TIRI – DCT Based Video Copy Detection System

Vaishali V. Sarbhukan*, Prof. V. B. Gaikwad**
*(Department of Computer Engineering, Mumbai University, Navi Mumbai)
** (Department of Computer Engineering, Mumbai University, Navi Mumbai)

Abstract

Nowadays thousands of videos are being uploaded to the internet and are shared every day. Out of these videos, considerable numbers of videos are illegal copies or some videos are manipulated versions of existing media. Due to these reasons copyright management on the internet becomes a complicated process. Copyright infringement and data piracy are major serious concerns. To detect such kind of infringements, there are two approaches. First is based on watermarking and other is based on content-based copy detection (CBCD). Existing video fingerprint extraction algorithms like color-space-based fingerprints, temporal fingerprints, spatial fingerprints, and spatio-temporal fingerprints have certain limitations. Therefore temporally informative representative images - discrete cosine transform (TIRI-DCT) is developed. TIRI-DCT is based on temporally informative representative images which contains spatial and temporal information of a short segment of a video sequence.TIRI-DCT has better performanance as compared to 3D-DCT. TIRI-DCT followed by a fast approximate search algorithms like inverted file-based similarity search and cluster-based similarity search. Existing exhaustive search method is time consuming process. Drawback of exhaustive search method is overcome in inverted file-based similarity search and cluster-based similarity search.

Keywords – Cluster-based similarity search, 3D-DCT, Inverted file-based similarity search, Fingerprinting system, TIRI-DCT.

1. Introduction

Due to growing broadcasting of digital video content on different media brings the search of copies in large video databases to a new critical issue. Digital videos can be found on TV Channels, Web-TV, Video Blogs and the public Video Web servers. The massive capacity of these sources makes the tracing of video content into a very hard problem for video professionals. Recently for increasing online video repositories copyright infringements and data piracy have become serious concerns. Copyright infringement occurs when someone other than the copyright holder copies the "expression" of a work. Copyright infringement is often associated with the terms piracy and theft. Today's widespread video copyright infringement calls for the development of fast and accurate copy-detection algorithms. To detect infringements, there are two approaches. First is based on watermarking and other is based on content-based copy detection (CBCD).The watermarking is a widely used technique in the photography field. It allows the owner to detect whether the image has been copied or not. The limitations [1] of watermarks are that if the original image is not watermarked, then it is not possible to know if other images are copied or not. The primary aim of content-based copy detection (CBCD) is "the media itself is the watermark". The key advantage of content-based copy detection over watermarking is the fact that the signature extraction can be done after the media has been distributed.

2. Strusture Of Fingerprinting System

Fig. 1 shows the overall structure of this fingerprinting system. Content-Based Copy Detection finds the duplicate by comparing the fingerprint of the query video with the fingerprints of the copyrighted videos. To find a copy of a query video in a video database, one can search for a close match of its fingerprint in the corresponding fingerprint database (extracted from the videos in the database). Closeness of two fingerprints represents a similarity between the corresponding videos; two perceptually different videos should have different fingerprints.

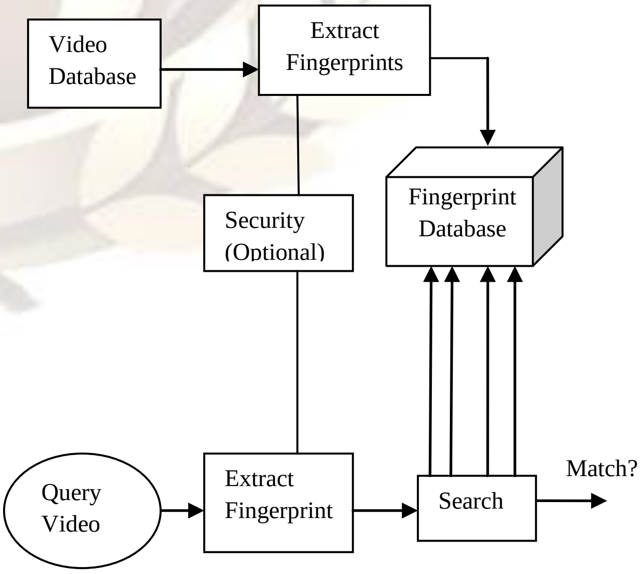

Figure 1 Overall Structure of Fingerprinting System.

3. Properties of Fingerprints

Ideally, a design of a video fingerprint should have the following characteristics that hold true for a large corpus of video content of diverse types**.**

3.1 Robustness

Robustness of a fingerprint requires that it changes as little as possible when the corresponding video is subjected to content-preserving distortions.

3.2 Discriminant

The video fingerprints for different video content should be distinctly different.

3.3 Easy to compute

The fingerprint should also be easy to compute. For online applications, a fingerprinting algorithm should be able to extract the signatures as the video is being uploaded.

3.4 Compact

If a fingerprint is not compact, finding a match for it in a very large database can become a time-consuming process.

3.5 Secure

The fingerprinting system should be secured, so as to prevent an adversary from tampering with it.

3.6 Low complexity

The algorithm for extracting video fingerprints should have low computational complexity so that a video fingerprint can be computed fast.

3.7 Efficient for matching and search

Although there are generic algorithms that treat all fingerprints as a string of bits in matching and search, a good design of video fingerprint should facilitate approximation and optimization to improve the efficiency in matching and search.

4. Traditional Video Fingerprint Extraction Algorithms

Existing methods for CBCD usually extract a small number of features like signatures or fingerprints from images or a video stream and then match them with the database according to a dedicated voting function. Different types of traditional video fingerprint extraction algorithms are color-space based fingerprints, temporal fingerprints, spatial fingerprints, and spatio-temporal fingerprints.

4.1 Color-space-based fingerprints

Color-space-based fingerprints are among the first feature extraction methods used for video fingerprinting. They are mostly derived from the histograms of the colors in specific regions in time and/or space within the video. According to Lienhart [2] the color coherence vector (CCV) differentiates between pixels of the same color depending on the size of the color region they belong to. Lienhart et al and Sanchez et al have been tested for the domain of TV commercials and are susceptible to color variations. Naphade el al proposed a technique that was also experimented by Hampapur el al. Naphade[3][4] propose to use YUV histograms as the signature of each frame in the sequence and the use of histogram intersection as a distance measure between two signatures (frames). First disadvantage of color-space-based fingerprint is that color features change with different video formats. Another drawback of color features is that they are not applicable to black and white videos.

4.2Temporal 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. According to Indyk and Shivkumar[5], first, a video sequence is segmented into shots. Then, the duration of each shot is taken as a temporal signature, and the sequence of concatenated shot durations form the fingerprint of the video. Temporal signatures are computed on adjacent frames in a video. Temporal fingerprints are extracted from the characteristics of a video sequence over time. These features usually work well with long video sequences, but do not perform well for short video clips since they do not contain sufficient discriminant temporal information.

4.3 Spatial fingerprints

Spatial fingerprint algorithm converts a video image into YUV color space in which the luminance (Y) component is kept and the chrominance components (U, V) are discarded. Spatial fingerprints are features derived from each frame or from a key frame. They are widely used for both video and image fingerprinting. Spatial fingerprints can be further subdivided into global and local fingerprints. Global fingerprints focus on the global properties of a frame or a subsection of it like image histograms, while local fingerprints usually represent local information around some interest points within a frame like edges, corners, etc. One shortcoming of spatial fingerprints is their inability to capture the video's temporal information, which is an important discriminating factor.Therefore spatio-temporal fingerprints are developed.

4.4 Spatio-temporal fingerprints

Spatio-temporal fingerprints that contain both spatial and temporal information about the video are thus expected to perform better than fingerprints that use only spatial or temporal fingerprints. Some spatio-temporal algorithms consider a video as a three-dimensional (3-D) matrix and extract 3-D transform-based features[6][7].

5. Proposed TIRI-DCT System

There are some disadvantages of existing fingerprint extraction systems .They are as follows

• 3-D transform to a video is a computationally demanding process.

• The computational bottleneck is the search time in the matching process rather than the fingerprint extraction time.

• Overlapping reduces the sensitivity of the fingerprints to the “synchronization problem” . The problem with the binarization scheme limits the number of coefficients.

These drawbacks are overcome in proposed Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT) system. As a Temporally informative representative Images (TIRI) contains spatial and temporal information of a short segment of a video sequence, the spatial feature extracted from a TIRI would also contain temporal information. Based on TIRIs; we propose an efficient fingerprinting algorithm called as Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT.TIRI-DCT is improved version of 3D-DCT.

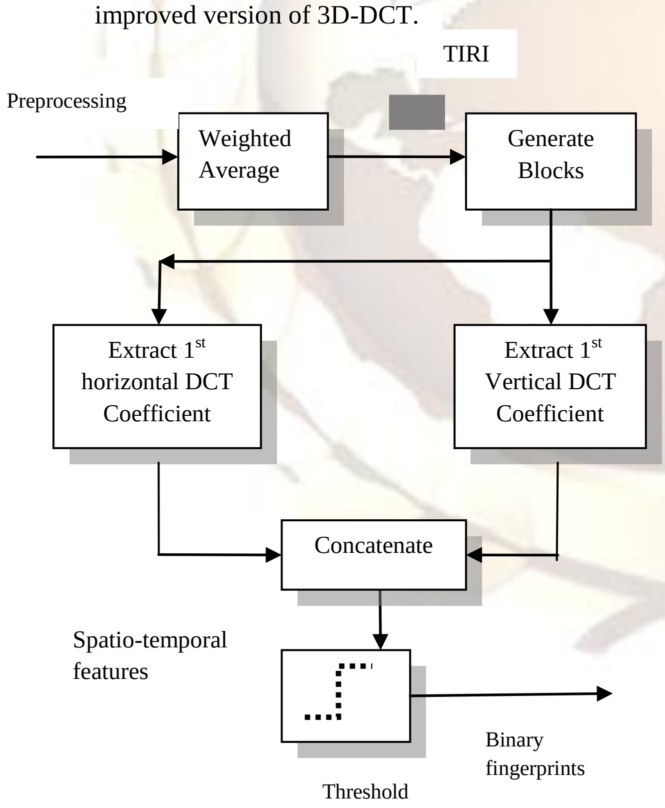

Figure 2 Schematic of the TIRI-DCT algorithm

Fig. 2 shows the block diagram of our proposed approach which is based on temporally informative representative images (TIRIs). In TIRI-DCT before extracting the fingerprints, the input video signal is processed. Copies of the same video with different frame sizes and frame rates usually exist in the same video database. As a result, a fingerprinting algorithm should be robust to changes in the frame size as well as the frame rate. Down-sampling[8] can increase the robustness of a fingerprinting algorithm to these changes. Prior to down-sampling, a Gaussian smoothing filter is applied in both domains to prevent aliasing. This down-sampling process provides the fingerprinting algorithm with inputs of fixed size (W X H) pixels and fixed rate (F frames/second). After preprocessing, the video frames are divided into overlapping segments of fixed-length, each containing J frames. The fingerprinting algorithms are applied to these segments. Overlapping reduces the sensitivity of the fingerprints to the “synchronization problem” which is called as “time shift”. As TIRI-DCT transform algorithm capture of the temporal information in a video using the feature extraction process. TIRI-DCT Algorithm includes following steps

Step 1: Generate TIRIs from each segment of J frames after preprocessing of input video .TIRIs are generated using $w_k = \gamma^k$.

Step 2: Segment each TIRI into overlapping blocks of size $2\omega \times 2\omega$, using

$$B^{i,j} = \{ l'_{x,y} \mid x \epsilon\, i\omega \pm \omega , y \epsilon\, j\omega \pm \omega \}$$

Where $i \epsilon \{0,1,2, \dots, W/\omega - 1\}$ and $j \epsilon \{0,1,2, \dots, H/\omega - 1\}$

When indexes are outside of boundary then TIRI image is padded with 0’s.

Step 3: Extract DCT coefficient from each TIRI block. These are first horizontal and first vertical DCT coefficient. First vertical frequency $\propto_{i,j}$ can be found for $B^{i,j}$ as

$$\propto_{i,j} = v^T B^{i,j}\, 1$$

Where

$$v = [\cos(0.5\pi/2\omega), \cos(1.5\pi/2\omega), \dots\dots \cos(1 - 0.5\pi/2\omega)]^T$$

And 1 is column vector of all ones. Similarly first horizontal frequency $\beta_{i,j}$ can be found for $B^{i,j}$ as

$$\beta_{i,j} = 1^T B^{i,j}\, v$$

Step 4: Concatenate all coefficients to form feature vector f.

Step 5: Find median m, using all elements of f.

Step 6: Generate binary hash h, using f

$$h_k = \begin{cases} 1, & f_k \geq m \\ 0, & f_k < m \end{cases}$$

6. Search methods for fast matching of video fingerprints

In real-world applications, the size of online video databases can reach tens of millions of videos, which translates into a very large fingerprint database size. This means that even if the fingerprint of a query video can be extracted very quickly, searching the fingerprint database to find a match may take a long time. For online applications, however, a reliable match should be found in almost real-time. Therefore fingerprint matching forms a practical bottleneck for online fingerprinting systems. A simple exhaustive search method which has a complexity of $O(N)$, where N is the number of the fingerprints in the database. Search methods inverted file-Based Similarity Search and Cluster-Based Similarity Search methods are modified version of search algorithm that was proposed by Oostveen [9].

6.1 Inverted-File-Based Similarity Search

The steps involved in Inverted-File-Based Similarity Search are as follows:
Step 1: Binary fingerprints are divided into n words of equal bits.
Step 2: The horizontal dimension of the table represents the position of a words.
Step 3: The vertical dimension of the table represents the possible values of words.
Step 4: Add index for each word of the fingerprint to the entry in column corresponding to the value of the word.
Step 5: Hamming distance is calculated between fingerprints in database and query fingerprint.
Step 6: If the distance is less than threshold value, then the query video will be announced as matching.
Step 7: Otherwise, it will be announced as not matching.

6.2 Cluster-Based Similarity Search

Cluster-Based Similarity Search is another similarity search algorithm for binary fingerprints. Our main idea is to use clustering to reduce the number of queries that are examined within the database. By assigning each fingerprint to one and only one cluster (out of k clusters), the fingerprints in the database will be clustered into k nonoverlapping groups. To do so, a centroid is chosen for each cluster, termed the cluster head. A fingerprint will be assigned to cluster if it is closest to this cluster's head . To determine if a query fingerprint matches a fingerprint in the database, the cluster head closest to the query is found. All the fingerprints (of the videos in the database) belonging to this cluster are then searched to find a match, i.e., the one which has the minimum Hamming distance (of less than a certain threshold) from the query. If a match is not found, the cluster that is the second closest to the query is examined. This process continues until a match is found or the farthest cluster is examined.

7. Conclusion

Due to increasing online video repositories, copyright infringements and data piracy have become serious concerns .Many videos uploaded on internet are illegal copies or manipulated versions of existing media. So today's widespread video copyright infringement calls for the development of fast and accurate copy-detection algorithms.Temporally Informative Representative Images (TIRI) - Discrete Cosine Transform (DCT) extracts compact content- based signatures from special images constructed from the video. Each image in TIRI represents a short segment of the video and contains temporal as well as spatial information about the video segment. To find whether a query video (or a part of it) is copied from a video in a video database, the fingerprints of all the videos in the database are extracted and stored in advance. The search algorithm searches the stored fingerprints to find close enough matches for the fingerprints of the query video. The disadvantages of existing fingerprint extraction systems are overcome in Proposed TIRI-DCT. TIRI-DCT is based on temporally informative representative Images. As a Temporally informative representative Images (TIRI) contains spatial and temporal information of a short segment of a video sequence, the spatial feature extracted from a TIRI would also contain temporal information.TIRI-DCT is faster than 3D-DCT while maintaining a very good performance over the range of the considered attacks. TIRI-DCT decreases false rejection rate (FPR) by increasing length of fingerprint. Another important property of TIRI-DCT is that it is computationally less demanding than 3D-DCT. TIRI-DCT followed by a fast approximate search algorithms like inverted file-Based Similarity Search and Cluster-Based Similarity Search. These search methods have better performance as compared to existing exhaustive search method. Proposed TIRI-DCT, Inverted-File-Based Similarity Search and Cluster-Based Similarity Search algorithms will improve performance by producing a high average true positive rate (TPR) and a low average false positive rate (FPR).

REFERENCES

[1] 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. Conf. Image Video Retrieval (CIVR), 2007.*

[2] C. K. R. Lienhart and W. Effelsberg, On the detection and recognition of television commercials, *in Proc. of the IEEE Conf. on Multimedia Computing and Systems, 1997.*

[3] A. Hampapur and R. Bolle ,Comparison of sequence matching techniques for video copy detection, *in Conference on Storage and Retrieval for Media Databases, 2002*.

[4] Naphade, M. R., Yeung, M. M. and Yeo, B.-L., A novel scheme for fast and efficient video sequence matching using compact signatures, *Proc. SPIE, Storage and Retrieval for Media Databases, vol. 3972, Jan. 2000*.

[5] P. Indyk, G. Iyengar, and N. Shivakumar. Finding pirated video sequences on the internet, *Technical report, Stanford University, 1999*.

[6] B. Coskun, B. Sankur, and N. Memon, Spatiotemporal transform based video hashing, *IEEE Trans. Multimedia, vol. 8, no. 6, Dec. 2006*.

[7] Radhakrishnan and C. Bauer, Robust video fingerprints based on Subspace embedding,*in Proc. ICASSP, Apr. 2008.*

[8] Mani Malek Esmaeili, Mehrdad Fatourechi and Rabab Kreidieh Ward ,Robust and Fast Video Copy Detection System Using Content Based Fingerprinting, *IEEE Trans On Information Forensics And Security,Vol.6 No.1,March 2011.*

[9] J. Oostveen, T. Kalker, and J. Haitsma, Feature extraction and a database strategy for video fingerprinting, *in Proc. Int. Conf. Recent Advances in Visual Information Systems (VISUAL), London, U.K., 2002, Springer-Verlag.*