

Fused Image Based Video Search Engine

Sheena Gupta[†] and Dr. R. K. Kulkarni^{††},
aries.sheen@gmail.com rk1_2002@yahoo.com
 VESIT, Mumbai University, India

Abstract

A video search engine is a web-based search engine which crawls the web to find videos stored on digital devices. Video search engine has given very little attention until the popularity of digital capturing devices and communication tools. Here a novel method is proposed for video search engine using fused image technique. To begin with, the system splits the video sequence into key-frames and then from the key-frames fused image is created of that sample video. The color and edge histogram features are extracted from that fused image of sample video. The experiment is performed on a database of 40 video samples spread across 4 categories. On the basis of query-clip the top 5 videos are retrieved from the database using Euclidean distance as a similarity measures. Performance evaluation is done with the help of average precision and average recall. The proposed method is also compared with existing video retrieval techniques on the basis of performance parameter.

Keywords:

Video retrieval, key-frames, fused-image, Color-moments, Edge histogram

1. Introduction

With the rapid advancement of network technologies and storage media leads to increase in the amount of video data. As video data possess lot of information for those using multimedia and various applications like entertainment, education, digital libraries, publications and broadcasting [1]. So an effective video retrieval system is required which search the relevant video from a large video database.

The main goal of search engine is to meet user's satisfaction and achieve high accuracy on top-ranked search documents with minimum time interval [4]. There are efficient search engine available today for text document and images such as Google, Bing etc. In the field of video, the retrieval is mainly based on similarity measurement. VisualSeek, Photobook, Blobworld, Virage along with VideoQ are few examples for video search engine [3].

Video search system uses different techniques for indexing and retrieving of videos from the video database. In Key-frame based retrieval first the key-frames are extracted and then the visual features are extracted from that key-frames [4]. Visual features of key-frames yield better result but huge computation is required and

computational cost is also high [2]. In pattern-based indexing and retrieval the shot detection is performed and then due to the relation or co-relation of these sequential shots videos can be retrieved based on temporal continuity. But the computational complexity increases rapidly as the amounts of shot increases [7]. In transformed visual content based retrieval the video is represented in transformed domain due to which the feature vector size is reduced [6]. Hybrid wavelet transform generated from orthogonal transform gives higher energy compaction and give much lesser feature extraction complexity [8]. In motion-based video retrieval the optical flow field of key-frames is computed. But it is still a challenge to find reliable motion descriptor of video [8].

The drawbacks like computational complexity and find a reliable motion descriptor of existing video retrieval techniques are overcome in a new proposed technique. In the proposed technique the retrieval system comprises of two phases. First phase is registration phase where each video is divided into key frames and then the fused image is created from key-frames and then the feature vector is built on the basis of fused image. In the second, execution phase the query video clip is processed for generation of feature vector and then this queried feature vector is compared with feature vector database using Euclidean distance as a similarity measure. Performance of the retrieval system is evaluated based on the Precision and Recall. The proposed method gives high retrieval efficiency and accuracy on top-ranked videos. The accuracy of proposed system can be increased further by extracting more visual features and by increasing the number of key-frames in fused image. But it increases the run-time and complexity of proposed system. The remaining part of the paper is organized as follows: The section 2, consists of proposed technique. The section 3, consists of modules of proposed fused image based video search engine. Section 4 illustrate the result. The conclusion is given in Section 5.

2. Proposed Fused Image Based Video Search Engine

Block diagram of proposed fused image based video search engine is as shown in Fig.1. This system consists of two phases as registration phase and query execution phase

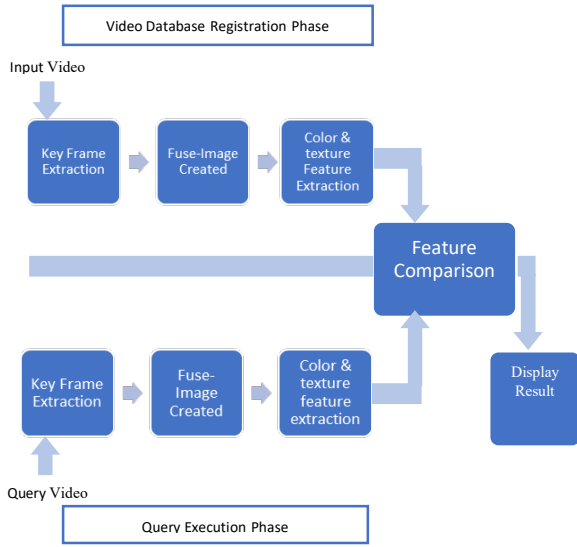


Figure.1. Block diagram of Proposed Fused Image based Video Search Engine

2.1 Registration Phase

In this phase a feature vector table is built up from the visual content of each fused image. Fused image of the respective sample video is to be stored and registered in video database. Algorithm for Registration Phase is given below

1. Select a video from stored video database
2. Extract the key-frames of that respective video
3. Fuse all the key-frames as a fused image which represent that respective video
4. Extract color and edge histogram features from that fused image
5. Concatenate all features of a fuse-image and stored as a complete feature vector of that respective video
6. Repeat step 1 to 5 for all videos in the database to get the feature vector table

2.2 Query Execution Phase

Algorithm for query-execution phase for a given query video is stated as below

1. Extract key-frames from a query video
2. Fuse all key-frames as a fused query image which represent that respective video
3. Extract color and edge histogram features from that fused query image
4. Concatenate all features of a fused query image and represent as a complete feature set of that respective query video

5. This feature set is compared with other feature set stored in feature database using Euclidean distance as a similarity measure.
- 6.

3. Modules Of Proposed Fused Image Based Video Search Engine

The framework designed retrieves effective and accurate videos .The overall framework is designed as follows.

3.1 Key-frame Extraction

First the total number of frames are calculated of that input video from the video database. Then our framework converts the video into key-frames by using Eq (1).

$$Kf = (n N / x)^{th} \quad (1)$$

Where Kf denotes the n^{th} key-frame, N denotes the number of frames for given input video, x is the total number of key-frames and n is a variable that vary from 1 to x. In proposed method 8 key-frames are used to represent the input video.

3.2 Fused Image Formation

After extracting all the key-frames they are fused to form a fused image. This fused image is basically combining the relevant information of all key-frames into a single image. This fuse-image is more informative and have high spatial and spectral resolution which is very efficient for fast video browsing.

3.3 Feature Extraction

Our framework uses color moments and edge histogram for extracting features. Extracted fused image of respective video is used for feature extraction.

3.3.1 Color Feature:

One of the color feature widely used by researchers are color moments. Color moments basically describes the color distribution of an image through its moment. First the fused image is converted from RGB to HSV color space. The HSV related fused image is as shown in Fig 4. The three central moments of an image color distribution are Mean, Std. deviation and Skewness. These three moments are considered for representing the feature vector [4]. An image is therefore characterized by 9 moments 3 moments for each 3 color channels. We will define the i^{th} color channel at the j^{th} image pixel as p_{ij} . The three color moments can be defined as

$$\text{MOMENT 1 - Mean} \\ E_i = \sum_{j=1}^N (1/N) p_{ij} \quad (2)$$

MOMENT 2 – Standard Deviation

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^J p_{ij} - E_i^2} \quad (3)$$

MOMENT 3 – Skewness

$$S_i = \sqrt{\frac{1}{N} \sum_{j=1}^J p_{ij} - E_i^3} \quad (4)$$

3.3.2 Edge Histograms:

The global feature composition of an image is most frequently represented by characteristics of the histogram. The representation of the content of the image contains edge in the image as an important feature [3]. In our system, initially the sample video sequence is converted into fused image and then that RGB fused image is converted into YCbCr color space. Y', Cb and Cr represents the luma components, blue difference chroma components and red-difference chroma components. The luma is differentiated from luminance which means light intensity by non-linearly gamma, it is denoted by using a prime (') on Y. The YCbCr created fused image is shown in Fig. 5. The YCbCr color space is extensively employed by the video and digital photography systems [4].

The image gradients are represented as follows Gx and Gy where,

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * A \quad \text{and} \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{bmatrix} * A \quad (5)$$

Where * denotes the 2-dimensional convolution operation and A denotes the source image. Then the edge angle is found using

$$\Theta = \arctan (G_x / G_y) \quad (6)$$

Where Θ denotes the edge angle. Edge histogram is generated where each bin value represents the percentage of a certain edge direction in an image.

4. Experimental Results And Evaluation

4.1 Experimental Results

Fused image based video search engine has been implemented using MATLAB R2017a and the performance of the proposed system is analyzed using evaluation metrics including precision and recall. Fig 3(a) and 3(b) shows one input video from the video database and the fused image creation from the key-frames of that sample video. We have performed experiments on a dataset of 40 videos. The collected videos contain the following categories such as plane, ocean, street and mountains. We

report here some results obtained on a part of a video sequence utilized for retrieval. The results of RGB fused image to HSV color space conversion for color histogram is shown in Fig. 4 and RGB fused image to YCbCr color space conversion output for edge histogram is shown in Fig.5. The Fig. 6 shows top 5 video retrieved from the video database of query clip shown in Fig. 3(a). For the entire query clips categories Q1, Q2, Q3, Q4, the average precision, average recall, average retrieval efficiency and average retrieval time have been determined and the values are tabulated in Table 1. Fig. 2 shows the graphical representation of average precision and average recall with query video categories. In Table 2. the average precision, recall, retrieval efficiency and retrieval time is calculated for existing different video retrieval techniques.

It is clear from Table 2. that video retrieval rate of proposed method is higher than existing techniques. It is also clear from Table 2. that retrieval time of proposed method is little bit higher than existing methods but when the accuracy is concerned this drawback can be neglected.

From Table 1. it is clear that average precision and recall percentage of proposed method is higher and retrieval efficiency is also higher in mountain (Q4) category. In Q4 category each frame are highly correlated and as the similarity between frames increases so retrieval of query image is more.

4.2 Evaluation Criteria

The performance of proposed system is evaluated using precision and recall. The precision and recall measurement are the more frequently using measurements to analyze the performance of a video retrieval system which can be defined as

The precision is the ability of the search engine to retrieve the best-ranked video that are mostly relevant

$$\text{Precision} = \frac{\text{No. of relevant videos retrieved}}{\text{Total No. of videos retrieved}} \quad (7)$$

The recall is the ability of the search engine to find all of the relevant videos in the database.

$$\text{Recall} = \frac{\text{No. of relevant videos retrieved}}{\text{Total No. of relevant videos in database}} \quad (8)$$

Table 1. Average Precision, Recall, Retrieval Efficiency and Retrieval Time For A Given Set Of Query Categories For Proposed Method

Query Categories	Avg. Precision (%)	Avg. Recall (%)	Avg. Retrieval Efficiency (%)	Avg. Retrieval Time (Sec)
Q1	36	18	90	58.9
Q2	48	24	80	56.51
Q3	58	29	90	58.9
Q4	78	39	100	52.11

Table 2. Average Precision, Recall, Retrieval Efficiency and Retrieval Time For Different Video Retrieval Methods And Proposed Method. Method 1 Is Key-Frame Based Retrieval, Method 2 Is Orthogonal Transform Based Feature Extraction, Method 3 Is Hybrid Transform Based Feature and Method 4 Is Motion Based Video Indexing And Retrieval.

Query Category Q4	Avg. Precision (%)	Avg. Recall (%)	Avg. Retrieval Efficiency (%)	Avg. Retrieval Time (Sec)
Method 1	60	30	80	33.04
Method 2	38	19	50	24.66
Method 3	32	16	60	27.51
Method 4	70	35	40	29.50
Proposed Method	78	39	100	52.11

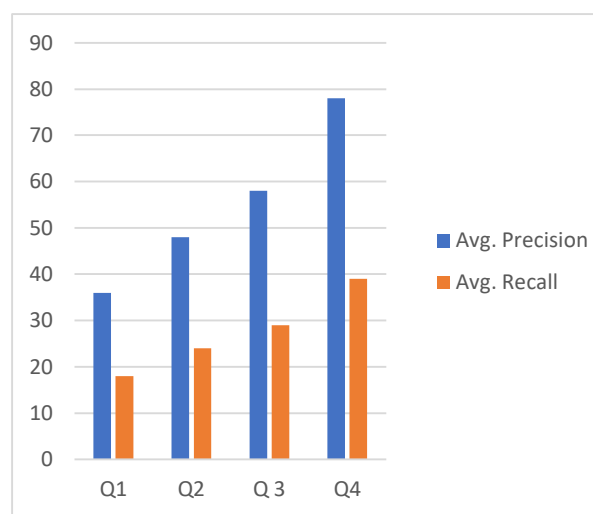
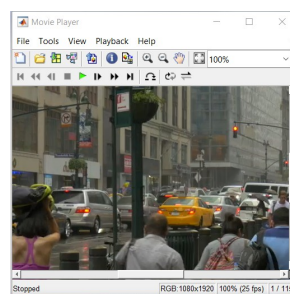


Figure 2 Shows the relation of average precision and average recall with different video categories for proposed method



(a)



(b)

Figure 3. Fused Image Extraction from Key-Frames (a) Query Video (b) Fused Image for the Video Clip in (a)



Fig 4. RGB to HSV color space conversion of Fused image to extract the color feature



Figure 5. RGB to YCbCr color space conversion of Fused image to extract the edge histogram feature

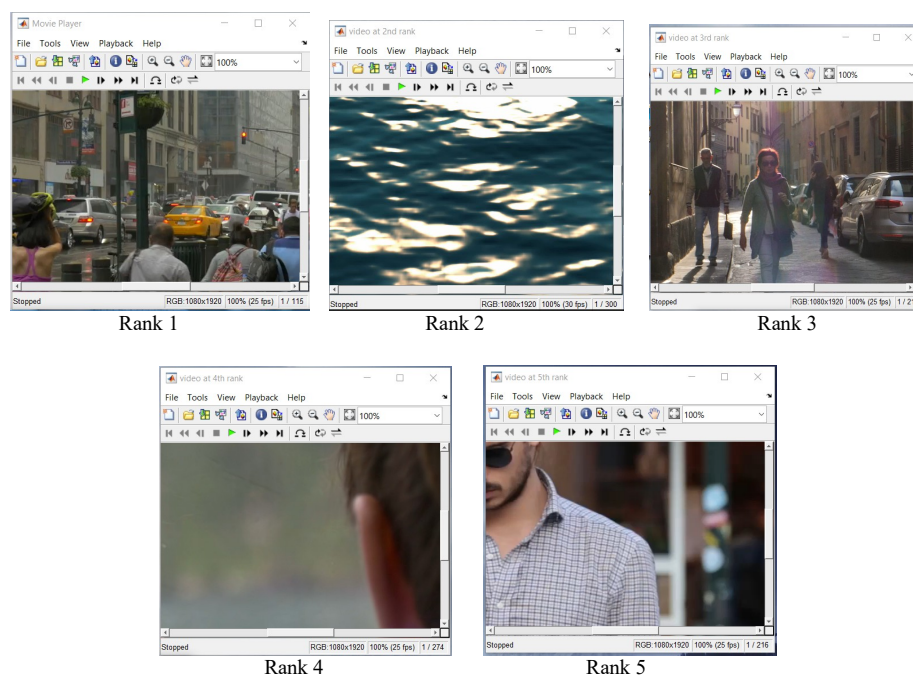


Figure 6. Shows the top 5 ranked videos of the input video in Fig. 3 (a)

5 CONCLUSION

The proposed fused image based video search engine has been tested and the average precision, recall, retrieval efficiency and retrieval time has been calculated for different categories of query video. It is clear from Table 1. that average retrieval time is constant irrespective of query categories and the percentage of precision and retrieval efficiency is more for mountain (Q4) query category. The proposed method is also compared with existing video retrieval techniques and the results are potted in Table 2. The proposed technique is predominately giving the best result in terms of precision and retrieval efficiency as compared to all existing techniques.

REFERENCES

- [1] Aasif Ansari and Muzammil H Mohammed "Content based Video Retrieval System- Methods, Techniques, Trends and Challenges" International Journal of Computer Applications, Vol. 112- No.7, Feb 2015.
- [2] B V Patel and B B Meshram, "Content Based Video Retrieval Systems", International Journal of UbiComp(IJU), vol. 3, No.2, April 2012.
- [3] T.N Shanmugam and Priya Rajendran, "An enhanced content-based video retrieval system based on query clip", International Journal of Research and Reviews in Applied Sciences, ISSN:2076-734X, EISSN:2076-7366, Volume 1, Issue 3:2009
- [4] P.Geetha and Vasumathi Narayanan, "An Effective Video Search Re-ranking for Content Based Video Retrieval", 978-1-4673-0131, 2011 IEEE.

- [5] H B Kekre, Sudeep D. Thepade and Saurabh Gupta, "Content Based Video Retrieval in Transformed Domain using Fractional Coefficients", International Journal of Image Processing (IJIP), Vol. 7 : Issue3: 2013.
- [6] Sudeep D. Thepade and Nalini Yadav, "Novel Efficient Content Based Video Retrieval Method using Cosine-Haar Hybrid Wavelet Transform with Energy Compaction", International Conference on Computing Communication Control and Automation, 978-1-4799-6892, 2015 IEEE.
- [7] Parag Kulkarni, Bhagyashri Patil and Bela Joglekar, "An Effective Content Based Video Analysis and Retrieval Using Pattern Indexing Techniques ", International Conference on Industrial Instrumentation and Control (ICIC), 978-1-4799-7165, 2015 IEEE
- [8] Ying Chen and Ou Wu, "Motion Video Retrieval Based on Optical Flow and Haar Wavelet", International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 978-1-61284-181, 2011 IEEE .
- [9] R.K Kulkarni and Sheena Gupta, " Video Search Performance Analysis: A Review Paper", International Journal Of Innovations and Advancement in Computer Science (IJACS), ISSN 2347-8616, Vol.6, Issue 8, August 2017
- [10] S. Padmakala, G. S. Anandha Mala and M. Shalini, "An Effective Content Based Video Retrieval Utilizing Texture, Color and Optimal Key Frame Features", International Conference on Image Information Processing (ICIIP 2011), 978-1-61284-861, 2011 IEEE.
- [11] Santhosh Kumar K and Mallikarjuna Lingam.K, "Content Based Video Retrieval Using Low and High Level Semantic Gap" International Journal for Research in emerging science and technology, Volume 2, Issue-7, July-2015
- [12] Zhang, Hong Jiang, et al. "An integrated system for content-based video retrieval and browsing." Pattern recognition, Vol. 30, No.4, 1997
- [13] Shikui Wei, Yao Zhao, Member, Zhenfeng Zhu, and Nan Liu, "Multimodal Fusion for Video Search Re-ranking", IEEE transaction on knowledge and data engineering ,Vol.20, No .5, August 2010
- [14] M.Petkovic "Content-based video retrieval", Centre for Telematics and Information Technology, University of Twente
- [15] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, and 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, November 2011