ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Ensemble of Distance Measures for Multimodal Video Retrieval System

Prof. D.D. Pukale[1], Neha Kumari[2], PalakVerma[3], Sahu Dhanalaxmi Indramohan[4], Shelke Chaitali Vikram[5]

Associate Professor, Department of Computer Engineering, Bharati Vidyapeeth's College of Engineering For Women, Savitribai Phule Pune University, Pune, India[1]

B.E. Student, Department of Computer Engineering, Bharati Vidyapeeth's College of Engineering For Women, Savitribai Phule Pune University, Pune, India[2,3,4,5]

ABSTRACT: Content-based retrieval of video data has become a challenging issue as video contains several types of audio and visual information which are difficult to extract. We present a model that retrieves the related videos from the video database on the basis of a query clip. The system splits the video into a sequence of frames and extracts a small number of key frames by employing Clustering and classifies them according to the various features like colour, texture, shape and audio and then they are stored in feature library. We propose six distance measure methods and spectrum matching with windowing for similarity measure. As we are using ensemble of matrix, we are getting multiple retrieved videos from six distance measure separately, then we calculate the frequency count of the retrieved videos and display the videos in the descending order of count, then the relevant video are displayed.

KEYWORDS: Content based video retrieval, key frame, clustering, feature extraction and similarity matching.

I. INTRODUCTION

In recent years the usage of multimedia information is increasing rapidly. Out of all the different media types present, video is the most challenging one as it combines all the other media information into a single data stream. Due to the higher transmission rates, decreasing cost of storage devices and improved techniques of compression, video is becoming available everywhere, however due to its length and its nature of unstructured format it is not easy to access [2].

In order to have an easier access to the large collections of video, our system will play an important role. In this system we are going to use different features of video like colour, shape, texture and audio to retrieve videos. It's a user friendly application in which a user can quickly and easily retrieve the video by using different type of query inputs as image query, audio query or a video query.

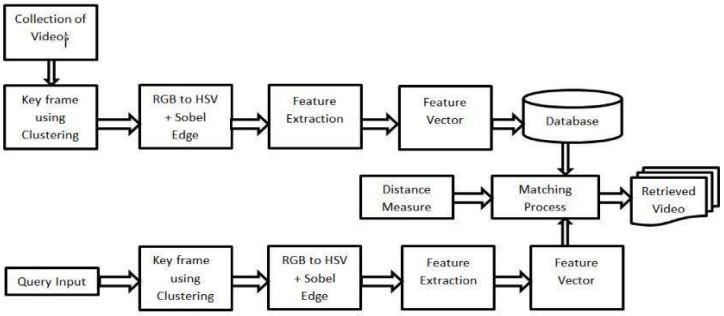

Fig 1. Architecture of MVR System

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

II. PREVIOUS WORK ON VIDEO ABSTRACTION

There have been numerous researches in video retrieval system based on text detection, colour, texture, shape feature and audio. Despite of these efforts the existing system is not powerful enough either due to algorithm for different feature extraction, similarity measure or type of input query. Existing system broadly use text and images for video retrieval from the database.

Content based Video Retrieval using Edge detection, Entropy and black and white colour feature [5], this system works only on entropy, edge and colour. In content based video retrieval using information theory[6], the system works on clustering but it doesn't cluster automatically.

In content based lecture video retrieval using speech and video text information[5], although it integrate all the features and gives better result but it doesn't work for audio query as it is based on speech recognition.

In video Retrieval using automatically extracted audio [7], it is based on metadata and audio.

III. PROPOSED SYSTEM

Our system is having four modules.

Module 1- It will convert the input query like video clip into frames and will take image and audio clip as it is.

Module 2- Select key frames from frames using clustering.

Module 3- Convert the key frames into features vectors based on colour, shape, texture and for audio based on audio features.

Module 4- Searches for similar videos based on feature vector values by comparing the histograms stored in database.

Our multimodal content based video retrieval system uses the contents of a video such as colour, shape, and texture, audio to index and represent the video for easy retrieval. Now the contents of the videos in the database are extracted and described by feature vectors which are used to form a feature vector database. In order to retrieve videos, users provide the system with different queries. Then these queries are converted into internal representation of feature vectors by the system. The indexing between the feature vectors of the query example and those in the database are then calculated by similarity matching algorithm and the retrieval is performed.

A. Keyframe

In a video as there are lots of redundancy present due to similar information in each of the frames of the video. Therefore, key frame helps to reduce the redundancy. The relevant content should be represented as much as possible by the extracted key frames. The key frame extraction based on features such as colours (particularly the colour histogram), texture and shapes [3].

We are using Clustering method for key frame extraction as this algorithm cluster the frames and then choose frames closest to the cluster centres as the key frames.

Clustering is a powerful technique used in various disciplines for example Pattern Recognition Speech Analysis and Information Retrieval etc. [13].

Unsupervised clustering approach was introduced to determine key frames within a shot boundary In this ,we using a different clustering approach to key frame extraction.

Given a video shot

s = { f1, f2 ……fn }

from a shot boundary detection algorithm we cluster the N frames into M clusters say

σ1,σ2,...,σn

The similarities between two frames is defined as the similarity of their visual content and these visual content could be colour, texture ,shape of an object of the frame or the combination of the above. In this, we select the colour histogram of a frame as our visual content, although other visual contents are readily integratable into the algorithm. The colour histogram we used is 16x8 2D HS colour histogram in the HSV colour space. The similarity

$$\sum_{h=1}^{16}\sum_{s=1}^{8} min(H_i(h,s), H_j(h,s)) \qquad (1)$$

Any clustering algorithm uses a threshold parameter, which controls the density of clustering. The higher threshold

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

means more the number of clusters. In human learning and recognition system we also have this threshold concept. For example, if the threshold is high, we will classify them into different categories. The threshold parameter provides us a control over the density of frame classification. Before a new frame is classified into a certain cluster, firstly the similarity between this node and the centroid of the cluster is computed. If this value is less than it mean that this node is not close enough to be added into the cluster.
The unsupervised clustering algorithm can be summarized as follows:

→

1. Initialization: f1 σ1
2. Get the next frame fi. if the frame pool is empty; go to 6.
3. Calculate the similarities between fi and existing cluster.
4. Determine which cluster is closest to fi and maxsim. let,

$$Maxsim = \max_{k=0}^{numCluster} sim(f_i, \sigma_k).$$

If the maxsim<threshold,it means that fi is not closed enough to be put in any of the cluster go to 5.otherwise put fi into the cluster which has maxims and go to 6.

5. numCluster=numCluster + 1.A new cluster if formed:

$$f_i \rightarrow \sigma_{numCluster}$$

B. Feature Extraction

1. **Colour Feature:** For video retrieval purpose colour is the mostly used visual content.

HSV Color Model: It defines a color space consist of three main components: Hue (color type ranges from 0to360).

Saturation(vibrancy of the color ranges from 0% to 100%). Value (brightness of the color ranges from 0 to 100%)[1].

The different HSV planes are shown as Figure 2.

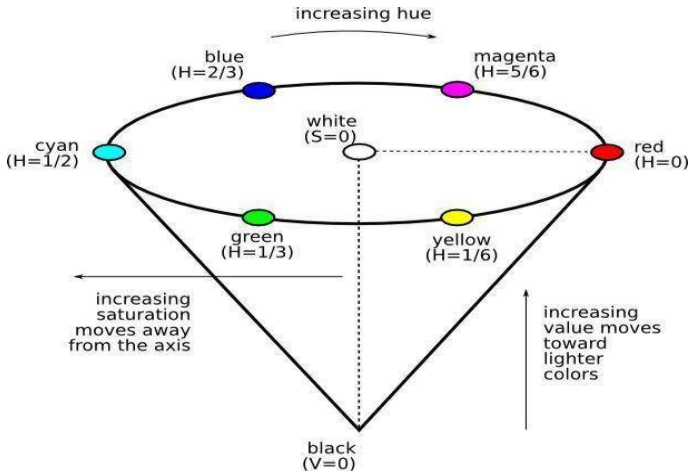

Fig 2: HSV Model

In it quantization of the number of colors into several bins is done so as to decrease the number of colors used in retrieval.
We propose the scheme to produce 15 non-uniform colours. The formula that transfers from RGB to HSV is defined as below:

$H = \cos^{-1}\{½[(R-G) + (R - B)]/\sqrt{(R-G)2 + (R-B)(G-B)}\}$
$S = 1 - 3/R + G + B[\min(R, G,B)]$ $V=1/3(R + G + B)$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

V=1/3(R + G + B)

The R, G, B represent red, green and blue components respectively with value between 0-255. In order to obtain the value of H from 0 to 360, the value of S and V from 0 to 1, we do execute the following formula:
H= ((H/255*360) mod 360
V= V/255
S= S/255

1. Shape Feature:

For shape feature Sobel edge detection algorithm is used. The advantage of Sobel edge operand is its smoothing effect to the random noises in the frame. And because it is the differential separated by two rows or two columns, so the edge elements on both sides have been enhanced and make the edge seems thick and bright. Sobel operator is a gradient operator. The first derivative of a frame is based on a variety of two-dimensional gradient approximation, and generates a peak on the first derivative of the image, or generates a zero-crossing point on the second derivative .Calculate the magnitude and the argument value of the frame horizontal and vertical first-order or second-order gradients, at last calculate modulus maxima along the angular direction and obtain the edge of the frame.

2. Texture feature:

For shape feature Gray Level Co- occurrence Matrix (GLCM).is used. The identification of specific textures in an frame is achieved primarily by modeling texture as a two-dimensional gray level variation. This two dimensional array is called as Gray
Level Co- occurrence Matrix (GLCM).
The steps for extracting texture feature of key frame using GLCM can be given as below:

1) Separate the R, G,B planes of frame.
2) Repeat steps 3_6 for each plane
3) Compute four GLCM matrices (direction for ∂=00, ∂=450, ∂=900, ∂=1350) as given by eq.(1)
4) For each GLCM matrix compute the statically feature Energy (Angular second moment),
5) Entropy (ENT), Correlation (COR), Contrast (CON) [18, 22] as follows where p(i,j) is probability density.

3. Audio Feature:

Audio is non-stationary signal where properties change quite rapidly over time. This is fully natural and nice thing but makes
the use of DFT or autocorrelation as such impossible. For most phonemes the properties of the audio remain invariant for a
short period of time (5-100 ms). Thus for a short window of time, traditional signal processing methods can be applied
relatively successfully
Most of audio processing in fact is done in this way: by taking short windows (overlapping possibly) and processing them.
Basic Functionality in Matlab for audio file:-

1. read an audio file (i.e., open a .wav audio file and read the audio sample into a MATLAB array).
2. write an audio file (i.e., write a MATLAB array of audio samples into a .wav audio file)
3. play a MATLAB array of audio samples as an audio file
4. record an audio file into a MATLAB array
5. plot an audio file (MATLAB array) as a waveform
6. plot an audio file (MATLAB array)
7. convert the sampling rate associated with an audio file (MATLAB array) to a different sampling rate

high pass filter an audio file (MATLAB array) to eliminate hum and low frequency noise

C. Similarity Measurements:

1) Video: We are going to use six distance measures in order to increase the accuracy of the result.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

1. Euclidean Distance: Deriving the distance between two data points involves computing the square root of the sum of the squares of the differences between corresponding values.

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = d$$

2. Spearman Distance: It is a measures the correlation between two sequences of values.

$$d_{st} = 1 - \frac{(r_s - \bar{r}_s)(r_t - \bar{r}_t)'}{\sqrt{(r_s - \bar{r}_s)(r_s - \bar{r}_s)'}\sqrt{(r_t - \bar{r}_t)(r_t - \bar{r}_t)'}}$$

3. Correlation: It is a measure of statistical dependence between two random variables or two random vectors of arbitrary, not necessarily equal dimension

$$d_{st} = 1 - \frac{(x_s - \bar{x}_s)(x_t - \bar{x}_t)'}{\sqrt{(x_s - \bar{x}_s)(x_s - \bar{x}_s)'}\sqrt{(x_t - \bar{x}_t)(x_t - \bar{x}_t)'}}$$

4. Cosine Distance: Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them.

$$d_{st} = 1 - \frac{x_s x'_t}{\sqrt{(x_s x'_s)(x_t x'_t)}}$$

5. Cityblock distance: It is the sum of distances along each dimension. This is equal to the distance a traveler would have to walk between two points in a city. The city-block distance is a metric, as it satisfies the triangle inequality.

$$d_{st} = \sum_{j=1}^{n} |x_{sj} - x_{tj}|$$

6. Minkowski distance: The Minkowski distance is a metric in a normed vector space which can be considered as a generalization of both the Euclidean distance and the Manhattan distance.

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n} |x_{sj} - x_{tj}|^p}$$

Ensemble of Matrix: In our system, ensemble of matrix is a group of six distance measure described above. Firstly, each distance measure calculates the distance between the query input and the database videos/audios and then displays the relevant videos. Then for the ensemble of matrix we calculate the frequency count of the retrieved videos. Frequency count is the number of occurrence of the videos in each retrieved results of different distance measures. Then we display the videos in the descending order according to the frequency count.

2) Audio:

For the similarity measurement of audio feature we are using inbuilt windowing method of Matlab.

Windowing: It reduces the amplitude of the discontinuities at the boundaries of each finite sequence acquired by the digitizer.

Windowing consists of multiplying the time record by a finite- length window with an amplitude that varies smoothly and gradually toward zero at the edges. This makes the endpoints of the waveform meet and, therefore, results in a continuous waveform without sharp transitions. This technique is also referred to as applying a window. It is the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

process of taking a small subset of a larger dataset, for processing and analysis. A naive approach, the rectangular window, involves simply truncating the dataset before and after the window, while not modifying the contents of the window at all.

1. Firstly, we read the query audio file in .wav file in order to convert it from time domain to frequency domain.
2. Then record the samples of query audio file and store it in .mat file.
3. Audio is extracted from stored videos present in the database and their samples are calculated and stored in .mat file.
4. Now the .mat files of query audio and stored audio samples are compared by using windowing.
5. (a) A small rectangular window is used to scan the query audio file with the stored audio file for similarity.
(b) If the comparisons of those signals are below a certain threshold**,** then we are retrieving videos of those audios files as the final results

IV. RESULTS OBTAINED

A. Dataset:

In our system, we have a database of 100 videos, which are divided into five categories like Bollywood, Sports, News, Educational and Documentary
Each category is having 20 videos.

B. Evaluation measure:

We use recall and precision as the standard evaluation measurement. They are defined as:-

Precision=Total number of retrieved relevant videos / Total number of retrieved videos

Recall= Total number of retrieved relevant videos / Total number of relevant videos

Table 1: Results Obtained For Video Query

Category / Results	Query Video	Total video in database	Retrieved video by system	Relevant video retrieved	Relevant video in database	Precision	Recall
Bollywood	vid1.mp4	20	5	5	7	1.00	0.71
Sports	Vid24.mp4	20	5	4	9	0.80	0.44
News	Vid43.mp4	20	8	7	9	0.87	0.78
Educational	Vid62.mp4	20	7	6	10	0.85	0.60
Documentary	Vid95.mp4	20	5	4	8	0.80	0.50

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

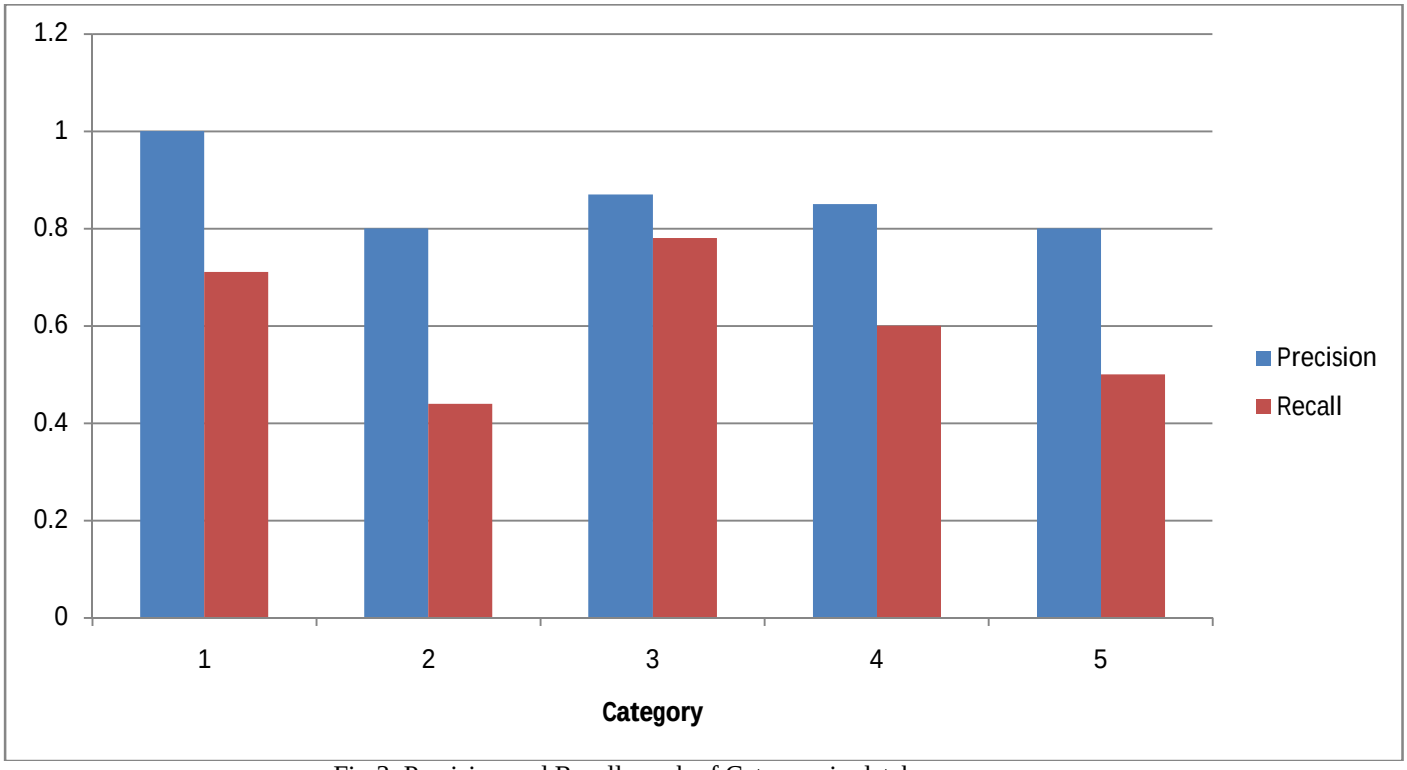

Fig 3: Precision and Recall graph of Category in database

In order to obtain results more accurately we are using six different distance measure and then the final output is been displayed through ensemble of matrix, as shown in table 2 and table 3.

Table 2: Results for the input video query "moon.mp4" based on different distance measure

Category / Results	1	2	3	4	5
Euclidean	35	57	11	23	68
City block	23	57	68	35	11
Minkowski	16	19	35	11	23
Correlation	11	68	35	16	57
Cosine	11	43	16	19	68
Spearman	35	19	11	68	57

The result obtained from each distance measure in table 2 and then frequency count of retrieved results are shown in table 3 .
In final result we displayed the result in descending order of frequency count.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Table 3: Results of Ensemble of matrix(final result by combining all distance measure) from frequency count

Image Index	Frequency count
11	6
35	5
68	5
57	4
23	3
16	3
19	3
43	1

Table 4: Results obtained from distance measure for input query “moon.mp4”

Distance method / Result	Query Video	Total video (searching in all category)	Relevant video in database	Retrieved video by system	Relevant video retrieved	Precision	Recall
Euclidean	moon.mp4	100	10	7	4	0.57	0.4
City block	moon.mp4	100	10	6	4	0.66	0.4
Minkowski	moon.mp4	100	10	8	5	0.65	0.5
Correlation	moon.mp4	100	10	9	3	0.33	0.3
Cosine	moon.mp4	100	10	8	2	0.25	0.2
Spearman	moon.mp4	100	10	7	2	0.28	0.2

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

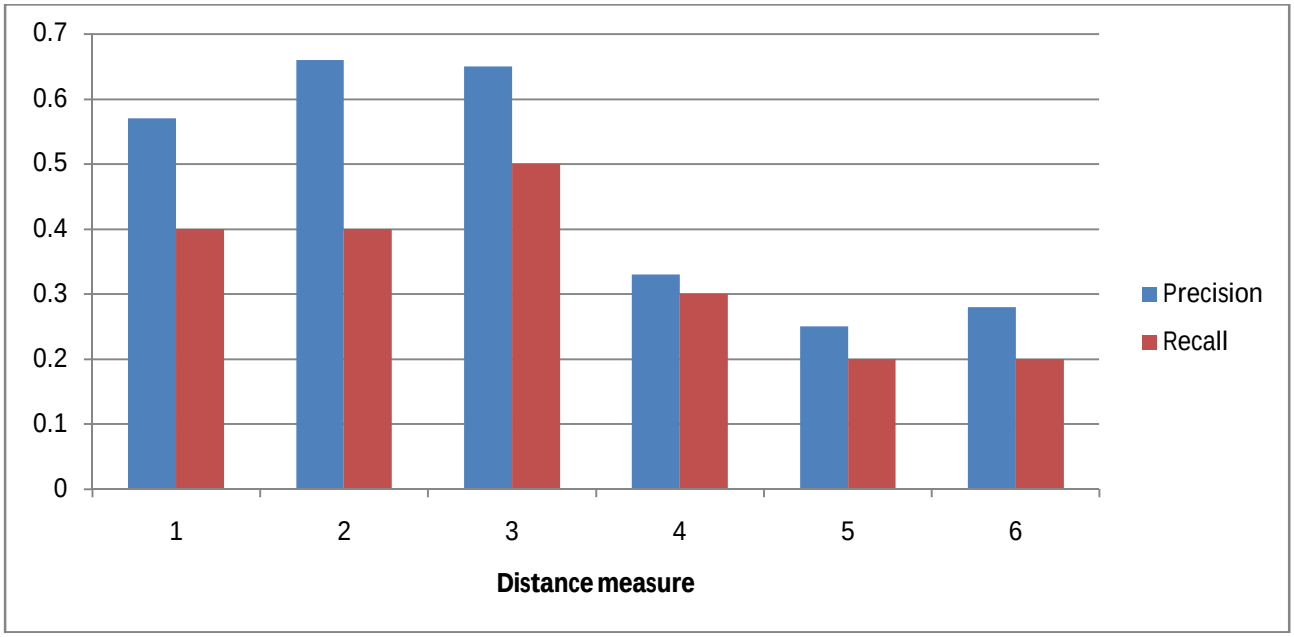

Fig 4: Precision and Recall graph of distance measure result

V. CONCLUSION

In this system we are implemented multimodal content based video retrieval system. Video Retrieval is area that integrates various fields such as machine learning, artificial intelligence, data base management systems, etc. There are large numbers of algorithms for retrieval of videos. In our proposed system we have implemented the retrieval system by integration of different video features and key frames extraction using clustering. Multiple features produces effective and efficient system as the time required for the retrieval is reduced. By introducing the ensemble of matrix of six distance measures, most relevant results are displayed, As each distance measure calculates the distance between the query input and the database videos/audios and then displays the relevant videos. Then for the ensemble of matrix we calculate the frequency count of the retrieved videos. Then we display the videos in the descending order according to the frequency count. The only constraint in our system is that if the video size is very large then the segmentation and indexing becomes time consuming.

REFERENCES

1. Prof D.D. Pukale, Miss. Laxmi P. Dabade, Miss. Varsha B. Patil, Miss. ChandaniP.Lodha, Miss. Nikita S. Ghode, "Image and annotation retrieval via image content and tags", International Journal of Scientific and Research Publications, Volume 4, Issue 4, April 2014 1 ISSN 2250-3153 .
2. ZhongQu,"AnImprovedKeyframe Extraction Method Based on HSV Colour Space", Journal Of Software, Vol.8, No.7, July 2013.
3. ShimnaBalakrishnan, Kalpana S. Thakre," VIDEO MATCH ANALYSIS:A Comprehensive Content based Video Retrieval System", International Journal of Computer Science and Application Issue 2010 ISSN 0974- 076752.
4. Shripd A.Bhat, OmkarV.Sardessai, PreeteshP.Kunde andSarveshS.Shirodkar," Overview of Existing Content Based Video Retrieval Systems",International Journal of Advanced Engineering and Global Technology Vol-2 Issue-2, February 2014ISSN No: 2309-4893
5. V. Patel and A.V. Deorankar, "Content Based Video Retrieval using Entropy, Edge Detection, Black and White colour Features", *in proc. IEEE Computer Engineering and Technology (ICCET), 2nd International Conference on Vol. No. 6 Page(s): 272 –276*, 2010
6. Hadi Yarmohammadi and Mohammad Rahmati, "Content Based Video Retrieval using Information Theory", *in proc. IEEE Iran Conf. ,Machine vision and Image Processing, pp. 214-218*, 2013.
7. Kale, A. and Wakde, D.G., "Video Retrieval Using Automatically Extracted Audio", *in proc. IEEE International Conference on Cloud & Ubiquitous Computing & Emerging Technologies (CUBE), DOI: 10.1109/CUBE.2013.32, Page(s): 133 - 136*, 2013
8. Padmakala, S., Anandha Mala, G.S. and Shalini, M, "An Effective Content Based Video Retrieval Utilizing Texture, Color, and Optimal Key frame Features", in IEEE International Conference on Image Information *Processing (ICIIP), DOI: 10.1109/ICIIP.2011.6108864, Page(s): 1 – 6*, 2011
9. Alan Hanjalic,"Shot-Boundary Detection: Unraveled and Resolved", IEEE Transactions On Circuits And Systems For Video Technology, Vol.12,February 2002.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

10. T.N.Shanmugam , PriyaRajendran, "An Enhanced Content-Based Video Retrieval System Based On Query Clip", International Journal of Research andReviews in Applied Sciences ISSN: 2076-734X, EISSN: 2076-7366 Volume 1, Issue 3(December 2009).

11. VakkalankaSuresh,C.Krishna Mohan, R. Kumaraswamy and B.Yegnanarayana, "Content-Based Video Classification using SVMs", in Int. Conf. Neural Information Processing (ICONIP-04), Calcutta, India, Nov. 22-25, 2004, pp. 726- 731.

12. Shruti Vaidya, Dr. Kamal Shah," Audio Denoising, RecognitionandRetrieval by Using Feature Vectors**",** *IOSR Journal of Computer Engineering.*