IR-MF-SVMe: Image Retrieval using Multiple Features Extraction with Support Vector Machine ensemble

Venkataravana Nayak K.^{*1}, Arunalatha J. S.¹, Venugopal K. R.², ¹University Visvesvaraya College of Engineering, Bengaluru, India. ²Bangalore University, Bengaluru, India. e-mail: venkatcsuvce@gmail.com, aruna veeresh@yahoo.co.in, venugopalkr@gmail.com

Received: June 16, 2021. Received: November 12, 2021. Accepted: December 7, 2021. Published: December 22, 2021.

Abstract— Image representation is a widespread strategy of image retrieval based on appearance, shape information. The traditional feature representation methods ignore hidden information that exists in the dataset samples; it reduces the discriminative performance of the classifier and excludes various geometric and photometric variations consideration in obtaining the features; these degrade retrieval performance. Hence, proposed multiple features fusion and Support Vector Machines Ensemble (IR-MF-SVMe); an Image Retrieval framework to enhance the performance of the retrieval process. The Color Histogram (CH), Color Auto-Correlogram (CAC), Color Moments (CM), Gabor Wavelet (GW), and Wavelet Moments (WM) descriptors are used to extract multiple features that separate the element vectors of images in representation. The multi-class classifier is constructed with the aggregation of binary Support Vector Machines, which decrease the count of false positives within the interrelated semantic classes. The proposed framework is validated on the WANG dataset and results in the accuracy of 84% for the individual features and 86% for the fused features related to the state-of-the-arts.

Keywords— Color Auto Correlogram, Color Moments, Retrieval Accuracy, Support Vector Machine ensemble, Wavelet Moments.

I. INTRODUCTION

Due to the development of advanced innovation, a huge increment in the number of images put away in digital format. The range from individual accumulations to medical, logical images are presently gathered in databases. Numerous associations, clients presently obtain huge quantities of images, and critical to retrieve important multimedia assets and to adequately find matching images in large archives. In this context, Content-Based Image Retrieval (CBIR) systems are prevalent to browse, search, and retrieve images through an enormous dataset with the least human intervention.

The task of the image retrieval model is to search and arrange the images from the archive containing visual semantic relation with the query image with minimum human interaction with the system. The search engines over the web retrieve images by relying on the captions and metadata, the query is the text or keywords matched with the database images, the resulted in most of the images are not relevant. Because of differences in manual annotation and human visual perception. The automatic annotation of images based on the contents of the image depends much on determining the color, shape, texture oriented details of the developed system. But due to visual perception differences mislead the process of obtaining the images [1].

During the searching process, the user sees high-level or semantic features instead of low-level features such as color, shape, and texture of the image. Because of this reason, retrieval process performance becomes low. The large quantity of features extraction and feature space difficulties are the important challenges in the retrieval process [2]. So the focus is on the representation of images through key features extraction for increasing the accuracy and reducing the time of retrieval.

The CBIR framework overcomes the specified problems by considering the image contents visual analysis part in the feature extraction stage. The size of the database also plays an important factor because the need for comparison of database images with the query image in finding the image similarity is one of the important phases in the retrieval process. The features selected for the retrieval system are user-dependent,

Volume 15, 2021

discriminative feature representation, size of the feature vectors, and feature mapping in finding image similarity, about these to be taken care of to enhance the retrieval system performance [3].

In the existing system, the descriptors of Color Histogram, Edge Directed Histogram, Color Auto-Correlogram, Wavelet Texture, Scale-Invariant Feature Transform (SIFT) are used to extract corresponding features for representing dataset images. The features are stored in matrices and are transformed into latent space and the weighted latent features are combined. Then the features are optimized using the modified Nesterov Optimal Gradient approach. To reduce the prediction error, on the training data the prediction weights and extracted matrices are learned [4]. The time of training and computation and hinge loss is relatively high.

In the proposed framework, the contents of color, texture, and shape of images are obtained with the descriptors of Color Histogram, Color Moments, Color Auto-Correlogram, Wavelet Moments, and Gabor Wavelet respectively. The obtained multiple features are merged and put in a vector. This multiple features combination provides discriminative details of the image. The features stored in the vector are trained using multi-binary SVM and the similarity distance is calculated using Weighted Euclidean Distance (WED) metrics. On the basis of similarity value, retrieved the relevant images and evaluated the process performance for individual and fused features using accuracy performance metric.

Objective: To retrieve the most relevant images through extracting visual features, creating distinctive and strong features for obtaining a consistent result.

Contributions:

• The optimal features combination obtained using the traditional approach in the feature extraction stage of the retrieval framework for image features representation upsurges the performance (86%) equated to [4] and [5].

• The ensemble strategy of SVM decreases the complexity involved in the optimization of features compared to [4]

• The ensemble strategy of the proposed approach affords a stable outcome compared to the existing approach 1-Nearest Neighbor classifier.

• It resolves the unhelpfulness of the Euclidean distance metric in computing similarity match between samples in the scenario of the presence of query sample at the center of the circle and distribution of training samples densely on the surface of the circle equiradius to the query sample.

Organization: The paper is organized as follows: Section-II delivers details of related work on image retrieval and optimization technique. Section-III describes anticipated method implementation functions. Section-IV deliberates performance study. Section-V presents the conclusion.

II. RELATED WORKS

Yildizer *et al.*, [6] formulated an image retrieval model using multiple SVMs ensembles. To extract features from images, the Daubechies Wavelet Transformation is used and it reduces the dimension of the features, helps to reduce distance computation cost, reduces search space. The model performance is less and does not give better performance for the increased number of input samples.

Marakakis *et al.*, [7] devised a Relevance Feedback approach to retrieve images. The Gaussian Mixture models to represent image and SVM for training and the Kullback-Leibler approximations to find similarity are used. Employed kernel function in each iteration of Relevance Feedback in order to carry out Relevance Feedback task. The approach provides good performance for datasets with a small number of samples but not for a large number of samples.

Gao *et al.*, [8] described saliency in the database to show significant improvement in visual image re-ranking. Computed the score of distinctiveness in every image salient window in a hierarchical way. The initial retrieved images are reranked with respect to the distinctiveness and online weight. The accuracy obtained is further increased by increasing the discriminative ability in the feature extraction stage.

Jin *et al.*, [9] developed a retrieval framework with color, texture, and shape features. The obtained features trained using an SVM classifier and similarity distance with respect to the regions is computed with the equal proportional scaling algorithm. The performance was improved further with features optimization.

Zhang *et al.*, [10] developed an image retrieval approach by the combination of the descriptor. The color, shape features are extracted and the similarity distance is calculated with traditional metrics. The performance obtained is average because of features high dimension and due to the lack of inclusion of variations in the features extraction stage.

Whang *et al.*, [11] developed an image retrieval with relevance feedback method of SVM. The color, shape, and texture features are obtained with CH, Edge Directed Histogram, and Pyramid WT descriptors and decreased dimension with Principal Component Analysis. The features probabilities were computed with an adapted Gaussian mixture model. The similarity distance is computed with the kullback-Leibler divergence metric. The performance is less due to the probabilistic nature of the obtained features and takes more time for retrieval.

Sugamya *et al.*, [12] proposed a two-phase approach for image retrieval. In the first stage, the color, texture, and shape features are extracted. In the second phase, to handle and reduce noise in the samples the SVM classifier is used. The retrieval system performance is average because of features with less discriminative ability. The representation of the features is not adequate.

Sotoodeh *et al.*, [13] developed an image retrieval framework with texture features. Using Radial Mean Local Binary Pattern, the features of texture are extracted and optimized with the Particle Swarm algorithm. Using the k-means clustering Algorithm with Iterated Local Search. The model gives a good performance but takes time in the retrieval process.

To overcome the inadequacies that exist in the state-of-the-art approaches, the proposed framework (IR-MF-SVMe) obtains

Author	Features	Technique	Merits	Demerits	
					Performance
Yildizer et al.,	Daubechies Wavelet	Multi-SVM	Reduce distance	Does not give better	75.89%
[2012] [6]	Transformatione		computation cost,	performance for the	accuracy
			reduces search space	increased number of input	
				samples	
Marakakis <i>et</i>	Gaussian Mixture	SVM	Good performance	But not for large number of	81%
al., [2011] [7]	models		for dataset with small	samples	accuracy
			number of samples		
Gao <i>et al.</i> ,	Image re-ranking	SVM	Simple to implement	Features discriminative	74.3%
[2015] [8]				power is less	accuracy
Jin et al.,	Color,Shape and	SVM	Features not	The performance can be	82.3%
[2017] [9]	Texture features		optimization	improved with features	accuracy
				optimization	
Zhang et al.,	Color and Shape	SVM	Features	Features high dimension and	78.93%
[2014] [10]	features, Visual		representation is	lack of inclusion of	accuracy
	Optimization model		inadequate	variations in the features	
				extraction phase	
Whang <i>et al.</i> ,	Color, Shape and	PCA, Kullback-	Performance is good	Retrieval time is high due to	83%
[2016] [11]	Texture features	Leibler	due to the features	features probabilistic nature	accuracy
		Divergence metric	combination		
Sotoodeh et al.,	Texture features,	k-means	Simple model	Performance is less	63.08%
[2019] [13]	Particle Swarm	clustering,			accuracy
	Optimization	Extended			
		Canberra metric			
Zenggang et al.,	Color and Shape	Euclidean	Simple model	Performance is average	68%
[2019] [14]	features	Distance metric			accuracy
Alsmadi et al.,	Color, Shape and	Meta-heuristic	Model performance	Takes time in the retrieval	88.83%
[2018] [1]	Texture features	Algorithm	is good	task	accuracy

Table 1: Related Works

method, appropriate feature vectors are selected. Then, the image similarity is computed with the Extended Canberra metric. The model performance is less due to the use of single modality features.

Zenggang *et al.*, [14] designed a retrieval algorithm with color and shape features. The closeness among the query image and achieved images is computed with the Euclidean distance metric. The algorithm performance is average due to the use of multi-modal features inadequate the combination in the image representation. The performance can be enhanced through the usage of appropriate features aggregation and its optimization process.

Alsmadi *et al.*, [15] developed a retrieval model using multimodel features. The color, texture, and shape features are obtained with Color Histogram, Canny Edge Detection, and GLCM descriptors compatibly. The similarity between the archive images and query images is calculated using a Genetic an optimal features combination *via* a traditional approach in the feature extraction stage, to make features representation more discriminative. Decreases the complexity involved in the features optimization phase through the use of ensemble strategy. It results in improved performance compared to the prevalent approaches.

III. PROPOSED METHOD

Problem Statement: Develop a model to retrieve relevant images *via* visual features by constructing distinct features for obtaining a stable result.

Objectives:

- To increase the retrieval accuracy
- To decrease the retrieval time

The proposed framework has two sections, the first section is the training stage and the second section is testing. The images are resized *i.e.*, image pixel values are scaled to make images clearer, forwarded to the feature extraction stage. The color, shape, texture features of the samples are extracted using corresponding features descriptors. The extracted features of all the samples are converted into a matrix format and generated a feature vector. The features in the vector are trained with an SVM classifier and assigned a label for each sample. The next stage is testing, a query sample is given as input to the proposed system and based on the similarity match, the relevant samples related to the query are retrieved. The performance of the process is evaluated. The proposed system framework is shown in Fig. 1.



Fig.1: The Proposed System Architecture

In the proposed work, the dataset samples are resampled, and created a number of datasets by taking care of the created number of datasets size does not become too small for avoiding overfitting. The binary SVM classifier retrieves images from the dataset, predicts from a single class. So, assigned a binary classifier to each of the small datasets created and created multiple binary SVM models. So multiple binary SVM models are used to predict the multiple classes for that initially all the features extracted from the samples along with the feature vector, the names of the images are cascaded at the end and generate labels. After that, the binary SVM models are trained by using each feature vector from the feature dataset. This will create an array of multiple binary SVM models. The trained binary SVM models are aggregated and the combined performance is evaluated to predict the outcome through the average weighted voting strategy.

Normalization

It is the process of preparing the data through different steps for the retrieval model to simplify the analysis and computational process and enhance accuracy, reduce complexity. The pattern's visibility in the texture image is affected by the variations of local illumination or from various sources in the process of image capturing. The undesirable artifacts that influence the quality of the images are eliminated through the process of linear normalization. The distribution of dark and light pixels of an image is the contrast. The contrast of an image is increased, to increase the differences among the dark and light pixels that enhance the image quality. In the color images more info is existing than grayscale, conversion from color images to greyscale images decreases the need for processing the number of pixels. Human vision is more sensitive towards contrast than pixel values [16].

Table 2: Notations used

Notations	Meaning
Н	Histogram
\tilde{H}_i	Normalized Histogram
CLi	Color level
P _i j	Color level pair
M _i	Mean
σ_i	Standard Deviation
$D_m(P,Q)$	similarity function among image P and image Q
M_{i}^{1}, M_{i}^{2}	First order moment
σ_i^1, σ_i^2	second order moment
s_{i}^{1}, s_{i}^{2}	Third order moment
	Gabor function through scaling and shifting
$g_{\alpha,\beta,a,b(x)}$	factors
ΨĊ	Mother wavelet
C _{pl}	positive label class
C _{nl}	Negative labels class
р	weight
9	input vector
r	bias
TP, TN	True positives and negatives
FP, FN	False positives and negatives

Feature Extraction

To enhance the features of an image, the dataset images are pre-processed by eliminating undesirable distortions before the process of features extraction. It is model construction first phase. The color, shape, and texture feature combination obtained for the training phase is as follows: The proposed system uses five various types of extracted features, are CH, CC, CM, GW, and WM, and are used to perform the training process. A brief explanation about features extraction is given below.

The images are primarily differentiated based on color difference, color feature is steady, less prone to geometrical variations. The entire color info of an image can be replaced by a small quantity of representing colors [17]. The choice of color space is important even though the usage of colors provides discriminative information [18]. In the Euclidean space, the color differences obtained through the distance metric do not match the color difference as perceived by the Human visual system. RGB space is the most generally used color space [19], but it is device-dependent, perceptually not uniform. So data transforming from RGB space to HSV perceptual uniform space is natural and fast before feature extraction. The conversion is given by Equation (1).

INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING DOI: 10.46300/9106.2021.15.192

$$H = \cos^{-1} \frac{\frac{1}{2} [(R-G) + (R-B)]}{\sqrt{((R-G)^2 + (R-G) + (R-B))}} S = 1 - \frac{3}{R+G+B} (min(R,G,B)) V = \frac{3}{R+G+B}$$

The images are quantized in HSV space into 8 * 2 * 2 equal bins and obtained 1 * 32 vector HSV histogram to make the distribution of quantitative data uniform. The color-based similarity analysis is carried out with Color Vectors or Histograms and bins.

Color Histogram

Represents color distribution in the image [20], for all color spaces the color histogram is generated. The three-dimensional HSV color space [21] is used, to show the pixel count for the colors that appeared. The CH is expressed as three luminance histograms in which each one shows the distribution of brightness of each one Hue/Saturation/Intensity Value color channel. The drawback is (i) based on only color histogram comparisons not able to differentiate similar images having a different color, (ii) for noise intervention it is highly sensitive, and (iii) the feature space occupies more dimensions for the case of bins of high dimension. Color Histogram represents colors distribution in images *via* a set of bins. Each bin corresponds to a color in the quantized color space. For a given image I, color histogram H is represented through a vector:

$$H = \sum_{i=1}^{n} H_{i}, H_{i} = H_{1}, H_{2}, \dots, H_{n}$$
(1)

where, i = 1,2...,n, represents a color bin, H_i represents a number of pixels of color i in the image, n represents total number of bins. Each bin contains a number of pixels having corresponding similar colors. For the comparison purpose of different size images, they are normalized. The normalized histogram is represented as:

$$\tilde{H} = \sum_{i=1}^{n} \tilde{H}_{i}, \tilde{H}_{i'} = \tilde{H}_{1'}, \tilde{H}_{2'}, \dots, \tilde{H}_{n'}$$
(2)

where, $\tilde{H}_i = H_i/N_p$, N_p is the total no of image pixels.

Histograms provide more information by effectively finding local regions of higher occurrence of parameters in identifying the difference. A histogram parts the image data into equalsized intervals called bins. The width of a bin shows the presence of a pattern in the data. Varying the size of the bins varies the quantity of data in every bin *i.e.*, data gets grouped differently. A bin width has to be chosen appropriately to find contributory features in the distribution of image details.

Color Auto Correlogram

Represents change of spatial correlation of pairs of colors with the distance, *i.e.*, seizures only spatial correlation among matching colors. Describes global distribution of the local spatial correlation of colors [22]. Takes into account local color spatial correlation and global distribution of local color spatial correlation. It is effective to appearance change hence suitable for retrieval tasks in the case of the large image data set. For a given image I, color level pair $P(CL_i, CL_j)$, similarity level S is defined through the equation (3).

$$S \sum_{c_i}^{s} \frac{d}{(CL_i, CL_j)} = P(p_1 - p_2 = s)$$
 (3)

where, $p_1 \in CL_i$ and $p_2 \in CL_i$

Color Moments

To differentiate the images based on color features [23], the Color Moments: Mean, Standard Deviation, and Skewness are used as measures. The moments give the measurement to the color similarity among the images. The similarity value is compared to the indexed images of the database in the case of the retrieval task. In an image, the color distribution is interpreted as a probability distribution and is considered by a number of unique moments. This assumption follows that "if the color in an image follows a certain probability distribution, the moments of that distribution are then used as features to identify that image on the basis of color". The following equations give color moments as:

Mean: is the average color value in an image

$$M_i = \sum_N^{J=1} \frac{1}{N} P_{ij}$$

Standard Deviation: is the square root of the variance of the distribution

$$\sigma_i = \sqrt{(\frac{1}{N} \sum_{N}^{j=1} (P_{ij} - M_i)^2)}$$
(5)

Skewness: is the measure of the degree of asymmetry in the distribution.

$$s_i = \sqrt{\left(\frac{1}{N} \sum_{N}^{j=1} (P_{ij} - M_i)^3\right)}$$
(6)

The weighted differences sum among the moments of the two distributions is the similarity function among two images distribution and is:

$$D_m(P,Q) = \sum_{i=1}^r W_{i1} |M_i^1 - M_i^2| + W_{i2} |\sigma_i^1 - \sigma_i^2| + W_{i3} |s_i^1 - s_i^2|$$

$$(7)$$

where, P, Q -two image distributions to be compared, i - index of the current channel (H = 1, S = 2, V = 3), n - channel number, M_i^1, M_i^2 - first, σ_i^1, σ_i^2 - second and s_i^1, s_i^2 - third moments of the two image distributions, W_i - weights for each moment and N- represents a number of pixels.

In image retrieval, the parameter values are attuned with respect to the preferences of extracted multi-modal features. With respect to D_m values, the images pairs are ranked in increasing order and considered similarity in the same order as more to less. For *HSV* color space, H value relates to the color type (red, green, blue) regularly measured more relevant in adjudging similarity observed than the *V* value relates to the intensity of the images.

Gabor Wavelet

In analyzing data, the Fourier transform represents the abrupt changes inefficiently. It epitomizes data as a sum of sinewave forms, not localized in space or time that oscillates forever. Wavelets are used for representing, analyzing multiresolution images, removing noise. The new class of functions *i.e.*, wavelets were well localized in time, and frequency is used to examine the abrupt fluctuations in images for image data analysis. Rapidly decaying wave-like oscillation having mean is zero extends to infinity, exists for a finite duration. Wavelets are the elements of a mutually similar Gabor functions family [24], created from one elementary Gabor function through scaling and shifting.

$$g_{\alpha,\beta,a,b(x)} = |a|^{-\frac{1}{2}} g_{\alpha,\beta}\left(\frac{x-b}{a}\right) \tag{8}$$

where, a, b are scaling, shifting factors. The mother wavelet energy is localized around x = 0 and all the wavelets are normalized |g| = 1.

Wavelet Moments

Constructing moment descriptors to increase pattern representation ability by the combination of wavelet and moment analysis. Wavelet moments belong to the orthogonal moment family, use orthogonal wavelet function as kernel and the $g_{m(r)}$ function is replaced with a wavelet basis functions.

$$M_{nm} = \int \int g_{m(r)} \exp^{-im\theta} f(r,\theta) r dr d\theta$$
(9)

wavelet basis functions of the following form are chosen

$$\Psi_{(a,b)}(r) = \frac{1}{\sqrt{a}}\Psi(\frac{r-b}{a}) \tag{10}$$

where, $\Psi(\mathbf{j})$ is mother wavelet used to generate the whole basis.

Multi-Support Vector Machine

Design ensemble of classifiers from which the final result is the result of all the individual classifiers [25]. The ensemble SVM performance relay on the representative ability of the features extracted and training dataset quality. In Ndimensional space, to reduce overfitting and increase efficiency, the optimization and generalization strategy is used. SVM simultaneously decreases model dimension and minimizes the estimation error [26].

In the case of One vs One SVM, the total number of classifiers required is $\frac{n(n-1)}{2}$. The total number of classifiers required for One vs One is more compared to one vs all. So it utilizes more space and time and if the data for one class is large, the training becomes expensive. In the case of a multiclass retrieval system, one vs all SVM is used and constructed n binary SVM models. In n^{th} binary class SVM problem, from the remaining classes C_m class is separated, considered as c_{pl} positive label class, considered all other classes as a single class C_c of negative labels class C_{nl} , this process is repeated for all m classes and combined all m binary SVM

classifiers in constructing final multi-class classifier. The plane that separates features from class C_m and combined class C_c is determined with the SVM approach as:

$$(p * q) + r = +1$$

 $(p * q) + r = -1$
 $(p * q) + r = 0$

where, p is weight, q is an input vector and r is bias. The values of p and r are computed with the above equation and linear algebra with a margin value of $2\sqrt{kk}$. Where, k-is the number of samples. The computed marginal value categorizes new sample points into a different category level.

The RBF kernel is a nonlinear kernel; it represents the data via a set of pairwise similarity comparisons. Uplifts dataset samples over to feature space for parting samples into classes with linear decision boundary which is of high dimension through the nonlinear combination of features. The relevance of the RBF kernel is high towards real-time datasets [27]. Thus, it provides increased accuracy with its speed of convergence.

IV. RESULT AND DISCUSSION

The proposed Image Retrieval model - IR-MF-SVMe is implemented in the MatLab version in an **i5** system with 4GB RAM. Performances of the model is assessed with respect to accuracy by varying number of features and features fusion. **1**) Datasets

The model is implemented and tested on the following datasets.

• WANG: It has colored images of different objects and they are classified into 10 classes each one has 100 images, totally it has 1000 images [5] [28] with dimension of 384 * 256; among 1000 images, 900 are used in training and 100 are for testing.

• CIFAR-10: It contains a total of 6000 images [4] [29] with dimension 32x32. It is categorized into 10 classes each class has 6000 images, among 6000 images; 5000 are used in training and 1000 for testing.

• Oxford Flower: It contains images of flowers that are commonly present in the U.K. It has 17 classes and each class has 80 images, totally 1360 images [5] [30], among 1360 images, 1020 are used in training and 340 for testing.

• Corel: It contains totally 1000 images [11] [31], divided into 10 classes each class has 100 images, among 1000 images; 900 are used in training and 100 for testing.

2) Performance Analysis

It consists of two sections: In the first section, the performance of the process is evaluated and in the second section the obtained result is compared with the prevailing methods. The resultant values are tabulated in the tables and the graphs for the tabulated values are shown.

(i) **Retrieval Accuracy:** Retrieval Accuracy is the closeness of measured value with respect to the known or standard value. It

is associated with precision and inverse of it and recall and inverse of it *i.e.*, computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(14)

where, the parameters *TP*, *TN*, *FP* and *FN* represents true and false positives and negatives used in evaluating the metric. The closeness of two visuals for the training samples gives stable result and enhance unified decision making [32].

Table 3: Retrieval Performance (Individual Class)

Training %	15%	30%	45%	60%	75%	Average
Africa	60	70	80	90	100	80
Beach	40	40	60	80	80	60
Buildings	45	60	75	80	90	70
Bus	100	100	100	100	100	100
Dinosaurs	100	100	100	100	100	100
Elephant	30	50	80	90	100	70
Flower	100	100	100	100	100	100
Horse	100	100	100	100	100	100
Mountain	60	70	80	90	100	80
Food	100	100	100	100	100	100
Average	73.50	79	87.50	93.50	97	84

Table 4: Performance of the system for each Class (Features Fusion)

Class	Cityblock	Manhattan	Euclidean	KNN	IR-MF-
	[33]	[34]	[34]	[35]	SVMe)
Africa	25	10	10	80	80
Beach	35	30	25	20	60
Buildings	25	05	05	60	70
Bus	95	35	70	100	100
Dinosaurs	100	100	100	90	100
Elephant	70	20	25	80	70
Flower	100	100	100	70	100
Horse	100	100	95	90	100
Mountain	10	20	15	50	80
Food	20	25	45	80	100
Average	58	44.50	49	72	86

(a) **Performance of individual features:** The performance is measured against each class of the dataset individually. Initially, 15% of the trained features are selected and the accuracy value is computed for all the classes. Then, similarly measured the accuracy by selecting 30%, 45%, 60%, 75% and 100% trained features from each class image features. The resulted values are tabulated and shown in Table III. It shows system accuracy of every class in the database among all the classes Bus, Dinosaurs, Elephant, Flower, and Horse classes produces the maximum accuracy because they are rich in color and they do not have any messy background. The tabular result shows that the number of features selected, the model feature extraction phase makes the extracted features less redundant and misleading, it helps to enhance accuracy, thus the accuracy of the model outperforms the other models. The extent of images used to train the SVM classifier plays a huge role. The

number of images used for training will have an impact on the performance *i.e.*, accuracy. It results in less accuracy if the number of images is less and it leads to high accuracy by providing more images. The performance of each class is shown in Fig. 2.



Fig. 2: Performance of individual features of each class

(b) Performance of features fusion: The trained features fusion performance is measured and resulted value is tabulated in Table IV. The weighted Euclidean Distance (WED) gives closeness of two visuals. The closeness of archive images and query images is computed with WED. It is observed that the performance of the multiple features fusion is better than the individual features performance. The graph for the features fusion is shown in Fig. 2. The discriminative features representation supports in enhancing the retrieval performance, thus the Low level visual features are fused in order to make the visual features more robust and unique for the appropriate representation, it necessitates high computational cost [36] and provides stable results. The Deep neural network boost retrieval performance at a high computational cost [37] and *necessitate* minimal preprocessing and feature extraction.



Fig. 3: Performance of features fusion

(ii) **Performance comparison:** The performance of the retrieval process is measured using the KNN classifier and City Block, Euclidean, Manhattan distance metrics and compared with the performance of the proposed method. The KNN classifier use WED for the similarity computation among archive images and the query image. It provides a stable outcome by managing the existence of noise and imbalance inputs [38] *i.e.*, takes care of the increase in the number of features of data sample greater than the training samples in the scenario wherein the case SVM fails. Fig. 3 shows a

comparison of the various methods with respect to retrieval accuracy. The proposed method performance is compared with [4][5], [9-11], and [13-15].

Author	Dataset	Performance	
		(Accuracy in %)	
Proposed(IR-MF-	Wang	86	
SVMe)	Cifar	85	
	Oxford flowers	86	
	Corel	84	
Zhang et al.,[2014]	Corel-1000	79.67	
[10]			
Luo et al.,[2015] [4]	Wang	65.67	
Wang et al., [2016]	Corel	83	
[11]			
Jin et al., [2017] [9]	Oxford flowers	82.3	
Sotoodeh et al.,	Wang	63.08	
[2019] [13]			
Zenggang et al.,	Traffic signs	68	
[2019] [14]	-		
Alsmadi et al.,	Corel	88.83	
[2018] [15]			

Table 5: Performance Comparison



Fig. 4 :Retrieval performance comparision

V. CONCLUSIONS

Proposed an image retrieval framework using multiple feature fusion and SVM ensemble (IR-MF-SVMe) and tested on WANG dataset. Obtained an accuracy of 84% and 86% accuracy for the individual and fused features, comparatively better than the earlier methods [4][5], [9-11], and [13-15]. The importance of different image features and combinations of these are perceived in the image pattern representation. It is clear through the result analysis, the proposed system is comparatively lightweight and flexible in retrieving the images from the archives. Useful in the analysis of multi-resolution References

- [1] Majaj, Pelli, Kurshan, and Palomares, "The role of Spatial
- Frequency Channels in Letter Identification," *Journal of Vision Research*, vol. 42, no. 9, pp. 1165–1184, 2002.
 [2] Guang Liu, Zuo Li, Lei, and Xu. "Image Retrieval based

images. The future work focus is on the use of parallelization

with **GPUs** to reduce computational complexity.

- on Micro-Structure Descriptor", Journal of Pattern Recognition, vol. 44, no. 9, pp. 2123–2133, 2011.
- [3] K. Venkataravana Nayak, M. Geetanjali, J. S. Arunalatha, and K. R. Venugopal, "IR-HF-WED: Image Retrieval using Hybrid Feature Extraction with Weighted Euclidean Distance," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 6, pp. 3987–3999, 2020.
- [4] Yong, Wen, Dacheng, Jie, and Xu, "Large Margin Multi-Modal Multi-Task Feature Extraction for Image Classification." *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 414–427, 2015.
- [5] Jiachen Yang, Bin Jiang, Baihua Li, Kun Tian, and Zhihan Lv, "A Fast Image Retrieval Method Designed for Network Big Data", *IEEE Transactions on Industrial Informatics*, vo. 13, no. 5, pp. 2350–2359, 2017.
- [6] Yildizer, Ali, Hassan, and Alhajj, "Efficient Contentbased Image Retrieval using Multiple Support Vector Machines Ensemble," *Journal of Expert Systems with Applications*, vol. 39, no. 3, pp. 2385–2396, 2012.
- [7] Marakakis, Siolas, Galatsanos, Likas, and Stafylopatis, "Relevance Feedback approach for Image Retrieval Combining Support Vector Machines and adapted Gaussian Mixture Models," *IET Image Processing*, vol. 5, no. 6, pp. 531–540, 2011.
- [8] Yuan Gao, Miaojing Shi, Dacheng, and Chao, "Database Saliency for Fast Image Retrieval," *IEEE Transactions* on Multimedia, vol. 17, no. 3, pp. 359–369, 2015.
- [9] Cong Jin and Shan Wu, "Content-based Image Retrieval based on Shape Similarity Calculation," *Journal of 3D Research*, vol. 8, no. 3, pp. 23, 2017.
- [10] Ming, Ke, Feng, Jianzhong, Kong, and Lu, "A novel Image Retrieval Method based on Hybrid Information Descriptors," *Journal of Visual Communication and Image Representation*, vol. 25, no. 7, pp. 1574–1587, 2014.
- [11] Xiang, Lin, Wei, Dong, and Yang, "A new SVM-based Relevance Feedback Image Retrieval using Probabilistic Feature and Weighted Kernel Function," *Journal of Visual Communication and Image Representation*, vol. 38, no. 1, pp. 256–275, 2016.
- [12] Sugamya, Suresh, and Vinaya, "A CBIR Classification using Support Vector Machines.," *International Conference on Advances in Human Machine Interaction* (*HMI*), pages 1–6, 2016.
- [13] Mahmood Sotoodeh, Mohammad Reza, and Reza Boostani, "A Novel Adaptive LBP-based Descriptor for

Color Image Retrieval," *Journal of Expert Systems with Applications*, vol. 127, no. 1, pp. 342–352, 2019.

- [14] Xiong, Tang, Chen, Xuemin, Kaibin, and Conghuan, "Research on Image Retrieval Algorithm based on Combination of Color and Shape Features," *Journal of Signal Processing Systems*, vol. 93, no. 1, pp. 1–8, 2019.
- [15] Mutasem Alsmadi. Query-sensitive Similarity Measure for Content-Based Image Retrieval using Meta-Heuristic Algorithm", *Journal of King Saud University-Computer* and Information Sciences, vol. 30, no. 3, pp. 373–381, 2018.
- [16] John. An Introduction to MPEG Video Compression. Internet: members. aol. com/symbandgrl, 1998.
- [17] Jing Ming Guo, Heri Prasetyo, and Jen Ho Chen, "Content-based Image Retrieval using Error Diffusion Block Truncation Coding Features," *IEEE Transactions* on Circuits and Systems for Video Technology, vol. 25, no. 3, pp. 466–481, 2014.
- [18] Gert Wouwer, Stefan, Paul, and Dirk, "Color Texture Classification by Wavelet Energy Correlation Signatures," *International Conference on Image Analysis and Processing*, pages 327–334, 1997.
- [19] Rokia, Sarifuddin, and Vaillancourt, "An Effective Approach towards Content-Based Image Retrieval," pages 335–343, 2004.
- [20] Michael Swain and D Ballard, "Color Indexing," International Journal of Computer Vision, vol. 7, no. 1, pp. 11–32, 1991.
- [21] A Raza, Hassan D, Hussain D, Shabbir, Mehboob, and Banjar, "Correlated Primary Visual Texton Histogram Features for Content Base Image Retrieval," *IEEE Access*, vol. 6, pp. 46595–46616, 2018.
- [22] H A Hassan, N M Tahir, I Yassin, C Yahaya, and S M Shafie, "Visualisation of Exudates in Fundus Images using radar Chart and Color Auto Correlogram Technique," *International Conference on Computer Vision and Image Analysis Applications*, pages 1–6, 2015.
- [23] H Yu, M Li, H Jiang, and J Feng, "Color Texture Moments for Content-Based Image Retrieval," In Proceedings of the International Conference on Image Processing, vol. 3, pp. 929–932, 2002.
- [24] C Li, G Duan, and F Zhong, "Rotation Invariant Texture Retrieval Considering the Scale Dependence of Gabor Wavelet," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2344–2354, 2015.
- [25] Lars Hansen and Salamon. "Neural Network Ensembls". IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(10):993–1001, 1990.
- [26] Kunwar P S, Nikita, and Gupta S, "Support Vector Machines in Water Quality Management," *Analytica Chimica Acta*, vol. 703, no. 2, pp. 152–162, 2011.
- [27] Avinanta Tarigan, Dewi A, Suhendra, and Budiman, "Determination of SVM-RBF Kernel Space Parameter to Optimize Accuracy Value of Indonesian Batik Images Classification," *Journal of Computer Science*, vol. 13, no. 11, pp. 590–599, 2017.
- [28] https://www.kaggle.com/ambarish/wangdataset.
- [29] http://www.cs.toronto.edu/~kriz/cifar.html.

- [30] https://www.kaggle.com/cantonioupao/ oxford-flower-17categories-labelled.
- [31] https://www.kaggle.com/elkamel/corel-images.
- [32] Chunjie Zhang, Jian C, and Qi Tian, "Multiview Semantic Representation for Visual Recognition," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 2038–2049, 2018.
- [33] F Malik and B Baharudin, "Analysis of Distance Metrics in Content-based Image Retrieval using Statistical Quantized Histogram Texture Features in the DCT Domain," *Journal of King Saud University-Computer* and Information Sciences, vol. 25, no. 2, pp. 207–218, 2013.
- [34] Sanjay P and Sanjay T, "Content Based Image Retrieval using Various Distance Metrics," *International Conference on Data Engineering and Management*, pages 154–161, 2010.
- [35] Daniel Pedronette, Filipe Goncalves, and Ivan Guilherme, "Unsupervised Manifold Learning through Reciprocal KNN-Graph and Connected-Components for Image Retrieval Tasks" *Journal of Pattern Recognition*, vol. 75, pp. 161–174, 2018.
- [36] Bushra Z, Rehan A, N Ali, Mudassar A, S Jabbar, Kashif, Awais, and Gwanggil, "Intelligent Image Classificationbased on Spatial Weighted Histograms of Concentric Circles," *Journal of Computer Science and Information Systems*, vol. 15, no. 3, pp. 615–633, 2018.
- [37] Guanqiu Qi, Huan Wang, Matthew Haner, Chenjie Weng, Sixin Chen, and Zhiqin Zhu, "Convolutional Neural Network based Detection and Judgement of Environmental Obstacle in Vehicle Operation," *Journal* of Chinese Association for Artificial Intelligence Transactions on Intelligence Technology, vol. 4, no. 2, pp. 80–91, 2019.
- [38] Thomas G D, "Machine Learning Research," Journal of Artificial Intelligence Magazine, vol. 18, no. 4, pp. 97– 136, 1997.

Author Contributions:

Venkataravana Nayak K has implemented the Model in python.

Arunalatha J S has analyzed the model outcome. Venugopal K R was substantiated the end result.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0 https://creativecommons.org/licenses/by/4.0/deed.en_US