Research Journal of Applied Sciences, Engineering and Technology 4(24): 5391-5396, 2012
ISSN: 2040-7467

Submitted: March 18, 2012 Accepted: June 15, 2012 Published: December 15, 2012

Copyright Detection System for Videos Using TIRI-DCT Algorithm

S. Nirmal, S. Shivsankar, S.R. Vignesh and Rajiv Vincent
Department of Computer Science and Engineering, Easwari Engineering College, Chennai, India

Abstract: The copyright detection system is used to detect whether a video is copyrighted or not by extracting the features or fingerprints of a video and matching them with fingerprints other videos. The system is mainly used for copyright applications of multimedia content. The copyright detection system depends on an algorithm to extract fingerprints which is the *TIRI-DCT Algorithm* of a video followed by an approximate search algorithm which is the *Inverted File Based Similarity Search.* To find whether a video is copyrighted or not, the query video is taken and the feature values of the video are extracted using the fingerprint extraction algorithm, it extracts feature values from special images called frames constructed from the video. Each frame represents a part or a segment of the video and contains both temporal and spatial information of the video segment. These images are called Temporally Informative Representative Images (TIRI). The fingerprints of all the videos in the database are extracted and stored in advance. The approximate search algorithm searches the fingerprints which is stored in the database and produces the closest matches to the fingerprint of the query video and based on the match the query video is found whether it is a copyrighted video or not.

Keywords: Binarization, DCT coefficients, filter mask, representative images, temporally informative threshold, words

INTRODUCTION

The number of videos that are being shared and uploaded everyday on the internet is uncountable. Some of these videos are illegal copies or modified versions of the existing videos. This makes copyright management a tedious and difficult process. A quick and accurate copyright detection algorithm of videos is required. Videos are a complex type of media and are available in many formats, it is better to base the detection system on the content and features of the video rather than its name or description. A fingerprint is a feature derived from the video which uniquely identifies the video. Finding the copy of the query video in a database is done by searching for a close match of the video's fingerprint in the corresponding fingerprint database. If the fingerprints are similar then it represents the similarity between the corresponding videos. Two different videos should have totally different fingerprints.

A fingerprint should be differentiable, simple to compute, concise and easy to search in the database. The fingerprint should be robust and should change as little as possible to content-preserving attacks such as changes in brightness or contrast of the video, logo insertion, rotation, compression, cropping etc. The fingerprints should be differentiable, to ensure that two different videos are distinguishable. They should be simple to compute because in online applications, an algorithm should extract the features as the video is being uploaded. An algorithm with many computations is not essential. Fingerprints should be concise because finding its match in a database would not be time consuming. The basic working of the system is that the query video is taken and its fingerprint is extracted and searched for a similar fingerprint from the fingerprint database then compared to obtain a match. If a match is obtained then the user is notified that his video cannot be uploaded, else the video is considered to be copyrighted and stored in the database.

Fingerprint extracting algorithms is divided into four groups based on the features they extract: Color-space-based, temporal, spatial and spatio-temporal. Color-space-based fingerprints are derived mostly from the histograms of colors in specific regions of time or in the video. The color-space-based does not apply for black and white videos (Hampapur and Bolle, 2001). Temporal fingerprints are features extracted from a video based on the time (Chen and Stentiford, 2008). It does not work for short video sequences because it does not have sufficient temporal information. Spatial fingerprints are features derived from a single frame or a key frame of the video. Spatial fingerprints may depict the global properties of a frame or a section of it (such as histograms of its contents) or local properties

Corresponding Author: S. Nirmal, Department of Computer Science and Engineering, Easwari Engineering College, Chennai, India

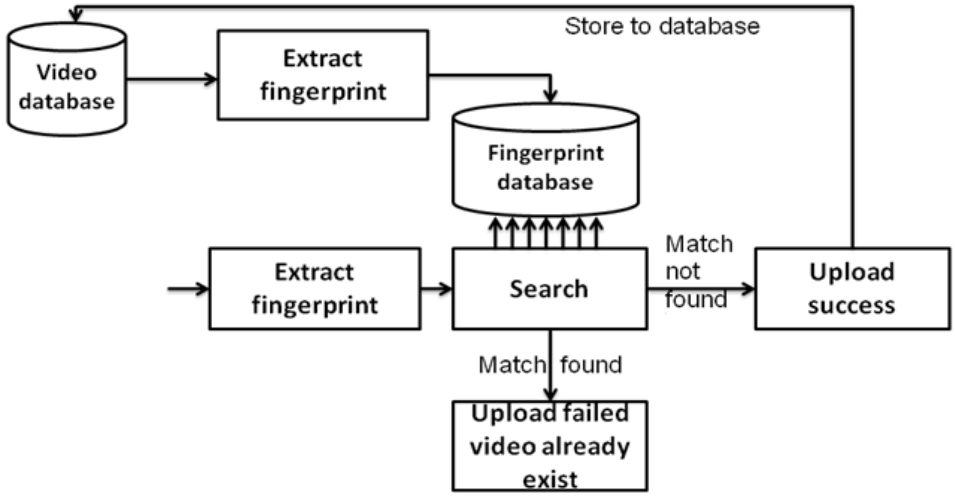

Fig. 1: Copyright detection system

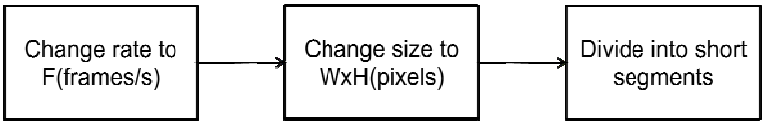

Fig. 2: Pre-processing of the video

which depict some interest points within the frame (edges, corners). Spatial fingerprints do not contain the temporal information of the video, which is an essential differentiating factor. Spatio-temporal fingerprints are used for copyright detection because they contain both spatial and temporal information of a video sequence.

The fingerprints extracted are of binary type. The extracted fingerprint contains all the necessary features that are needed to be compared with another fingerprint. When a video arrives the fingerprint is extracted based on the TIRI-DCT algorithm. Then the extracted fingerprint is compared with the existing fingerprints in the fingerprint database. If the fingerprint is available in the database then the video is declared as non-copyrighted video. If the fingerprint is not available then the extracted fingerprint is uploaded in the database for future use. Figure 1 shows the architecture of the copyright detection system.

METHODOLOGY

Pre-processing: Before extracting the fingerprints, Pre-processing of the video is performed. The video is converted from RGB to YUV color space to generate the features more accurately. The YUV color space is used to encode a color image or video taking human perception into account, the Y'UV model defines a color space in terms of one luminance (Y') and two chrominance (UV) components. The Y'UV colour model is used in the PAL and SECAM composite color video standards. There is a possibility of same video with different frame sizes and frame rates to usually exist in the same video database. Down-sampling the video can increase the effectiveness of the fingerprinting algorithm to the changes in frame rate and frame size (Coskun *et al*., 2006a, b). This process of down-sampling provides the fingerprinting algorithm with inputs of fixed size and fixed rate so that the same videos with the different frame size and frame rate can be detected. After pre-processing, the frames of the video are divided into overlapping segments of some fixed-length and each segment contains a set of frames. The fingerprinting algorithms are applied to these segments. Overlapping the segments reduces the sensitivity of the fingerprints to the time shifts in the video. Figure 2 shows the pre-processing steps of the video.

Extracting fingerprints: The method of extracting TIRIs is done by calculating the weighted average of the frames. The result is a blurred image containing information about the possible existing motions in a video.

TIRI-DCT algorithm: The TIRI is generated from the luminance values or intensity values of the pixels in a frame from a set of frames (Gonzalez, 2006). These values are obtained by performing filtering operations directly on pixels of a frame from the video. It consists of multiplying each pixel in the neighborhood by a

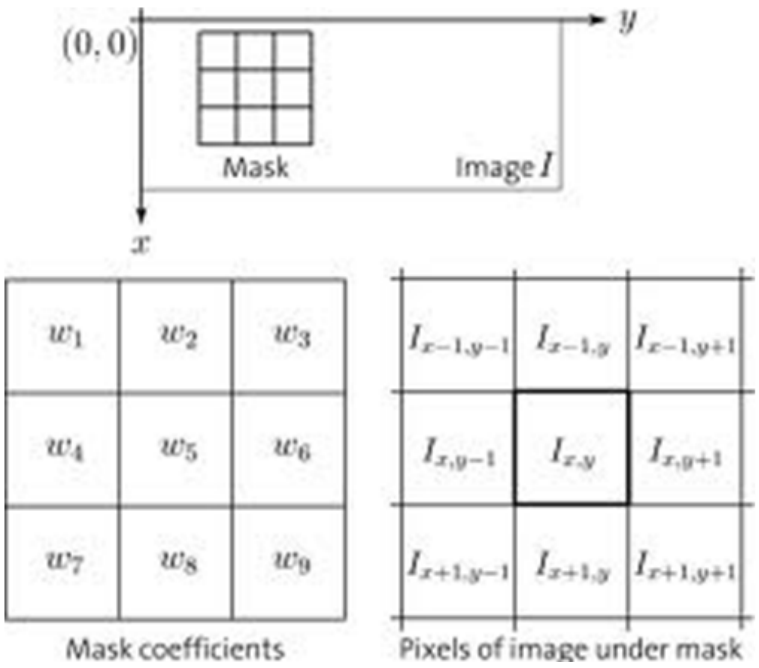

Fig. 3: Mechanism for extracting pixel intensity values

corresponding coefficient and summing the results to obtain each point (x, y). If the neighbourhood is of size Figure 3 illustrates the mechanics of the filtering process. It consists of moving the centre of the filter mask w_5 from point to point in image I. At each point (x, y), the response of the filter at that point is the sum of products of the filter coefficients and the corresponding neighbourhood pixels in the area spanned by the filter mask. For a mask of size m×n, assume that m = 2a+1 and where n = 2b+1 a and b are nonnegative integers. Figure 3 shows the method of extracting the pixels values.

The features are derived by applying 2D-DCT (Discrete Cosine Transform) on the overlapping blocks from each of the TIRI. DCT is used to transform a signal from spatial representation into a frequency representation. In a frame of the video, the energy will be concentrated in the lower frequencies, so if a frame is transformed into its frequency components and the higher frequency coefficients are thrown away, the amount of data needed to describe the frame can be reduced without sacrificing much of the quality of the video. The basic operation of 2D-DCT consists of an input frame m x n, let f (i, j) be the intensity of pixel in row I and column j, let F (u, v) is the DCT coefficient in row K1 and column K2 of the DCT matrix. For most frames or images, much of the signal energy lies at the low frequencies; these appear in the upper left corner of the DCT. Compression is achieved since the lower right values represent the higher frequencies and are often small enough to be neglected with little visible m×n, mn coefficients are required. The coefficients are arranged as a matrix, called filter, mask or window. distortion. The DCT input is an 8x8 array of integers. This array contains each pixel's gray scale level; 8 bit pixels have levels from 0 to 255. The Fig. 4 shows the TIRI blocks from which the DCT coefficients are derived. The DCT coefficients are divided into "DC coefficient" and "AC coefficients". DC coefficient is the coefficient with zero frequency in both dimensions and AC coefficients are remaining 63 coefficients with non-zero frequencies.

The first set of horizontal and vertical DCT coefficients (features) are extracted from each block. The values of features from all blocks are concatenated to form the feature vector. Each of the feature vectors is compared with a threshold. A threshold is of two types constant threshold and variable threshold. In case of variable threshold, the threshold T can be found by selecting an initial estimate. The set of frames are segmented based on T. This will produce two groups of pixels: R1, consisting of all pixels with intensity greater than T and R2, consisting of pixel values less than or equal to T. The average of the pixel intensities M1 and M2 in regions R1 and R2 is computed. The new threshold is computed as:

$$T = ½(M1 + M2)$$

The process is repeated until the difference in T in successive iterations is smaller than a predefined

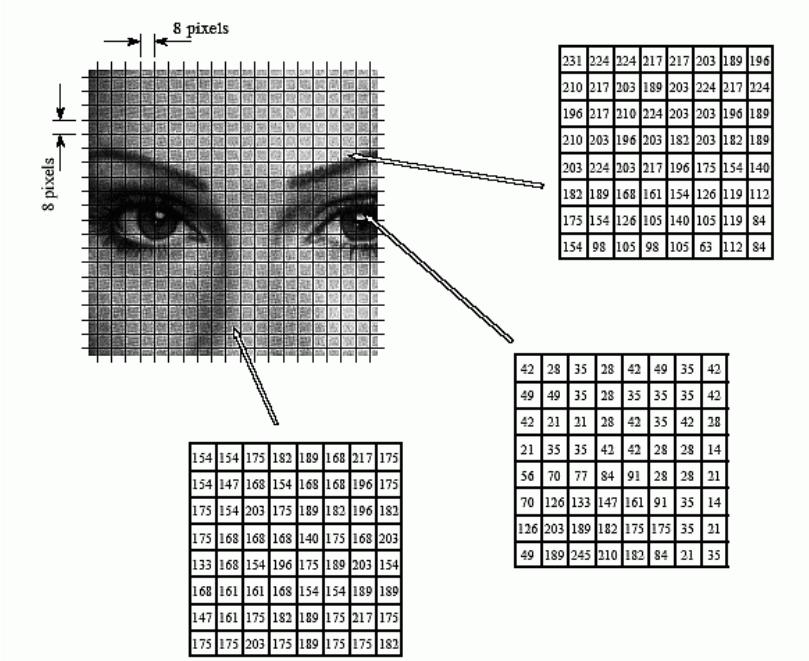

Fig. 4: TIRI blocks

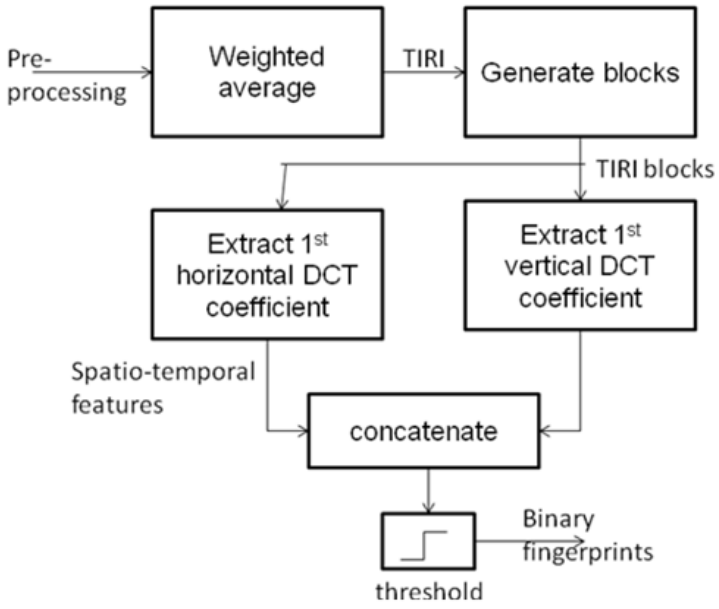

Fig. 5: The schematic diagram of the TIRI algorithm

value.of T. A binary fingerprint is then generated. Figure 5 shows the scheme of the TIRI algorithm

Matching of fingerprints within the database: In order to determine whether a query video is a modified one of a video in a database, the fingerprint of the query video is first extracted. The fingerprint database is searched for the closest fingerprint to the extracted query video's fingerprint. In copy detection, the task is to determine if a specific query video is a copied or modified version of a video database.

Inverted-file-based similarity search: The idea behind the method is, if two fingerprints are similar to be

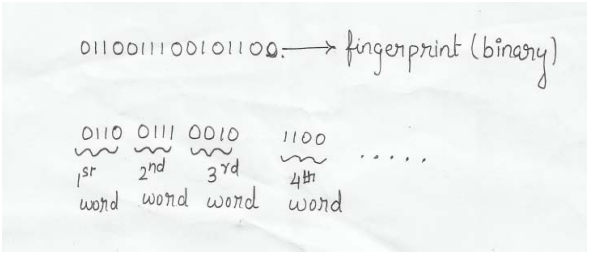

Fig. 6: Splitting fingerprints into words

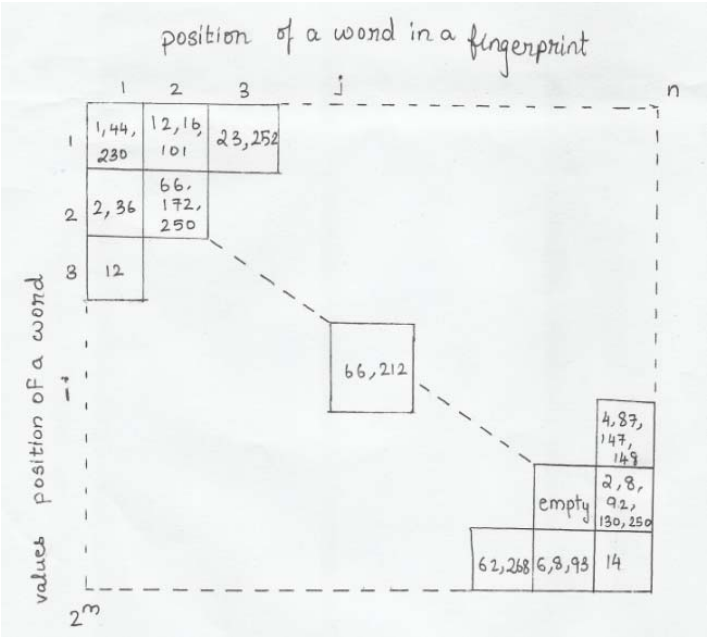

Fig. 7: Inverted-file-based similarity search

considered as matches then the probability of exact matches in the sub blocks of the fingerprint is high (Oostveen *et al.*, 2002). Here the fingerprint is divided into non-overlapping blocks. Each block maybe of m bits. These blocks are called Words. The words create an inverted file from t he fingerprint database. All the fingerprints have equal lengths, so inverted files can be as table of size $2^m \times n$, where n represents the number of words in a fingerprint. The horizontal part of the table indicates the position of the word inside a fingerprint. The vertical part of the table indicates directions to the corresponding value of words possible. The table is generated starting with the first word of each fingerprint then add an index of the fingerprint to the first column entry corresponding to the value of the word. The process is continued for all the words in each fingerprint and all the columns in the inverted file table.

To find the query video's fingerprint in the database, the fingerprint is first divided into n words (having m bits).The query is compared to all the fingerprints which start with the same word. The indices of the fingerprints are found from the entry in the first column of the inverted file table. The Hamming distance between these fingerprints and the query is then calculated. If the fingerprint has a hamming distance less than some predefined threshold, it is announced as a match. If no match is found, the procedure is repeated for the fingerprints that have the same second word as the query's second word. The procedure will be continued until the match is found or the last word is examined. Figure 6 and 7 shows the method of splitting the fingerprints to Words and the scheme of inverted-file-based similarity search algorithm.

RESULTS AND DISCUSSION

For getting accurate results, a performance comparison is made between TIRI-DCT and 3D-DCT. 3D-DCT unlike TIRI-DCT considers the video as a

three-dimensional matrix of luminance values and extracts 3D transform based features. The video is segmented into nine parts and the DCT coefficients for each part in the video are extracted. After pre-processing the low frequency coefficients of the transform are extracted and thresholding is done and the fingerprints are generated. The generated fingerprints have equal number of 0's and 1's. This increases the robustness of fingerprints, but decreases the discrimination. The drawback in 3D-DCT is that different coefficients have different ranges thus a common threshold for the process of binarizing the different frequencies is not optimal. This problem is not present in TIRI-DCT since all the features are in the same frequency range, binarization using a common threshold can be done.

Example:

5, 18, 70, 32, 56, 67, 45, 81, 97

The threshold for the sequence can be 56, which is a single number and thus a common threshold.
The binary fingerprint based on threshold will be as:

0 0 1 0 0 1 0 1 1

A common threshold is not possible in the case of 3D-DCT.

CONCLUSION

This study proposes a method for detecting duplicated or modified versions of copyright videos. This method is suitable for detecting videos with content-preserving changes such as changes in brightness or contrast of the video, logo insertion, rotation, cropping and compression. The another type of attack apart from the content-preserving attacks are content-changing attacks such as changing background or picture in picture etc. These changes cannot be detected using global fingerprints. Such changes can only be handled by using local fingerprints via interest point-based algorithms. Combining both global and local fingerprints can result in a complete copyright detection system.

REFERENCES

Chen, L. and F.W.M. Stentiford, 2008. Video sequence matching based on temporal ordinal measurement. Pattern Recogn. Lett., 29(13): 1824-1831.

Coskun and N. Memon, 2006a. Confusion/diffusion capabilities of some robust hash functions. In Proceeding of Conference Information Sciences and Systems (CISS), pp: 1188-1193.

Coskun, S.B. and N. Memon, 2006b. Spatiotemporal transform based video hashing. IEEE T. Multimedia, 8(6): 1190-1208.

Gonzalez, B. and M. Charbit, 2006. Digital Signal and Image Processing using Matlab. ISTE Ltd., London, pp: 763, ISBN: 1905209134.

Hampapur and R.M. Bolle, 2001. Video copy detection using inverted file indices IBM research division thomas. J. Watson Res. Center, Tech. Rep., pp: 9.

Oostveen, J., T. Kaller and J. Haitsma, 2002. Feature extraction and a database strategy for video fingerprinting. In proceedings of International Conference Recent Advances in Visual Information Systems (VISUAL), London, U.K., pp: 117-128.