1-7, 2009.

[22] Q.J. Wang and R.B. Zhang, “LPP-HOG: A new local image descriptor for fast human detection”, Proc. of IEEE Int. Sym. on Knowl. Acquis. Model. Workshop, pp. 640-643, 2008.

[23] H. Yang, Z. Song and R. Chen, “An incremental PCA-HOG descriptor for robust visual hand tracking”, Int. Symp. on Vis. Comput., pp. 687-695, 2010.

[24] M.M. Asha and J.J. Ranjani. "Secure image retrieval using pyramid histogram of oriented gradient descriptor", Proc. of IEEE Int. Conf. Adv. Comput. Comm. Syst. (ICACCS2013), pp. 1-5, 2013.

[25] R.W. Sun,T. Rui, J.L. Zhang and Y. Zhou, “Pedestrian detection by PCA-based mixed HOG-LBP features”, Proc. of the 4th Int. Conf. on Internet Multimed. Comput. Serv., pp. 92-95, 2012.

[26] S. Chen and C. Liu, "Precise eye detection using discriminating hog features.", Proc. of Comput. Anal. Images Patterns, pp. 443-450, 2011.

[27] K. Mihreteab, M. Iwahashi and M. Yamamoto, "Crow birds detection using HOG and CS-LBP.", Int. Symp. on Inteligent Signal Process. Comm. Syst. (ISPACS), pp. 406-409, 2012.

[28] A.A. Fathima,V. Vaidehi, N. Rastogi, R.M. Kumar and S. Sivasubramaniam, “Performance analysis of multiclass object detection using SVM classifier”, Proc. of IEEE Int. Conf. on Recent Trends in Inf. Technol. (ICRTIT), pp. 157-162, 2013.

[29] K.V. Suresh, "HOG-PCA descriptor with optical flow based human detection and tracking.", Proc. of IEEE Int. Conf. on Comm. and Signal Process. (ICCSP 2014), pp. 900-904, 2014.

[30] J. Meng and S. Li, “Pedestrian detection based on the improved HOG features”, Proc. of 2nd Int. Workshop on Mater. Eng. Comput. Sciences. Atlantis Press, (IWMECS, 2015), vol. 1, pp. 701-704, 2015.

[31] S. Wold, K. Esbensen and P. Geladi, "Principal component analysis", Chemometr. intell. Lab. syst., vol. 2, no. 1-3, pp. 37-52, 1987.

[32] T. Joachims, "Transductive inference for text classification using support vector machines", ICML, vol. 99, pp. 200-209. 1999.

[33] J. Meng and Y. Yang, “Symmetrical Two-Dimensional PCA with Image Measures in Face Recognition”, Int. J. Adv. Robotic Syst., vol. 9, pp. 1-10, 2012.

[34] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, “Overview of the Face Recognition Grand Challenge”, Proc. of IEEE Comput. Society Conf. on Comput. Vis. pattern recognit. (CVPR 2005), pp. 947-954, 2005.

[35] W.J. Krzanowski "Selection of variables to preserve multivariate data structure, using principal components," Appl. Stat., pp. 22-33, 1987.

[36] J. Li and J.Z. Wang, "Automatic linguistic indexing of pictures by astatistical modeling approach", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1075-1088, 2003.

[37] J.Z. Wang, J. Li and G. Wiederhold, "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture libraries", IEEE Trans. on Pattern Anal. Mach. Intell., vol. 23, no.9, pp. 947-963, 2001.

[38] P.S. Hiremath and J. Pujari, "Content based image retrieval using color, texture and shape features.", Proc. of IEEE Int. Conf. on Adv. Comput. Comm. (ADCOM 2007), pp. 780-784, 2007.

[39] F. Jing, M. Li, H.J. Zhang and B. Zhang, "Support vector machines for region-based image retrieval.", Proc. of IEEE Int. Conf. on Multimedia and Expo, (ICME'03), vol. 2, pp. 11-21, 2003.

[40] S. Chatzichristofis and Y. Boutalis, "A hybrid scheme for fast and accurate image retrieval based on color descriptors", Proc. of Int. Conf. on artif. intell. soft comput. (ASC 2007), pp. 280-285, 2007.

[41] Y. Zhu, W. Mio and X. Liu, "Optimal factor analysis and applications to content-based image retrieval", Proc. Int. Conf. Comput. Vis. Comput. Graph., pp. 164-176, 2008.

[42] R.E.G. Valenzuela, W.R. Schwartz and H. Pedrini, “Dimensionality reduction through PCA over SIFT and SURF descriptors”, Proc. IEEE 11th Int. Conf. Cybernetic Intell. Syst. (CIS 2012), pp. 58-63, 2012.

[43] I. Jolliffe, "Principal component analysis", Int. encyclopedia stat. sci., pp. 1094-1096, 2011.

[44] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, “Overview of the Face Recognition Grand Challenge”, Proc. of IEEE Comput. Society Conf. Comput. Vis. pattern recognit., vol. 1, pp. 947-954, 2005.

[45] P. Nomikos and J.F. MacGregor, "Monitoring batch processes using multiway principal component analysis", Am. Inst. Che. Eng. j., vol. 40, no. 8, pp. 1361-1375, 1994.

[46] J. Porter, A. Guirao, I.G. Cox and D.R. Williams, “Monochromatic aberrations of the human eye in a large population”, J. Opt. Soc. Am. A, vol. 18, no. 8, pp. 1793-1803, 2001.

[47] K.J. Friston, C.D. Frith, P.F. Liddle, and R.S.J. Frackowiak. "Functional connectivity: the principal-component analysis of large (PET) data sets", J. Cereb. Blood Flow Metab., vol. 13, no. 1, pp. 5-14, 1993.

[48] T.A. Houweling, A.E. Kunst and J.P. Mackenbach, “Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter?”, Int. J. Equity Health, vol. 2, no. 1, pp. 1-12, 2003.

[49] M.E. Wall, A. Rechtsteiner and L.M. Rocha, "Singular value decomposition and principal component analysis", A practical approach to microarray data anal., pp. 91-109, Springer US, 2003.

[50] M. Hubert, P.J. Rousseeuw and K. Vanden Branden, “ROBPCA: a new approach to robust principal component analysis”, Technometrics, vol. 47, no. 1, pp. 64-79, 2005.

[51] B.D. Van Veen, "Eigenstructure based partially adaptive array design", IEEE Trans. Antennas Propag., vol. 36, no. 3, pp. 357-362, 1988.

[52] B. Schölkopf, A. Smola and K.R. Müller, “Kernel principal component analysis”, Proc. Int. Conf. artif. neural Netw., pp. 583-588, 1997.