Robust Video Copy Detection System

Saleem Sayyed[1] Arshad Shaikh[2] Farooque Shaikh[3] Kavita Bhangale[4]
[1,2,3]Student of B.E [4]Assistant Professor
[1,2,3,4]Department of Electronics & Telecommunication Engineering
[1,2,3,4,5]M. H. Saboo Siddik C.O.E, Mumbai, India

Abstract— The video copy detection system based on content fingerprinting, is used for video indexing and copyright applications. This system consists of two algorithms: a fingerprint extraction algorithm succeeded by a fast approximate search algorithm. The first algorithm extracts compact content-based signatures from the images captured from the query video. Each such image signifies a short sector of the input video and contains temporal as well as spatial information about the video sector. These images are denoted by temporally informative representative images. To find whether a query video (or a part of it) is copied from a video in a system database, fingerprints of all the existing videos in the system database are extracted and kept in advance. The search algorithm searches the retrieved fingerprints to find matches for the fingerprints of the input video. The proposed approximate and quick search algorithm facilitates the online application of the system to a large video database of tens of millions of fingerprints, so that a match (if it exists) is found in a few seconds.

Key words: Content-based fingerprinting, multimedia fingerprinting, video hashing, video copy detection, video copy retrieval

I. INTRODUCTION

In recent times, lots of videos are being uploaded on the internet and are shared. Among these videos, a large numbers of videos are there that are illegal copies or some videos are manipulated version of an already existing video. As a result, copyright management on the internet becomes a complex process. Today's worldwide video copyright infringement calls for the development of accurate and fast copy detection algorithm. To find infringements, we can have two approaches. The first approach is based on watermarking and the other one is based on Content Based Copy Detection (CBCD).The watermarking is a widely used technique in the photography field. The purpose of a watermark is to identify the work and discourage its unauthorized use. With the help of watermarking, we can find if images are copied or not. Digital watermarking is the process of embedding a certain piece of information called as watermark into multimedia content including text documents, images, audio or video streams. [1] The first limitation of watermark is that if the original image is not watermarked, then it is impossible to realise whether other images are copied or not. Another drawback of watermarking is that the degree of robustness is not adequate for some of the attacks that are encountered frequently. Areas of bright highlights or dark shadows do not accept watermarks well. In highlights, where coverage of color approaches 0%, there are very few pixels to modify to add the signal. In the shadows, where coverage of color approaches 100%, the capability to modify pixels is essentially halved because of ink saturation. There may be a need to increase the strength setting using the slider to improve watermark visibility. The watermark can also be strengthening in highlights, but this generally negatively impacts visibility. The limitations of watermarking can be overcome by using another technique developed called as Content Based Copy Detection (CBCD).

Content Based Copy Detection (CBCD) is an emerging and active research area due to various improvements witnessed in multimedia and communication technologies. The main aim of Content Based Copy Detection (CBCD) can be expressed by the quote that says "the media itself is the watermark", i.e., the media comprises of distinct information that is used for detecting their copies. Content Based Copy Detection (CBCD) techniques offer an alternate approach to watermarking in order to identify video series from the same source. Watermarking depends on inserting a unique pattern in the video stream; video copy detection techniques compare content based signatures for detecting copies of video. The principle advantage of Content Based Copy Detection (CBCD) over watermarking is the fact that the signature extraction can be done after the media has been distributed. Content Based Copy Detection detects the duplicate by comparing the fingerprint of the query video with the copyrighted videos.

II. BRIEF HISTORY

A. History of Video Fingerprinting:

Research that began a decade ago in video copy detection has developed into a technology known as "video fingerprinting". A video fingerprint is a quantifier that is extracted from a piece of video content. The process of choosing a fingerprint from the original video content is referred to as video fingerprinting. There is an obvious correlation between human fingerprint and video fingerprinting. Just as human fingerprint can uniquely identify a human being, video fingerprint can uniquely identify a piece of video content. The parallelism expands to the process of subject identification by fingerprint: At the start, known fingerprints are stored in a database, and then a subject's fingerprint is queried against the database to match.

B. Video Fingerprints Used Previously:

1) Color-Space-Based Fingerprints:

Color-space-based fingerprints are among the first feature extraction methods used for video fingerprinting. They are normally derived from the histograms of colors in certain regions in time and/or space within the video. Advantages of color histograms are efficiency and insensitivity to small changes in camera viewpoint. Color histograms are frequently used to compare images. Color histograms have some limitations such as it provides no spatial information and it rarely describes what all colors are there in the image and their exact quantities.

2) Temporal Fingerprints:

To overcome drawback of color-space-based fingerprints new video fingerprint extraction algorithm is developed that can be applied to the luminance (the gray level) value of the frames. In order to find pirate videos on the Internet, Indyk et al. [2] use temporal fingerprints based on the shot boundaries of a video sequence and the time distance between boundaries is its signature. The feature of a boundary image isn't used since the content of this image is unreliable, usually black or hard to be identified by human eyes. This technique needs a shot-boundary-detection algorithm, and it can be adequate for finding a full movie, but it might not work well for short episodes with a few boundaries. The common way for shot detection is to evaluate difference value between consecutive frames represented by a given feature.

3) Spatial Fingerprints:

Spatial fingerprint algorithm [3] converts a video image into YUV color space in which the luminance (Y) component is kept and the chrominance components (U, V) are discarded. The luminance image is further subdivided into a fixed-sized grid of blocks independent of frame resolutions. Spatial fingerprints are features derived from each frame or a key frame and are mostly used for both image and video fingerprinting. There is a huge research in the area of image fingerprinting and a large no of researchers have extended the concepts developed for image fingerprinting to the video fingerprinting field.

The spatial fingerprints are further subdivided into global and local fingerprints.

1) Global Fingerprints: It is the method to obtain signatures. Global fingerprints concentrate on the global properties of a frame or a subsection of it like image histograms.
2) Local Fingerprints: It usually represents local information around some interest points within a frame like corners. The key point detection is performed by extracting the extrema in high-pass difference of Gaussian images (DOG), at different scales.

4) Spatio-Temporal Fingerprints:

One shortcoming of spatial fingerprints is their inability to capture the video's temporal information, which is an powerful discriminating factor. Hence the new fingerprinting approach is designed to call as spatio-temporal fingerprints. A Spatio-temporal fingerprints that consists of both spatial and temporal information about the video are thus expected to perform better than fingerprints that use only temporal or spatial fingerprints. Temporal information is related to time. Spatial information describes the physical location of objects and metric relationship between objects.

III. LITERATURE SURVEY

A. Video Fingerprint Using 3-D DCT Algorithm:

Coskun consider a video as a three-dimensional (3-D) matrix of luminance values based on hash function. According to Coskun, 3D-DCT hash algorithm [4] is based on the low-pass coefficients of 3-D transformations of the luminance component of a video sequence. Coskun consider 3D-DCT but several other transforms such as Discrete Wavelet Transforms can be used. The final hash string is generated from the relative magnitude relationship among selected coefficients. The 3D-DCT robust hashing algorithm has three major steps as shown in fig 1.

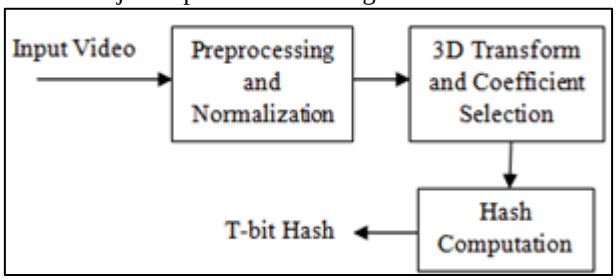


Fig. 1: 3 D-DCT Robust Hashing Algorithm

B. Preprocessing and Normalization:

Pre-processing pre-process the video signals. The pre-processed video can still have an infinite variety of appearances due to innocent or malicious modifications, which range from MPEG compression to frame drops in the transmission. According to Coskun [4], the input video sequence is first converted to a standard video signal in terms of frame dimensions and of the number of frames via smoothing and sub sampling.

In normalization the original video signal with arbitrary dimensions is temporally smoothened and sub sampled. In involves two types of normalization: Temporal and Spatial. In case of temporal normalization, the array of pixels having the same location in successive frames is referred as a pixel tube and the (m, n) th pixel tube is defined as {V (m, n, f): f=1, 2,……., f total} where f total is the total original number of frames, and there is one such tube for each of the pixels in the frame. In spatial normalization, the frames are spatially smoothened and sub sampled to yield the goal sequence. The smoothened video signal is afterwards sub sampled in time, to reduce the input clip video the target number of frames "F".

C. 3D-Transforms and Coefficient Selection:

Most 3D-transform techniques with good compacting characteristics can serve the purpose of summarizing and capturing the video content as they collect and embed in their coefficients the information concurrently from time and space. Coskun describes two types of transforming [4] 3D-DCT and 3D-RBT.

D. 3D-DCT Transform Case:

After applying the DCT transform to the normalized sequence, a 3-D array of DCT coefficients is obtained. Typically low frequency DCT coefficients contain the predominant part of the energy and they are also robust against most of the signal processing attacks, such as filtering and compression. To satisfy the uniqueness or discrimination property, one must judiciously enroll coefficients from mid- to high-frequency regions. We exclude the lowest frequency coefficients in each direction to enhance the uniqueness among video shots that have similar, but not identical content. The weakness of 3D DCT method is that, the universal coefficient set whose random subset is used in hash extraction is known to an attacker. Thus, he may forge the original video without disturbing the rank order of those coefficients. In other words, he may modify the forged video by slightly altering candidate DCT

coefficients and making them to have the same rank order as the original video. Then, no matter which subset of the coefficients is used, the extracted hash would be exactly the same.

E. Hash computation

Once the 3D-transform is applied and the specific coefficients are chosen, hash computation procedure is the same, regardless of which transformation is used. The selected transforms coefficients are binarized using the median of the rank-ordered coefficients.

IV. PROPOSED METHOD

In TIRI-DCT method, two fast searching methods are used for matching process of fingerprint. These two methods are cluster based similarity search and inverted file based similarity search. As Temporal Informative Representative Image (TIRI) contains both spatial and temporal information on a short segment of a video clip, the spatial feature selected from a TIRI would also contain temporal information. Based on TIRIs; an effective fingerprinting algorithm is proposed to call as Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT) and compared with Discrete Cosine Transform (DCT). TIRI-DCT is improved version of 3D-DCT.

A. Video Fingerprint Generation:

Fig 2 shows the block diagram of proposed approach which is based on temporally informative representative images (TIRIs). In TIRI-DCT before extracting the fingerprints, the input signal from the video is processed. Preprocessing is mostly used for improving coding efficiency and visual quality of the video compression system. Copies of the same video with multiple frame rates and sizes generally exist in the same database of video because of which, fingerprinting algorithm should be resilient to changes in the frame rate as well as the frame size. Down sampling usually increase the resilient of a fingerprinting algorithm to these changes. This is generally done to minimize the data rate or the data size.

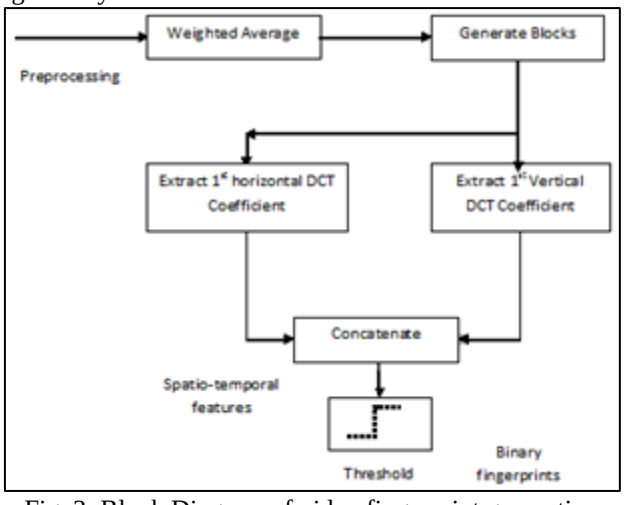


Fig. 2: Block Diagram of video fingerprint generation

Down sampling factor, M is either an integer or a rational fraction more than unity. Since down sampling decreases the sampling rate, we should be careful enough so that the Shannon-Nyquist sampling theorem criterion is sustained. If the sampling theorem is not fulfilled then the resulting digital signal may have aliasing. Therefore, to make sure that the sampling theorem is fulfilled, we make use of a low pass filter as an anti-aliasing filter to decrease the bandwidth of the signal before the signal is down sampled; the overall process (low pass filter then down sample) is called as decimation. The only reason to filter the bandwidth is to avoid the case where the new sample rate would become lower than the Nyquist requirement and then cause the aliasing by being below Nyquist minimum.

As shown in fig 3, each video is down-sampled both in time and space. Prior to down-sampling, a Gaussian smoothing filter is applied in both domains to stop aliasing. This down-sampling process gives the fingerprinting algorithm with inputs of fixed size (WxH) pixels and fixed rate (F frames/second). After preprocessing, the video frames are divided into overlapping sectors of constant length, each consisting of K frames. The fingerprinting algorithms are applied to these sectors. Overlapping minimises the sensitivity of the fingerprints to the "synchronization problem" which is called as "time shift".

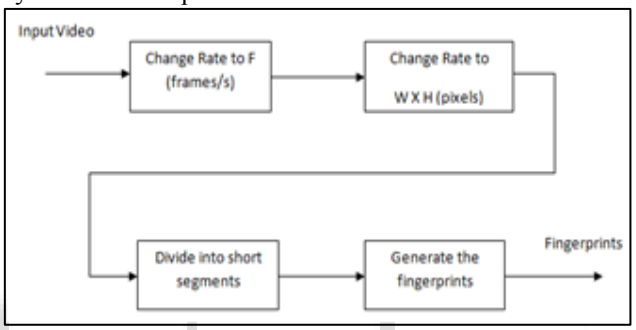
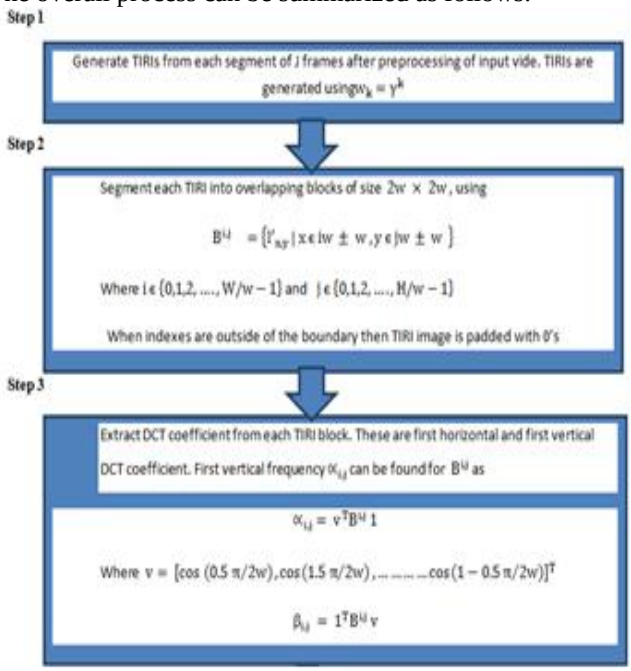
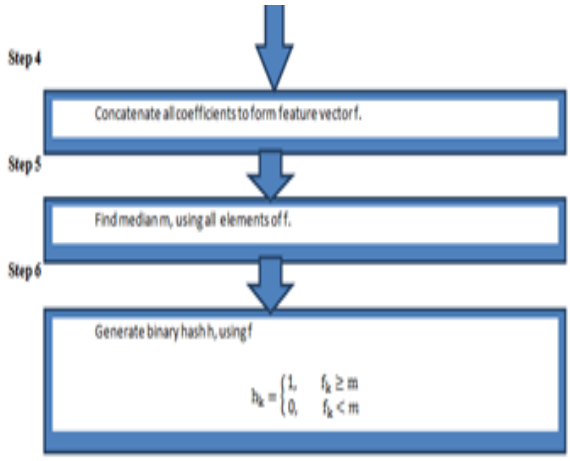


Fig. 3: Preprocessing Steps

As TIRI-DCT transforms algorithm capture of the temporal information in a video using the same feature extraction process. The features are obtained by applying a 2D-DCT on overlapping blocks of size from each TIRI. As shown in figure 2, the first horizontal and the first vertical Discrete Cosine Transform (DCT) coefficients (features) are extracted from every block. The value of the coefficients from all the blocks is concatenated to form the feature vector.

The overall process can be summarized as follows:





B. Video Copy Detection:

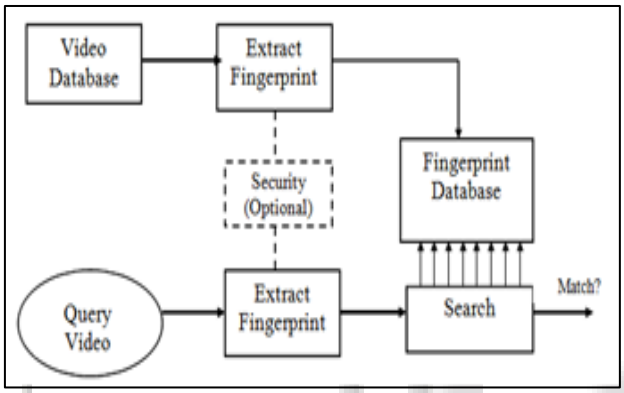


Fig. 4: Block Diagram of video copy detection

C. Cluster Based Similarity Search:

Cluster based similarity search is a similarity search algorithm for binary fingerprints. The main idea of this algorithm is to minimize the number of queries which are examined within the database by using clustering. By allocating each fingerprint to only one cluster (out of m clusters), the fingerprints in the database will be clustered into 'm' non overlapping groups. In order to achieve this, a centroid is chosen corresponding to each cluster, which will be termed as the cluster head. A fingerprint will be assigned to cluster if it is nearest to this cluster's head as shown in fig 5.

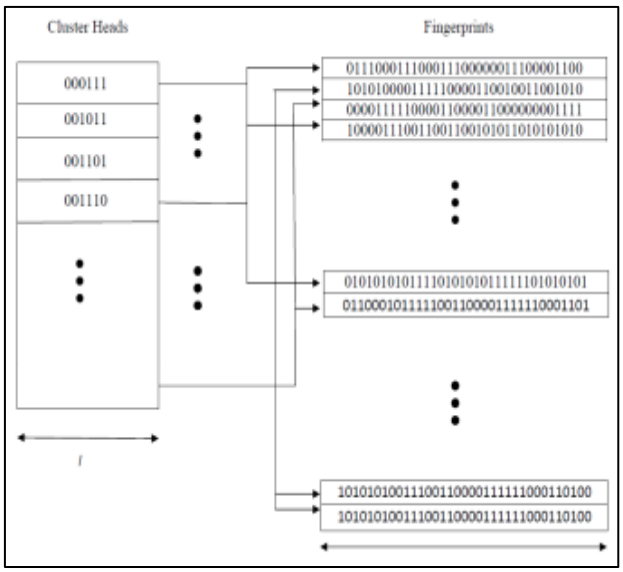


Fig. 5: Cluster Based Similarity Search

To find whether a query fingerprint matches a fingerprint present in the database, the cluster head nearest to the input is found. All the fingerprints of the videos stored in the database, belonging to this cluster are then searched to find a match, i.e., the one which is having the smallest Hamming distance (less than a certain threshold) from the query. If a match is not found, the cluster which is the second nearest to the query is examined. The process carries on until a match is found or the farthest cluster is examined. If still the match is not found, the input is declared to be out of the database. These cluster heads should be chosen such that a minute change in the fingerprint does not lead to a situation where the fingerprint being assigned to another cluster. In general setting, cluster centers with all the binary vectors with length < L are chosen. To allocate a fingerprint to a particular cluster, the fingerprint needs to be first divided into words or segments of length s. Each word is then indicated by one bit in the -bit cluster head, depending on the majority of word's bit values; for example, it is represented by 1, if it has more than s/2 1's and it is represented by 0, if it has less than s/2 1's. Equivalently, each bit of the cluster head can be replicated s times and the Hamming distance between the expanded bit version of all the cluster heads and the fingerprint is calculated. The cluster head nearest to the fingerprint is then allocated to that fingerprint. Thus the worst case complexity of cluster based similarity search is O(N).

V. Conclusion

This paper proposes a fingerprinting system for video copy detection. It can be used for copyright management and indexing applications. The system consists of a fingerprint extraction algorithm followed by an approximate search method. The proposed fingerprinting algorithm (TIRI-DCT) extracts robust, discriminate, and compact fingerprints from videos in a fast and reliable fashion. These fingerprints are extracted from TIRIs containing both spatial and temporal information about a video segment. We demonstrate that TIRI-DCT generally outperforms the well-established (3D-DCT) algorithm and maintains a good performance for different attacks on video signals, including noise addition, changes in brightness/contrast, rotation, spatial/temporal shift and frame loss. It is shown experimentally that TIRI-DCT has a high average true positive rate of 98.2% and a low average false positive rate of 0.97%.

As part of our future work

1) We will conduct a detailed analytical study of the security of fingerprinting algorithms.
2) We will carry an extensive comparison study to compare our fingerprinting algorithms to other state-of-the-art algorithms.
3) We also plan to study the performance of the system in the presence of some other attacks, such as cropping, and logo insertion.

VI. Acknowledgment

We would like to thank Er. Abdul Sayeed for his support throughout the time of preparation. We would also like to thank all the professors concerned, for their useful guidance.

REFERENCES

[1] J. Oostveen, T. Kalker, and J. Haitsma, "Feature extraction and a database strategy for video fingerprinting," in Proc. Int. Conf. Recent Advance in Visual Information Systems (VISUAL), London, U.K., 2002, pp. 117–128, Springer-Verlag.
[2] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in Proc. Int. Conf. Very Large Data Bases (VLDB), San Francisco, CA, 1999, pp. 518–529, Morgan Kaufmann Publishers Inc..
[3] Mrunalini Parag Bhogle and Anil Chhangani, "Content Based Copy Detection Using TIRI-DCT Method", in international journal of engineering sciences & research technology, pp 449-454, July, 2014.
[4] B. Coskun, B. Sankur, and N. Memon, "Spatiotemporal transform based video hashing," IEEE Trans. Multimedia, vol. 8, no. 6, pp. 1190–1208, Dec. 2006.

