Color Content Based Video Retrieval Using Discrete Cosine Transform Applied On Rows and Columns of Video Frames with RGB Color Space

Dr. Sudeep D. Thepade  
Dean(R&D) and Professor  
Department of Computer Engineering  
Pimpri-Chinchwad College of Engineering

Ajay A. Narvekar  
B.E.Student  
Department of Computer Engineering  
Pimpri-Chinchwad College of Engineering

Ameya V. Nawale  
B.E.Student  
Department of Computer Engineering  
Pimpri-Chinchwad College of Engineering

Abstract—The volume of video data is increasing because of advanced technology in imaging communication. This tends to increase the popularity of the videos. The conventional textual description based video retrieval which is based on text entered by user, so, suffers from user perception. Hence, content based video retrieval is more effective than text based retrieval, as it gives result closer to semantic view. Here, a novel content based video retrieval based on discrete cosine transform (DCT) applied on rows and columns of video frames to extract features as selected high energy is proposed with the help of RGB colour space. Video database of 500 video samples spread across 10 categories is considered as a test bed for experimentation. Performance comparison is done with the help of cross-over point of average precision and average recall values of 500 query videos fired on the whole database. The experimental results have shown that the video retrieval with DCT applied on columns gives better performance.

Index Terms—Content Based Video Retrieval (CBVR), Color Spaces, Discrete Cosine Transform (DCT), Feature coefficient Extraction.

I. INTRODUCTION

The advancement in imaging technology holds a very high attraction towards the videos rather than still images. Videos are considered to be more expressive in human communication. Due to advanced digital technology, these videos can be created, stored, deleted, updated, and edited by the user. Also video is gaining more popularity among users on internet. The user wants any video to be searched and retrieved for relevant matches from huge video pool. Due to ease of technology numbers of videos are being generated and stored resulting into voluminous video databases.

In conventional text-based video retrieval, the query is processed on the textual properties of the videos in database. In this retrieval, the irrelevant videos are also retrieved which are not required by the user. So, if the keywords related to video is stored and accordingly the query is processed, a problem occurs in searching and retrieving the videos due to more keywords stored regarding each video in database. Also, if user’s perspective is considered then, each video will have different keywords. This will create an ambiguity in video retrieval using this technique. To overcome this lacuna content based video retrieval is proposed, where the video is described with the help of its contents as in color, texture or shape information of the video frames.

Here, the novel idea is proposed as content based video retrieval (CBVR) using discrete cosine transform (dct) with earlier color spaces in content based image retrieval [9]. In this approach, the color content of each video is considered to form feature vector.

As proposed, the feature vector for each video is generated using discrete cosine transform technique. Ten frames are extracted from each video sample (here every fifth frame is considered). For every frame, the pixels are divided into two clusters based on the threshold value as upper cluster (having pixels greater than threshold) and lower clusters (having values lesser than threshold). The averages of these upper and lower clusters are considered as elements of the feature vector.

II. DISCRETE COSINE TRANSFORM ON ROWS AND COLUMNS

A. DISCRETE COSINE TRANSFORM

The DCT block computes the unitary discrete cosine transform (DCT) of each channel in the M-by-N input matrix, u [9].

\[ y = dct(u) \] % Equivalent MATLAB code

When the input is a sample-based row vector, the DCT block computes the discrete cosine transform across the vector dimension of the input. For all other N-D input arrays, the block computes the DCT across the first dimension [9]. The size of the first dimension (frame size), must be a power of two. To work with other frame sizes, use the Pad block to pad or truncate the frame size to a power-of-two length.

When the input to the DCT block is an M-by-N matrix, the block treats each input column as an independent channel containing M consecutive samples. The block outputs an
M-by-N matrix whose lth column contains the length-M DCT of the corresponding input column.

\[
y(k, l) = w(k) \sum_{m=1}^{M} u(m, l) \cos \left( \frac{(2m-1)(k-1)}{2M} \right)
\]

\[k = 1, ..., M\]

Where

\[w(k) = \begin{cases} \frac{1}{M}, & k = 1 \\ \frac{2}{M}, & 2 \leq k \leq M \end{cases}\]

### III. CONTENT BASED VIDEO RETRIEVAL

In modern video retrieval system, the basic need is to handle the faster growing database as well as efficiency of retrieval. Many developments are undergoing to improvise the efficiency and accuracy video retrieval. The major technique is by using contents of videos as feature vector and storing those vectors in database. These feature vectors are used for processing of query fired by user and comparing the query’s feature vector and feature vector of each video. The relevant or closest match videos are displayed as relevant results.

This paper talks about content based video retrieval with color content using discrete cosine transform on rows and columns. The frame is extracted by 5th frame frequency. In this approach of system, dct is applied on the rows and columns. After feature extraction module, the similarity check is done by using absolute difference between query video and each video from data set. The similar videos of smaller difference are relevant videos to the query fired. These relevant videos are displayed.

Figure 1 shows the basic system model of content based video retrieval of this paper.

![Fig 1 Basic Content Based Video Retrieval System](image)

### IV. IMPLEMENTATION

#### A. Platform Experimentation

The implementation of model is done in MATLAB with basic system of Intel core 2 duo (2.93GHz) with 2GB RAM and minimum of 250GB hard disk for storage. The modules of model are run under MATLAB compiler. The operating system used is windows 7 for matlab environment.

#### B. Video Dataset

The videos are categorized in 10 assorted categories containing 50 similar videos. So, in total there are 500 videos in data set for experimentation. All the videos are normalized in each category. The videos are chosen such that those of same categories differ very minutely in color content so as to find accurate result.

![Fig 2 Sample videos from Video Dataset](image)

#### C. Feature Extraction

In this, feature vectors are generated by using 10 frames of each video (in multiples of 5 starting from 1 i.e., 1, 6, 11, 16 and so on till 46) and by applying dct on them in RGB color space [9]. After taking DCT vectors, we are limiting the values as only first 5 values per color matrix that is R, G and B. This will result into 15 values per frame and 150 values per video. This will help in boosting performance by saving comparison time and vector storage space. Then, these feature vectors are used for query processing and performance comparison.

#### D. Performance Comparison Parameters

The performances of the proposed approaches are evaluated on the video dataset using the precision-recall cross over points. Precision-recall are given by equations 2 and 3.

\[
\text{precision} = \frac{\text{relevant } \cap \text{ retrieved }}{\text{retrieved }}
\]

\[
\text{recall} = \frac{\text{relevant } \cap \text{ retrieved }}{\text{relevant }}
\]
E. Similarity Measurements

Here absolute difference (AD) is used for similarity measurement of feature vectors of the videos in content based video retrieval. In this method, the feature vectors of a query video are compared with feature vectors of all the videos stored in the database by taking absolute difference between them. The differences will be ordered in increasing fashion so as to show most relevant searches on top, as the videos with which the absolute difference will be lower will be most relevant ones. Like this first 50 search results are taken into consideration and shown thereby. By this way these results will be taken to find precision and recall for each video.

V. RESULTS AND DISCUSSION

Both the variations of the proposed CBVR method are experimented with 500 queries fired on video database for each variation.

Fig 3: Average precision-recall cross over point values of proposed DCT based CBVR using AD for RGB color space

RGB color space and AD similarity measurement criteria are considered. The precision-recall cross over points of each query is computed to get the average precision-recall cross point of respective variation of proposed CBVR. Higher the cross over point value better is the CBVR performance. Figure 3 shows average cross-over points of precision-recall values for respective DCT methods obtained using absolute difference in DCT based CBVR. DCT on columns have shown better results in each case than that on rows.

VI. CONCLUSION

Content based video retrieval is gaining momentum over text based video retrieval. This paper focuses mainly on selection of features by applying DCT on rows and columns of video frames and selecting first 15 high energy coefficients. In all, two variations of the proposed CBVR methods are experimented using RGB color space and one similarity measurement criteria (AD). With the video dataset having 500 videos spread across 10 assorted categories. The average precision recall crossover point value of 500 queries fired on dataset is considered for performance comparison. DCT applied on columns is found to be better method to extract contents of videos and represent it as a feature vector.

REFERENCES

[1] M. Petkovic “Content-based video retrieval”, Centre for Telematics and Information Technology, University of Twente.