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. 2, Issue 5, May 2014

Video Retrieval Using Fusion of Visual Features and Latent Semantic Indexing

Rashmi M, Roshan Fernandes

Dept. of Computer Science and Engineering, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Nitte,
Udupi Dist., Karnataka, India

ABSTRACT: It is commonly acknowledged that ever-increasing video databases should be efficiently indexed to facilitate fast video retrieval. Categorizing web-based videos is an important yet challenging task. The difficulties arise from large data diversity within a category, lack of labelled data etc. Similarity matching algorithm plays an important role in Video Retrieval System. Most of the video retrieval systems are designed using traditional similarity matching algorithms that are based on distance measures. The Accuracy of retrieval system depends on the method used for detecting shots, kind of video features used for retrieval. Semantic video indexing is a step towards automatic video indexing and retrieval. Here a Latent semantic indexing (LSI) technique, based on Singular Value Decomposition and fusion of visual features like color and edge is proposed for video indexing and retrieval. A key feature of LSI is its ability to establish associations between similar kinds of information, so the probability of producing accurate index is very high.

KEYWORDS: Latent Semantic Indexing, Singular Value Decomposition, Video Visual Feature, Similarity Matching Algorithm, Video Retrieval System.

I. INTRODUCTION

Video is a collation of still images presented to the viewer at a faster rate so as to give the illusion of motion. With the recent advances in processing capability of computer systems, internet, multimedia technologies, the number of multimedia files and archives increase dramatically. Therefore, it becomes an important research topic to mine and cluster the multimedia data, especially to accommodate the requirements of video retrieval in a distributed environment. Assisting a human operator to receive a required video sequence within a potentially large database is called as Video Retrieval System [1].

Dynamic video is an important form of multimedia information. A video may have an auditory channel as well as a visual channel. The available information from videos includes the following[2]: 1) video metadata, which are tagged texts embedded in videos, usually including title, summary, date, actors, producer, broadcast duration, file size, video format, copyright, etc.; 2) audio information from the auditory channel; 3) transcripts: Speech transcripts can be obtained by speech recognition and caption texts can be read using optical character recognition techniques; 4) visual information contained in the images themselves from the visual channel.

Semantic video indexing is the process of attaching concept terms to segments of a video [3].Which enables user to access videos according to their interests and preferences regarding video content. In most of the existing approaches to retrieve the videos, low level features such as color, texture are used for queries. Here users usually have a more abstract notion of what will satisfy them using low level features to correspond to high level abstractions is one aspect of semantic gap. The proposed method concentrates on the visual contents of videos like color and edge for video indexing and retrieval.

Content-based video indexing and retrieval have a wide range of applications such as quick browsing of video folders, analysis of visual electronic commerce (such as analysis of correlations between advertisements and their effects), remote instruction, digital museums, news event analysis [2].

II. RELATED WORK

Multimedia information indexing and retrieval are required to describe, store, and organize multimedia information and to assist people in finding multimedia resources conveniently and quickly [2]. Traditional content-based approaches for deriving semantics, purely based on low-level features, such as color and texture, have shown their limitations in conquering the so-called “semantic gap”[5].

In [1] authors used visual features like color, motion and edge of all the frames of all shots of a video, and then using LSI for finding similarity between clip and the videos in the database.

The basic idea of author in [6] is, semantic concept vectors are defined to represent the semantics of shot and scene. The retrieval is achieved at low level feature layer and high level semantic concept layer. The results from different layers are integrated to obtain a better result.

In [11] authors presented a generic semi-automatic text based approach for the development of a semantic video annotation and retrieval system. The system is a semi-automatic one as it is based on manually obtained speech transcripts or overlay texts of the videos.

Authors have developed a probabilistic contextual fusion method for improving the performance of semantic concept detection in images and videos. The method considers the reliability of individual detectors and refines the detection scores. Using the refined scores each detector computes a probabilistic estimate for the existence of each concept in [7].

Authors proposed a set of methods to organize and manage locatable videos, including splitting locatable video into geographic semantic features of video clips with roads and employing regular grid to build an index engine. In addition, a location-based information retrieval algorithm, with which the rapid and efficient retrieval of the video clips in the chosen region from an electronic map is presented in [12].

Authors in[13] proposed a content-based video copy detection using discrete wavelet transform . Where a three level decomposition is performed on the video frames using Daubechies wavelet transform (Db4) to obtain the feature descriptors. In which computation required for the similarity search is reduced by detecting the corresponding video first and then the original segment of the copy.

A. *Latent semantic indexing (LSI):*

LSI is a technique used for intelligent information retrieval (IR). In [8] it is stated that LSI is a method that exploits the idea of vector space model and Singular Value Decomposition (SVD). LSI uses SVD to reduce noise and dimensionality in the initial term-document representation and to capture latent relationships between the terms and the document.

Singular value decomposition (SVD) is already known for its ability to derive a low-dimensional refined feature space from a high-dimensional raw feature space, and capturing the essential structure of a data set within a feature set, and several studies have already focused on the use of SVD for texture analysis in image processing. SVD is based on a theorem from linear algebra which says that a rectangular matrix W can be broken down into the product of three matrices - an orthogonal matrix *U*, a diagonal matrix *S*, and the transpose of an orthogonal matrix *V*. The theorem is usually presented something like this:

$$W = U\,S\,V^T$$

Where $U^TU = I$, $V^TV = I$; the columns of *U* are orthonormal eigenvectors of WW^T , the columns of *V* are orthonormal eigenvectors of W^TW, and *S* is a diagonal matrix containing the square roots of eigenvalues from *U* or *V* in descending order[14]. In the LSI method the matrix *W* is approximated using *k* largest singular values. Other singular values are discarded.

$$W = U\,S_k V^T \quad \text{where k=min (size (W));}$$

Based on these we propose a system which will extract visual features from video and then apply SVD and LSI for indexing and retrieval.

III. Methodology

The proposed system works by analysing the key frames in video shots and extracting the different visual features from these frames. Then the feature matrix is formed by combining the different types of features from all shots of a video. Then the Latent Semantic Indexing is performed on the feature matrix. The result of LSI on all videos is stored in feature database. Finally the similarity measure is calculated for the given query video clip and the features in the feature database and the most similar videos are retrieved from the database.

Different steps in proposed system are as follows.

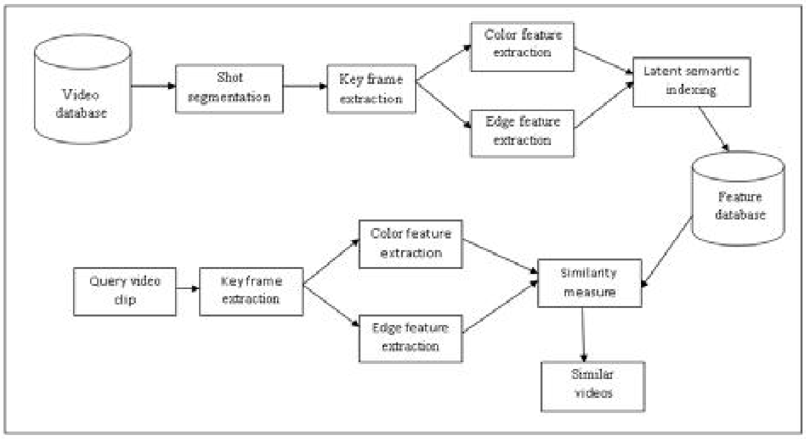


Fig.1. Framework for video retrieval system

A. *Shot Segmentation:*

Each video from video database are divided into number of shots. To accomplish this, the video is first converted into sequence of frames. Then compare the color histogram difference between the first frame and other frames sequentially. If the difference is within the threshold then that group of frames forms a shot. If the difference exceeds the threshold then take that frame as first frame for the next shot, compare it with remaining frames sequentially, and continue the process.

B. *Visual Feature Extraction:*

The application uses fusion of two visual features color and edge feature of all key frames of all shots of a video.

i. Color Feature Extraction:

Color quantization or color image quantization is a process that minimizes the number of individual colors employed in an image or frame of a video clip, generally with the intent that the new image should be as visually identical to the original image [1].

To accomplish this, firstly, all the key frames of a shot of the video clip are resampled. The resampled image is converted to $L * a * b$ color space. The color space converted frames are then divided into blocks of size Nc x Nc. Then, each block of every frame is subjected to DCT transformation. From the obtained result the second and third components of each block which are in DCT domain are scanned in zigzag fashion. While, zigzag scanning of a block, the first $Nc /2$ elements of the both the component are extracted and concatenated to form quantized color feature of the block. Likewise color feature vector of all the key frames are formed. Then concatenation of all color vector of key frames of a shot forms a color feature vector of the shot.

ii. Edge Feature Extraction:

To extract edge feature, the key frames of a shot is resampled. Then by applying ‘edge detection algorithm’ the edges are generated. The edge matrix ‘E’ *is* divided into number of blocks of size Nc x Nc. Then edge feature matrix is generated as given below.

Edge (i, j) = sum (all elements in block (i, j));

Then convert Edge matrix into a vector . The concatenation of edge vectors of all key frames of a shot forms the edge feature of the shot.

C. Video Retrieval Using Similarity Matching:

i. Form Feature Matrix A:

For each shot of a video, feature vectors of color and edge are collected from the previous step. The transpose of each feature vector is then computed for every video shot. Then for each shot, append color feature column below the edge feature column to form feature of that shot. Features of all the shots of a video are concatenated horizontally by applying zero wherever necessary to form feature matrix '*A*' of size M X N the video, where 'N' is number of shots in the video.

ii. Apply Latent Semantic Indexing:

Feature matrix *A* obtained from the above step *is* subjected to SVD decomposition. Using the SVD theorem, the matrix A is decomposed to compute '*k*' largest singular values and singular vector, as

$$\mathbf{A}_{M \times N} = \mathbf{U}_{M \times k} \,.\, \mathbf{S}_{k \times k} \,.\, \mathbf{V}^{T}_{k \times N}. \textbf{ Where } k = \min (M, N);$$

The U,S, V component of all videos are stored in the feature database.

iii. Similarity Measure:

When a query clip is given by user, extract its features form query feature 'q' of size 'Q x 1' and using each feature from feature database a query vector co-ordinate is formed as follows,

$$\mathbf{q_{co} = q^{T} \, x \, U \, x \, S^{-1}}$$

The Similarity between query clip and the each video in the database is measured as follows,

$$\mathbf{Sim = max\ ((q_{co}\ x\ V^{T}(j,:))\ /\ (norm(q_{co})\ X\ norm(V^{T}(j,:))))} \text{ where j is number of shots in the video.}$$

If 'Sim' is greater than the specified threshold video is similar to the query clip and is retrieved.

IV. Result Analysis

The most common evaluation measures used in IR are 'precision' and 'recall'. The same is used to measure the performance of proposed system.

Where

$$\text{Precision} = \frac{\text{No. of relevant videos retrieved}}{\text{Total number of videos retrieved}}$$

and

$$\text{Recall} = \frac{\text{No. of relevant videos retrieved}}{\text{Total No. of relevant videos in the collection}}$$

The proposed system is implemented in the MATLAB platform (version 7.14) and tested using the database video clips of MPEG-2, MP4, and AVI format. As there is no standard dataset specified in the literature, we prepared a dataset which contains collection of 60 videos of varying size. The sample retrieval results are shown in table 1.

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. 2, Issue 5, May 2014

Query Video	Precision	Recall	Time Consumed(Seconds)
Movie10.mp4	0.666	0.5714	89.45
Movie8.mpeg	0.666	0.666	44.88
Movie15.mp4	0.5	0.5	42.65
Movie33.mp4	0.5	0.44	50.56

Table1: Sample Performance Measurement of Results

The sample output of the experiment:

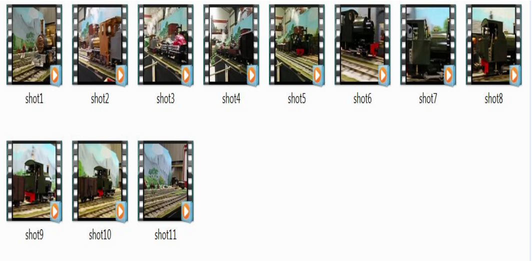
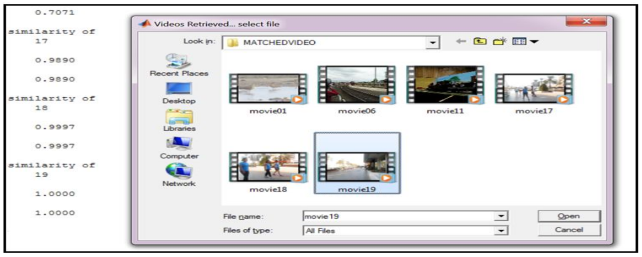


Fig.2. Shots generated for a video



```
      0.7071

similarity of
    17

      0.9890

      0.9890

similarity of
    18

      0.9997

      0.9997

similarity of
    19

      1.0000

      1.0000
```

Videos Retrieved... select file

Look in: MATCHEDVIDEO

File name: movie19

Files of type: All Files

Fig.3. Videos retrieved for a query

V. CONCLUSION AND FUTURE WORK

The application introduces content based video retrieval by making use of visual information present in the video, and then applying latent semantic indexing on it for indexing and retrieval of videos. The experimental results show that the proposed algorithm gives a very good accuracy while retrieving the video from the database. One can also observe that time taken to retrieve the video from the database is less. As the processing of features is computationally expensive the system can be modified to parallel version for efficient performance. Because there are number of visual features are available in the video, the performance of the system can be further improved by considering maximum number of these features. The performance can be further improved by using a efficient key frame extraction technique to extract the key frames.

REFERENCES.

1. Kalpana S. Thakare, Ramchandra Manthalkar, Archana M.Rajurkar, Deepa Desha-pande, 'Video Retrieval using Singular Value Decomposition and Latent Semantic Indexing', International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, pp.1-5, 2012.

2. Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, Stephen Maybank, 'A Survey on Visual Content-Based Video Indexing and Retrieval' , IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 41, No. 6, pp. 797 - 819, 2011.
3. Yong Wei, Suchendra M. Bhandarkar and Kang Li, 'semantics-based video indexing using a stochastic modeling approach', IEEE International Conference on Image Processing, Vol.4, pp. 313-316, 2007.
4. F. Idris and S . Panchanathan, 'Image/Video Indexing in the Compressed Domain', Canadian Conference on Electrical and Computer Engineering, Vol.2, pp. 903-906, 1996.
5. Xiaoming Nan, Zhicheng Zhao, Anni Cai, Xiaohui Xie, 'A Novel Framework for Semantic-based Video Retrieval', IEEE International Conference on Intelligent Computing and Intelligent Systems, Vol.4, pp. 415-419, 2009.
6. Jianrong Cao , Jinan, 'Semantic-Based Video Data Modeling for Scenery Documentary', Fourth International Conference on Natural Computation, Vol.2, pp.158-161, 2008.
7. YusufAytar, Bilal Orhan and Mubarak Shah,' improving semantic concept detection and retrieval using contextual estimates', IEEE International Conference on Multimedia and Expo, pp. 536-539, 2007.
8. Konstantin Biatov, Joachim Koehler, Daniel Schneider, 'Audio Clips Content Comparison Using Latent Semantic Indexing', IEEE International Conference on Semantic Computing, pp. 509 - 512, 2009.
9. Na Zhao, Shu-Ching Chen, Mei-Ling Shyu, Stuart H. Rubin, 'An Integrated and Interactive Video Retrieval Framework with Hierarchical Learning Models and Semantic Clustering Strategy', IEEE International Conference on Information Reuse and Integration, pp. 438-443, 2006.
10. Ali Kazemi, Ali Nasirzadeh Azizkandi, Ali Rostampour, Hassan Haghighi, Pooyan Jamshidi and Fereidoon Shams,' Measuring the Conceptual Coupling of Services Using Latent Semantic Indexing', IEEE International Conference on Services Computing (SCC), pp. 504-511, 2011.
11. Dilek Küçük a, Adnan Yazıcı ,'A semi-automatic text-based semantic video annotation system for Turkish facilitating multilingual retrieval', Expert Systems with Applications: An International Journal, Vol.40, Issue 9, pp. 3398-3411, 2013.
12. Yong Wu, Xuejun Liu, Guangfa Lin, 'The locatable video: Acquisition, Segmentation, Retrieval',19th International Conference on Geoinformatics, pp. 229-233, 2011.
13. Gitto George Thampi, D. Abraham Chandy, ' Content-Based Video Copy Detection Using Discrete Wavelet Transform', IEEE Conference on Information & Communication Technologies (ICT), pp. 998-1002, 2013.

BIOGRAPHY

Rashmi M. is a PG scholar in Computer Science and Engineering Department, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Nitte, Udupi Dist. Karnataka, India. She received Bachelor of Engineering degree in Computer Science and Engineering in the year 2004 from Kurunji Venkatramana Gowda College of Engineering, Sullia, D.K., Karnataka, India. She worked as a faculty in the department of computer science and engineering, Vivekananda College of Engineering and Technology, Puttur, DK. from July 2005-June 2008. From August 2008 onwards she is a faculty in CSE Department, Srinivas Institute of Technology, Mangalore. Her research interests are Image processing, Algorithms, Distributed Computing and Database Management Systems.

Roshan Fernandes is working as an Assistant Professor in Computer Science and Engineering Department, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Nitte, Udupi Dist. Karnataka, India. He received Bachelor of Engineering degree in Computer Science and Engineering in the year 2001 from MIT, Manipal, Udupi dist., Karnataka, India. He pursued Master of Technology in Computer Engineering at SJCE, Mysore, with first rank and a gold medal awardee. He Worked as an Assistant Lecturer in the Department of Information Technology at NITK, Surathkal, from August 2001 to March 2002. Then worked as a Lecturer in the Department of Computer Science and Engineering at P. A. College of Engineering, Mangalore from March 2002 to February 2004. He served as a Project Trainee in Motorola India Pvt. Ltd., Bangalore, from July 2006 – July 2007. He published number of papers in various national and international conferences and journals. He is expertised in the following subjects: Algorithms, Java and J2EE, Web 2.0 Programming (AJAX, Flex, PHP, JavaScript), Data Structures, RDBMS and Advanced DBMS, System Software, Computer Graphics using OpenGL, Microprocessors (8086), Embedded Systems.