# SIFT and DWT based Video Sequence Matching Method for Videocopy Detection

Kanchana Rani G
Department of CSE
SCT College of Engineering
Trivandrum

Chitharanjan K
Department of CSE
SCT College of Engineering
Trivandrum

## ABSTRACT

This paper introduces SIFT and DWT based video sequence matching method for video copy detection. Since local features have good stability and discriminative ability SIFT descriptors are used here for video content description. Content matching using SIFT takes large amount of time and it is computationally expensive for high dimensions and large number of points. These difficulties are solved by using dual threshold method which divides videos into segments having homogeneous content and by performing keyframe extraction on each of these segments. SIFT features are then extracted from these keyframes and SIFT feature sets of two video frames are matched using SVD based method. It has the problem of high processing time proportional to the length of video content. So we proposed DWT based fingerprint generation technique to reduce the processing time. Fingerprints of videos are generated and fingerprint matching is performed in the preprocessing step. So based on these results, it decides whether the SIFT feature matching has to be performed or not. Experimental results shows that SIFT and DWT based video sequence matching method for video copy detection can effectively detect video copies. Proposed system has following advantages such as, based on the spatial features it can effectively find optimal sequence matching result from the disordered matching results, it can effectively reduce the processing time and it is adaptive to video frame rate changes. Experimental results also demonstrate that the proposed method can obtain a better tradeoff between the effectiveness and the efficiency of video copy detection.

**Keywords-** Video copy Detection, Scale Invariant Feature Transform, Singular Value Decomposition, Discrete wavelet transform.

## 1. INTRODUCTION

Videos are the most effective means for communication. Each and every day thousands of videos are getting generated and published. Among these, most of the videos are near duplicates. According to the statistics of [1], there exist 27 percent redundant videos which are duplicates or near duplicates [7] of some popular videos in the search results from YouTube, Google video and Yahoo! Video search engines. So, efficient and effective methods are needed to detect the video duplication [7]. A valid video copy detection method is based on the concept of [2] that “video itself is watermark” and makes entire video content to detect copies.To facilitate the discussion of “video copy” this paper uses the definition of video copy in TRECVID 2008 tasks.

Definition of copy video: A video V1, by means of various transformations such as addition, deletion, modification (of aspect, color, contrast, encoding, and so on), camcording, and so on, is transformed into another video V2, then video V2 is called a copy of video V1. Ten transformations [3] are defined in content-based copy detection task of TRECVID 2008. These 10 transformations are as below, see [4] for detail. Table 1 shows five transformations that can be performed in a video.

T1. Cam-cording; T2. Picture in picture; T3. Insertions of pattern: Different patterns are inserted randomly: captions, subtitles, logo, sliding captions; T4. Strong re-encoding; T5. Change of gamma; T6, T7. Decrease in quality: Blur, change of gamma (T5), frame dropping, contrast, compression (T4), ratio, white noise; T8, T9. Post production: Crop, Shift, Contrast, caption (text insertion), flip (vertical mirroring), Insertion of pattern (T3), Picture in picture (the original video is in the background); T10. Combination of random five transformations among all the transformations described above.

**Table 1 : Video Content Transformations**

| Type | Example |
|---|---|
| T1-Cam-cording: Cam-coding is done by filming a movie on a screen. It can be done manually. | |
| T2-Picture in picture: In this transformation a video is inserted in another video, the special location and scale of the inserted video can be changed. | |
| T3-Insertion of patterns: Different patterns such as captions, subtitles, logo, sliding captions are inserted randomly. | |

| | |
|---|---|
| T4-Strong re-encoding: the bit rate is changed, video resolution is reduced, and the video can be also encoded with a different codec. | |
| T5-Change of gamma: Randomly changed the gamma value for each color. | |

A video copy can be generated by a number of transformations. So the objective of this video copy detection system is to check whether the query video is a copy of any of the video in the video database. If this copy detection system finds a copied content it has to return the name of copy video from the database. As in [5] Fig 1 shows the video copy detection system frame work which has two parts:

- Offline Step: This system process the reference video in this step. Divide the video into segments of homogeneous contents, extracts keyframes from each segment and performs feature extraction on each of the keywords. These features must be robust and effective to all possible transformations that can be done in a video.
- Online Step: Here the query video is analyzed and processed. Keyframe extraction is performed after the segmentation and then performs feature extraction. Resulting features are compared with that of reference videos and matching result is analyzed. Then returns the detection result.

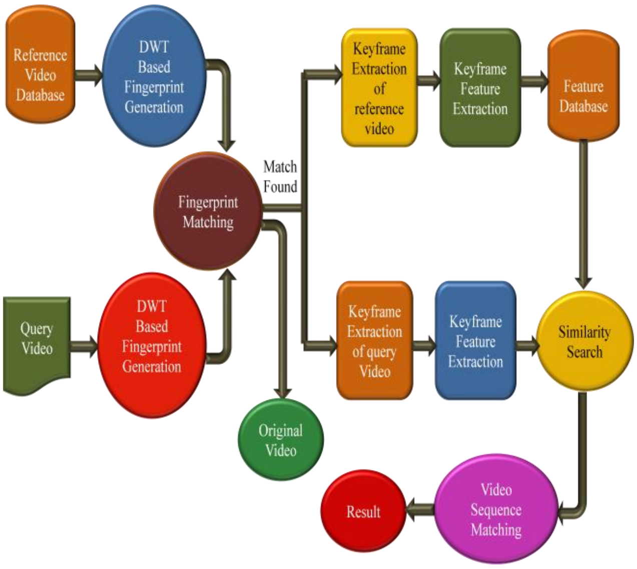
Reference Video Database
DWT Based Fingerprint Generation
Match Found
Fingerprint Matching
Keyframe Extraction of reference video
Keyframe Feature Extraction
Feature Database
Query Video
DWT Based Fingerprint Generation
Original Video
Keyframe Extraction of query Video
Keyframe Feature Extraction
Similarity Search
Result
Video Sequence Matching

**Fig 1 : Video Copy Detection System Framework**

## 2. RELATED WORK

As reviewed in [5] [6], many methods have been proposed in the area of video copy detection. By the comparative study it is clear that *copy* is the subset of *near duplication.* Copies have an origin where the near duplicate [7] may not. One type of the copy detection methods uses global descriptors where, others use local descriptors. Methods based on global descriptors are performed by using the low-level spatio-temporal features of the entire image. It requires only simple computation but the problem is that its performance is not satisfactory while detecting the duplicate videos which are generated by complicated transformations. Processing based on the local descriptors first detect the local spatio-temporal feature points, which are the keypoints and then used the content around those points. But these local descriptors have high computational cost. Based on [24] SIFT (Scale Invariant Feature Transform) feature points are very efficient to identify the objects in the video. It not only has good tolerance to image rotations, scale changes and illumination variations, but also is robust to additive noise, affine distortion [8] and change of viewpoints.

Video copy detection system based on SIFT is implemented to perform the following steps for both query video and the reference video. Steps include conversion of videos into segments of homogeneous content, conversion of segments into frames, identifying SIFT feature points from each of the frames and matching the feature points of both query video frames and reference video frames. The problem is that even if both videos are original, it is needed to perform all the above steps. SIFT feature extraction takes most of the processing time because of its number of processing steps such as scale space extrema detection, keypoint localization, orientation assignment and keypoint descriptor. In addition to that, keyframe extraction and feature point matching takes more time. The system with SIFT has to perform all these time consuming steps for all videos, since the system can only identify the possibility of matching in the last step. So the entire system has high processing time and computation cost. These problems can be avoided by the proposed system, named Discrete Wavelet Transform (DWT) [10] based Fingerprint Generation.

## 3. PROPOSED SYSTEM

Video copy detection system implemented based on the SIFT technology has certain problems such as high processing time and computational cost. These problems can be avoided by the proposed system which uses DWT based finger print generation technique [10]. This system has mainly two steps. One is the *Preprocessing* step and next is the *Copy detection step*. DWT fingerprint generation is implemented in the preprocessing step. DWT is any wavelet transform for which the wavelets are discretely sampled.

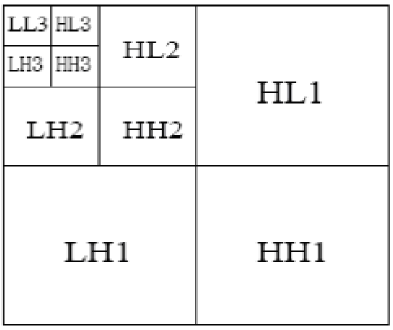

**Fig 2 : DWT transformed 2D Image**

In image processing, the basic idea of DWT is to decompose the image into sub-images of independent frequency district and different spatial domain. After performing the DWT transformation, the original image is decomposed into four frequency districts which are one low frequency district (LL) and three high frequency districts(HH,LH,HL). L and H represents low pass filter and high pass filter respectively. Then co-efficient of sub-images are transformed. Sub-level frequency district information is obtained by performing DWT transformation on low frequency district. Fig 2 shows the two dimensional image on which DWT transformation is performed three times. By performing DWT, original image is decomposed into four frequency districts, HH1, LH1, HL1, LL1. Then LL1 is again decomposed to get HH2, LH2, HL2, LL2. By repeating the process *n*- level decomposition can be performed on original image. The information of low frequency is an image which is close to the original image. The frequency districts of HH, LH and HL respectively represents diagonal detail, upright detail and level detail of the original image.

Haar wavelet [6] is the simplest possible wavelet. Technical difficulty of the Haar wavelet is , it is not continuous and therefore not differentiable. But this property becomes an advantage for signal analysis with sudden transactions. As in [6], Haar wavlet is implemented as follows.

$2^n$ numbers are the input of Haar, the paired up input values, stored the difference and passed the sum. This process is repeated recursively by pairing up the sums to get the next scale: finally resulting in $2^n - 1$ differences and one final sum.

The 2×2 Haar matrix that is associated with the Haar wavelet is $H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.Using DWT, one can transform input sequence $(a_0, a_1, \ldots, a_{2n}, a_{2n+1})$ of even length into a sequence of two-component-vectors $((a_0, a_1)\ldots(a_{2n}, a_{2n+1}))$ . By performing right-multiplication on each vector with the matrix $H_2$ , one gets the result $((s_0, d_0)\ldots(s_n, d_n))$ of one stage of the fast Haar-wavelet transform. The process is repeated by forming sequence of *s*. *s* is the averages part and d is the details part.

The 2N×2N Haar matrix is derived by the following equations.

$$H_{2N} = \begin{pmatrix} H_N & \otimes & \begin{bmatrix} 1 & 1 \end{bmatrix} \\ & & \\ I_N & \otimes & \begin{bmatrix} 1 & -1 \end{bmatrix} \end{pmatrix}$$

$$I_N = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 0 & \cdots & 1 \end{bmatrix}$$

$\otimes$ is the Kronecker product. The Kronecker product of $A \otimes B$, where $A$ is an m×n matrix and $B$ is a p×q matrix, is expressed as,

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

Above procedure is applied on both query video and reference video. Fig 3. Is a sample of resulting finger print. Implementation of this return fingerprints for each of the video. Then twelve basic features are extracted from both fingerprints and using these features fingerprints has compared. Since fingerprints related to the video content, if the matching value of fingerprints exceeds threshold value then it indicates the probability of video copying. Otherwise, after the preprocessing step the system display a message that both videos are original.

If the fingerprints are matching, then it is needed to perform following operations as in [5].

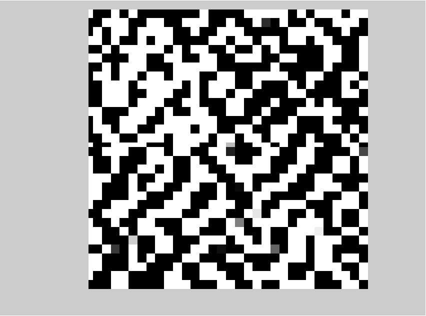

**Fig 3 : Image of DWT Fingerprint**

## 3.1 Frame Coversion

For the processing of a video, first it is needed to be converted into frames. So this module is designed to convert the input video files into frames. This frame conversion can be performed using video file reader and the resulting frames are needed to be stored in different folders, one for query video and other for reference video. Fig 4 shows the resulting frames.

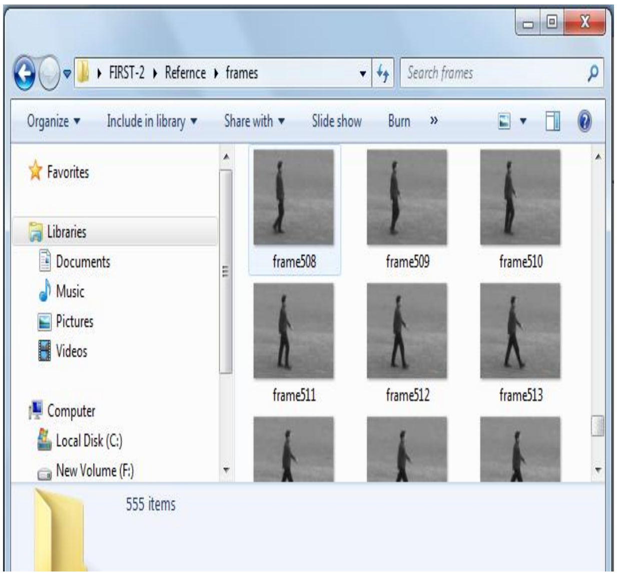

**Fig 4 : Video Frames**

## 3.2 Key Frame Extraction

Key frames have information which can explain most of the video content. As done in [5], *Auto dual threshold method* is used to eliminate the redundant frames. First it converts video into segments of frames having homogeneous content and then the first and last frames of each segment are selected as the key frames.

## 3.3 Feature Extraction

Here SIFT features are extracted from each of the keyframes. As in [5] SIFT feature extraction is implemented in four steps, which are Scale –Space extrema detection, Keypoint localization, Orientation assignment and plotting of keypoint descriptors. Fig 5 shows the feature points in keyframes of both query video and reference video.

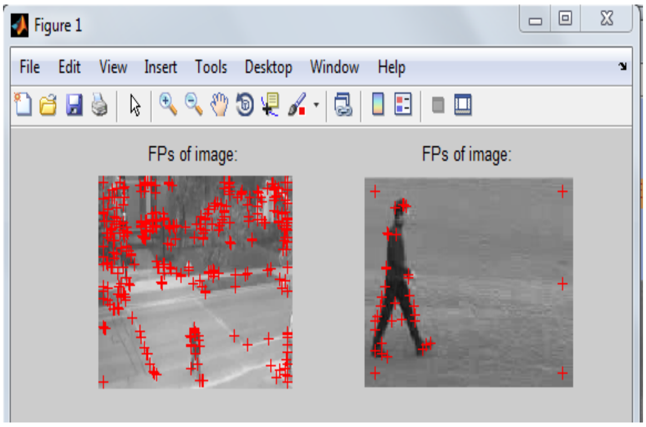

**Fig 5 : SIFT Feature points**

## 3.4 Feature Comparison

It performs matching of feature points of both videos and draw lines between matching feature points. SVD [Singular value Decomposition] technique [5] is used here to calculate the matching between points. Fig 6 shows the output of feature comparison step.

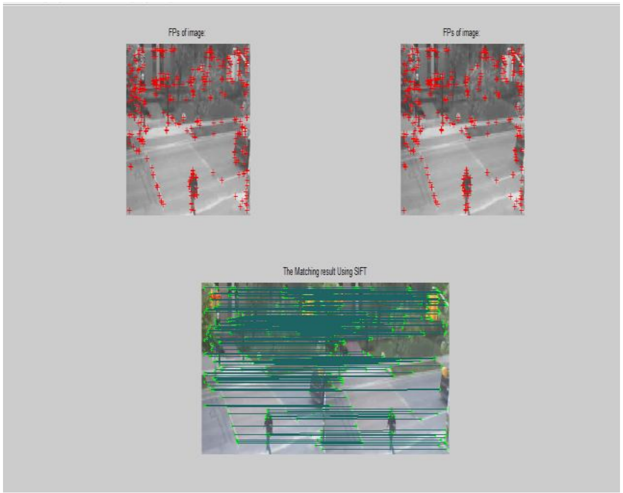

**Fig 6 : Matching feature points**

## 4. RESULT

Comparison of video copy detection systems based on SIFT and SIFT with DWT shows that later one is three times faster than the system based on SIFT. Comparison is performed by checking both the systems with maximum inputs. That is the videos which are generated by applying various possible transformations on an original video are given as input for both the systems. Then their processing time is calculated and average value is computed, which shows that the system based on SIFT with DWT is better than the previous system. This is because processing time and the computational cost of the proposed system is very low as compared to the existing system.

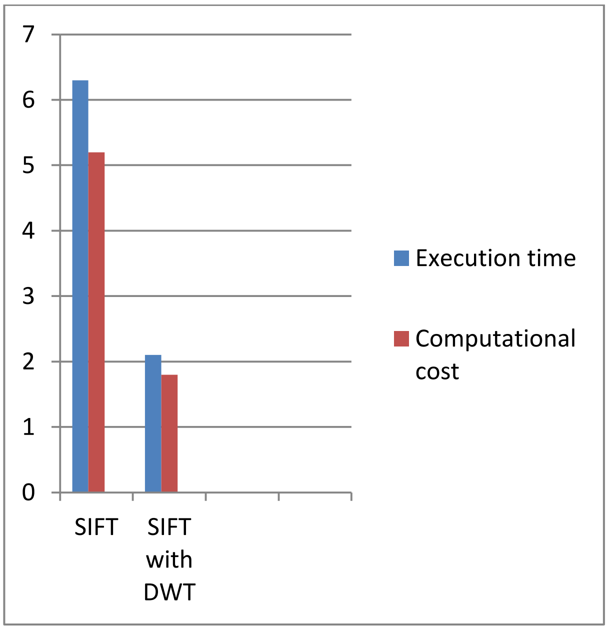

**Fig 7: Comparison between SIFT and SIFT with DWT based systems.**

## 5. CONCLUSION

Video copy detection system based on SIFT method has the problems of high computational cost and processing time. These problems are eliminated by combining DWT fingerprint generation technique with the SIFT. Comparison of video copy detection systems based on SIFT and SIFT with DWT shows that later one is three times faster than the system based on SIFT. This is because during the comparison of two videos, which are original ones, the system based on SIFT with DWT only takes time for the DWT fingerprint generation and there comparison from which it can decide the originality of the videos. Whereas the system based on SIFT has to perform the entire processing of the video. So, proposed system is better as compared to the existing system.

## 5. REFERENCES

[1] X. Wu, C.-W. Ngo, A. Hauptmann, and H.-K. Tan, "Real-Time Near-Duplicate Elimination for Web Video Search with Content and Context," IEEE Trans. Multimedia, vol. 11, no. 2, pp. 196-207, Feb. 2009.

[2] A. Hampapur and R. Bolle, "Comparison of Distance Measures for Video Copy Detection," Proc. IEEE Int'l Conf. Multimedia and Expo (ICME), pp. 188-192, 2001.

[3] TRECVID 2008 Final List of Transformations, http://www-Nlpir.nist.gov/projects/tv2008/active/copy. detection/final.cbcd. video.transformations.pdf, 2008.

[4] Final CBCD Evaluation Plan TRECVID 2008 (v1.3), http://www- nlpir.nist.gov/ projects/ tv2008/ Evaluation-cbcd- v1.3.htm,2008.

[5] Hong Liu, Hong Lu and Xiangyang Xue, "A Segmentation and Graph-Based Video Sequence Matching Method for Video Copy Detection." IEEE transactions on knowledge and data engineering, vol. 25, no. 8, August 2013.

[6] Wikipedia: www.wikipedia.com.

[7] Xiao Wu, Chong-Wah Ngo, Alexander G. Hauptmann, "Real-Time Near-Duplicate Elimination for Web Video Search With Content and Context," IEEE Transactions on Multimedia, vol. 11, no. 2, February 2009.

[8] 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.

[9] L. Liu, W. Lai, X.-S. Hua, and S.-Q. Yang, "Video histogram: a novel video signature for efficient web video duplicate detection," in *Proc.Multimedia Modeling Conf.*, Jan. 2007.

[10] Mei Jiansheng, Li Sukang and Tan Xiaomei "A Digital Watermarking Algorithm Based On DCT and DWT," Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA'09) *Nanchang, P. R. China, May 22-24, 2009, pp. 104-107*