RESEARCH ARTICLE OPEN ACCESS

Baseline Collaborative Image Retrieval Using Euclidean Distances Metric Learning

Shubhangi Durgi [1], Prof. Pravinkumar Badadapure [2]
Department of Electronics and Telecommunication
Imperial college of Engineering
Pune - India

ABSTARCT

Content-Based Image Retrieval (CBIR) has been a long standing research challenge in the world of Image Processing. In CBIR, relevant images are identified based on their similarities to query images. Most CBIR algorithms are hindered by the semantic gap between the low-level image features used for computing image similarity and the high-level semantic concepts conveyed in images. One way to reduce the semantic gap is to utilize the log data of users' feedback that has been collected by CBIR systems in history, which is also called "collaborative image retrieval". We consider the problem of learning a mapping function from low-level feature space to high-level semantic space. To solve this problem, we propose semantically motivated Image Manifold Learning (IML) algorithm. Under the assumption that the data lie on a sub manifold embedded in a high dimensional Euclidean space, we propose a relevance feedback scheme which is naturally conducted only on the image manifold in question rather than the total ambient space. We then develop an algorithmic framework to approximate the optimal mapping function by a Laplaican Eigenmap. The semantics of a new image can be inferred by the" Laplaican Eigen maps". We compare results of our algorithm with traditional Euclidian Distance algorithm. We also developed an algorithm for outside database images by "Artificial Neural Network (ANN)" We use precision and accuracy parameters to compare IML with Euclidean distance learning &ANN. Experimental results show that our approach is effective in improving the performance of content-based image retrieval systems. The experiment also indicates that the IML algorithm is more effective, which exploit log data for image retrieval.

Keywords:- semantic gap, log data of users, sub manifold, relevance feedback, Laplaican Eigenmap, ANN, precision, accuracy.

I. INTRODUTION

CBIR has been very challenging topic, because CBIR [1],[5] is based on high level feature & low level feature. Low level features visualize color, texture, shape & so on. High level feature express emotions meaning association of feature expression with combination of perceptual feature. Thus, It is difficult to extract high level features like emotions, or what are the activities present in that image. But they give relatively more important meanings of objects and scenes in the images that are perceived by human beings. So generally low level features like color, texture, shape & edge are used for retrieval of the image.

Fig. 1 shows the architecture of a typical CBIR system. Each image in the image database is in standard form. For all images in database, first, features are extracted and the obtained feature space is stored in the feature database. When a query image is selected, its feature space will be compared with those in the feature database one by one and the similar images with the smallest feature distance will be retrieved.

CBIR can be divided into mainly two stages:

• Preprocessing: First step is to extract a feature, which describes its contents .In this processing, we perform feature filtration, normalization, segmentation (i.e. divide the image

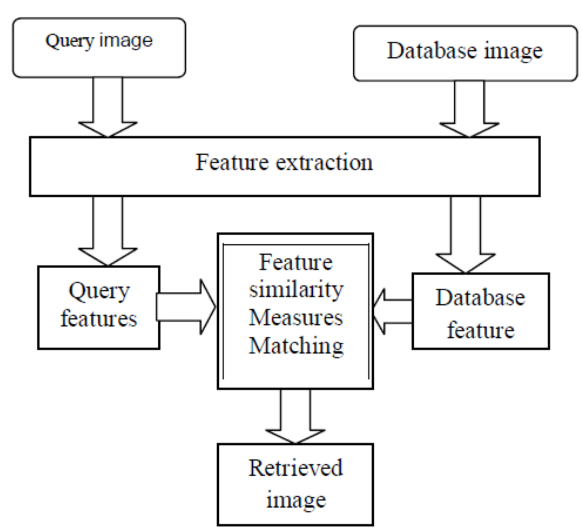

Fig. 1: Image Retrieval processing

• Feature extraction: Lew level features are used to describe the content of the image. Image Features can be classified into primitives

In Euclidean distance, image retrieval techniques build on the assumption that the image space is Euclidean. However, in many cases, the image space might be a non-linear sub-manifold which is embedded in the ambient space. Intrinsically, there are two fundamental problems in image retrieval: 1) how do we represent an image? 2) How do we judge similarity?

One possible solution to these two problems is to learn a mapping Function from the low-level feature space to the high-level semantic space. The former is not always consistent with human perception while the latter is what image retrieval system desires to have. Specifically, if two images are semantically similar, then they are close to each other in semantic space. In this paper, our approach is to recover semantic structures hidden in the image feature space such as color, texture, etc.

As we point out, the choice of the similarity measure is a deep question that lies at the core of image retrieval. In recent years, manifold learning [3] has received lots of attention and been applied to face recognition, graphics, document representation, etc. These research efforts show that manifold Structure is more powerful than Euclidean structure for data representation.

It is worthwhile to highlight several aspects of the framework of analysis presented here:

(1) Throughout this paper, we denote by image space the set of all the images. Different from most of previous geometry-based Works which assume that the image space is a Euclidean space in this paper, we make a much weaker assumption that the image space is a Riemannian manifold embedded in the feature space. Particularly, we call it image manifold. Generally, the image manifold has a lower dimensionality than the feature space. The metric structure of the image manifold is induced but different from the metric structure of the feature space. Thus, a new algorithm for image retrieval which takes into account the intrinsic metric structure of the image manifold is needed.

(2) There are two key algorithms in this framework. One is the retrieval algorithm on image manifold, and the other is an algorithm for learning a mapping function from feature space (color, texture, etc.) to high-level semantic space. The learning algorithm will gradually "flat" the image manifold, and make it better consistent with human perception. That is, if two images are close (in the sense of Euclidean metric) to each other, they are semantically similar to each other.

In this we describe the proposed framework for learning a semantic space to represent the underlying image manifold.

A. Objective of the System:-

There is large number of images present in the image database. We have used WANG Database of 90 images for our project which contains images in 'jpeg' format. Initially query image is given, and then low level features like color, texture and shape are extracted from the query image. For color feature extraction three color moments are used in three color channels (H, S,V), So there are 9 color features. For texture feature extraction we have used 3 Level DWT, So there are total 9 texture feature and. For shape feature extraction canny edge detection method is used. There are 18 shape features. Total 36 features of Query image are extracted. Then feature vector is calculated. Same features are extracted from the images present in the image database. The database is made to store the feature vectors calculated for the images present in the database. After Feature extraction next step is similarity measurement. For similarity measurement different algorithms are used like Image Manifold Learning (IML), Euclidean distance (ED) &ANN for outside database images.

The top closest images to our query image are retrieved. The search is usually based on similarity rather than exact match. Then user gives the feedback in the form of 'relevance judgments'. Relevant images are the images obtained in first iteration which are from the same class as that of Query. In first iteration these values are relevant and non-relevant .Relevant means the image relevant to the user and non-relevant means the image is definitely not relevant. If the user is satisfied with the obtained results, then feedback loop stops otherwise it continues until user gets satisfied with results. Finally, obtained results are compared using certain parameters like Accuracy, Precision, Recall rate etc

II. IMPLEMENTATION AND DESIGN DETAILS

B. Feature Extraction

Low-level Image Feature Representation:

Low –level image feature representation is one of the key components for CBIR system .Three types of visual features were used in this work, including color, shape and texture. The same set of image features have been used in the previous research on image retrieval.

1. Color

Color is one of the most widely used visual features in content-based image retrieval. While we can perceive only a limited number of gray levels, our eyes are able to distinguish thousands of colors and a computer can represent even millions of distinguishable colors. Color has been successfully applied to retrieve images, because it has very strong correlations with the underlying objects in an image. Moreover, color feature is robust to background complications, scaling, orientation, perspective, and size of an image. Although we can use any color space for computation of a color histogram HSV (hue, saturation, value), HLS (hue, lightness, saturation), and CIE color spaces (such as CIELAB, CIELUV) have been found to produce better results as compared to the RGB space. Since these color spaces are visually (or perceptually) uniform compared to the RGB, they are found to be more effective to measure color similarities between images.

RGB Color space is perceptually not similar to human color vision. So it is necessary to convert RGB color space into other (Perceptually close to human color vision).HSV, CIE, LUV color spaces are there. Above Color spaces can be obtained by Non-Linear transformation of RGB color space.

CIE color space is Inconvenient, because of calculation complexities of the transformation to and from RGB color space. So we have used HSV color space, which is perceptually uniform.

$$H= \text{a Cos} \left\{ \frac{1/2(R-G)(R-B)}{\sqrt{(R-G)^2}(R-B)(G-B)} \right. \quad (1)$$

$$S=1- \frac{3}{(R+G+B)}[min(R,G,B)] \quad (2)$$

$$V = \frac{1}{3} \quad (R+G+B) \quad (3)$$

There are different color feature extractions methods like:-

1. Color Histogram

2. Color Coherence Vector

3. Color Moments

Color Histogram method is relatively insensitive to position and orientation changes and they are sufficiently accurate. But they do not capture spatial relationship of color region. So they are limited to discriminating power. Color Coherence Vector method is better than color histogram method. This method combines the special correlation of color regions as well as the global distribution of local special correlation of colors. But there is one disadvantage of this method it requires very expensive computations. So in our project we have used color moment method [6] for color feature extraction. This method is more robust and runs faster than histogram based methods. So, Color feature index size=No. Of color channels * 3 moments = 9 color features .We are extracting 9 color features for CBIR. We have used color mean, color variance and color skewness in 3 different color channels (H, S, V)

2. Texture

Texture is another popular features used in CBIR. We used texture features based on wavelet transformation. The Discrete Wavelet Transformation (DWT) [8] was first applied to images with a Daubechies-4 wavelet filter. 3-levels of wavelet decomposition are used to obtain ten subimages in different scales and orientations. One of the subimages is a subsample average image of the original one and was discarded be-cause it contains less useful information. Then entropies of the other nine subimages are used texture feature of an image. Major characteristic of texture is the repetition of a pat-tern or patterns over a region in an image. The elements of pattern are called as textons. The difference between two textures can be due to degree of variation of the textons. It can also be due to spatial distribution of the textons in the image. So 9 texture features are used here.

$$E_i = \sum_{N}^{j=1} \frac{1}{N} P_{ij} \quad (4)$$

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

$$S_i = \sqrt[3]{\left(\frac{1}{N}\sum_{N}^{j=1}\left(P_{ij} - E_{ij}\right)^3\right)} \quad (6)$$

3. Shape

Edge features have been shown to be effective in CBIR since it provides information about shapes of different objects. Canny edge detection is used to obtain the histogram for edge direction. Then, the edge direction histogram was quantized into 18 bins of each of 20 degrees. So there are 18 different shape features are used to extract shape feature from an image .Shape can roughly be defined as the description of an object minus its position, orientation and size. Therefore, shape features should be invariant to translation, rotation, and scale, for an effective CBIR, when the arrangement of the objects in the image is not known in advance. To use shape as an image feature, it is essential to segment the image to detect object or region boundaries; and this is a challenge. Techniques for shape characterization can be divided into two categories.

The first category is boundary-based, using the outer contour of the shape of an object and the second category is region-based, using the whole shape region of the object. The most prominent representatives of these two categories are Fourier descriptors and moment invariants. The main idea behind the Fourier descriptors is to use the Fourier-transformed boundaries of the objects as the shape features, whereas the idea behind moment invariants is to use region-based geometric moments that are invariant to translation and rotation.

III. SIMILARITY MEASUREMENT

There are two models which we'll be using for similarity measurement:

1. Euclidean Distance Metric
2. Image Manifold Learning (IML)
   - Laplaican Eigen map (LE) for inside images.
   - Artificial Neural Network (ANN) for outside images.

Above two models are described in the following subsection:

C. Euclidean Distance Metric

1. Load the query image.
2. Calculate the Feature Vector for query & database images using Color, Texture & Shape properties.

3. Calculate Euclidean distances between two images, using below formula

$$D = \sqrt{\sum_{i=1}^{N}[F_Q - F_{DB}(i)]^2} \quad (7)$$

4. Depending on distance we rank the images & display the top rank images.

D .Image Manifold Learning (IML)

Manifold Learning:

We propose a long-term learning [4] approach to discover the true topology of the image manifold using user interactions. To be specific, we aim at mapping each image into a semantic space in which the distances between the images are consistent with human perception. The problem we are going to solve can be simply stated below:

4 Inferring a Distance Matrix in Semantic Space from User Interactions

In this we describe how to infer a distance matrix in semantic space from user interactions. Here, we present a simple method to update the distance matrix gradually [2]
.

Let *B* denote the distance matrix, $B_{ij} = \|x_i - x_j\|$. Intuitively, the images marked by the user as positive examples in a query session share some common semantics. Therefore, we can shorten the distances between them. Let *S* denote the set of positive examples, $S = \{s_1, s_2, \ldots, s_k\}$. We can adjust the distance matrix as follows:

$$B_{s_i s_j} \leftarrow B_{s_i s_j} / \alpha \quad (s_i, s_j \in S) \quad (8)$$

Where α is a suitable constant greater than 1. Similarly, we can lengthen the distances between the positive examples and negative examples, as follows:

$$R_{s_i s_j} \leftarrow R_{s_i s_j} \times \beta \quad (s_i \in S, t_j \in T) \quad (9)$$

Where $T = \{t_1, t_2, \ldots, t_k\}$ is the set of negative examples, and β is a suitable constant greater than 1. As the user interacts with the retrieval system, the distance matrix will gradually reflect the distances between the images in semantic space which is consistent with human perception.

5. Using Manifold Structure for Image Representation

We have obtained a distance matrix in semantic space. In this subsection, we discuss how to find the semantic representation for each image in database, while the distances are preserved. Recently, there has been some renewed interest [3] in the problem of developing low dimensional representations when data arises from sampling a probability distribution on a manifold. To choose a proper mapping algorithm, the following two requirements should be satisfied:

1) Since the image distribution in feature space is highly irregular and inconsistent with human perception, the mapping algorithm must have the locality preserving property.

2) The mapping algorithm should explicitly take into account the manifold structure.

Based on these two considerations, we use Laplaican Eigen maps [3] to find such a mapping. We first compute the similarity matrix as follows:

$$W_{ij} = \begin{cases} exp\left({}^{B_{ij}}/_{t}\right) & if\ B_{ij} < \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Where t and ε is a suitable constant and B is the distance matrix obtained in the previous subsection. Note that, the weight matrix has locality preserving property, which is the key feature of Laplaican Eigenmap.

Suppose y= $\{y_1, y_2, \ldots, y_m\}$ is a one-dimensional map of $\{x_1, x_2, \ldots, x_m\}$ in the LE semantic space. A reasonable criterion for choosing a "good" map is to minimize the following objective function under appropriate constraints:

$$\min_y \sum_{i,j}(y_i - y_j)^2 W_{ij} \quad (11)$$

The objective function with our choice of weights W_{ij} incurs a heavy penalty if neighboring points x_i and x_j are mapped far apart. Therefore, minimizing it is an attempt to ensure that if x_i and x_j is "close" then y_i and y_j is close as well. To minimize this objective function, it is equivalent to solve the following eigenvector problem:

$$Ly = \lambda D y \quad (12)$$

Where D is a diagonal matrix, whose entry is column sum (also row sum, since W is symmetric) of matrix W,

$$D_{ij} = \sum_j W_{ij} \quad (13)$$

L is called Laplacian matrix, L = D – W. Let $y_{(0)}$, $y_{(1)}$, …, $y_{(n)}$ be the solutions of the above eigenvector problem, ordered according to their eigenvalues, $\lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_n$. It is easy to show that $\lambda_0 = 0$, and $y(0) = (1,\ldots 1)$. We leave out y_0 and use the next k eigenvectors for embedding in k-dimensional Euclidean space.

$$x_i \rightarrow z_i = \left(y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(k)}\right) \quad (14)$$

where $y_i^{(j)}$ is the i^{th} entry of the eigenvector $y^{(j)}$. z_i is a k dimensional map of image xi in the LE semantic space.

E. Artificial Neural Network (ANN)

The nearest images obtained using feature extraction techniques are routed to Neural Network classification [8]. Neural Networks are very effective in case of classification problems where detection and recognition of target is required. It is preferred over other techniques due to its dynamic nature of adjusting the weights according to final output and applied input data. This adjustment of weights takes place iteratively until desired output is obtained. And this weight adjustment of network is known as learning of neural network.

The architecture of neural network consists of a large number of nodes and interconnection of nodes. A multiple-input neuron with multiple inputs 'R' is shown in Figure 1. The individual inputs P_1, P_2…., P_R are weighted by corresponding elements $W_{1,1}$, $W_{1,2}$, … , $W_{1,R}$ • of the weight matrix W.

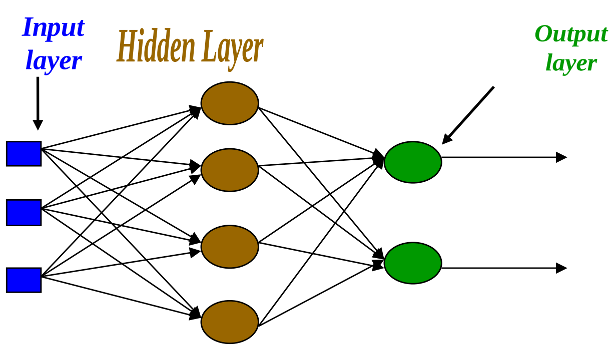

Fig2. Multi-Feed Forward Neural Network

The neuron also has a bias 'b', which is summed with the weighted inputs to form the net input

'n':

$$n = W_{1,1} P_1 + W_{1,2} P_2 + \cdots W_{1,R} \bullet P_R + b \quad (15)$$

In matrix form, this can be rewritten as

$$n = W. P + b \quad (16)$$

Now, the neuron output is given as,

$$a = f\,(W. P + b) \quad (17)$$

The transfer function used above is a log-sigmoid transfer function. This transfer function takes the input (which may have any value between plus and minus infinity) and squashes the output in between 0 to 1 range, according to the expression

$$y = \frac{1}{(1+e^{-n})} \quad (18)$$

The nodes at a particular stage constitute a layer. The first layer is called input layer and last layer is called output layer. The layers in between output and input layer are called hidden layers. As the number of hidden layers in the network increases, the performance of network increases. Each node in a network serves the purpose of summation of all its inputs. The output of a node is further applied to the next node.

IV. EXPERIMENTAL RESULTS

A general purpose image database consists of 90 images which used for experimentation. T he database consists of different categories such as Africans and villages, Beaches, Buildings, Buses, Dinosaurs, Animals, Flowers, Horses, Food and Natural scenes. All the categories are used for retrieval. These images are stored in JPEG format with size 256x256 and each image is represented with HSV color space. One limitation of the LE semantic space is that, it only contains those images in database, i.e., training set. It is unclear how to evaluate the map in the LE semantic space for new test data. To overcome this limitation, we are learning an algorithm for query image outside the database by Artificial Neural Network (ANN).

In order to measure retrieval effectiveness for an image retrieval system, precision and recall values are used, the ratio of relevant retrieved images to the total number of retrieved images (precision) and the ratio of retrieved relevant images to the total number of relevant images in the database (Recall).Table1 summarized the experiment result based on different scope with Euclidean distance, Image Manifold, Artificial Neural Network. Table2&3 summarizes the experiment results compared with the Euclidean &IML algorithm.

Precision = (the no. of retrieved images that are relevant) / (The no. of retrieved images).

Recall = (the no. of retrieved images that are relevant) / (The total no. of relevant images).

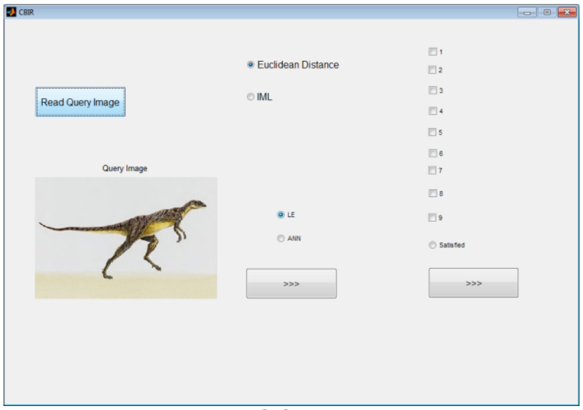

Fig3.Query image

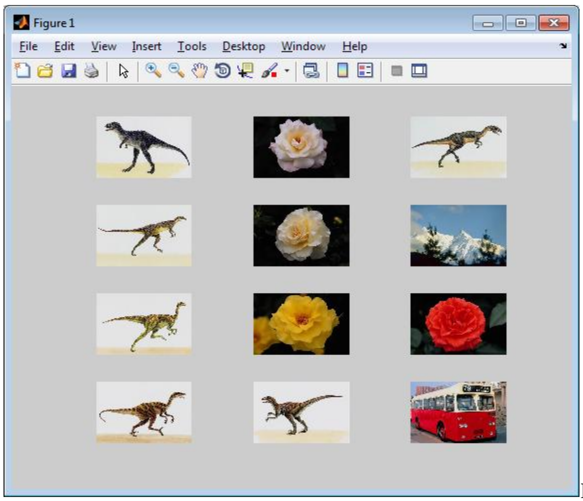

Fig3
. Image retrieval by Euclidean Distance

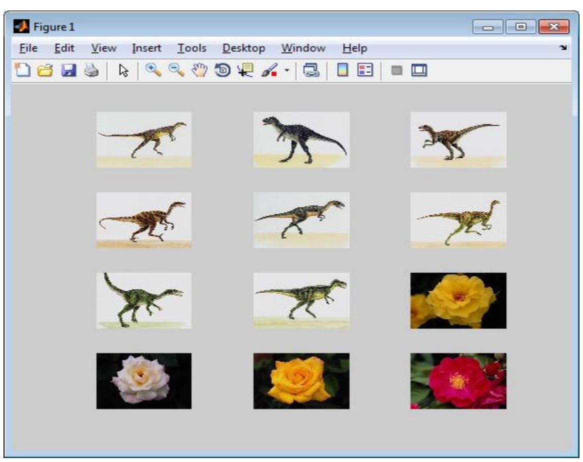

Fig4 Image Retrieval by IML algorithm without f/b

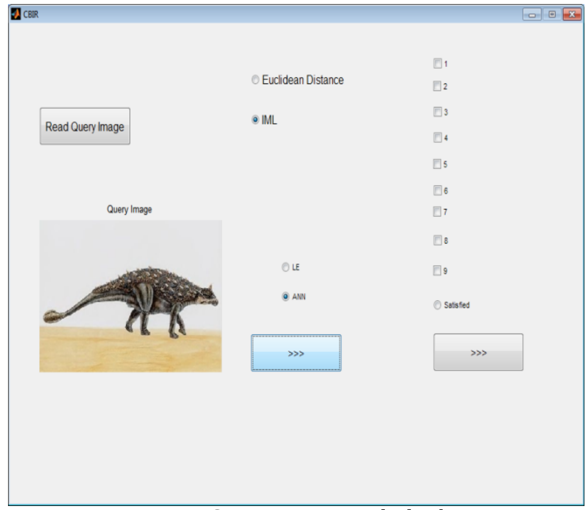

Fig 5 Query image outside database

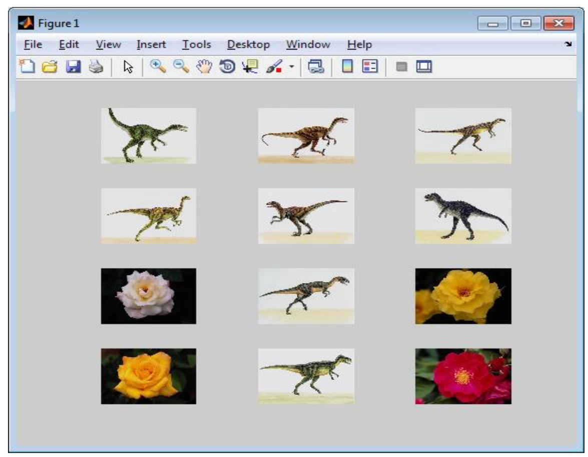

Fig 6.Image Retrieval by ANN For outside database

Top Images	**5**	**10**	**15**	**20**
ED	49.43	39.14	29.36	21.99
IML	74.88	48.65	39.43	31.70
ANN (outside database)	70.88	40.99	31.44	25.78

Table1: Average precision for Top Images: Euclidean Distance(ED), Image Manifold Learning (IML) without f/b, Artificial Neural Network (ANN)

Class	Average precision using		
	ED	IML (w/of/b)	IML (with f/b)
Bus	52.38	53.57	84.33
African	29.76	39.49	63.49
Beaches	35.71	46.86	76.39
Building	29.76	33.33	57.14
Dinosaur	47.62	58.33	74.60
Elephant	23.81	30.95	71.43
Flower	60.71	52.38	69.84
Horse	30.95	32.14	58.73
Mountain	23.81	23.81	57.14
Food	17.86	33.33	58.73

Table2: Average precision of: Euclidean Distance (ED), Image Manifold Learning(IML) With f/b & without f/b

Class	Average Recall using		
	ED	IML (w/of/b)	IML (with f/b)
Bus	69.84	71.43	84.13
African	39.68	52.38	63.49
Beaches	47.62	57.14	76.19
Building	39.68	44.43	57.14
Dinosaur	69.49	77.38	74.60
Elephant	31.75	41.27	71.43
Flower	80.95	69.84	69.8
Horse	41.27	42.86	58.73
Mountain	31.75	31.76	57.14
Food	23.81	44.44	58.73

Table3: Average Recall of: Euclidean Distance (ED), Image Manifold Learning (IML) With f/b & without f/b

V. CONCLUSION

In this paper, under the assumption that the data lie on a sub manifold hidden in a high dimensional feature space, we developed an Image Manifold Learning (IML) algorithmic to learn the mapping between low-level image features and high-level semantics. It utilizes relevance feedback to enhance the performance of image retrieval system from both short- and long-term perspectives. In IML algorithm we consider the user's relevant feedback. Hence experiment result shows that IML algorithm is effective than Euclidean distance. Retrieval using ANN algorithm provides accurate result which matches with human perception.

REFERENCES

[1] Luo Si · Rong Jin · Steven C.H. Hoi · Michael R. Lyu-*Collaborative image retrieval via regularized metric learning.*

[2] He, X., Ma, W.-Y., *Zhang, H.-J.: Learning an image manifold for retrieval*

[3] M. Belkin and P. Niyogi, "*Laplacian eigenmaps for dimensionality reduction and data representation*", *Advances in Neural Information Processing Systems*, 2001.

[4] X. He, O. King, W.-Y. Ma, M.-J. Li, and H.-J. Zhang,"*Learning a semantic space from user's relevance feedback for image retrieval*", *IEEE Trans. on Circuit and System for Video Technology*, Jan, 2003.

[5] Ashwin, T.V., Navendu, J., Ghosal, S.," *Improving image retrieval performance with negative relevance feedback*" In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP'01), UT (2001)

[6] S. Mangijao Singh , K. Hemachandran," *Content-Based Image Retrieval using Color Moment and Gabor Texture Feature*"IJCSI International_Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012.

[7] Dhanya Bibin, Punitha P." *Image Classification using ArtificiNeural Networks: An experimental study on corel database*,"In International Journal of Machine Intelligence ISSN: 0975–2927 & E-ISSN: 0975–9166, Volume 3, Issue 4, 2011, pp-225-229

[8] Shereena V.B., Julie M. David" *Content Based Image Retrieval Classification Using Neural Network*" Vol.6, No.5, October 2014

[9] Syed Akhter Hussain ,Prof.A.N.Holambe" *A Comparative Study Of Diffrent Transformation Techniques For Cbir*" October - 2012.

[10] S. Mangijao Singh , K. Hemachandran, "*Content-Based Image Retrieval using Color Moment and Gabor Texture Feature*" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012.