

# A Survey on Video Content Identification Tool

Juhi S. Patil, Rachana A. Satao  
Department of Computer Engineering,  
STES's Smt.Kashibai Navale College of Engineering,  
Pune – 411041, India

## ABSTRACT

Now-a-days, the search engines available are text-based search engines. Thus using text-based search engines one can efficiently search for the desired video. Many times it happens like video name that has fired as a query, contains irrelevant data. Even recently it is found that some illegal information is communicated via video by embedding it into a longer video. And also it is found that broadcast channels and IPTV services many times use same digital videos. An efficient method of consuming, storing and retrieving such vast amounts of videos is essential. This has led to the emergence of video copy detection as an active area of research. In this survey, a study of different MPEG standard, challenges in video copy detection, brief idea about video fingerprint and its application are discussed.

## Keywords

Video copy detection, video fingerprint, MPEG standards

## 1. INTRODUCTION

The upload and download of videos has been increased now a day. A survey found that in 2010 more than 2 billion videos were watched on YouTube and numerous videos are uploading daily at a rate of 24 hour of content each minute. In half a decade over video identification is studied primarily for the design of complete retrieval systems, as well as fingerprint extraction and information categorization schemes, for the retrieval of short near-duplicate clips exhibiting a whole or vital temporal overlap. [1] Brisk development in the area of multimedia technology, lead to easy access and store of video data of massive volume. It is well reflected in the availability of such video data on different sites like video blogs and Web-TV. Sharing and distribution of video over all such sites resulted into exponential growth of data volume. Moreover, the technology has made possible duplication and editing of video data that may lead to breach of digital rights. Hence, copyright protection becomes a fundamental issue and the huge volume of video data makes the task more difficult.

## 2. MPEG

[2]Many people would like to use audiovisual information for various purposes; this information becomes available from many sources around the world. This scenario has raised the issue of efficient retrieval of multimedia material that's interesting to the user. For example, finding information by rich-spoken queries, hand-drawn images and humming improves the user-friendliness of computer systems and lastly addresses what most people have been expecting from computers.

### 2.1 MPEG Standards

The Moving Picture Coding Experts Group (MPEG) is an active group of the Geneva-based ISO/IEC standards organization, (International Standards Organization/International Electro-technical Committee) accountable for the development of international standards for processing, compression, decompression, and coded representation of moving pictures, audio, and a combination

of the two. MPEG has developed standards like MPEG-1,-2,-3. These standards have won Emmy Award.

- MPEG-1: Used for the storage and retrieval of moving pictures and audio on storage media [2].
- MPEG-2: For digital television, it's the timely response for the satellite broadcasting and cable television industries in their transition from analog to digital formats [2].
- MPEG-4: Codes content as objects and enables those objects to be manipulated individually or collectively on an audiovisual scene[2].

The main function of MPEG-1, -2, and -4 standards is to make content available where as MPEG-7 let one to find the content one is searching for.

MPEG-7: MPEG-7 is a standard for describing features of multimedia content. The world's richest set of audio-visual descriptions is provided by MPEG-7.

### 2.2 The Key Role of MPEG-7

MPEG-7 is officially named as "Multimedia Content Description Interface". It describes multimedia content so that users can search, browse, and retrieve that content more efficiently and effectively than they could using today's mainly text-based search engines. Multimedia features are described by MPEG-7.

[2]Typical MPEG-7 includes:

- Video: Allow mobile phone access to video clips of goals scored in a soccer game, or automatically search and retrieve any unusual movements from surveillance videos.
- Audio: One wants to search for songs by humming or whistling a tune or, using an excerpt of Pavarotti's voice, get a list of Pavarotti's records and video clips in which Pavarotti sings or simply makes an appearance.
- Image: describe objects, containing color patches or textures, and get examples from which one select item to compose your image. Or check if your company logo was advertised on a TV channel as contracted.
- Graphics: Sketch a few lines on a screen and get a set of images containing similar graphics, logos, and ideograms.

### 2.3 MPEG-7 Application Domains

[2]All domains making use of multimedia will benefit from MPEG-7 including,

- 1) Digital libraries, Education (image catalogue, musical dictionary, Bio-medical imaging catalogue)
- 2) Multimedia editing (personalized electronic news service, media authoring)
- 3) Cultural services (history museums, art galleries, etc.),
- 4) Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems)

- 5) Broadcast media selection (radio and TV channel)
- 6) Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face),
- 7) E-Commerce (personalized advertising, on-line catalogues, directories of e-shops)
- 8) Surveillance (traffic control, surface transportation, non-destructive testing in hostile environments, etc.),
- 9) Investigation services (human characteristics recognition, forensics),
- 10) Home Entertainment (systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game, karaoke)
- 11) Social (e.g. dating services),

There are mainly two terminologies used in video processing

**CBVR:** In case of content-based video retrieval (CBVR) system, aim is to retrieve similar videos in the same category

**CBVD:** Content-Based Video Copy detection (CBVD) system it is to be detected whether a query sequence is a copied version of reference sequence or not. Instead of being an identical or near-replicated video sequence, a copy may be a transformed video sequence [3].

A content based video copy detection system consists of two major modules namely, fingerprint generation and sequence matching technique as shown in Fig. 1. [1]

Fingerprint can be used as a unique feature of any multimedia object. Two multimedia objects can't have same fingerprint. Basically it is used to conclude does two video have same contents or not even under quality-preserving distortions like resizing, frame rate change, lossy compression. Sequence matching technique detects whether a query sequence is copied version of referenced one or not based on their fingerprints.

### 3. APPLICATIONS

The video signature has a large number of applications like

- 1) Rights management and monetization: video fingerprint helps to identify who is real owner.
- 2) Distribution management: Using video signature one can stop the leakage/distribution of some sensitive data over a network or via email by coding video identification tool at firewall.
- 3) Usage monitoring: It helps to maintain statistics For example; an advertising agent can check that does its material have distributed as expected using video fingerprint.
- 4) Video content-based linking: Video content in the web page can be linked to some other video.

- 5) Database management and deduplication: Mainly video content identification application is useful for large volume of data creators and owners like studios, personal libraries.

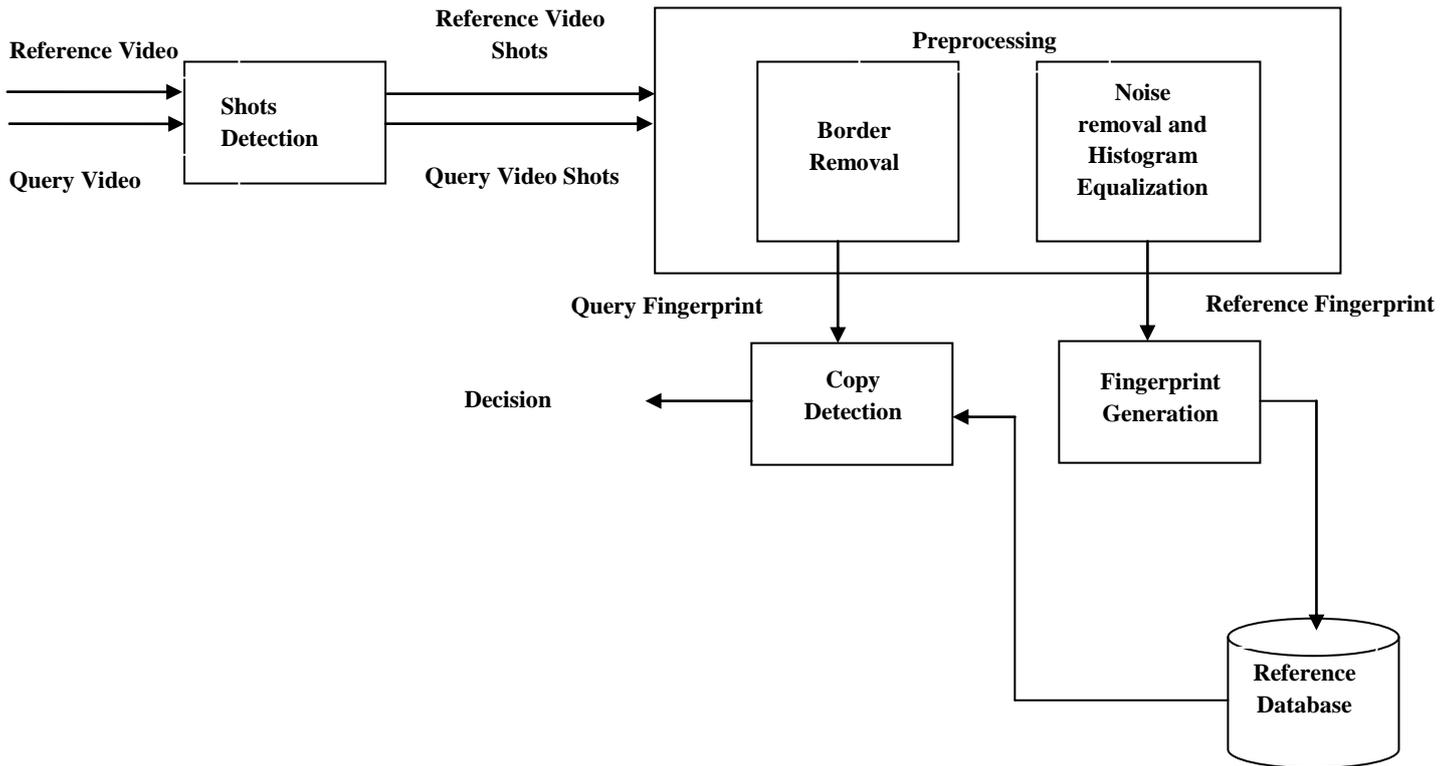
### 4. COPYRIGHT PROTECTION

[1]To address the issue of copyright protection basically there are two approaches. First one is Water-Marking and second is Content-based Copy Detection (CBCD).

In water-marking, non-visible information called watermark is set into the content. This watermark is the sign of ownership. Main disadvantage of watermarking is that it's not applicable for video sequences already in distribution without any watermark. Whereas in content-based method, no additional information i.e. watermark is embedded. It is said that "Video itself is the watermark" [4]. So in CBCD unique signatures are extracted from the contents of the video. Similarly questioned videos are also extracted to signature extraction and that signature is compared with those of the original media stored in the database. If both signatures match, then it is concluded that two videos have same content. Videos are prone to attacks like photometric attack i.e. changing brightness /contrast, contamination by noise, post-production attack i.e. changing display format, logo insertion etc. To handle such attacks, attempts have been made to design robust signatures or different post-processing techniques are adopted [1].

### 5. FINGERPRINT

Fingerprint is nothing but a unique identity of a video. It must satisfy properties like Uniqueness, Robustness to editing operations, Independence, Fast matching, Fast extraction, Compactness, Nonalteration of the content, Self-containment of the signatures, Coding independence, Partial matching, Accurate temporal localization of duplicated and embedded content. Video Fingerprint must satisfy above properties, detailed outlined in [5]. There are two types of video fingerprints global and local. Global ones are obtained from whole video or from a subset of sequences whereas local ones are obtained from each frame in the video. Global fingerprint is obtained from local one. To generate the fingerprints, researchers have been tried a wide variety of frame level features. Colour histogram [6,7] is most commonly used. But, it lacks in terms of discriminability as the spatial distribution of colour is not retained in the histogram. Binary signature based on colour histogram has been deployed in [8]. Luminance based descriptors [9], dominant colour [10], gradient based features [11] and texture feature [12] are also attempted. A graph based technique has been explained in [13] to define a spatial correlation descriptor where edges represent the content proximity of the regions in the frame and a node in the graph denotes a region in the frame.



**Fig 1: Block Diagram For CBCD**

In [14, 15, 16, and 17], global fingerprints are calculated using ordinal measures. In [15], a single video frame is divided into  $2 \times 2$  sub-images and they are ranked based on their average intensity. Then  $2 \times 2$  rank matrix is calculated. This rank matrix is nothing but the fingerprint of the frame. For matching a query and reference video sequence, spatial and temporal measures are calculated using rank matrix. In [16], the sub-images are calculated in the same way as that of [15], but in [16] rank matrix is calculated for whole video instead of single frame. Ranking is done for Corresponding sub-images of all the frames in the sequence. Then  $n \times m$  matrix is obtained where,  $n (= 4)$  is the number of sub-images in each frame and  $m$  is the number of frames in the sequence. Lastly for matching query and reference video sequence, a spatio-temporal measure where the rank matrix over the video sequence is formed based on the number of SURF interest points in the sub-images is proposed.

## 6. SEQUENCE MATCHING

On the basis of extracted signature, the query video sequence and reference sequences in the database are to be matched. Researchers have tried out various types of matching techniques, broadly classified as (1) Dense matching technique and (2) Sparse matching technique.

In Dense scheme all the frames are considered for comparison. In a sparse technique only key frames are considered. Thus a sparse technique is faster but a dense one is more robust.

In sparse matching technique, selection of key frame is an important task. In [18] sequence matching technique based on a set of key frames (or sub-sampled frames). Similar approaches have also been used in [19, 20].

*Maani et al.* [17], in their technique, corresponding to each keyframe in the query sequence have selected a set of

matched keyframes from the database. From the matched set of keyframes, it is tried to find out continuous sub-sequence. If the length of such sub-sequence exceeds a threshold then query sequence is considered as a copy. Selection of the threshold here is an important issue.

## 7. CHALLENGES IN VIDEO COPY DETECTION

A video clip can be off different type depending on its type of use like RealVideo™ for the Internet and MPEG1 for an intranet. Now a days, tapes are used to store the source material and its is digitized and encoded by digitizer/encoder devices. Several distortions are raised by the process of digitizing and encoding like change in contrast, changes in brightness, shifts in hue, changes in saturation and spatial shifts in the picture.

In addition to the digitizer artifacts, lossy encoding processes introduce artifacts like the blocking effects in MPEG. Figure 1 show frames obtained from a set of video clips. The clips are created from source material on VHS tape. The frames are approximately the same frames from each of these clips.

[21]The figure 2 shows six corresponding frames taken from different sources, namely, MPEG1, an AVI, a RealVideo 28k (for a 28k modem), a RealVideo 512k (for a 512k Face Image RealVideo-28k (160£120), Face Image RealVideo-512k (160£120), Two People Image MPEG1 176£112, Two People Image AVI 160£120 connection), a MPEG1 and an AVI sequence, respectively. The resolution of all the frames is 160£120, except the MPEG-1 frames, which are 176£112. Copyright Detection is main challenge in multimedia management. Editing operations are very common in case of video data. A robust system needs to be developed against common editing operations.



**Fig. 2: Images taken from different sources. Left to Right:  
Face Image 176×112 MPEG1, Face Image 160×120 AVI**

Management and storage of such a large volume of multimedia data is also a challenging area of research. Even fast retrieval of data from such a huge amount of data is also brain hammering task.

## 8. CONCLUSION

This survey explored detailed idea about MPEG and its various standards. The survey has addressed detailed video processing techniques. Moreover, copyright protection for video, challenges in video copy detection is discussed. This paper gives brief idea about various works done by researchers in field of video processing and video protection. The survey suggests directions for further research to develop the robust system for video content identification. The survey can be extended to the concept of video linking. In video linking, video content identification is done and video classification is done according to contents. This will make fast retrieval of video data.

## 9. REFERENCES

- [1] Content Based Video Copy Detection: Issues and Practices, Sanjoy Kumar Saha CSE Department, Jadavpur University Kolkata, India
- [2] International Organization for Standardization Organization Internationale De Normalisation Iso/Iec Jtc1/Sc29/Wg11 Coding Of Moving Pictures And Audio.
- [3] J. Law-To, , L. Chen, A. Joly, I. Laptev, O. Buisson, V. Gouet-Brunet, N. Boujemaa, and F. Stentiford. Video copy detection: A comparative study. In Proc. CIVR, 2007.
- [4] A. Hampapur and R. Bolle. Comparison of sequence matching techniques for video copy detection. In Proc. Intl. Conf. on Multimedia and Expo, pages 188 – 192, 2001.
- [5] S. Lee and C. D. Yoo. Video fingerprinting based on centroids of gradient orientations. In Proc. ICASSP, pages 401 – 404, 2006.
- [6] S-C. S. Cheung and A. Zakhor. Efficient video similarity measurement with video signature. IEEE Trans. CSVT, 13 No. 1:59–74, 2003.
- [7] A. M. Ferman, A. M. Tekalp, and R. Mehrotra. Robust color histogram descriptors for video segment retrieval and identification. IEEE Trans. on IP, 11(5):497 – 508, 2002.
- [8] Y. Li, J. S. Jin, and X. Zhou. Video matching using binary signature. In Proc. Intl.Symp. on Intelligent Signal Processing and Comm. Systems, pages 317 – 320, 2005.
- [9] R. Radhakrishnan and C. Bauer. Robust video fingerprints based on subspace embed-ding. In Proc. ICASSP, pages 2245–2248, 2008.
- [10] A. Hampapur and R. Bolle. Comparison of sequence matching techniques for video copy detection. In Proc. Intl. Conf. on Multimedia and Expo, pages 188 – 192, 2001.
- [11] S. Lee and C. D. Yoo. Video fingerprinting based on centroids of gradient orientations. In Proc. ICASSP, pages 401 – 404, 2006.
- [12] L. Chen and T. S. Chua. A match and tiling approach to content-based video retrieval. In Proc. Intl. Conf. on Multimedia and Expo, 2001.
- [13] M. C. Yeh and K.-Y. Cheng. A compact, effective descriptor for video copy detection. In Proc. of ACM Multimedia, 2009.
- [14] R. Mohan. Video sequence matching. In Proc. ICASSP, pages 3697 – 3700, 1998.
- [15] C. Kim and B. Vasudev. Spatiotemporal sequence matching for efficient video copy detection. IEEE Trans. on CSVT, 15, No. 1:127 – 132, 2005.
- [16] L. Chen and F. W. M. Stentiford. Video sequence matching based on temporal ordinal measurement. Pattern Recognition Letters, 29:1824–1831, 2008.
- [17] R. C. Harvey and M. Heefeda. Spatio-temporal video copy detection. In Proc. Multi-media System conference, pages 35 – 46, 2012.
- [18] A. K. Jain, A. Vailaya, and W. Xiong. Query by clip. Multimedia System Journal, 7,No. 5:369 – 384, 1999.
- [19] S-C. S. Cheung and A. Zakhor. Efficient video similarity measurement with video signature. IEEE Trans. CSVT, 13 No. 1:59–74, 2003.
- [20] K. W. Sze, K. M. Lam, and G. Qiu. A new key frame representation for video segment retrieval. IEEE Trans. CSVT, 15 No. 9:1148 – 1155, 2005.
- [21] A. Hampapur, K. Hyun, and R. Bolle, Comparison of sequence matching techniques for video copy detection, nProc. Conf. Storage Retrieval Media Databases, 2002, pp. 194–201.