Video Content Identification Using Video Signature: Survey

Tejashri Shinkar[1], D. B. Hanchate[2]

[1] Student of ME-II, Department of Computer Engineering,
VPCOE, Baramati, Savitribai Phule Pune University, Maharashtra, India
[2]Department of Computer Engineering,
VPCOE, Baramati, Savitribai Phule Pune University Maharashtra, India

---

Abstract – *The amounts of videos generated by people are increasing day by day. At each minute billions of videos are uploaded over the network. The videos are different types duplicate, edited, pirated etc. There are very few tools are available to find out near duplicate videos. Existing methods are not finding out video segment from larger unrelated video. These methods generate a signature of three types like spatial, temporal and spatio-temporal. This paper proposed a method which extract frame level features through this create a temporal signature. Using this signature it accurately detects a video segment which is embedded into larger unrelated video.*

Key Words: **Content identification, Video Signature, Content localization, Video Identification, CBCD technology, BCS, LBP, Spatio-temporal features**

1. INTRODUCTION

In the mass media industries; the data volume is growing enormously. There is also advancement in the telecommunication and in the internet data transfer. Also infrastructure of multi-platform content delivery is advanced which accelerate the transmitting speed and increased volumes. Due to these conditions, video contents are generally managed using metadata like keywords, thumbnail images, preview videos, etc. Generally this provides significant volume of metadata search results through the increase in the data volume and the circulation of video contents.

Due to this it is difficult to identify video content or to search a specific scene manually by visual inspection. To solve this problem, paper proposed a technology that identifies video content automatically and efficiently by managing it as metadata. The proposed system focuses on the first to design tools for video fingerprinting that will provide high robustness over common editing operations like cropping, labelling, morphing etc. Also detect the particular video content that will be embedded in longer unrelated video.

Video identification technology analyses each frame of the video content and extracts a descriptor which is unique. This unique descriptor is called video signature which is used to identify identical video scenes.

High precision and high speed video identification is achieved through the video signature without embedding ID information in the content.

The Video Signature Tools, which standardizes an interoperable descriptor for video identification. This system performs in three steps: 1) Video Signature Extraction 2) Video Signature Compression and 3) Video signature matching.

2. PREVIOUS WORK

Literature survey includes different design issues for creating video fingerprint/video signature. It is also focus on different strategies for localization and identification of a video embedded in unrelated video content. The Video signature can be used for near duplicate clip detection. Basically the video signatures are based on the frame level features such as keypoint based, block based or global.

The problem of identification of video content embedded in longer video content is studied by Hampapur et. al. [2]. In this paper, ordinal signature which is based on blocks achieves highest performance. To calculate signature first frame is divided into blocks. The comparative study between different types of signature done by Hampapur et. al[2]. In his work he considered three types of signatures as follows:

1. Ordinal Signature:

For signature calculation first divide each frame into blocks. Then calculate mean intensity for each block and sort them into ascending format. Rank vector will be calculated that is used as the feature of the frame. The procedure will be as mentioned in the paper [8][9]. Best performance is reported for shorter queries. This paper did not report the matching segment localization accuracy.

2. Motion Signature:

For each block in frame , select image patch at block center. Find the SAPD (Sum of absolute Pixel Differences) at each point in search neighborhood frame. To detect match location minimum SAPD is considered. For best

match highest correlation factor is considered.

3. Color Histogram Signature:

In this [10] technique author uses the YUV histogram as the signature. The distance measure between two frames is histogram intersection. Video signature is represented as a sequence of YUV histogram of each frame. The matching factor will be calculated by NHI (Normalized Histogram Intersection [10].

Law-To et al. [3] was performed a comparative study for video copy detection techniques. Different state of the art techniques are described using various kinds of descriptors and voting functions. Techniques compared are:

1. Global Descriptor:

In this technique three types of signature measurement like Temporal, Ordinal and Temporal-ordinal. In Temporal activity defined depending on intensity of each pixel. A signature is computed around maxima of temporal activity and spectral analysis is performed through classical FFT. In Ordinal measurement signature is computed by dividing the image into number of blocks and sorting will be done using their average grey level. Signature uses the rank of each block and distance will be calculated for checking the similarity between two videos.

In Temporal Ordinal measurement instead of using rank of regions in the image use the rank of regions along the time proposed by L. Chen and F. Stentiford [11].

2. Local Descriptor:

The spatial, temporal and spatio-temporal information is considered in local descriptor. New version of Harris Interest point detector described by [12] A. Joly st. al. To increase compression only key-frames are considered for feature extraction.

Finding Video copies system called video Copy Tracking is developed. Where Harris points of interest are extracted and signal description is as described in [12]. In temporal information for CBCD two labels have been selected:

a. Label Background: motionless and persistent points along frames.
b. Label Motion: moving and persistent points.

In temporal information to build trajectory motion properties are adapted from label behaviour. Final signature for each trajectory is composed of 20 dimensional vector, trajectory properties and label behaviour.

Space Time Interest Points (STIP) technique was developed to detect spatio-temporal events. Here interest points are used for CBCD. So Values of spatio-temporal Gaussian derivatives are normalized by the spatial detection scale and temporal detection scale. And L2 distance is metric for the local signatures.

3. Voting Functions:

After performing similarity search to finding more specific similarity according to some criteria there voting functions is used. Signature is depending upon temporal, ordinal or temporal ordinal features. According to that voting functions are decided. If transformation like resize, rotation and translation for spatial and slow/fast motion from the temporal point of view then point selection will be dependent on criteria. The author Law-To et al. provide a block diagram of CBCD overview.

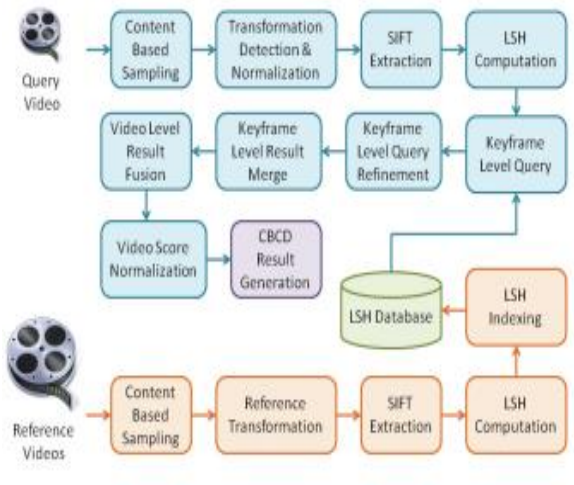

Fig.1. Overview of CBCD[3]

Liu et al. proposed the best performing methods in TRECVid 2009 for video signature. In this method sparse temporal sampling is performed. For each shot single key-frame will be extracted which describes the contents of frame using SIFT features. Two transformations- half resolution and strong re-encoding are applied to each original reference key-frame. It is performed to cope with transformations of the query video. These additional versions of the reference key-frame are helpful to detect query key-frames with PiP and

strong re-encoding transformations. All three sets of key-frames are processed independently. SIFT features are extracted to establish image quality. Indexing is performed through locality sensitive hashing efficiently. Copy detection is based on a gradual key-frame level query process, followed by video-level merging.

In [5] Shen, presents a duplicate video clip detection system. The videos are near about similar. The author focuses on comparison between two detection schemes. In a first scheme technique is used as a bounded coordinate system (BCS), while second one utilizes frame symbolization. In BCS compact representation of features where temporal information is nullified and each entire video is summarized by a single vector. In the second scheme symbolization of frame will be performed. Where mapping each video to a sequence of symbols will be done and temporal order taking into account. To reduce the search space database indexing is used which also improves matching speeds. The experiment was performed on TV commercials where BCS system give better result comparatively other one. So final conclusion is system is better for matching short clips but not efficient for localizing the clip embedded into larger unrelated video.

In [6], another duplicate video clip detection system is presented by Shang. In this sytem also video clips are nearly similar. Two problems are tried solve here. One is how to represent video in a compact way and second is how to match frames efficiently. To describe spatial information two different strategies are used. First is automatic and based on CE i.e. Conditional entropy while second one is heuristic and based on LBP i.e. Local Binary Pattern. The given figure depicts the ordinal measure where image is divided blocks. The dimension of the block is 3*3. An average of grey level values in each block is computed. The set of average grey level values (intensities) is arranged in ascending order and allocate a rank to each one. Color degradation is the robust factor for ordinal measure.

1. CE based spatiotemporal feature:

It is based on conditional entropy so here is concepts from information theory are used. Feature selection is important for conditional entropy. Select a feature subset which carries as much information as possible. It is also called as minimize conditional entropy. The procedure of ordinal relation selection is given into algorithm [6]. In this method function used mapping is a hash function. Where hash function is useful to convert frames into binary numbers and spatio-temporal information encoding.

2. LBP based spatiotemporal feature:

The given figure shows the procedure of LBP (Local Binary Pattern). LBP computes in a simple way so the features can be extracted in challenging real time system. It also provides tolerance for illumination changes.

To extract the temporal features *w-shingling* concept is modelled which is originally used for text retrieval to measure document similarity. The efficient retrieval is achieved by fast intersection kernel method [13] [14] into the inverted file. Experiment will be performed on two web video datasets (One is constructed by author team and other is provided by CMU and CityU)

This method efficiently find out near duplicate web video retrieval but not successful for locating segment in larger video.

In [7] provides an approach for detecting a near duplicate videos from large database. The detection will be performed on fingerprint or signature based. Because in signature the features are temporal in other words spaces remain mostly unchanged even after various noise attacks. Frame level features are extracted using MPEG-7 Color Layout descriptor (CLD). For fast searching speed and reduction in descriptor size clustering using k-means is performed over frame level descriptor. For video signature/fingerprint centroids of the cluster considered as a key feature.

The author proposed a two phase procedure as shown in above figure. In first phase will find out nearest neighbours (NN) using coarse search. To generate a signature/fingerprint vector quantization will be performed on the individual vectors in the model using a codebook. To improve the performance different techniques are proposed here like information which is already calculated (precomputed) for VQ symbols, partial distance pruning and dataset pruning etc. In second phase to find out nearest neighbors unquantized features are considered. Either approach of distance thresholding or based on registration will be considered for the matching pupose. If the query contains multiple video portions then this method is not effective.

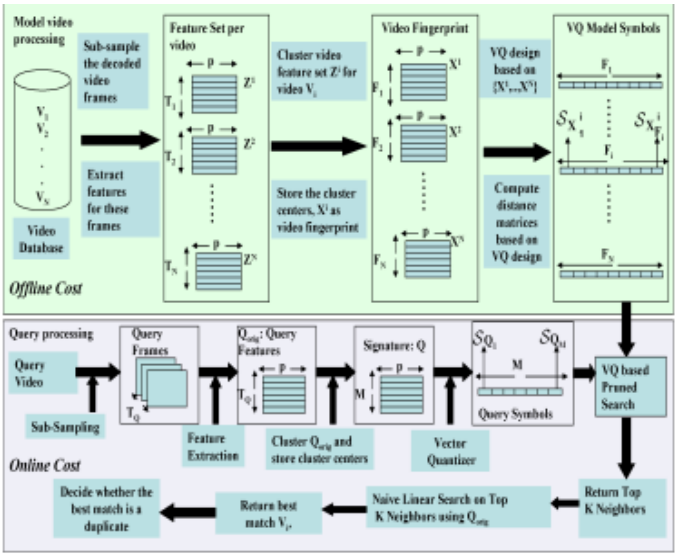

Fig.2 Block diagram duplicate detection framework [7].

The author [16] of this proposed inverted file based similarity search for fingerprint matching. The method was proposed for detection of near duplicate videos. So the videos considered are nearly similar. Each fingerprint is divided into small non overlapping blocks, called as words. These words are useful for creating an inverted file from database of fingerprints which is represented as a table. From the table word position inside a fingerprint is represented horizontally and possible values of the words are shown vertically.

The fingerprint is divided into number of words for the detection of query fingerprint. The query fingerprint is compared with all fingerprints from the database which started from the same word. If from the first column of the inverted file found fingerprint with the same starting word then hamming distance is calculated for the query and the fingerprints. If hamming distance is less than some predefined threshold it will be declared as a match. If it is not found then procedure will be repeated. The length of the word should be 16 bits. For larger than 16 bits length word have to be complete inverted file is not practically possible. Similarity search method based on header is proposed which reduces complexity. For that purpose centroid is chosen from each fingerprint called as header. Dummy fingerprints are generated through the header of the fingerprint present in the database. Each query fingerprint is compare with the dummy fingerprint. Hamming distance is calculated for the fingerprint and find out the fingerprint having minimum hamming distance. To reduce the search time actual comparison is performed with that fingerprint.

The author Saddam Bekhet and Amr Ahmed proposed technique [17] for compressed video shot matching. Matching is done through compact signatures extracted directly without decompression. Key technique is used as Dominant Color profile (DCP). DCP represents information of color in a similar way to scene representation by the human's retina, in the form of spikes. DCP having quality of it encodes both spatial and temporal information. So it is more efficient way for real time matching.

There are different parameters are considered are quantization factor, number of blocks and number of dominant colors. The results are computed against various and challenging datasets, that proves robustness of DCP. There are efficient computations which efficiently retrieve an initial maximum set of matching videos. The author work on different layers like re-ranking of the videos for further semantic analysis as mentioned in [18].

Future work can be performed over more compact and fixed length signature regardless of video shot length.

3. PROPOSED METHOD

The method proposed by the system is to detect video content from larger unrelated video using a video Signature. Temporal based signature is generated. The block diagram of the system as shown fig . The system is divided into three steps –

1. Signature Extraction:

a. Video signature is made up of fine signature and coarse signature.
b. To extract fine signature first extract frame level features. Luminance is considered as a key feature for signature.
c. By averaging luminance value frame signature will be formed. Combining frame signature and confidence value fine signature will be generated.
d. By applying bag of words algorithm over fine signature coarse signature is created.
e. Coarse signature followed by fine signature generates the required video signature.

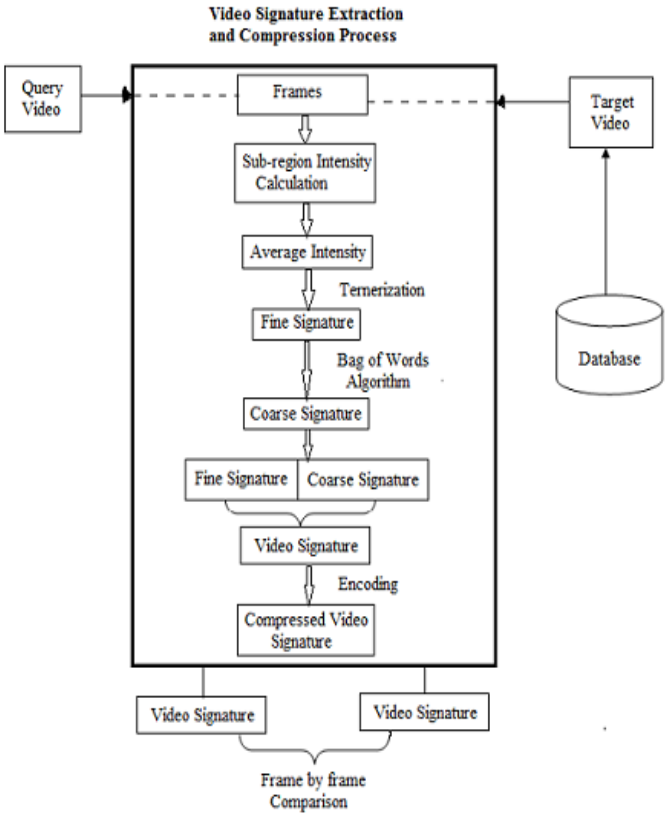

Fig.3. Block diagram of Video signature system

Signature Compression:

To speed up the system compression will be applied over video signature.

2. Matching:

Signature is made up of the frames. Query video and target video (Video in which find out the segment of query video) for both signature will be generated. Frame by frame comparison will be performed. Distance factors are considered while matching. If match will found matched segment displayed otherwise there is no matching segment in the video.

3. APPLICATIONS

- **Rights Management and monetization:**

To detect copyright infringement it is useful for content owners and to identify content owner is the requirement for content consumer. So it is useful for both content owners and consumers.

- **Distribution Management:**

It is useful in organization by applying video fingerprint i.e signature database that contains sensitive data and automatically detects and stops accidental transmission of this content via email, unauthorized copy to external device or other.

- **Video-Content based linking:**

Video content can become a linking content like a text and it can be used to infer association with other pages.

- **Database management and duplication:**

The system is useful for high volume owners and creators like studios and archives also personal video libraries to manage the data and avoid duplication.

4. CONCLUSIONS

The paper presents survey of content identification from videos to find out near duplicate videos. In this paper learnt that video signature of different types. The various kinds of features are extracted from images according to that signature will be generated like ordinal, motion etc. The survey also includes different types of descriptors (local and global) and methods available for generating various types of signature with their pros and cons.

Method proposed by this paper is not just useful for finding duplicate videos but also it recognizes the video segment which is embedded into larger unrelated video. For that purpose it will extract frame level feature through this it will create video signature. Using this signature matching of two video will be performed.

Further enhancement can be performed by extracting other type of features and different type of matching strategies.

ACKNOWLEDGEMENT

This is to acknowledge and thank all the individuals who played defining role in shaping this paper. We avail this opportunity to express our deep sense of gratitude and whole hearted thanks to our friends and family for giving their valuable guidance, inspiration and encouragement to embark this paper.

REFERENCES

[1] Paschalakis et al, "MPEG-7 Video Signature Tools for Content Identification" in *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 22, No. 7, July 2012.

[2] A. Hampapur, K. Hyun, and R. Bolle, "Comparison of sequence matching techniques for video copy detection", in Proc. Conf. Storage Retrieval Media Databases, 2002, pp. 194-201.
[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. 6th ACM Int. Conf. Image Video Retrieval, Jul. 2007, pp. 371-378.
[4] Z. Liu, T. Liu, and B. Shahraray, "AT&T research at TRECVID 2009 content-based copy detection," in Proc. TRECVid Workshop.
[5] H. T. Shen, X. Zhou, Z. Huang, J. Shao, and X. Zhou, "UQLIPS: A real-time near-duplicate video clip detection system," in Proc. 33rd Int. Conf. Very Large Data Bases, Sep. 2007, pp. 1374-1377.
[6] L. Shang, L. Yang, F. Wang, K.-P. Chan, and X.-S. Hua, "Real-time large scale near-duplicate web video retrieval," in Proc. ACM Int. Conf. Multimedia, Oct. 2010, pp. 531-540.
[7] A. Sarkar, V. Singh, P. Ghosh, B. S. Manjunath, and A. Singh, "Efficient and robust detection of duplicate videos in a large database," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 6, pp. 870-885, Jun. 2010
[8] D. Bhat and S. Nayar, "Ordinal measures for image correspondence," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 Issue: 4, pp. 415–423, April 1998.
[9] R. Mohan., "Video sequence matching.," in *Proceedings of the International Conference on Audio, speech and Signal Processing, IEEE Signal Processing Society*, 1998.
[10] M. Y. M. Naphade and B.-L. Yeo, "A novel scheme for fast and efficient video sequence matching using compact signatures.," in *Proc. SPIE, Storage and Retrieval for Media Databases 2000*, Vol. 3972, pp. 564–572, Jan. 2000.
[11] L. Chen and F. W. M. Stentiford. Video sequence matching based on temporal ordinal measurement. Technical report no. 1, UCL Adastral, 2006.
[12] A. Joly and C. Frelicot. Content-based copy detection using distortion-based probabilistic similarity search. IEEE transaction on Multimedia, 2007.
[13] S. Maji, A. Berg, and J. Malik. Classification using intersection kernel support vector machine is efficient. In *Proc. CVPR*, 2008
[14] J.X. Wu, and J. M. Rehg. Beyond the Euclidean distance: Creating effective visual codebooks using the histogram intersection kernel. In *Proc. ICCV*, 2009.
[15] Chou, Chien-Li, Hua-Tsung Chen, and Suh-Yin Lee. "Pattern-Based Near Duplicate Video Retrieval and Localization on Web-Scale Videos." Multimedia, IEEE Transactions on 17.3 ,382 395, 2015.
[16] Divya Devan, Gopu Darsan, "A Compact SpatioTemporal Fingerprint for Video Copy Detection System" IJSER ISSN (Online): 2347-3878, Volume 1 Issue 3, November 2013.
[17] Saddam Bekhet, Amr Ahmed, "Compact Signature-Based Compressed Video Matching Using Dominant Color Profiles (DCP)." IEEE Transaction on Pattern recognition, 2014.
[18] A. Altadmri and A. Ahmed, "A framework for automatic semantic video annotation," *Multimedia Applications and Tools,* vol. 64, pp. 1-25, 2013.