AN ANALYSIS OF INDEPENDENCE OF VIDEO SIGNATURES BASED ON TOMOGRAPHY

Sebastian Possos, Adriana Garcia, Marilyne Mendolla, Jonathan Schwartz, and Hari Kalva
Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431

ABSTRACT

The current technological advances allow for the easy distribution and duplication of video content through various venues, the most popular being the World Wide Web. Upholding copyright laws has thus become an important task; the result to this problem needs to be implemented in such a way to provide automatic video identification. This paper presents a novel approach to video identification based on the video tomography technique. The proposed video signature was designed and evaluated based on its ability to uniquely identify videos. A video signature should be both independent and robust. This paper focuses on evaluating the independence of the proposed video signatures. The signatures were evaluated using a large test set developed for the MPEG activity on digital video signatures. The results show that the proposed signature works well and exhibits strong independence necessary to support general purpose video identification systems.

Index Terms— *video tomography, video signature, shot detection, independence*

1. INTRODUCTION

Due to the great technological developments for producing, processing, editing and copying digital multimedia data, video identification has become a hot topic for copyright holders and media distributors. The growing popularity of web sites like YouTube, where multimedia content is uploaded and broadcasted by anyone who has an Internet connection, has intensified the need for the tracing and controlling of video content and copyright for huge amounts of videos. This challenge gives birth to new areas of research like copy detection, multimedia indexing and multimedia content retrieval. MPEG has recently issued a call for proposals to standardize technology for digital video signatures [1].

The key requirements identified in [1] include: uniqueness, robustness, independence, fast matching, fast extraction, compactness, non-alteration, self-contained, and coding independence. Of these, the first three requirements affect the functionality of a video signature and the remaining requirements affect the implementation of video identification systems. Uniqueness and independence are related attributes and are necessary for the broad applicability of content. Robustness, on the other hand, affects the operating conditions for video identification (i.e. resolution changes, brightness changes, etc.). In this paper we focus on uniqueness and independence of video signatures and evaluate the signatures for these attributes.

The proposed approach to video signatures is based on video tomography. The signature has two components: a shot or segment level signature which is globally unique, and a frame signature that is locally unique. The combination of shot and frame signatures is used to identify video clips. We evaluated this signature for independence and reported on the performance.

The rest of the paper is organized as follows: first, the related work is discussed in section 2. Second, the new algorithm based on shot signatures is presented along with its implemented results. Finally, conclusions are drawn and future work is suggested.

2. RELATED WORK

2.1. Video Identification

There are two main ways to identify a video [2]: 1) Based on a digital watermark and 2) Based on the video content

2.1.1. Digital Watermarks

Digital watermarking relies on extracting an embedded watermark in a video in order to determine the video source. Watermarking was first proposed as a solution for identification and tamper detection in video and images by G. Doer et al [3]. However, they are not designed to identify unique clips from the same video source.

Two major drawbacks to digital watermarking are: the embedding of a robust watermark in the video source, and the large collection of "un-watermarked" files that already exist.

2.1.2. Video content based identification

Content based identification, on the other hand, uses the content of the video to compute a unique signature based on various video features. A content based video identification system survey is presented by X. Fang at el. [4] and by J. Law-To at el. [5].

A proposal for copy detection in streaming videos is presented by Y. Yan at el. [6], where a video sequence similarity measure is used, which is a composite of the frame fingerprints extracted for individual frames. Partial decoding of the incoming video is performed and the DC coefficients

for key frames are used to extract and compute frame features.

[7] and [8] are also based on key frame analysis: [7] proposes a clustering technique where the authors take key frames for each cluster of the query video and perform a key frame based search for similarity regions in the target videos. [8], on the other hand, uses local features; it extracts key frames to match against a database and then uses the local spatial-temporal features to match the videos.

As can be inferred from the above, many of the content based video identification methods use video signatures generated from individual frame content. This adds incredible overhead and complexity, especially in long videos, as they require feature extraction and comparison on a frame basis.

3. SIGNATURE DESIGN

The video signature for the proposed video identification system is based on video tomography.

3.1. Video Tomography

Video tomography was first presented as way to extract lens zoom, camera pan and camera tilt information using modified motion analysis [9], by introducing tomographic techniques into a motion estimation algorithm. The images generated from this method resemble the flow patterns of ridges in human fingerprints, hence the idea of using them as an identification method.

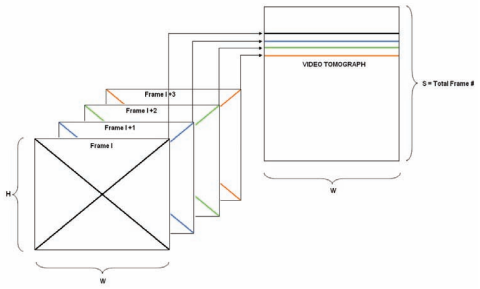

Figure 1.Video Tomography Extraction for 1 of 3 components.

In video tomography, a single line is extracted from each frame of the Y component of a video. These single lines are sequentially transposed to create a new tomography image. Figure 1 shows the process of generating a tomography image. This image is then processed through a Canny edge detector which extracts the edges to reveal patterns in the spatio-temporal changes. The research reported in [9] determined the video's camera work based on these patterns.

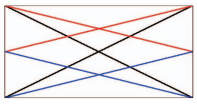

Figure 2. Tomography line pattern extraction over 1 frame

This tomography generation was modified slightly for our design. First, the video is scaled to a resolution of 360x240. Then, the tomography images are generated from three different scan patterns as indicated in Figure 2: two upper diagonals (from the upper corners to the middle), two lower diagonals (from the middle to the lower corners) and two regular diagonals (joining opposite corners). Both diagonals of each set are superimposed and a composite signature image is created using the OR operation.

The amount of level changes (edges) on these three composites is counted on 8 specific vertical and 8 specific horizontal lines, which are evenly distributed along the edge tomography. This produces 16 counts on each of the three diagonal composites. These are then combined to form a 48 byte signature for each shot. The signature size is always 48 bytes regardless of the number of frames in the shot. Figure 3 shows the signature generation process.

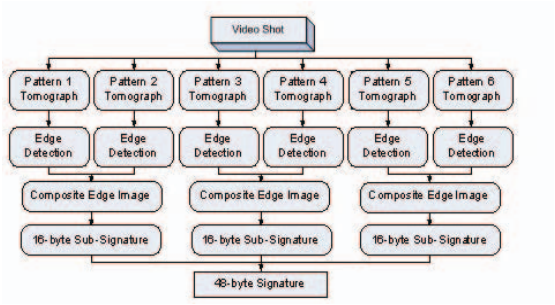

Figure 3. Shot Signature Process

Matching two signatures is achieved by finding the minimum Euclidean Distance between the points in a 48 byte 48-dimensional space:

$$D = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2}$$

The tomography signature method has been applied to "whole to whole" video matching scenarios [2] but has never been implemented in a "part to whole" setting (like matching a clip to a movie).

3.1.1. Frame tomography signature

The shot signatures described in the previous section can be used to locate all shot signatures in the database that closely match the query video. To identify the precise location of the query video in the shot, a local signature is necessary. This local signature can also be used to detect shot boundaries in order to speed up the shot identification process.

The frame tomography signature is extracted from the same tomography image (pre-composition image) used in shot signature generation. Since these tomography images are generated using a specific pattern, each shot and frame signature can easily be extracted as the analysis progresses

through the video, requiring only one complete video analysis.

After retrieving the tomography lines for a frame, the frame tomography signature is obtained by dividing these lines into 4 segments and counting the edges among each of them. The result is a 48 byte 24-dimensional signature.

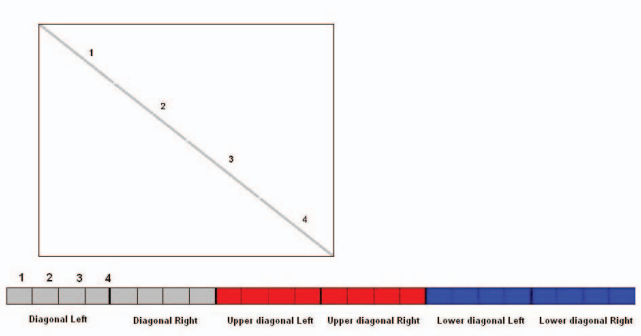

Figure 4. Frame Tomography signature

The distance between frames can also be successfully calculated using Euclidean distance.

4. PROPOSED APPROACH

The proposed signatures are used to identify short clips (query videos) in a given longer video. The independence tests are designed to verify that a video signature can uniquely identify videos. Independence is verified by comparing all possible clip pairs in the database. If a match is detected between an unrelated pair, it is marked as a false positive.

4.1. Query matching using shot and frame signatures

Since the tomography signature does not reflect the length of the video and cannot be taken apart to identify specific portions of its source, identification is done in two stages: 1) identifying the closest shot and 2) identifying the closest frame.

The search for the closest shot begins with determining a shot pattern. A shot pattern for a video consists of a list of shot boundaries and the duration of the shot. When searching for matches, the first step is to identify shot patterns that are close to the query shot pattern. The second step then computes the shot distance between the query and the original video using the proposed video signatures. A set of shots with 30 smallest distances is selected. From this set a frame evaluation is required. This evaluation is accomplished by the extraction of a frame selection for the query and the video clip. The distance between frame signatures is then averaged for every candidate. The matching video clip should have the lowest average value.

4.2. Shot detection

The first step is identifying the shots in the database and the query clip. A shot pattern is stored for all videos in the database. For query clips, shot patterns are generated at the time of the query. Shot detection is the first step in the algorithm and the most important task in the clip identification. Two techniques were explored 1) tomography based shot detection and 2) crater distance method – a method based on frame signature differences. Both of these approaches use the same data extracted by video signature generation and therefore do not incur additional overhead. The crater distance method was found to perform better and was chosen for shot detection. Details on the crater distance method are purposely omitted since the shot detection algorithm used has no effect on the results.

4.3. Process Summary

The steps in video identification are as follows:

1) generate shot and frame signatures and shot pattern for the videos in the database
2) for each query, generate shot pattern, video signatures, and frame signatures
3) using shot patterns and video signatures, search the database for candidate videos
4) for each candidate shot, use frame signatures to localize the frame matches.

Accurate comparison is accomplished by using the Euclidean distance between the two given signatures. This approach allows fast queries, as the shot patterns and shot level signatures can localize video matches. The frame signature can then be used to identify and match the frames in the query from the average distance and standard deviation of the frame set.

5. PERFORMANCE EVALUATION

Performance is evaluated by testing the independence of the video signatures. The independence test defined in [1] works as follows. The database has 1883 three minute clips. Each clip is divided into six 30 second segments and three query videos are created using the first 2, 5, and 10 seconds of each 30 second segment. Three additional queries are created by inserting these 2, 5, and 10 second segments in a 30 second video that is not in the database. This process is depicted in Figure 5. Each 30 second segment results in 6 queries, and each three minute clip generates 36 queries total. These 36 queries are compared against all the videos in the database. The total number of video pairs compared is c = 127, 644,804. This test was designed to test independence precisely.

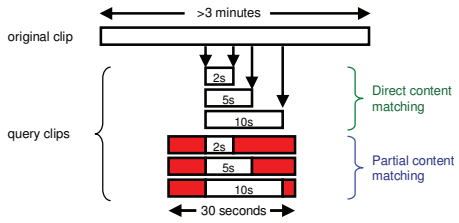

Figure 5. Query construction for independence tests [1]

The query generation and evaluation is a complex and time consuming process; thus, for the test case featured in this article 1640 original clips were randomly selected out of the 1883 original database to constitute a new subset database. From the 2s, 5s and 10s query clips, 54 clips were selected. The total number of clip pairs compared for this test case is c= 85'560. The goal of the algorithm is to identify clip pairs that are independent. The algorithm's tuning parameters were empirically obtained and are presented in Table 1.

Table 1: Algorithm parameters

Description	Value
Detection threshold for crater distance shot method	12.6

5.1. Metrics and Results

The performance of the algorithm was measured in the following ways: 1) Recall, 2) Precision, 3) Prediction precision – for the clips correctly identified as dependent, prediction precision takes the ratio of clips found within one second of the ground truth and the total number of clips.

Table 2: Performance Summary

Description	2 Sec	5 Sec	10 Sec
Total Instances	88'560	88'560	88'560
Independent clip pairs	88'506	88'506	88'506
Identified as Independent	87'716	87'521	85'146
True Positive (tp)	87'713	87'518	85'144
False Positive (fp)	3	3	2
False Negative (fn)	739	934	3'308
True Negative (tn)	51	51	52
Identified as dependent	790	985	3360
Recall: tp/(tp+fn)	99.16%	98.94%	96.26%
Precision: tp/(tp+fp)	100%	100%	100%
Prediction Presicion	94.44%	94.44%	96.30%

Additionally, the algorithm is evaluated in terms of its speed, specifically the time required for a query clip to complete steps 2 through 4 of section 4.3.

Table 3: Time complexity

Step	Description	Value
2	Shot pattern and signature generation	156ms
3	Database search	15ms
4	Frame level comparison	185ms
	Final result for database narrowing search	13ms

6. CONCLUSIONS

This paper presents a novel approach to video identification using video tomography. The approach uses globally unique shot signatures and locally unique frame signatures to quickly identify videos. The proposed approach is evaluated using the independence criteria defined in the MPEG Video Signatures CFP. The results show that the proposed video signatures exhibit strong independence required of video signatures. The algorithm has low complexity and can be integrated into real-time copy detection systems. This work is being extended to evaluate the robustness of the proposed signatures.

7. REFERENCES

[1] MPEG Video Subgroup, "Updated Call for Proposals on Video Signature Tools," MPEG2008/N10155, October 2008, Busan, KR.

[2] G. Leon, "Content Identification using video tomography", M.Sc. Thesis, College of Engineering and Computer Science, Florida Atlantic University, August 2008

[3] G. Doerr and J.L. Dugelay, "A guide tour of video watermarking," Signal Processing: Image Communication, Volume 18, Issue 4, April 2003, Pages 263-282.

[4] X. Fang, Q. Sun, and Q. Tian, "Content-based video identification: a survey," Proceedings of the Information Technology: Research and Education, 2003. ITRE2003. pp. 50-54.

[5] 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 Proceedings of the 6th ACM international Conference on Image and Video Retrieval, CIVR '07, pp. 371-378.

[6] Y. Yan, B.C.Ooi, and A. Zhou, "Continuous Content-Based Copy Detection over Streaming Videos," 24th IEEE International Conference on Data Engineering (ICDE), 2008

[7] N. Guil, J.M. Gonzalez-Linares, J.R. Cozar, and E.L. Zapata, "A Clustering Technique for Video Copy Detection," Pattern Recognition and Image Analysis, LNCS, Vol. 4477/2007, pp. 451-458.

[8] G. Singh, M. Puri, J. Lubin, and H. Sawhney, "Content-Based Matching of Videos Using Local Spatio-temporal Fingerprints," Computer Vision – ACCV 2007, LNCS vol. 4844/2007, Nov. 2007, pp. 414-423.

[9] Akutsu and Y. Tonomura, "Video tomography: An efficient method for camera work extraction and motion analysis," Proceedings of the 2nd international Conference on Multimedia, ACM Multimedia 94, pp. 349-356, 1994.

[10] Manjuntah, P. Salembier and T. Sikora "Introduction to MPEG-7: Multimedia content Description Interface", John Wiley and Sons, 2002.

[11] M.Bertini, A. Del Bimbo, W.Nunziaty, "Video Clip Matching Using MPEG-7 Descriptors and Edit Distance", Image and video retrieval, pp .133-142, 2006.