ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Content Based Copy Detection Using TIRI-DCT Method

Mrunalini Parag Bhogle [*1], Anil Chhangani [2]
[*1] P.G Student LTOCE Koparkhairane, Department Computer Engg,Navi Mumbai, India
[2] Research Scholar, Shri Jagdishprasad Jhabarmal Tibrewala University, India
mrunalini_t@rediffmail.com

Abstract

Water marking relies on inserting information into the video stream in order to detect copies. This paper presents an approach to watermarking called content base copy detection(CBCD) which uses a combination of feature based matching and inverted index to detect copies video clips This paper presents a study of methods for video copy detection .Different techniques are described: spatial temporal , spatio- temporal information. Various properties are required for fingerprints. In proposed Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT) system. TIRI-DCT method; two fast searching methods are used for matching process of fingerprint.

Keywords: fingerprints, content base copy detection, Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT), robust

Introduction

In this paper, two approaches are used . First is based on watermarking and other is Content Based Copy Detection (CBCD).Watermarking technique is used in the photography field[1].

Limitation of watermarking

1. not possible to know whether other image are copied or not.
2. Degree of robustness is not adequate for some of the attacks that encounter frequently (Areas of bright highlights or dark shadows do not accept watermarks well).

To overcome limitations of watermarking another technique is 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, such as adoption of more [2]efficient multimedia coding standards and the astounding increase in data transfer rates. The primary aim of Content Based Copy Detection (CBCD) is "the media itself is the watermark", that is, the media (image ,video, audio) contains enough information that can be used for detecting copies. Content Based Copy Detection (CBCD) techniques provide an alternative approach to watermarking for identifying video sequences from the same source. Watermarking relies on inserting a distinct pattern into the video stream; video copy detection techniques match content based signatures to detect copies of video. The key 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 finds the duplicate by comparing the fingerprint of the query video with the fingerprints of the copyrighted videos.

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.

- **Robustness**

The robustness of a fingerprint require that it changes as little as possible when the corresponding video is subjected to content-preserving operations or a video fingerprint should stay largely invariant for the same video content under various types of processing, transformations and manipulations, such as format conversion, transcoding, and content editing. Content-preserving attacks (distortions) are changes that are made to the video unintentionally or intentionally by users of video-sharing websites. These changes can include format changes, signal processing operations, changes in brightness/contrast,

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

added noise, rotation, cropping, logo insertion, compression, etc.

- **Discriminant**

The video fingerprints for different video content should be distinctly different. In short two perceptually different videos have two different fingerprints.

- **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.

- **Compact**

A computationally demanding algorithm is not suitable for online applications, where thousands of videos need to be examined simultaneously in order to find possible copyright infringements therefore fingerprints should be compact. If a fingerprint is not compact, finding a match for it in a very large database can become a time-consuming process. Therefore fingerprint should be compact.

- **Secure**

The fingerprinting system should be secured, so as to prevent an adversary from tampering with it. Security is specifically important for copy-detection applications.

- **Low complexity**

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

- **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.

- **Scalability**

Different fingerprinting application scenarios are usually associated with different robustness requirements and, thus, different optimal "operating points" in the trade–off between the compactness of the signature and its robustness. Using MPEG-7 audio signatures, this can be accounted for by scaling several parameters of the signature in response to the application requirements and thus controlling the recognition strength and, thereby also, the data rate.

Background and Literature Survey

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. The luminance image is further subdivided into a fixed-sized grid of blocks independent of frame resolutions. One popular block based luminance signature is based on ordinal ranking. It was designed by Bhat and Nayar for image identification. After frame subdivision, the average pixel value for each block is computed, and an abstraction follows by ranking the blocks by their average pixel values. The rank of each block in ordinal position is assigned to the block as its signature. signature is more robust than some temporal. For a frame subdivision containing Mblocks, the required number of bits for a frame fingerprint is M* ceiling (log2M).R. Mohan[6] uses block-based luminance signature called as ordinal ranking in video fingerprinting and matching. Spatial fingerprints are features derived from each frame or from a key frame. Widely used for both video and image fingerprinting. There is a large body of research in the area of image fingerprinting and many researchers have extended the concepts developed for image fingerprinting to the video fingerprinting field. Spatial fingerprints can be further subdivided into global and local fingerprints.

- Global Fingerprints

It is the method to extract signatures. Global fingerprints focus on the global properties of a frame or a subsection of it like image histograms.

- Local Fingerprints

It usually represents local information around some interest points within a frame like edges, corners etc. The key point detection is performed by extracting the extrema in high-pass difference of Gaussian images (DOG), at different scales.. There is a large body of research in the area of image fingerprinting and many researchers have extended the concepts developed for image fingerprinting to the video fingerprinting field and uses scale invariant feature transform (SIFT) because it performs the best in terms of region representation specificity and robustness to image transformations. SIFT approach is widely used in computer vision. The scale invariant feature transform (SIFT) is an algorithm which is used to detect and describe local features in images. Key stages of SIFT includes feature matching and indexing, Cluster identification by the Hough transform voting, Model verification by linear least squares, Outlier detection. SIFT approach transforms an image into a large collection of local feature vectors, each of which is invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection.

Spatio-Temporal Fingerprints

One shortcoming of spatial fingerprints is their inability to capture the video's temporal information, which is an important discriminating factor. Therefore the new fingerprinting approach is designed to call as 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 finger prints. Spatial information describes the physical location of objects and metric relationship between objects. Applications of spatial information include satellite images, global positioning system (GPS),real estate sales etc. satellite images bring daily weather reports and provide farmers with information for precision agriculture. Global positioning system (GPS) monitors the locations of thousands of trucks and taxies. Real estate sales use geographic information systems. Spatial information is broader information technology sector and contains scientific and technological links such as environmental science, health delivery, computer science, resource management, planning, engineering and logistics. Temporal information is related to time. Some spatio-temporal algorithms consider a video as a three-dimensional (3-D) matrix and extract 3-D transform-based features. Applying a 3-D transform to a video is a computationally demanding process and may pose problems in online applications. Coskun and Sankur [3] [4] have presented two algorithms for computing a robust hash from video clips for the purposes. These are Discrete Cosine Transform (DCT) based hash algorithm and Randomized Basis Set Transform (RBT) algorithm[8]. According to Coskun the DCT hash is more robust, but lacks the security aspect, as it is easy to find different video clips with the same hash value. The RBT based hash, being secret key based, does not allow this and is more secure at the cost of a slight loss in the receiver operating curves. Others use spatio-temporal interest-point descriptors to generate the fingerprints [4][5].

Proposed 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 and thus the fingerprint length that can be used.

- 3D-DCT is resistant to different types of distortions that can happen to video signals.

These drawbacks are overcome in proposed Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT) system .In proposed TIRI-DCT method; two fast searching methods are used for matching process of fingerprint. These two methods are **inverted file based** similarity search and **cluster based** similarity search. As a Temporally Informative Representative Image (TIRI) contains spatial and temporal information on a short segment of a video sequence, the spatial feature extracted from a TIRI would also contain temporal information. Based on TIRIs; an efficient 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. In TIRI – DCT, first step is generation of temporally informative representative images (TIRI).So TIRIs generation is discussed which form the basis of the TIRI-DCT fingerprinting algorithm and then TIRI-DCT is discussed in detail. Most of fingerprint extraction algorithms focus only on extraction of robust, compact and discriminant fingerprint. But the search time in the matching process is serious concerns rather than the fingerprint extraction time. So proposed TIRI-DCT method along with fast search algorithm outperforms than 3D-DCT since it is more robust, discriminant, and fast. Also 3D-DCT is resistant to different types of distortions like noise, brightness, contrast, rotation, time shift, spatial shift, and frame loss. These problems are eliminated in TIRI-DCT[9][10].

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

Proposed TIRI-DCT Algorithm

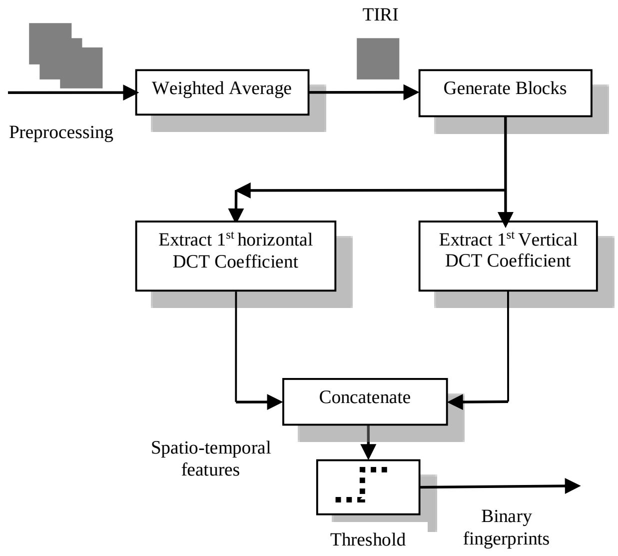

Figure. 1.4.1 Schematic of the TIRI-DCT Algorithm

Figure 1.4.1 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 video signal is processed. Preprocessing is often used to improve visual quality and coding efficiency of video compression systems. 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 can increase the robustness of a fingerprinting algorithm to these changes. In signal processing, down sampling or sub- sampling is the process of reducing sampling rate of the signal. This is usually done to reduce the data rate or the size of data. Normally down sampling factor is denoted by M[6]. M is an integer or a rational fraction greater than unity. This factor multiplies sampling time or equivalently, divides sampling rate. Since down sampling reduces the sampling rate, we must be careful to make sure the Shannon-Nyquist sampling theorem criterion is maintained. If the sampling theorem is not satisfied then resulting digital signal will have aliasing. Aliasing is an effect that causes different signals to become indistinguishable when samples. Aliasing is also known as distortion or artifact that results when the signal is reconstructed from the samples is different from the original continuous signal. Aliasing may be a temporal type or spatial aliasing. To ensure that the sampling theorem is satisfied, a low pass filter is used as an anti-aliasing filter to reduce the bandwidth of the signal before the signal is down sampled; the overall process (low pass filter then down sample) is called as decimation.Anti-aliasing means removing signal components that have a higher frequency than is able to be properly resolved by recording (or sampling) device. This removal is done before (re) sampling at a lower resolution[8]. When sampling is performed without removing this part of the signal, it causes undesirable artifacts such as black and white noise. If the original signal had been bandwidth limited, and then first sampled at a rate higher than Nyquist minimum, then the down sampled signal may already be Nyquist compliant, so the down sampling can be done directly without any additional filtering. Down sampling only changes the sample rate not the bandwidth of the signal.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.

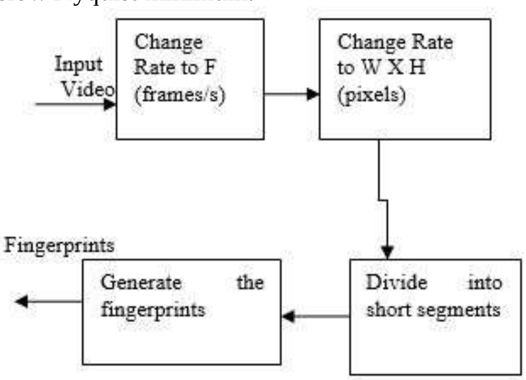

Figure 1.4.2 Preprocessing Steps

As shown in figure 1.4.2, each video is down-sampled both in time and space. 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 × H) pixels and fixed rate (F frames/second). Gaussian filter is a low pass filter, attenuating high frequency signals. Gaussian smoothing is used as preprocessing stage in computer vision algorithms in order to enhance image structures at different scales.After preprocessing, the video frames are divided into

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

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 transforms algorithm capture of the temporal information in a video using the same feature extraction process. Features are derived by applying a 2D-DCT on overlapping blocks of size from each TIRI[7]. As shown in figure 1.4.1 the first horizontal and the first vertical Discrete Cosine Transform (DCT) coefficients (features) are extracted from each block. The value of the features from all the blocks is concatenated to form the feature vector. Each feature is then compared to a threshold (which is the median value of the feature vector) and a binary fingerprint is generated. TIRI-DCT algorithm includes the 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 $2w \times 2w$, using

$$B^{i,j} = \{l'_{x,y} \mid x \in iw \pm w, y \in jw \pm w\}$$

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

When indexes are outside of the 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 $v = [\cos(0.5\pi/2w), \cos(1.5\pi/2w), \dots\dots\dots \cos(1 - 0.5\pi/2w)]^T$

And 1 is a 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

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

Advantages of proposed system are as follows.

• A Spatio-temporal fingerprint is adopted because of their comprehensiveness.

• The TIRI-DCT method introduces a fingerprinting system that is robust, discriminant, and fast.

• The TIRI-DCT outperforms the well-established (3D-DCT) algorithm and maintains a good performance for different attacks like noise, time shift, spatialshift, brightness/contrast, rotation, frame loss that normally occurs on video signals.

• TIRI –DCT algorithm is applied to color videos as well as black and white videos.

Conclusion

The system consist of a fingerprint extraction algorithm followed by an approximate search method.The proposed fingerprint algorthim (TIRI-DCT) extracts robust,discriminant and compact fingerprint from video in a fast and reliable fashion

The use of the inverted file index for video copy detection can tolerate changes in resolution of the video stream and changes in the frame rate.

Acknowledgements

This research has been conducted with support of the Anil Chhangani for help with discussion, selecting algorithm.

References

1. *Ahmet Saracoglu, Ersin Esen, Tugrul K. Ateş, Banu Oskay Acar,Unal Zubari, Ezgi C. Ozan,Egemen Ozalp,A. Aydn Alatan, Tolga Çiloglu, "Content Based Copy Detection with Coarse Audio-Visual Fingerprints" , inSeventh International Workshop on ContentBased Multimedia Indexing 2009*
2. *B.Coskun,B.Sankur, and N. Memon,"Spatiotemporal transform based video hashing" ,IEEE Trans. Multimedia, vol. 8, no. 6, Dec. 2006*
3. *R. Radhakrishnan and C. Bauer, "Robust video fingerprints based on Subspace embedding",in Proc. ICASSP, Apr. 2008*
4. *G.Willems, T. Tuytelaars, and L. Van Gool, "Spatio-temporal features for robust*

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

content-based video copy detection", in Proc. ACMInt. Conf.Multimedia Information Retrieval, New York, NY, 2008, ACM.

5. *A. Joly, O. Buisson, and C. Frelicot, "Content- based copy retrieval using distortion-basedprobabilistic similarity search",IEEE Trans.Multimedia, vol. 9, no. 2, Feb. 2007.*
6. *R. Mohan, "Video sequence matching",in Proceedings of the International Conference on Audio, Speech and Signal Processing, IEEE Signal Processing Society, 1998.*
7. *J. Oostveen, T. Kalker, and J. Haitsma, "Feature extraction and a database strategy for video fingerprinting",in Proc. Int. Conf. Recent Advancesin Visual InformationSystems (VISUAL), London, U.K., 2002, Springer-Verlag.*
8. *Mani Malek Esmaeili, Mehrdad Fatourechi, and Rabab Kreidieh Ward, "A 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. *Chen, L. and Stentiford, F. W. M., Video sequence matching based on temporal ordinal measurement" ,Pattern Recognition Letters, vol. 29, no. 13, Oct. 2008.*
10. *Maguelonne H´eritier, Vishwa Gupta, Langis Gagnon, Gilles Boulianne, SamuelFoucher, Patrick Cardinal "CRIM'S Content Based Copy Detection System forTRECVID", in 2009*