Efficient and Scalable Content- based Video Copy Detection System

M.Ramya
PG Student, Dept. of CSE
GCE-Tirunelveli

R.Kanthvel
Department of CSE
GCE-Tirunelveli

ABSTRACT
A video copy detection system is an emerging research area that has received a considerable amount of attention in recent years. The main goal of content based video copy detection system is to find whether a query video is copied from video in a video database or not. This paper uses strengths of TIRI-DCT algorithm, content based features for finger-print generation of a particular video, and two fast search methods for efficient match of finger-prints within a large database. The contribution of this paper include, extracting compact content-based signatures from TIRI image constructed from the video. To detect query video is pirated video or not, the finger-prints 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 finger-prints of the query video. The proposed system can be used for video indexing and copyright applications.

Index Terms
Content based finger-printing, multimedia duplicate detection, multimedia finger-printing, video copy detection, video copy retrieval.

1. INTRODUCTION
Today, video finger-printing is an essential and enabling tool adopted by the industry for video content identification and management in online video distribution. Tens of thousands of videos are being uploaded to the internet and shared every day. A considerable number of these videos are illegal copies or manipulated versions of existing media, making copyright management on the internet a complicated process. As video is the most complex type of digital media, it has so far received the least attention regarding copyright management. Because videos are available in different formats, it is more efficient to base the copy detection process on the content of the video rather than its name, description, or binary representation. Multimedia finger-printing also known as robust hashing has been recently proposed for this purpose [1].

A finger-print is a content-based signature derived from a video or other form of a multimedia asset. So, that it specifically represents the video or asset. Extracting video finger-prints, which are a set of short feature vectors that are unique to video clips can be used for video identification. Video finger-prints based on visual attention regions [14] which remain the same for the perceptually same scenes with different types of distortions and also different for different scenes.

To find a copy of a query video in a video database, one can search for a close match of its finger-print in the corresponding finger-print database extracted from the videos in the database. Closeness of two finger-prints represents a similarity between the corresponding videos; two perceptually different videos should have different finger-prints.

This paper is organized into 4 parts. In part 2 existing approaches in video copy detection system and its limitations are discussed. In part 3 the details of proposed system and its functionalities are explained. In part 4 experimental results are presented. In part 5, we conclude with the discussion of future work.

1.1 Video Finger-Printing:
A video finger-print is an identifier that is extracted from a piece of video content. The process of extracting a finger-print from the video content is referred to as finger-printing the video or video finger-printing. There is an obvious analogy to human finger-print and finger-printing. Just like human finger-print that can uniquely identify a human being, video finger-print can uniquely identify a piece of video content. The analogy extends to the process of subject identification by finger-print. First, known finger-prints must be stored in a database, and then a subjected finger-print is queried against the database for match.

Video finger-printing and finger-print based video identification are commonly known as video copy detection or more generally Content-Based Copy Detection (CBCD). Copy detection is the application that motivated the development of video finger-printing [10].

2. EXISTING APPROACHES
The existing approaches for video copy detection. They are,

- Watermarking Approach
- Image based approach to video copy detection

2.1 Watermarking Approach
Water marking relies on inserting information into the video stream in order to detect copies. Copy detection for digital media is critical to preventing copyright violations and enforcing copy right. Watermarking [3] is the most widely accepted form of copy detection. This approach utilizes 3-D DCT algorithm.

The drawback of 3-D DCT algorithm is its binarization phase. To obtain common threshold for the binarization phase is far from optimal, because different frames are having different

frequencies. However, there are two significant limitations in water marking approach. They are,

Legacy Content
Attacks

LEGACY CONTENT: Since watermarks must be introduced into the original content before copies are made, it cannot be applied to content which is already in circulation. For example, watermarking approach would not provide a solution to find all clips of star wars posted on the web.
ATTACKS: The watermark on a particular piece of content is compromised, there is no alternative approach to copyright enforcement.

2.2 Image Based Approach To Video Copy Detection

This approach introduces a video copy detection system which efficiently matches individual frames and then verifies their spatio-temporal consistency [7]. The approach for matching frames relies on a local feature indexing method, which is at the same time robust to significant video transformations and efficient in terms of memory usage and computation time. It matches either key frames or uniformly sampled frames.

2.2.1 Key Frames

Key frames are characteristic frames of the video. Key frames are used to obtain similar frames for two matching videos. To extract key frames, first detect shot boundaries by measuring gray level changes on a spatio-temporal slice of the video, and thresholding them [3]. The video hashing technique [15] is applied to every frame of a video sequence.

2.2.2 Limitations

The image based approach to video copy detection system efficiently matches individual frames and then verifies their spatio-temporal consistency. The approach for matching frames relies on a recent local feature indexing method, which is at the same time robust to significant video transformations and efficient in terms of memory usage and computation time. This image based approach matches either key frames or uniformly sampled frames. To further improve the results, a verification step has been taken to robustly estimates a spatio-temporal model between the query video and the potentially corresponding video segments. The video matching is based on individual key frames, so the searching time is very low. The image based approach is not tolerable to all the attacks on video signals such as, changes in brightness/contrast, rotation, frame loss, noise addition and spatial/temporal shift.

3. PROPOSED CONTENT BASED VIDEO COPY DETECTION SYSTEM (CBCD)

CBCD is a complimentary approach to watermarking technique. Because watermarks must be introduced into the original content before copies are made, it cannot be applied to content which is already in circulation. CBCD provides solution to the problem of legacy content and attacks [10]. Suppose the watermark on a particular piece of content is compromised there is no alternative approach to copy-right enforcement.

The primary thesis of content based video copy detection system is to detect the copies of video clips to find whether it is a pirated copy or not. The proposed video copy detection system based on content based finger-printing can be used for video indexing and copyright applications. The system depends on a finger-print extraction algorithm followed by a fast approximate search algorithm.

The finger-print extraction algorithm extracts compact content- based signatures from special images constructed from the video. Each such image represents a short segment of the video and contains temporal as well as spatial information about the video segment. These images are denoted as temporally informative representative images (TIRI).

TIRI is a blurred image that contains all possible motions in a particular video sequence. Each TIRI image represents a short segment of the video and contains spatial and temporal information about the video segment. The TIRI image plays a vital role in generation of finger-print for any input video because of its compactness [13]. The TIRI image is generated by extracting the features from all the short segments of a particular video.

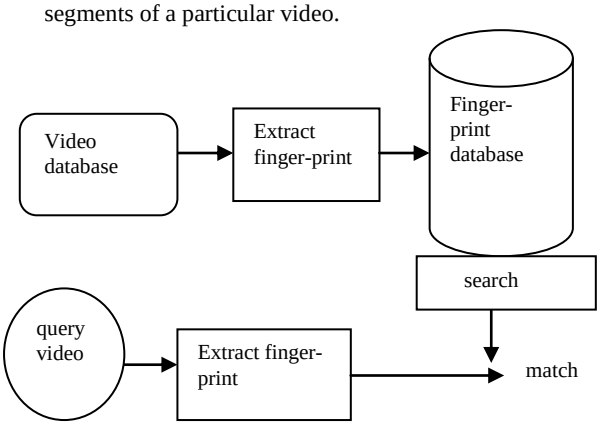

Figure 1: Schematic representation of video copy detection system

The fig 1 shows, all the finger-prints are stored in advance, If a query video is given, extract finger-print for that particular query video then searches all the finger-prints in a finger-print database .If a match is found the query video is a pirated video, otherwise it is not a copied video. The proposed system consists of following steps.

- Extracting robust and discriminant finger-prints.
- Fast matching of finger-prints within a large database.

3.1 Extracting Robust And Discriminant Finger-Prints

In this module, generate the finger-prints based on TIRI image. Before extracting the finger-prints, first pre-process the video signals. Because copies of the same video with different frame sizes and frame rates usually exist in the same video

database. The finger-printing algorithm should be robust to changes in the frame size as well as the frame rate. Therefore down-sampling can be applied to increase the robustness of a finger-printing algorithm to these changes of frame rate and frame size. The extraction of finger-prints includes the following steps.

3.1.1 Pre-processing
3.1.2 Short segment formation
3.1.3TIRI image generation
3.1.4 Finger-print generation

3.1.1 Pre-Processing

In this step, first select the video from the video database, and then split the particular video into frames. After splitting the video into frames, convert the frames into Luminance and Chrominance Valued (LUV) frames.

3.1.2 Short Segment Formation

After pre-processing, the video frames are divided into overlapping segments of fixed-length, each containing j frames. Then the finger-printing algorithm is applied to these segments. The amount of overlapping is experimentally chosen to be 50%. Overlapping reduces the sensitivity of the finger-prints to the "synchronization problem" which is referred as "time shift".

3.1.3 Tiri Image Generation

For TIRI image generation, spaio-temporal finger-printing method [2] is adopted. This method calculates a weighted average of the frames to generate a Temporally Informative Representative Image (TIRI). The TIRI image is a special image that plays best to capture the temporal information in a video.

The resulting TIRI image is basically a blurred image that contains information about all possible existing motions in a video sequence. The TIRI image is generated as follows.

Let $l_{m,n,k}$ be the luminance value of the (m,n) the pixel of the k_{th} frame in a set of frames. The pixels of TIRI are then obtained as a weighted sum of the frames.

$$\mathbf{l_{m,n}=\sum_{k=1}^{j} w_k l_{m,n,k}}$$

$l_{m,n,k}$ -luminance value of the (m,n)th pixel of the k_{th} frame in a set of j frames

w_k - exponential weighting function($w_k = \gamma_k$)

γ- values describes the performance of the fingerprinting system.Then examine different weight factors (constant, linear, and exponential) and observed that exponential weighting function generate images that best capture the motion.

PROPOSED TIRI-DCT ALGORITHM

By using proposed TIRI-DCT algorithm, features are derived by applying a 2D-DCT on overlapping blocks of size from each TIRI with 50% overlap. The first horizontal and the first vertical DCT coefficients (features) are then extracted from each block.

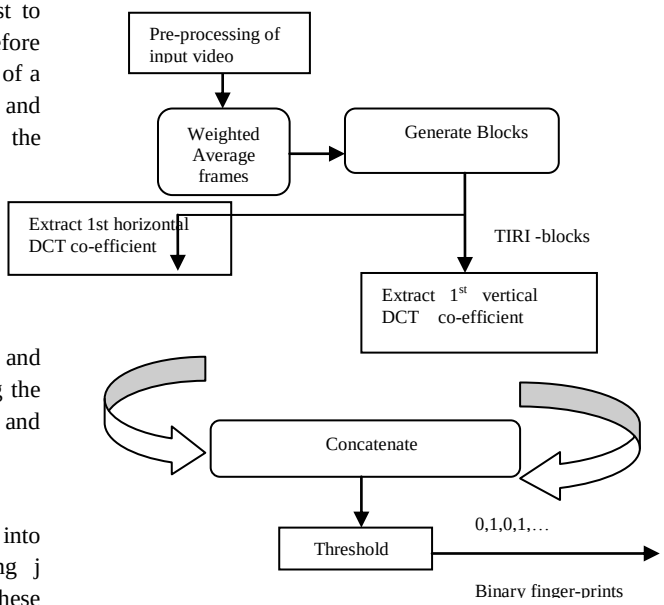

Figure 2: Schematic representation of proposed TIRI-DCT algorithm

The figure 2 shows pre-processing step is applied to weighted average of the frames. Then generate TIRI-blocks and the first horizontal and vertical DCT co-efficient are extracted from each block. The values of the features from all the blocks are concatenated to form the feature vector. Each feature is then compared to a threshold and binary fingerprint is generated.

3.1.4 Finger-Print Generation

The features are derived from the TIRI image, and the finger-print is generated by using TIRI-DCT algorithm .The resulting finger-print is just a string of bits that represents the signatures of the video data.

In order to determine whether a query video is an attacked version of a video in a database or not, its fingerprint is first extracted. The fingerprint database is then searched for the closest fingerprint to the extracted query fingerprint.

It should be mentioned that in copy detection, the problem is to determine if a specific query video is a pirated version of a video in the database. If the finger-print match is found, the query video is a pirated or copied video , otherwise it is not a copied video.

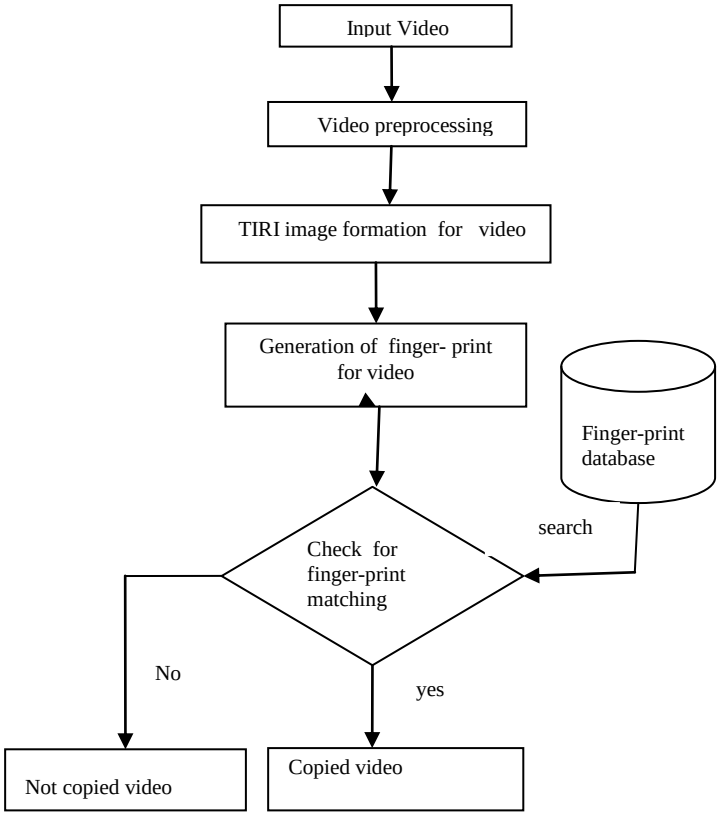

Figure 3: Flow chart for finger-print extraction of video

The above flowchart for finger-print generation of video shows that, the generated finger-print for any input video is matched against the finger-print database. If the match is found, the query video is copied video. Otherwise it is not a copied video.

3.2 FAST MATCHING OF FINGER-PRINTS WITHIN A LARGE DATABASE

After the generation of finger-prints for a particular query video, to examine whether a query video is an attacked version of video in a database or not, The fast matching of finger-prints within a large database by using two fast search methods have been done.

4. EXPERIMENTAL RESULTS

This section demonstrate some experimental results for the generation of finger-print for any particular query video.Any number of videos in AVI format can be examined based on the features extracted from TIRI Image and as a result, generate the finger-print for any particular query video.

Figure 4: Query Video in AVI format

The figure 4 shows that choosing the video from the video collections, the video must be in avi format, and it should be a trained video.

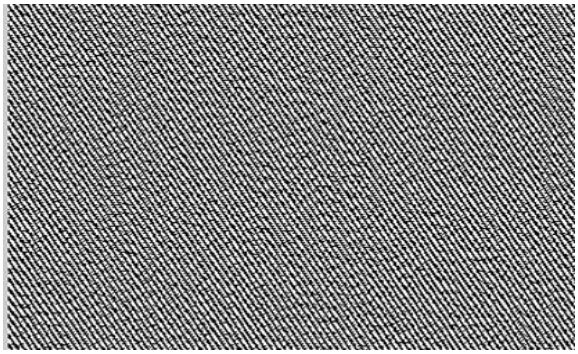

Figure 5: Finger-print generation for query video

The figure 5 shows that the features are derived from the TIRI image, with the derived features the finger-print is generated by using proposed TIRI-DCT algorithm. The resulting finger-print is just a string of bits that represents the signatures of the video data.

5. FUTURE WORK AND CONCLUSION

The proposed video copy detection system utilizes simple but efficient TIRI-DCT algorithm for generating a finger-print of any input video. The generated finger-print for a particular video is robust, discriminant, compact and tolerable for different attacks on video signals, including noise addition, changes in brightness/contrast, rotation, spatial/temporal shift and frame loss.

After the generation of finger-prints, the future work is to be efficient search of finger-print matching against the video database by using fast search methods to find whether an input or query video is pirated or not, in a fast and reliable fashion. The part of future work includes an extensive comparison of the fingerprinting algorithms to other state of the algorithms. To evaluate the system performance, the proposed fast search methods is applied to other fingerprinting methods in the presence of some other attacks, such as cropping, and logo insertion.

6. REFERENCES

[1] 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., pp. 117–128, 2002.

[2] C. Kim and B. Vasudev, "Spatiotemporal sequence matching for efficient video copy detection," IEEE Trans. Circuits Syst. Video Technol. vol. 15, no. 1, pp. 127–132, Jan. 2005.

[3] C. De Roover, C. De Vleeschouwer, F. Lefebvre, and B. Macq, "Robust video hashing based on radial projections of key frames," IEEE Trans. Signal Process., vol. 53, no. 10, pp. 4020–4037, Oct. 2005.

[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.

[5]B. Coskun and N. Memon, "Confusion/diffusion capabilities of some robust hash functions," in Proc. Conf. Information Sciences and Systems (CISS), pp. 1188–1193, Mar. 2006

[6] A. Joly, O. Buisson, and C. Frelicot, "Content-based copy retrieval using distortion-based probabilistic similarity search," IEEE Trans. Multimedia, vol. 9, no. 2, pp. 293–306, Feb. 2007.

[7] R. Radha krishnan and C. Bauer, "Content-based video signatures based on projections of difference images," in Proc. MMSP , pp. 341–344, Oct.2007.

[8] G.Willems, T. Tuytelaars, and L. Van Gool, "Spatio-temporal features for robust content-based video copy detection," in Proc. ACMInt. Conf. Multimedia Information Retrieval, in NEW YORK, pp. 283–290, 2008.

[9] L. Chen and F. W. M. Stentiford, "Video sequence matching based on temporal ordinal measurement," Pattern Recogn. Lett., vol. 29, no. 13, pp. 1824–1831, 2008.

[10] S. Lee and C. Yoo, "Robust video fingerprinting for content-based video identification," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 7, pp. 983–988, Jul. 2008.

[11] A. L. Varna and M. Wu, "Modeling and analysis of content identification," in Proc. ICME, pp. 1528–1531, 2009.

[12] J. Lu, "Video fingerprinting for copy identification: From research to industry applications," E. J. Delp, J. Dittmann, N. D. Memon, and P. W. Wong, Eds., SPIE vol.7254,no.1, pp.725402, 2009

[13] M. Malekesmaeili, M. Fatourechi, and R. K.Ward, "Video copy detection using temporally informative representative images," in Proc. Int.Conf. Machine Learning and Applications, pp. 69–74, Dec. 2009.

[14] X. Su, T. Huang, and W. Gao, "Robust video fingerprinting based on visual attention regions," in Proc. ICASSP, Washington, IEEE Computer Society, pp.1525–1528, Dec. 2009.

[15] M. Malekesmaeili and R. K. Ward, "Robust video hashing based on temporally informative representative images," in Proc. IEEE Int. Conf. Consumer Electronics, pp. 179–180, Jan. 2010.