Content Multimodal Based Video Copy Detection Method for Streaming Applications

Zobeida Jezabel Guzmán Zavaleta
Claudia Feregrino Uribe

Reporte Técnico No. CCC-14-001
20 de enero de 2014

Luis Enrique Erro 1
Sta. Ma. Tonantzintla,
72840, Puebla, México.

Content Multimodal Based Video Copy Detection Method for Streaming Applications

Zobeida Jezabel Guzmán Zavaleta[1], Claudia Feregrino Uribe[2]

[1,2] Coordinación de Ciencias Computacionales,
Instituto Nacional de Astrofísica, Óptica y Electrónica,
Luis Enrique Erro 1, Sta. Ma. Tonantzintla,
72840, Puebla, México
[1]zguzman@inaoep.mx
[2] cferegrino@inaoep.mx

Abstract. Piracy industry severely affects to motion pictures production and distribution companies, specially, using popular Internet applications as video on-demand (streaming) and P2P (peer to peer). It is therefore necessary to propose new robust and precise video copy detection methods suitable for these kinds of applications that, embedded in web monitoring systems may provide additional tools for protection and administration of video contents. In this document content based video copy detection (CBVCD) methods state-of-the-art is reported. CBVCD is also called robust hashes or fingerprinting. Additionally, two initial experiments to improve some robust state-of-the-art methods are showed.

Keywords. Digital Video Piracy, Video Fingerprinting, Content Based Video Copy Detection, Video Robust Hashing, Feature Extraction

Resumen. La piratería afecta severamente a las compañías productoras y distribuidoras de videos digitales, especialmente, mediante aplicaciones populares de Internet de video bajo demanda y de P2P (igual a igual). Por ello es necesario proponer métodos robustos y precisos de detección de copias adecuados para este tipo de aplicaciones que, insertados en sistemas de monitoreo en la red proporcionan herramientas adicionales para proteger o administrar los contenidos. En este documento se reporta la investigación del estado del arte de los métodos de detección de videos con base en contenido, también llamados con hashes robustos o de huellas digitales. También se muestran algunos experimentos iniciales con mejoras a algunos métodos sobresalientes del estado del arte.

Palabras clave. Piratería de videos digitales, Detección de Copias de Videos, Hashes Robustos de Video, Huellas Digitales de Videos, Extracción de Características

1. Introduction

Motion picture piracy damages cinematographic industry by billionaire losses every year [1]- [2]. Hence, international associations have the urgent necessity of new security schemes to avoid, decrease or manage the illegal video distributions. Illegal video distribution is reached mainly through Internet with P2P systems (peer-to-peer), UGC (user generated content) and streaming. The next distribution source is with hard copies. It has been estimated that profit margins generated by trafficking DVDs illegal copies are greater than drugs trafficking gains [3]. A study released by the monitoring firm Envisional found that 23.8% of global Internet traffic involves digital theft, with BitTorrent accounting for almost half of it, that is, 11.4% [4]. Video-based applications, and video streaming in particular, have become utterly popular generating more than half of the aggregate Internet traffic [5]. And over the next few years, 90 percent of the bits carried on the Internet will be video related and consumed by more than 1 billion users [6].

In streaming stored applications, clients request on-demand compressed audio or video files from servers. In streaming applications the client plays audio/video in a continuous playout a few seconds after it begins receiving the file from the server. Stored media applications have continuous playout requirements, nonetheless are less stringent than those for live, interactive applications such as Internet telephony and video conferencing [7].

An important video streaming on-demand corporation is YouTube, which is the worldwide dominant application on mobile and fixed networks, accounting for essentially a quarter of all traffic on the network during peak period [8]. YouTube uses Adobe Flash Video and HTML5 [9] technology to display a wide variety of video content. In order to optimize the bit rate and quality for the available network, YouTube uses Dynamic Adaptive Streaming over HTTP (DASH) and Adobe Dynamic Streaming for Flash (adaptive over RTMP) [10].

YouTube is able to support several original video resolutions from 360p all the way up to 4K [11], also smaller resolutions are supported; those are showed on figure 1 and also are listed on table 1. The default size for upload videos is either 480x385 if 4:3 video, or 640x360 for 16:9 content [12]. When a user uploads a video, the system automatically generates different supported video formats [13] and makes them available to download according to the user preferences and resources. Default video resolution offered by YouTube is 360p, however, users may choose between other available resolutions. In a study realized by [14], authors show that less than 5% of users perform a switch resolution, that is, the majority of users stick with the default video format. Thus, the most popular video format is 360p FLV for PC-players and 360p MP4 for mobile-players.

Although there are available different video resolutions, majority of users (approximately 95%) prefer to visualize video content in default size. Notwithstanding, the increase of available HD cameras and the increase of Internet speed, broadband to UFB, make possible and preferable that users prefers better video resolutions.

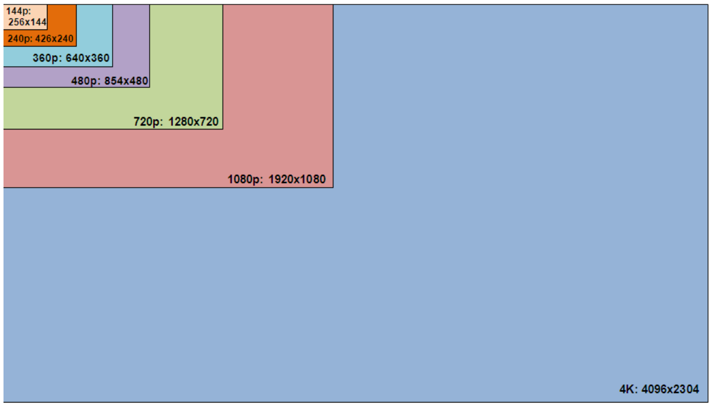


Figure 1: Video resolutions

Table 1: Video quality supported

	YouTube	**Netflix**
144p (256x144)	*	
240p (426x240)	*	
360p (640x360)	*	
DVD (720x480)		*
D1 480p (854x480)	*	
HD / 720p (1280x720)	*	*
SuperHD / 1080p (1920x1080)	*	*
4K (4096x2304)	*	

1.1 Bit rate recommendations

Adobe recommends using the bit rates given in table 2 for dynamic streaming on demand. The frame rate for videos below a bit rate of 100 Kbps could be set to lower values such as 15 fps, but at bit rates higher than 300 Kbps, a frame rate of at least 25 fps and ideally 30 fps is recommended. Additionally, Adobe recommends that the optimal keyframe interval is 5 seconds and the client side buffer is between 6 to 10 seconds [15]. On the other hand, requirements of YouTube for video transmission are listed on table 3 [14].

Table 2: Recommended bit rates for dynamic streaming on demand

Video Size Types	Video size	4:3 aspect size	16:9 aspect size	Total bit rate (Kbps)
QCIF	176x144	144x108 192x144	192x108 256x144	48 96
CIF	352x288	288x216 320x240	384x216 384x216	268 372
D1	720x486	640x480 640x480	852x480 852x480	800 (0.78Mbps) 1200 (1.17 Mbps)
HD	1280x720	- -	1280x720 1280x720	1800 (1.75 Mbps) 2400 (2.34 Mbps)

Table 3: YouTube Internet speed requirement

Resolution	360p (640x360)	480p (854x480)	720p (1280x720)	1080p (1920x1080)	4K (4096x2304)
Bit rate	1 Mbps	1.5 Mbps	3 Mbps	6 Mbps	UFB: Ultra-fast broadband >>10Mbps

1.2 Fingerprinting definition

Additional to other security techniques to avoid illegal video distributions (such as copyright protection) are necessary monitoring and identification systems to detect videos illegally distributed on the web. These efforts contribute to give the control of the video copies to their proprietary for monetization purposes. Actually, these systems are based on watermarking and fingerprinting techniques.

Basically, the definition of fingerprinting varies according to the source of the fingerprint. In general, a digital fingerprint represents a short, robust and distinctive content description allowing fast and privacy-preserving operations [16]. Fingerprinting refers to the process of adding fingerprints to an object or identifying those intrinsic to an object [17]. The uniqueness of the fingerprint is the key concept that enables a data owner to uniquely link a data customer to a specific file [18]. Moreover, fingerprinting refers to detecting and recognizing human fingerprints.

In the context of tracing illegal multimedia content redistribution, fingerprinting (also called transaction tracking or traitor tracing) describes a subtype of watermarking where a unique watermark (i.e. the fingerprint) is added to each copy of the target data. With this, each copy of a multimedia content can be uniquely identified by the watermark, analogously to a human fingerprint that uniquely identifies a person [19]. In Appendix A and Appendix B some references to traitor tracing codes and watermarking systems for traitor tracing are listed.

Additionally, fingerprinting may refer to using intrinsic data properties to uniquely differentiate data copies. In this context, fingerprinting is also called robust hashing or content based copy detection. In this approach fingerprinting does not embed any information; it analyzes the content (image, video, audio or text) to determine their unique characteristics. The identified pattern is stored in a database and can be used for recognizing the content in the future. Some applications of fingerprinting in this context are: broadcast and general media monitoring, copyright control, metadata (store all sorts of useful tracking information associated with content), behavioral modeling advertising, copy protection, forensics to detect whether video footage has been manipulated, additional business opportunities, as maintaining, licensing and managing access to large-scale fingerprint database [20].

In a content-based copy detection method (CBVCD), the fingerprint is extracted by computing a feature vector from the multimedia content, which can represent the content in a unique way. This fingerprint should be 1) distinguishable between different media contents even if these items are similar (avoiding false positive errors) and 2) robust, that is, the fingerprint must survive against various content transformations (avoiding false negative errors) [21].

In this document are utilized fingerprint extraction techniques used by monitoring systems to identify video copies on streaming on-demand applications. Fingerprints are first extracted from videos and stored on a database for its posterior search and matching. In order to avoid confusion, is used the term content based video copy detection (CBVCD).

1.3 Typical video attacks

Multimedia content is susceptible to different modifications, intentional or not. Current CBVCD methods are not able to resist a large number of attacks or to a diverse combination of them. According to [21], common intentional video attacks are:

1. Camera recording
2. Picture in picture
3. Insertion of patterns
4. Recompression (bit rate changes, frame frequency changes)

5. Pixel-level changes (blur, gamma, contrast, noise, filtering, etc.)
6. Geometric changes (resize, shift, rotation, etc.)
7. Temporal domain changes (frame dropping, insertion, resampling, etc.)

Camcorder theft is one of the biggest problems facing the film industry [22]. Illegal recordings from movies in the theater are the single largest source of fake DVDs sold on the street and unauthorized copies of movies distributed on the Internet. For this reason, camcording is a serious offense [23].

Similarly to video attacks, audio intentional degradations, according to [24], can be classified in:

- Numerical: can be simulated numerically.
- Acoustic: involve somehow a conversion to acoustic waves. Their simulation requires more equipment (microphones, loudspeakers, etc.).
- ‘‘Real-world’’: combines numerous degradations and requires a whole sound production chain, e.g., broadcast radio production and transmission.

According to [25] the two most challenging audio distortions are time stretching and pitch shifting. In literature experiments, synthetic distortions are strictly controlled and studied independently; whereas real-world video/audio signals provide a varied set of complex combinations among all these distortions [24]. In real distribution applications, videos are susceptible to many non-intentional attacks, mainly caused by signal processing or transmission errors. In the case of streaming on demand applications, videos are extremely vulnerable to data losses [26] and the audio streaming constraint induces the loss of alignment between the original audio excerpts and the observed audio frames [24].

Additionally, camera recording is an aggressive attack that involves many other attacks that distort the video and audio signal. In [27], authors compare the camera recording attack to the more aggressive compression strategies normally employed to eliminate forgery footprints. Camera recording, depending on the recording conditions, distorts the video in many ways. Modifications in visual component may include all signal processing attacks, resizing, rotation, cropping, zoom, etc. In audio component modifications involve D/A and A/D conversion, re-quantizing, re-sampling, noise addition including real noise and speech, time stretching, jitter, and pitch shifting.

1.4 Video copy detection

Current companies offer security services to try to stop piracy, such as Verimatrix [28], Civolution [29] and MarkAny [30], members of the DWA-Digital Watermarking Alliance [31]. Solutions that these companies offer include watermarking, transaction tracing and fingerprinting, together with web monitoring; but yet pirates can severely manipulate or attack a video in order to avoid detection.

Google and Rhozet have integrated the YouTube fingerprint creation software right into Carbon Coder software [32]. YouTube fingerprinting is unique to the YouTube website. Carbon Coder handles an array of critical operations including SD/HD and PAL/NTSC conversion, logo insertion, color space conversion, color correction and Closed Captioning extraction. [33] YouTube manage 24 hours of videos every minute, Content ID technology scans 100 years of video every day [34].

Other commercial technology is SmartID and CopySense automated content recognition (ARC) from Audible Magic. Identification is based on the perceptual characteristics of the audio itself which allows it to accurately identify content across file formats, codecs, bitrates, and compression techniques. The company affirms that identification is possible with audio clips as short as 10 seconds with identification rates of 99% with zero false positives, additionally, transaction requests can achieve sub-second response time, enabling massive scaling, even with reference databases in excess of 1 million hours of content [35].

Monitoring and identification systems require a robust matching method and fast enough to identify precisely a video among a huge amount of videos [36]. That is, it is necessary to count with less complex algorithms but at the same time more robust (to identify severely attacked videos); also they have to be able to process a large amount of data fast enough for video monitoring applications on the web. Other important factors in video identification systems are precision and recall or sensitivity metrics, pertaining to a confusion matrix. Precision refers to the positive predicted value (PPV), that is, the true positives obtained cases over all positive obtained cases. Recall is the true positive rate (TPR) and means the fraction of true positives out the positives (see equations 1 and 2).

$$PPV = \frac{TP}{TP + FP} \qquad (1)$$

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \qquad (2)$$

Where: P is the number of positive cases, TP and FP are the true positive and false positive obtained cases respectively and FN represents the false negative obtained cases.

2. State-of-the-art

Several attempts have been made to design robust fingerprints. As it can be seen in figure 2, the CBVCD methods are classified according to the way they obtain the video identifier.

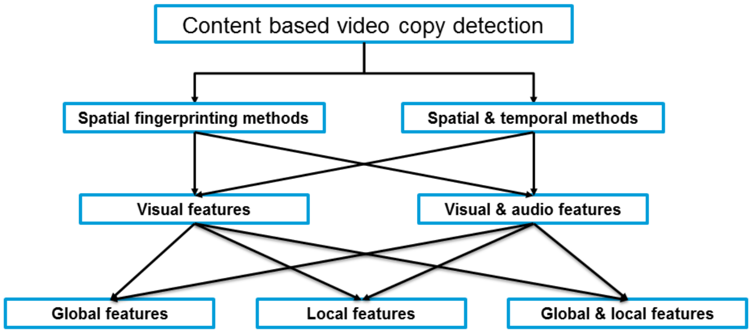

Content based video copy detection
Spatial fingerprinting methods
Spatial & temporal methods
Visual features
Visual & audio features
Global features
Local features
Global & local features

Figure 2 CBVCD taxonomy

State-of-the-art methods show robustness against different modifications, as it can be seen in table 4, most current CBVCD methods are focused in being robust against compression, random noise addition, resizing (or scaling), rotation, cropping and frame rate changes. Nevertheless, just a few methods are robust against camera recording or frame losses. In [27], authors compare the camera recording attack to the more aggressive compression strategies normally employed to eliminate forgery footprints. Camera recording, depending on the recording conditions, distorts the video in many ways. Modifications in visual component may include all signal processing attacks (noise, brightness and contrast change, color space format, etc.), resizing, rotation, cropping, zoom and projective transformations. In audio component modifications involve D/A and A/D conversion, re-quantizing, re-sampling, noise addition including real noise and speech, time stretching, jitter, and pitch shifting. Moreover, state-of-the-art methods are not robust (or not reported robustness in their results) against audio component attacks that diminish the audio quality and consequently, diminish the audio identification capacity of the copy detection method.

Similarly to CBVCD, in audio copy detection (CBACD) state-of-the-art, authors show different audio degradations in their results to reproduce the audio attacks (see table 5); being the most common compression (mp3), companding, filtering (low-pass), white noise addition and speech addition. According to [25] the two most challenging audio distortions are time stretching and pitch shifting. Yet again, the majority of methods are not robust against re-recording attack and time stretching.

Table 4 Common video attacks on the state-of-the-art

		Method													
	Attacks	(Roopalakshmi, 2013)	(Nie, 2013)	(Honghai, 2012)	(Li, 2012)	(Lei, 2012)	(Deng, 2012)	(Wei, 2011)	(Varna, 2011)	(Paschalakis, 2011)	(Esmaeli, 2011)	(Chiu, 2010)	(Lee, 2009)	(Lee, 2008)	Total
Signal processing	Compression	*		*	*					*		*	*	*	54%
	Re-encoding					*									8%
	Resolution reduction									*					8%
	Luminance histogram equalization					*			*					*	23%
	Brigntness	*								*	*	*		*	38%
	Contrast				*			*			*		*		31%
	Global color change												*	*	15%
	Motion blurring	*				*									15%
	Gaussian blurring (pixel ratio)		*		*	*								*	31%
	Random noise add	*	*			*	*	*				*			46%
	AWGN				*	*	*				*			*	38%
	Gamma correction					*		*						*	23%
	Filters: gaussian, median, average					*	*								15%
Geometric	Resizing /scaling			*		*	*					*	*	*	46%
	Rotation (degrees)	*	*	*	*	*		*			*			*	62%
	Cropping	*		*	*	*	*	*				*		*	62%
	Zoom	*										*			15%
	Letter box/pillar box					*	*								15%
	Moving caption	*													8%
	Insertion of pattern	*				*		*		*					31%
	Picture in picture	*				*		*							23%
	Flip (vertical mirror)			*		*		*							23%
	Bending				*										8%
	Camera recording	*								*					15%
Desynchronization	Interlaced/progresive conversion									*					8%
	Time shift										*				8%
	Spatial shift		*	*		*					*				31%
	Slow motion	*										*			15%
	Fast forward	*										*			15%
	Frame rate (fps)	*		*	*					*			*	*	46%
	Frame loss (dropped)				*		*				*				23%

Table 5 Common audio attacks on the state-of-the-art

Attacks			Method								Total
			[25]	[24]	[37]	[38]	[39]	[40]	[41]	[42]	
Numerical	Audio encodings	MP3	*	*	*	*	*	*		*	87 %
		WMA	*	*							25%
		GSM	*	*							25%
		Real Media		*							12.5%
		Re-quantizing						*		*	25%
		Re-sampling		*						*	25%
	Dynamic changes	Multiband companding		*	*	*	*				50%
		Amplitude dynamic compression		*							12.5%
		Volume change		*							12.5%
		Bandwidth limit /single-band companding				*	*				25%
		Amplitude boosting						*		*	25%
		Amplitude cutting						*		*	25%
		Normalizing						*		*	25%
		Invert								*	12.5%
	Filtering	All-pass filtering		*							12.5%
		Low-pass filtering	*					*		*	37.5%
		Band-pass filtering		*			*				25%
		Telephone band-pass		*							12.5%
		Echo filter		*							12.5%
		Hiss reduction								*	12.5%
	Noise addition	Eco-addition	*							*	25%
		White noise addition	*	*						*	37.5%
		Real-world noise addition				*			*		25%
		Speech addition		*	*		*				37.5%
		Dithering								*	12.5%
	Temporal	Time shift		*							12.5%
		Time stretching	*	*							25%
		Linear speed change		*							12.5%
		Jitter	*								12.5%
		Pitch shifting	*								12.5%
Acoustic		D/A A/D conversion		*							12.5%
		Re-recording		*					*		25%

In the table 6, some representative state-of-the-art approaches are listed, summarizing their classification and disadvantages. Features column refers to which characteristics to obtain the fingerprint are employed, that is, global (G) or local (L), spatial (S) or temporal (T) or visual (V) or acoustics (A) components, or a combination between them. Fingerprinting vector column shows the length of the fingerprint in bits per frame (f) or bits per keyframe (kf).

Table 6 Main CBVCD methods from the state-of-the-art

Features						Method	Fingerprint vector	Disadvantages
G	L	S	T	V	A			
x		x		x		CGO [43]	12 bits/f	No robust against geometric transformations
x		x		x		CGO + SPB [44]	52 bits/f	Additional training and classification increase the computational cost
x		x		x		Color correlation histogram [45]	35 bits/kf	Weakness against color correlation changes
x		x	x	x		2D-DCT of TIRI [46]	128 bits/TIRI	Performs a fast approximate search method, however, the fingerprint is only robust against signal processing attacks
x		x	x	x		LRTA [47]	126 bits/kf	Method robustness decrease against combined temporal and spatial attacks
x	x	x	x	x		R-D frames+ graph model [48]	256 bits/kf	Authors do not show robustness against geometric transformations
x	x	x	x	x	x	SURF+ spectral centroid [49]	16 bits/f	The descriptors are extremely summarized, that affects the detection precision with a huge amount of false negative cases
x	x	x	x	x	x	CST-SURF + MFCC [50]	16 bits/f	This method is more robust than its previous version, however, still has precision deficiency
x	x	x	x	x	x	SIFT+SURF+2D-DCT+ WASF [51]	nd	This method demands a considerably computational complexity

State-of-the-art methods show deficient performance, with not enough robustness against common and severe attacks or with high computational cost not suitable for streaming applications or with low precision, that is, high false negative cases or low true positive. CBVCD methods have to be robust against the most common attacks to both visual and audio components, that is:

1. Data losses: including frame dropping, bit rate change, frame rate change, compression, band pass filtering and jitter.

2. Camera recording and audio re-recording: that includes several modifications as rotation, cropping, color space transformations, projective transformations and signal processing attacks.

3. Mix audio and visual content: including insertion of patterns, subtitles, mix audio with speech and picture in picture.

Moreover, the utilization of a single video identification method is not sufficient for a system where videos are exposed to severe and diverse attacks. For that reason, it is indispensable to complement different feature extraction methods in order to provide a robust and precise video identification method.

Since there is a tradeoff between the size of a video descriptor and its robustness, video features extraction and description techniques have to be convenient selected to provide the high robustness against video attacks specifics on a selected application. In streaming distribution applications, additionally, it is necessary to uses a computationally less complex method and fast enough.

According to the state-of-the-art, the most robust fingerprinting types are those that combine spatial and temporal, global and local and visual and audio information of the video to extract a secure video identifier. However, due to the large number of operations to obtain a video descriptor, is not adequate for applications with a large amount of videos and/or streaming applications. Nevertheless, the convenient combination of those fingerprinting techniques in processing blocks for a multilevel search can improve the performance in a CBVCD method for these kinds of applications. In that way, in a multilevel copy detection method, a simplified global video descriptor can filter the query results and local descriptors can refine the matching. Additionally, matching metrics have to be optimized in according to the copy detection method to enhance the identification precision. Likewise, the search process plays an important role in the method performance, for that reason, optimized search methods have to be utilized.

3. Problem statement

3.1 Problem

Digital videos are vulnerable against severe intentional attacks affecting the precision of current video detection methods, for that reason, it is necessary to design CBVCD methods more robust and precise against more severe attacks such as camera recording. Additionally, popular video streaming on-demand distribution applications need copy detection methods with high processing effectiveness suitable for this kind of applications in the web.

3.2 Objectives

3.2.1 General Objective

To design a multimodal CBVCD method, that is, based on both visual and audio components, precise and robust enough, suitable for illegal video monitoring in streaming on-demand applications.

3.2.2 Specific Objectives

1. To identify an adequate combination of CBVCD methods, for both visual and audio components, in order to provide high identification precision, low computational cost and robustness against most common video attacks.
2. To design a multilevel and multimodal CBVCD method robust enough and effective, suitable for video monitoring systems in streaming on-demand applications.
3. To develop a robust video descriptor that complements to more precise matching metrics with less computational cost.

3.3 Research questions

- What are the most robust and less computational cost CBVCD methods for both, visual and audio features? How can they be combined to balance their robustness and computational cost in order to be suitable for video streaming on-demand applications?
- In order to design a multilevel extraction and detection, which feature extractors are computationally less expensive and which are more robust?
- What are the most precise metrics to match two video descriptors (the suspected video copy descriptor with a master video descriptor)?

3.4 Hypothesis

H1: Combining multimodal, global and local video features increase precision in video identification and improves robustness against severe video attacks in CBVCD methods.

H2: A multilevel video descriptor extraction improves the speed searching and matching in streaming CBVCD applications.

H3: Using better video descriptors based on simplified persistent features improves matching process in CBVCD methods.

3.5 Methodology

In order to reach the objectives this methodology is proposed:

- **To get a video dataset from internet open projects.** An important platform for researchers is TRECVID (Text REtrieval Conference - Video Retrieval Evaluation) conference series whose goal is to encourage research in information retrieval by providing a large test collection, uniform scoring procedures, and a forum for organizations interested in comparing their results [52]. Content based copy detection task has utilized the MUSCLE-VCD-2007 [53], a video list obtained from the internet archive [54] and open video project [55]. For practical purposes, videos in dataset will be composed of different durations (from 30 seconds to 30 minutes aproximately) and different categories including sports, educational, news, TV commercials and animated. Additionally, in order to test CBVCD methods robustness, each video in dataset will be modified with the most common video attacks that affect both visual and audio components:1) data losses: including frame dropping, bit rate change, frame rate change, compression, band pass filtering, jitter; 2) camera recording: that includes several modifications as rotation, cropping, color space transformations, projective transformations, signal processing attacks and audio re-recording; 3) mix audio and visual content: including insertion of patterns, subtitles, mix audio with speech and picture in picture; 4) Decrease of quality: noise addition (in visual and audio components), brightness/contrast change.
- **To identify and select the most robust, fastest and precise methods of the state-of-the-art to extract and describe the global and local video features in both visual and audio components.** The CBVCD methods on the state-of-the-art use very different techniques with different effectiveness and most of them use only visual or audio content. Additionally, there exists a tradeoff between size and robustness of the video descriptor (that is, the obtained fingerprint from the video). For those reasons it is necessary to identify which are the most appropriate methods for streaming on-demand applications. According

with state-of-the-art reported results about precision, robustness and computational cost, the best methods will be selected.

- **To identify advantages and disadvantages of selected methods and analyze possible improvements by complementing them with other approaches.** Each CBVCD method utilizes different video descriptor extractions, search methods and matching metrics. Every method has its own advantages and disadvantages and it is necessary to analyze them in order to get the best combination.
- **Benchmark testing and redesign of proposed improved approaches.**
- **To design multilevel and multimodal CBVCD using the proposed video descriptor methods.** Multilevel and multimodal designs are complementing each other in order to provide a CBVCD method with high precision, low computational cost and enough robustness. In a multilevel method, first level will help to filter the most similar videos using the less computational cost feature extraction, descriptor searching and matching, based on audio or video or both components, this level does not provide robustness enough. In second level, based on a more robust feature extraction, the search and matching computational cost will be decreased with a refined search only over the most similar video descriptors. And in a possible third level, the goal will be increased precision.
- **Test the multimodal method with the video dataset and redesign if it is necessary.**
- **Analyze and choose the matching video descriptor technique that harmonizes with the method designed, suitable for streaming applications.** The CBVCD method is the main block for the video identification in a monitoring system. However, when the video identifier or fingerprint is extracted, it is necessary to match it with the master video identifier previously stored in a database. For that reason, the matching metric has an important role for the exact identification. The searching of the video identifiers in a database is out of scope of this project.

3.6 Justification

Current digital video distribution applications require security schemes. Mainly the entertainment industry has the necessity of identifying if a distributed video copy is illegal or not in order to avoid or decrease the piracy industry attacks. To detect illegal video copies, it is essential a video identification and monitoring system robust and precise; that is, robust to severe and diverse video attacks with minimum errors. However, proposed video identification schemes are robust against only some moderate attacks or show poor precision, for example, with high false identification rates. Additionally, to ensure robustness a large number of operations are utilized, complex in some cases. This processing decreases the method performance in applications with large amounts of data, in broadcasting or web streaming distributions. For that reason, multimodal design based on visual and audio components, will provide more robustness against different typical attacks and multilevel design will help to decrease computational cost whereas it will increase robustness and precision.

The applications that benefit from this solution of robust and precise digital video identification are diverse. For example, in web monitoring, the benefited applications are on line distributions, downloads and streaming, copyright in P2P and UGC platforms. Secondarily, video identification could add robustness to transaction tracking or traitor tracing and watermarking applications. That is, video identification methods are used as a base for temporal and spatial alignments as a preprocessing step previous to watermark extraction.

3.7 Contributions

- A method to identify digital video copies for large amounts of data in data stream applications. The method will be robust against severe attacks and will improve the identification precision compared with the current methods.
- An improved video descriptor, multimodal and multilevel, adequate for video streaming applications.

4. Experimental results

Following the proposed methodology, two experiments were performed in order to improve some selected CBVCD methods proposed in the state-of-the-art. First, two different approaches from the state-of-the-art were selected: 1) a global based fingerprint and 2) local and global visual and acoustic spatiotemporal fingerprint; both showed good robustness to diverse attacks. These approaches were tested with a selected group of original and attacked videos to prove their performance. Some modifications were made for improving them in robustness characteristic. Second, with a more severe group of attacks, including both visual and acoustic components, a global acoustic fingerprinting method were tested and improved. Video dataset, the selected methods, results and concluding remarks are showed in following subsections.

4.1 Video dataset

The video dataset is composed by 21 open videos of different categories, that is, documental, TV commercials, animated, sports and movies. All of them are in color with audio component, in different compression formats and different duration, almost 2 hours in total. Table 7 enumerates the video dataset. Video references are listed at the end of this document. In experiment 1, only the first 18 videos were tested, for experiment 2, the last 3 videos were incorporated to the video dataset.

Table 7 Video dataset

#	Name	Format	Frame size (pixels)	Frame rate (fps)	Duration (mm:ss)	Size (MB)	Video bit rate (kbps)	Audio bit rate (kbps)	Audio sampled rate (kHz)	Channels
1	Documental1	MPEG	320x240	29	00:30	5.27	1428	128	32	1
2	Documental2	MPEG	320x240	29	02:20	24.50	1428	128	32	1
3	Documental3	MPEG	320x240	29	09:13	95.30	1428	128	32	1
4	Documental4	MPEG	320x240	29	06:50	70.70	1428	128	32	1
5	Documental5	MPEG	352x240	29	28:07	245.00	1200	192	44	2
6	Documental7	MPEG	320x240	30	06:04	48.70	1086	128	44	2
7	Animated1	MP4	320x240	29	06:10	25.70	578	64	48	2
8	Animated2	MP4	640x480	29	08:39	48.80	784	87	44	2
9	Animated3	MP4	320x240	29	06:52	28.50	577	64	48	2
10	Animated4	MP4	320x240	29	02:02	8.51	575	63	48	2
11	Sports2	MP4	320x240	29	03:11	13.20	577	63	44	2
12	TVComm1	MP4	640x360	30	01:01	4.20	575	96	44	2
13	TVComm2	MP4	640x360	24	00:30	2.20	594	94	44	2
14	TVComm3	MP4	432x360	29	00:30	1.37	380	96	44	2
15	TVComm4	MP4	1280x720	24	01:00	19.40	2712	151	44	2
16	TVComm5	MP4	640x360	25	02:03	9.61	648	95	44	2
17	TVComm6	MP4	640x360	25	02:17	9.00	548	94	44	2
18	TVComm7	MP4	640x352	24	00:30	1.98	547	96	44	2
19	OpenMovie1	AVI	1920x1080	24	09:56	885	12455	448	48	5
20	OpenMovie2	MP4	426x240	24	10:53	44.8	571	64	48	2

21	OpenMovie4	MOV	1280x534	24	12:14	354	4049	192	44	2

4.1 Experiment 1

Selected Method 1 (SURF + MFCC):

This method proposes a framework for estimating geometric distortions in video copies by employing visual-audio fingerprints [50]. With the visual and acoustic features extracted, they perform temporal and geometric frame alignments and then estimate the distortion. For the scope of this experiment, only the visual and acoustic features extraction is presented.

- **Visual fingerprint extraction**: Each frame of a video sequence is divided into 4 regions, for each region its SURF key points are extracted. The differences between the counts of these key points of subsequent frames, named CST-SURF, are the compact visual fingerprint.
- **Acoustic fingerprint extraction**: Compact representations of MFCCs (Mel-Frequency Cepstral Coefficients) are extracted from the audio profile. The audio signal is downsampled and segmented with a Hamming window function, then the MFCCs are calculated, this coefficient matrix is summarized using Singular Value Decomposition (SVD) technique and they employ 4 to 6 singular values for extracting acoustic signatures.

Method 1, proposes that it is possible to identify a copy of a video even if it was severely attacked using both the audio and visual components. This method was tested with the video dataset attacked with the following 5 attacks:

1. Contrast and bright change + Gaussian white noise addition + Rotation 1°+ Cropping 10% of pixels of borders
2. Pattern insertion (👁) + Text insertion: “Peppers are good!”
3. Rotation 1°+ Frame dropping, drop 1 frame of every 10 + Cropping 2% of pixels of borders
4. Attack num. 2 + Frame rate change to 25 fps
5. Attack num. 2 + Picture in picture, copy video inserted into TVComm3 (size 1.2x).

Distance values between the original video and its attacked versions are presented in table 8, values equal to ‘0’ means a perfect match. These results showing excellent rates for true positive cases using a threshold α=0.45 (TPR equal to 1 means a 100% of true positive cases); however, the compact representation of visual and acoustic fingerprints also shows a high number of false positive identifications.

Additionally, this method is not suitable for streaming applications due to all operations involved on each video frame. To improve this method, it is necessary to perform fewer operations and to explore another more informative representation for the fingerprints.

Table 8 Method 1 testing results

Video	**Attack**				
	1	2	3	4	5
1	0.25	0.19	0.20	0.17	0.16
2	0.26	0.14	0.16	0.13	0.11
3	0.29	0.20	0.23	0.18	0.15
4	0.24	0.27	0.28	0.26	0.17
5	0.40	0.41	0.41	0.40	0.29
6	0.33	0.34	0.34	0.33	0.24
7	0.16	0.09	0.10	0.10	0.08
8	0.24	0.28	0.27	0.28	0.25
9	0.26	0.23	0.25	0.24	0.22
10	0.24	0.25	0.28	0.25	0.24
11	0.30	0.35	0.37	0.36	0.33
12	0.23	0.19	0.21	0.19	0.20
13	0.12	0.13	0.10	0.13	0.12
14	0.23	0.22	0.24	0.22	0.22
15	0.21	0.21	0.21	0.21	0.21
16	0.14	0.22	0.22	0.22	0.19
17	0.27	0.39	0.39	0.39	0.32
18	0.15	0.01	0.02	0.01	0.08
TPR	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

The hypothesis to prove is: “Extract SURF of representative short images of selected key frames allows a more informative fingerprint that CST-SURF”

In order to improve SURF coefficients extraction are proposed two options:

Option P1:

First, it is necessary to preprocess the video: the video is downsampled to 4 fps, each keyframe is converted to gray scale and downsize to 30x30 pixels. In this representative small image the SURF coefficients are extracted. In order to summary the SURF coefficients, the SVD is obtained.

Option P2:

The extraction of SURF is over TIRIs and the description with a binary hash, similar to [46]. The binary hash value of SURF coefficients simplifies their description. For each coefficient *f* with *x=64* elements, the median value *m* is found and the hash is performed by equation (3).

$$h = \begin{cases} 1\,, & f_x \ \geq m \\ 0, & f_x < m \end{cases} \tag{3}$$

For practical purposes, in table 9 are only presented the visual fingerprint matching from the videos. Values equal to '0' means a perfect match. The columns M1, P1, and P2, show the matching distances obtained with approach 1, option P1 and option P2, respectively.

Table 9 improved method 1 testing results

#	**Attack**														
	1			2			3			4			5		
	M1	P1	P2	M1	P1	P2	M1	P1	P2	M1	P1	P2	M1	P1	P2
1	0.99	0	0	0.75	0.0	0	0.81	0	0	0.68	0	0	0.65	0.01	0
2	1.02	0	0	0.56	0.2	0	0.65	0	0	0.50	0.15	0	0.42	2	0
3	1.17	0	0	0.79	0.0	0	0.91	0	0	0.73	0	0	0.60	0	0
4	0.94	0	0	1.10	0.0	0	1.10	0	0	1.05	0	0	0.67	0	0
5	1.59	0	0	1.64	0.0	0	1.64	0.1	0	1.60	0	0	1.15	2	0
6	1.34	0	0	1.36	0.0	0	1.36	0	0	1.33	0	0	0.95	0	0
7	0.66	0	0	0.37	0.0	0	0.42	0	0	0.40	0	0	0.33	0	0
8	0.97	0.1	0	1.12	0.3	0	1.06	0.1	0	1.12	0.26	0	1.01	0.07	0
9	1.03	0	0	0.94	0.0	0	0.99	0	0	0.94	0.01	0	0.90	0.01	0
10	0.95	2	0	1.01	0.3	0	1.12	2	0	1.01	0.30	0	0.96	0.02	0
11	0.93	0.1	0	0.77	0.0	0	0.84	0	0	0.78	0	0	0.82	0.01	0
12	0.64	0	0	0.56	0.1	0	0.59	0	0	0.57	0.11	0	0.57	0.08	0

13	0.49	0	0	0.51	0.0	0	0.39	0	0	0.51	0	0	0.49	0.01	0
14	0.94	0.1	0	0.90	0.0	0	0.94	0.1	0	0.89	0	0	0.87	0.07	0
15	0.82	0	0	0.83	0.0	0	0.83	0	0	0.82	0	0	0.84	0	0
16	0.57	0	0	0.88	0.0	0	0.89	0	0	0.87	0	0	0.78	0.02	0
17	1.06	0	0	1.55	0.0		1.54	0	0	1.55	0	0	1.28	0	0
18	0.59	0	0	0.03	0.0	0	0.07	0	0	0.03	0.02	0	0.32	0.02	0
TPR	**0.00**	**0.95**	**1**	**0.10**		**1**	**0.10**	**1**	**1**	**0.05**	**1**	**1**	**0.15**	**0.90**	**1**

According to presented results, the SURF coefficients over TIRIs and binary hash representation perform the best results for the visual fingerprints. However, this kind of descriptor is not sufficiently informative due to identification errors, such as high false positive rates. In this case audio component features can complement it providing more information.

Selected Method 2 (Color correlation):

This method is based on invariance of color correlation histogram [45]. The process of feature extraction for a color keyframe (1 fps), involves three steps. First, the keyframe of size (*wxh*) is transformed to RGB (red, green and blue) color model and divided into 16x16 non overlapping blocks, for each block is calculated the average intensities of RGB components, generating a lower resolution image (*mxn*, where *m = [w/16]*, *n = [h/16]* and *[x]* is the nearest integer to *x*.). Second, color correlation is extracted from the lower resolution image and the percentage of pixels belonging to their corresponding color correlations is calculated, after that, are obtained six normalized real values for each image (keyframe). Third, first five truncated numbers are stored in a binary form as the feature for the input image. Color correlation is denoted by 6 cases, where the tuple (R_{xy},G_{xy},B_{xy}) represents red, green and blue channels for a pixel with coordinates (*x*,*y*) in a video frame:

1) $R_{xy} \geq G_{xy} \geq B_{xy}$
2) $R_{xy} \geq B_{xy} \geq G_{xy}$
3) $G_{xy} \geq R_{xy} \geq B_{xy}$
4) $G_{xy} \geq B_{xy} \geq R_{xy}$
5) $B_{xy} \geq R_{xy} \geq G_{xy}$
6) $B_{xy} \geq G_{xy} \geq R_{xy}$

Authors of method 2, reports a high robustness to different attacks, much more than other global fingerprinting methods and faster enough for real-time applications.

Results of the performed test to method 2 are shown in table 10. First column shows the video number and subsequent columns show Manhattan distances between the master fingerprint and the fingerprints of five different attacked video copies (same attacks as previous method). A distance equal to ‘0’, means that the two video fingerprints are the same.

Table 10 Method 2 testing results

Video	**Attack**				
	1	2	3	4	5
1	0.99	0.75	0.81	0.68	0.65
2	1.02	0.56	0.65	0.50	0.42
3	1.17	0.79	0.91	0.73	0.60
4	0.94	1.10	1.10	1.05	0.67
5	1.59	1.64	1.64	1.60	1.15
6	1.34	1.36	1.36	1.33	0.95
7	0.66	0.37	0.42	0.40	0.33
8	0.97	1.12	1.06	1.12	1.01
9	1.03	0.94	0.99	0.94	0.90
10	0.95	1.01	1.12	1.01	0.96
11	0.93	0.77	0.84	0.78	0.82
12	0.64	0.56	0.59	0.57	0.57
13	0.49	0.51	0.39	0.51	0.49
14	0.94	0.90	0.94	0.89	0.87
15	0.82	0.83	0.83	0.82	0.84
16	0.57	0.88	0.89	0.87	0.78
17	1.06	1.55	1.54	1.55	1.28
18	0.59	0.03	0.07	0.03	0.32
TPR	**0.00**	**0.05**	**0.10**	**0.10**	**0.15**

These results show the weakness of the method against color correlation changes. Trying to strengthen this method, the improvement proposed is to generate an acoustic fingerprint and combine it with the visual fingerprint.

Hypothesis to prove is: "Adding an additional acoustic fingerprint is possible to resist to color correlation change attack".

Acoustic fingerprint was obtained using MFCC, with the same parameters proposed by method 1. The matching is performed using the weighted average of the matching values of visual fingerprints and acoustic fingerprints. Table 11 shows the obtained results.

Table 11 Improved method 2 testing results

Video	**Attack**				
	1	2	3	4	5
1	0.25	0.19	0.20	0.17	0.16
2	0.26	0.14	0.16	0.13	0.11
3	0.29	0.20	0.23	0.18	0.15
4	0.24	0.27	0.28	0.26	0.17
5	0.40	0.41	0.41	0.40	0.29
6	0.33	0.34	0.34	0.33	0.24
7	0.16	0.09	0.10	0.10	0.08
8	0.24	0.28	0.27	0.28	0.25
9	0.26	0.23	0.25	0.24	0.22
10	0.24	0.25	0.28	0.25	0.24
11	0.23	0.19	0.21	0.19	0.20
12	0.16	0.14	0.15	0.14	0.14
13	0.12	0.13	0.10	0.13	0.12
14	0.23	0.22	0.24	0.22	0.22
15	0.21	0.21	0.21	0.21	0.21
16	0.14	0.22	0.22	0.22	0.19
17	0.27	0.39	0.39	0.39	0.32
18	0.15	0.01	0.02	0.01	0.08
TPR	**1.00**	**1.00**	**0.95**	**0.95**	**1.00**

In conclusion, the main advantages of method 2 are: the less complex operations, a few number of operations and the fingerprint representation is short (less storage necessity). The main disadvantage is the lack of robustness against intentional attacks that affect the color correlation in the video. Also, similar color correlation on different videos increases false positive error. This can be eliminated with an additional acoustic fingerprint.

4.2 Experiment 2

4.2.1 Intentional attacks

For experiment 2, intentional attacks combination was re-categorized. Intentional visual modifications were performed using MATLAB R2013a and Adobe Audition v6.0 compilation 732 (64bits) for audio attacks. Attacks to video dataset were:

Attack 1 (Data losses):

1. Frame dropping: 10% of frames were dropped.
2. Frame rate: changed to 20 fps.
3. Audio sampled rate: resampled to 32 kHz.
4. Audio compression: mp3 compression with 75% of quality.
5. Simulated jitter: addition of white noise (40dB) for 1 second at the beginning of audio signal.

Attack 2 (Camera and audio recording simulation):

1. Rotation: by 5°
2. Cropping: after rotation to maintain the original frame size.
3. Color space transformations: from YCrCb to RGB.
4. Projective transformations: theta=2°
5. Audio Compression: mp3 compression with 75% of quality.
6. Audio sampled rate: resampled to 41 kHz.

Attack 3 (Mix audio and visual content[1])

1. Insertion of patterns: standard image Baboon.jpg in gray scale of size 50x50 pixels, inserted at the left top of frames.

[1] References are at the end of this document

2. Picture in picture: Open1 video at background with size m+100 x n+50, where m x n is the size of attacked video.
3. Subtitles: are inserted 35 characters in the first frame of each second of the video, text was obtained from the Alice text file inserted at location [50 rTxt], where rTxt is the height of the frame - 30), inserted text was in color white and font size 18.
4. Mix audio with speech: Speech1 mixed with audio every 20 seconds.

Attack 4 (Decrease of quality):

1. Rotation: by 3°
2. Contrast and brightness change: decreased 10%
3. Visual noise addition: random Gaussian noise
4. Audio noise addition: white noise (20 dB)

Attack 5 (Real camera recording):

Videos were recorder in a dark room, simulating a cinema ambience. Each video was projected in a white wall with a Panasonic LCD projector model PT-LB2 [56] and recorded with a Sony DCR-SX43 camcorder [57] using output video format listed in table 12. Camera recording scenario is represented in figure 3.

Table 12 output video format

Format	Frame Size	Frame rate	Bit rate	Audio bit rate	Channels	Audio sample rate
MPEG	720 x 480	29 fps	9356 kbps	256 kbps	2 (stereo)	48 kHz

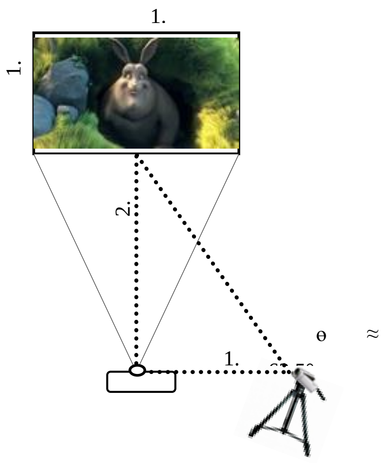


Figure 3 Camera recording scenario

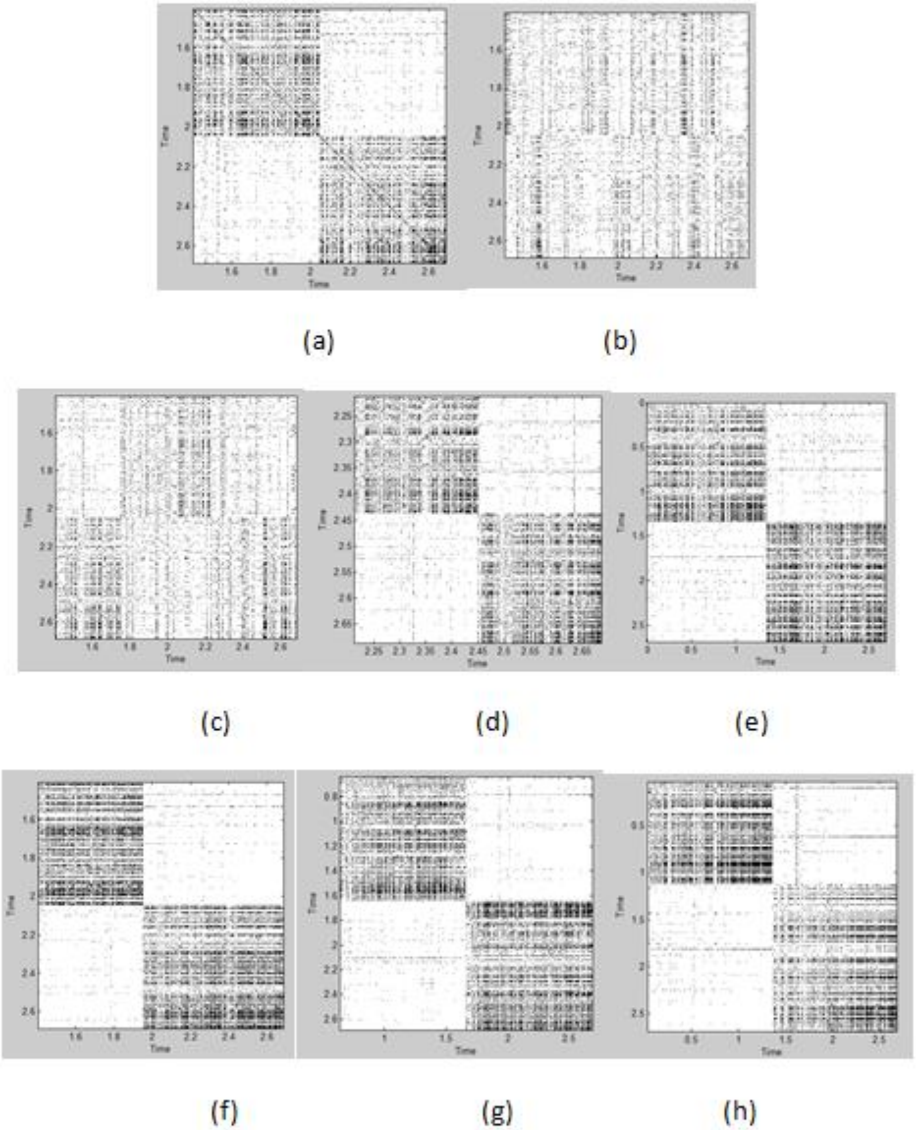
In experiment 2 the selected method was [25]. Authors suggest that their method is robust against time stretching, pitch shifting, compression and noise addition. To generate an acoustic fingerprint, audio signal is first transformed into a cochleagram using gammatone filterbank (64 filters whose centre frequency ranges from 50 Hz to 8 kHz). Second, similar to an image, SURF features are extracted from the cochleagram. Then, non-negative matrix factorisation (NMF) is performed to the SURF descriptor matrix to reduce the feature's dimension. After that, they construct a time series of delay co-ordinate state space vector, with embedding dimension 3, and time delay τ=3. For matching process, the cross recurrence plot (CRP) of time series delayed vectors, is evaluated with a threshold of 0.1. A CRP shows equivalences between two systems in different times; in this case, any diagonal path represents similar state sequences exhibited by both systems The percentage of black cells included in the main diagonal path is used by authors as the similarity measurement.

Authors showed experimental results with high identification rate and low computational complexity using short segments of audio clips (4 seconds long). However, when the audio size increases, computational cost increases dramatically due to operations of large matrices. In order to test the method with entire audio component of video dataset, computations are made by blocks; each audio signal is first divided into 16 blocks. Table 13 enumerates the length of each video fingerprint, execution time and audio duration in seconds for reference. In all cases, execution time of tested method is shorter than audio duration.

Table 13 Experimet 2 results

Audio	Fingerprint length (bits)	Execution time (sec)	Audio Duration (sec)
1	14058	17.48	30
2	48598	66.20	80
3	235408	273.45	553
4	363494	211.67	410
5	1040334	1205.14	1687
6	6406	236.44	364
7	299114	277.72	370
8	247074	349.08	519
9	102468	294.45	412
10	12018	76.13	122
11	192816	127.22	191
12	7002	38.36	61
13	1586	23.22	30
14	792	22.20	30
15	65252	39.31	60
16	42190	84.03	123
17	88064	92.08	137
18	19398	24.28	30
19	112368	430.03	596
20	82136	428.16	653
21	119770	502.03	734

Figure 4 shows the graphical results (CRPs) from the matching of audio1 with (a) original Audio1, (b) Audio 2, (c) Audio 3, (d) Audio1 with attack1, (e) Audio1 with attack2, (f) Audio1 with attack3, (g) Audio1 with attack4, (h) Audio1 with attack5. In plots, diagonal paths show equivalences between original audio and its modified versions, non-like to compare two fingerprints from different audio signals.



(a) (b)

(c) (d) (e)

(f) (g) (h)

Figure 4 matching Results for audio1.

Additional to CRP, in table 14 are presented Euclidean distances between original audio fingerprint and distorted audio fingerprints. This measure is simpler but less precise.

Table 14 Euclidean distance between original audios and attacked copies

Audio	Attack1	Attack2	Attack3	Attack4	Attack5
1	0.6965	0.6349	0.5634	0.5697	0.6268
2	1.2399	1.1870	1.1713	1.1963	1.1609
3	1.1597	1.1142	0.9929	1.1176	NA
4	0.8920	0.5977	0.4961	0.5306	NA
5	0.3195	0.3240	0.3003	0.3073	NA
6	1.3707	1.2146	1.4332	1.3233	NA
7	0.4890	0.4910	0.3198	0.3437	0.3437
8	0.3878	0.3936	0.9893	0.3933	0.9926
9	0.3684	0.3313	0.4388	0.4657	0.4005
10	0.9644	1.3121	1.3177	0.6860	1.3963
11	0.6721	0.7986	1.3403	0.4532	NA
12	1.4128	1.1333	1.1749	0.8788	2.2308
13	1.5784	0.9565	0.9507	0.6922	3.1112
14	1.2389	1.1647	1.1955	0.7143	2.6972
15	1.2017	0.6405	0.9855	0.8793	2.5718
16	1.4334	1.5160	1.0918	0.9131	0.6334
17	1.1989	1.2293	1.4593	0.6218	2.0846
18	1.1275	0.8461	0.7709	0.7941	2.7424
19	1.2387	1.1696	1.0509	1.1691	2.2087
20	0.7999	0.5654	0.4680	0.4495	3.0798
21	0.4926	0.4140	0.3862	0.3839	2.8339

5. Concluding remarks

- Color correlation histogram is a fast and simple form to generate a global fingerprint. However, is not enough descriptive for segments of different videos were the color correlation is near the same. Additionally, is not robust to common processing attacks that involve color correlation changes.
- Multimodal feature extraction strength the robustness against common processing video attacks.
- Extract the local features over preprocessed keyframes instead of the entire frame or frames sequence, decreases the computational cost. Decrement of robustness is compensated with the audio feature descriptor.
- SURF detection and extraction consumes the majority of total execution time in selected methods. For that reason, it is necessary to prove other faster robust feature extractors and descriptors.
- Cross recurrence plot is a widely used tool for compare two time series, however, it is still necessary to interpret this visual information in an automatically method.

6. Future work

The next step on this investigation is:

- Analyze other feature extractor and descriptor faster and more (or similar) robust to SURF.
- Test this other feature extractor and descriptor in both visual and audio components improving the obtained results in previous experiments.
- Design a fingerprint method, multilevel and multimodal, that utilizes the selected and improved approaches.
- Report results.

7. References

[1] WIPO, "Which products are typically affected? on Program Activities," 2013. [Online]. Available: http://www.wipo.int/enforcement/es/faq/counterfeiting/faq03.html. [Accessed 01 March 2013].

[2] S. E. Siwek, "The True Cost of Motion Picture Piracy to the U.S. Economy," IPI-Institute for Policy Innovation, 2006.

[3] B. Monnet and P. Véry, Les nouveaux pirates de l'entreprise. Mafias et terrorisme, Paris: CNRS, 2010.

[4] MPAA, "New study finds 23.8% of global Internet traffic invloves the illegal distribution of copyrighted work," 31 Jan 2011. [Online]. Available: http://www.mpa-canada.org/press/MPAA_January-31-2011.pdf. [Accessed 20 May 2013].

[5] S. Akhshabi, A. C. Begen and C. Dovrolis, "An Experimental Evaluation of Rate-Adaptation Algorithms in Adaptive Streaming over HTTP," *Proceedings of the second annual ACM conference on Multimedia systems,* pp. 157-168 , 2011.

[6] A. C. Begen, T. Akgul and M. Baugher, "Watching video over the web. Part 1: Streaming Protocols," *IEEE Internet Computing,* vol. 15, no. 2, pp. 54-63, 2011.

[7] J. F. Kurose and K. W. Ross, Computer networking. A top-down approach, Fifth ed., Addison-Wesley, 2010.

[8] Sandvine, "Global Broadband Trends 1H 2013," 16 07 2013. [Online]. Available: https://www.sandvine.com/trends/global-internet-phenomena/.

[9] YouTube, 2010. [Online]. Available: http://www.youtube.com/html5. [Accessed 2013].

[10] YouTube, "Google I/O - Adaptive Streaming for You and YouTube," 2013. [Online]. Available: https://www.youtube.com/watch?v=UklDSMG9ffU. [Accessed 7

October 2013].

[11] YouTube, "YouTube Official Blog," 2010. [Online]. Available: http://youtube-global.blogspot.mx/2010/07/whats-bigger-than-1080p-4k-video-comes.html. [Accessed 3 October 2013].

[12] G. Stearns, "YouTube Official Blog," 4 March 2010. [Online]. Available: http://youtube-global.blogspot.mx/2010/03/new-default-size-for-embedded-videos.html. [Accessed 11 october 2013].

[13] YouTube, "Supported YouTube file formats," 2013. [Online]. Available: https://support.google.com/youtube/troubleshooter/2888402?rd=1. [Accessed 15 October 2013].

[14] A. Finamore, M. Mellia, M. M. Munafò, R. Torres and S. G. Rao, "YouTube Everywhere: Impact of Device and Infrastructure Synergies on User Experience," *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pp. 345-360, 2011.

[15] A. Kapoor, "Adobe Developer Connection," 12 January 2009. [Online]. Available: http://www.adobe.com/devnet/adobe-media-server/articles/dynstream_on_demand.html. [Accessed 7 October 2013].

[16] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh and T. Holotyak, "Information-Theoretical Analysis of Private Content," in *IEEE Information Theory Workshop - ITW*, Dublin, 2010.

[17] J.-H. Lee, "Fingerprinting," in *Information Hiding. Techniques for Steganography and Digital Watermarking*, Artech House. Inc., 2000.

[18] S. Schrittwieser, P. Kieseberg, I. Echizen, S. Wohlgemuth, N. Sonehara and E. Weippl, "An Algorithm for k-anonymity-based," in *10th International Workshop on Digital-forensics and Watermarking IWDW 2011*, Atlantic City, New Jersey, USA, 2011.

[19] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich and T. Kalker, Digital Watermarking and Steganography, MA, USA: Morgan Kauffman Publishers. Elsevier, 2008.

[20] D. Milano, "White Paper: Video Watermarking and Fingerprinting. Rhozet A Business Unit of Harmonic Inc," 2010. [Online]. Available: http://www.digimarc.com/docs/technology-resources/rhozet_wp_fingerprinting_watermarking.pdf. [Accessed 2013].

[21] H. T. Sencar, S. Lian and N. Nikolaidis, "Content-based video copy detection – A survey," in *IntelligentMultimedia Analysis for Security Applications*, Berlin Heidelberg, Springer-Verlag, 2010, pp. 253-273.

[22] MPAA-Types of content theft, "Motion Picture Assosiation of America," 2013. [Online]. Available: http://www.mpaa.org/contentprotection/types-of-content-theft. [Accessed 28 Feb 2013].

[23] MPAA-Camcorder laws, "Motion Picture Assosiation of America," 2013. [Online]. Available: http://www.mpaa.org/contentprotection/camcorder-laws. [Accessed 23 September 2013].

[24] M. Ramona, S. Fenet, R. Blouet, H. Bredin, T. Fillon and G. Peeters, "A public audio identification framework for broadcast monitoring," *Applied Artificial Intelligence: An International Journal*, vol. 26, no. 1-2, pp. 119-136, 2012.

[25] N. Chen, H. D. Xiao, J. Zhu, J. J. Lin, Y. Wang and W. H. Yuan, "Robust audio hashing scheme based on cochleagram and cross recurrence analysis," *Electronic Letters*, vol. 49, no. 1, pp. 7-8, 2013.

[26] J. F. Kurose and K. W. Ross, Computer networking. A top-down approach, Fifth ed., Addison-Wesley, 2010.

[27] M. Visentini-Scarzanella and P. L. Dragotti, "Video jitter analysis for automatic bootleg detection," *Multimedia Signal Processing International Workshop on*, pp. 101-106, 2012.

[28] Verimatrix, 2012. [Online]. Available: http://www.verimatrix.com/solutions/forensic-watermarking.

[29] Civolution, 2013. [Online]. Available: http://www.civolution.com/applications/media-protection/.

[30] MarkAny, 2013. [Online]. Available: http://www.markany.com/en/?page_id=2123.

[31] DWA, "Digital Watermarking Alliance," 2013. [Online]. Available: http://www.digitalwatermarkingalliance.org/. [Accessed 25 Jul 2013].

[32] D. Milano, "White paper: Content Control- Digital Watermarking and Fingerprinting," [Online]. Available: http://www.digimarc.com/docs/technology-resources/rhozet_wp_fingerprinting_watermarking.pdf. [Accessed 19 september 2013].

[33] Harmonic, "Harmonic's Video Transcoding Solution Integrates YouTube Fingerprinting Technology," 20 April 2009. [Online]. Available: http://www.harmonicinc.com/news/harmonics-video-transcoding-solution-integrates-youtube-fingerprinting-technology. [Accessed 19 September 2013].

[34] YouTube, "YouTube," 2014. [Online]. Available: http://www.youtube.com/t/contentid. [Accessed 9 January 2014].

[35] Audible Magic, "Automated Content Recognition. Technology overview," 2014. [Online]. Available: http://www.audiblemagic.com/technology.php. [Accessed 10 January 2014].

[36] O. Pribula, J. Pohanka and J. Fischer, "Real-time Video Sequences Matching using the Spatio-Temporal Fingerprint," in *15th IEEE Mediterranean Electrotechnical Conference MELECON*, Valleta, Malta, 2010.

[37] L. Wang, Y. Dong, H. Bai, J. Zhang, C. Huang and W. Liu, "Content-based large scale web audio copy detection," *Multimedia and Expo IEEE Interntional Conference on* , vol. DOI: 10.1109/ICME.2012.17, pp. 961-966, 2012.

[38] Y. Zhang, M. Xu and E. Pratt, "Energy classification-assisted fingerprint system for content-based audio copy detection," *Communications (COMM), 9th International Conference on,* vol. DOI: 10.1109/ICComm.2012.6262598, pp. 35 - 38, 2012.

[39] Y. Tian, M. Jiang, L. Mou, X. Fang and T. Huang, "A multimodal video copy detection approach with sequential pyramid matching," *Image Processing (ICIP), 2011 18th IEEE International Conference on,* vol. DOI:10.1109/ICIP.2011.6116504, pp. 3629-3632, 2011.

[40] N. Chen, H. -D. Xiao and W. Wan, "Audio hash function based on non-negative matrix factorisation of mel-frequency cepstral coefficients," *IET Information Security,* vol. 5, no. 1, pp. 19-25, 2011.

[41] W. Son, H.-T. Cho, K. Yoon and S.-P. Lee, "Sub-fingerprint masking for a robust audio fingerprinting system in a real-noise environment for portable consumer devices," *IEEE Transactions on Consumer Electronics,* vol. 56, no. 1, pp. 156-160, 2010.

[42] N. Chen, W. Wan and H. -D. Xiao, "Robust audio hashing based on on discrete-wavelet-transform and non-negative matrix factorisation," *IET Communications,* vol. 4, no. 14, pp. 1722 - 1731, 2010.

[43] S. Lee and C. D. Yoo, "Robust Video Fingerprinting for Content-Based Video Identification," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 18, no. 7, pp. 983-988, 2008.

[44] S. Lee, C. D. Yoo and T. Kalker, "Robust Video Fingerprinting Based on Symmetric Pairwise Boosting," *IEEE Transactions on Circuits Systems for Video Technology,* vol. 19, no. 9, pp. 1379-1388, 2009.

[45] Y. Lei, W. Luo, Y. Wang and J. Huang, "Video Sequence Matching Based on the Invariance of Color Correlation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 22, no. 9, pp. 1332-1343, 2012.

[46] M. M. Esmaeili, M. Fatourechi and R. K. Ward, "A Robust and Fast Video Copy Detection System Using Content-Based Fingerprinting," *IEEE Transactions on*

Information Forensics and Security, vol. 6, no. 1, pp. 213-226, 2011.

[47] X. Li, B. Guo, F. Meng and L. Li, "A novel fingerprinting algorithm with blind detection in DCT domain for images," *AEU - International Journal of Electronics and Communications,* vol. 65, no. 11 DOI:10.1016/j.aeue.2011.03.005, pp. 942-948, 2011.

[48] X. S. Nie, J. Liu, J. D. Sun and e. al., "Robust video hashing based on representative-dispersive frames," *Science China Information Science,* vol. 062104(11), no. DOI: 10.1007/s11432-012-4760-y, p. 56, 2013.

[49] R. Roopalakshmi and G. Ram Mohana Reddy, "A novel spatio-temporal registration framework for video copy localization based on multimodal features," *Signal Processing,* p. http://dx.doi.org/10.1016/, 2012.

[50] R. Roopalakshmi and G. R. M. Reddy, "A framework for estimating geometric distortions in video copies based on visual-audio fingerprintings," *Signal Image and Video Processing. Springer-Verlag London,* no. DOI 10.1007/s11760-013-0424-7, 2013.

[51] Y. Tian, M. Jiang, L. Mou, X. Fang and T. Huang, "A Multimodal Video Copy Detection Approach with Sequential Pyramid Matching," *Image Processing (ICIP), 2011 18th IEEE International Conference on,* vol. DOI:10.1109/ICIP.2011.6116504, pp. 3629-3632, 2011.

[52] TRECVID, "Digital Video Retrieval at NIST," 2013. [Online]. Available: http://trecvid.nist.gov/. [Accessed 1 Aug 2013].

[53] MUSCLE, "Video Copy Detection Evaluation Showcase," 2007. [Online]. Available: https://www.rocq.inria.fr/imedia/civr-bench/data.html. [Accessed 1 Aug 2013].

[54] Internet Archive, "Internet Archive- Movie Archive," 2013. [Online]. Available: http://archive.org/details/movies. [Accessed 1 Aug 2013].

[55] Open Video Project, "The Open Video Project- A shared digital video collection,"

2013. [Online]. Available: http://www.open-video.org/. [Accessed 1 Aug 2013].

[56] Panasonic, "PT-LB2 Series LCD Projector," July 2010. [Online]. Available: https://panasonic.ca/english/broadcast/presentation/projector/pdf/B_PT-LB2_LB1U_Final.pdf. [Accessed 10 November 2013].

[57] Sony, "eSupport Sony," 2013. [Online]. Available: http://esupport.sony.com/LA/p/model-home.pl?mdl=DCRSX43R&template_id=2&region_id=2&tab=manuals#/manualsTab. [Accessed 10 November 2013].

[58] Internet, Portaltic, "La tasa de piratería de contenidos digitales en España roza el 80%," *Portaltic Europapress,* pp. http://www.europapress.es/portaltic/internet/noticia-tasa-pirateria-contenidos-digitales-espana-roza-80-20111108141034.html, 8 Nov 2011.

[59] J. Jenks, "M.P.A.A. Motion Picture Association of America. "2011 Theatrical Market Statistics. Latest box office and movie attendance trends"," 21 March 2012. [Online]. Available: http://www.mpaa.org/Resources/5bec4ac9-a95e-443b-987b-bff6fb545. [Accessed 01 March 2013].

[60] L.E.K., "The Cost of Movie Piracy. An analysis prepared by LEK for the Motion Picture Association. L.E.K. and MPA," 2006. [Online]. Available: http://austg.com/include/downloads/PirateProfile.pdf. [Accessed 23 feb. 2013].

[61] CMPDA, "IPSOS/OXFORD Economics study reveals $1.8 billion in losses across the canadian economy due to movie piracy," 17 Feb 2011. [Online]. Available: http://www.mpa-canada.org/press/CMPDA-CAFDE_News-Release_Ottawa-ON_February-17-2011_EN.pdf. [Accessed 20 May 2013].

[62] MPAA, "Types of Content Theft," 2013. [Online]. Available: http://www.mpaa.org/contentprotection/types-of-content-theft. [Accessed 28 Feb 2013].

[63] J. F. Kurose and K. W. Ross, Computer Networking. A Top-Down Approach, Addison-Wesley, 2010.

[64] P. Over, "Guidelines for TRECVID 2011," 4 Jan 2011. [Online]. Available: http://www-nlpir.nist.gov/projects/tv2011/tv2011.html#ccd. [Accessed 21 September 2013].

[65] C. Deng, Y. Zhang and X. Gao, "Robust Video Fingerprinting using Local Spatio-Temporal Features," in *International Conference on Computing, Networking and Communications, Cognitive Computing and Networking Symposium*, Maui, Hawaii, 2012.

[66] M. Li and V. Monga, "Robust Video Hashing via Multilinear Subspace Projections," *IEEE Transactions on Image Processing,* vol. 21, no. 10, pp. 4397-4409, 2012.

[67] S. Wei, Y. Zhao, C. Zhu, C. Xu and Z. Zhu, "Frame Fusion for Video Copy Detection," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 21, no. 1, pp. 15-28, 2011.

[68] X. Nie, J. Sun, Z. Xing and X. Liu, "Video fingerprinting based on graph model," *Multimedia Tools and Applications,* no. DOI 10.1007/s11042-012-1341-4, p. [On Line First], 2013.

[69] R. Roopalakshmi and G. R. M. Reddy, "Robust Features for Accurate Spatio-Temporal Registration of Video Copies," in *International Conference on Signal Processing and Communications (SPCOM)*, Bangalore, Karnataka, India, 2012.

[70] J. Xu, H. Zhao, K. Long and X. Yang, "An audio-video mixed fingerprint generation algorithm based on key frames," in *Communication Technology (ICCT), 2011 IEEE 13th International Conference on*, China, 2011.

[71] B. Coskun, B. Sankur and N. Memon, "Spatio–Temporal Transform Based Video Hashing," *IEEE Transactions on Multimedia,* vol. 8, no. 6, pp. 1190- 1208, 2006.

[72] X. Nie, J. Liu, J. Sun and W. Liu, "Robust video hashing based on double-layer embedding," *IEEE Signal Processing Letters,* vol. 18, no. 5, pp. 307-310, 2011.

[73] C. De Roover, C. De Vleeschouwer, F. Lefebvre and B. Macq, "Robust video hashing based on radial projections of key frames," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4020- 4037, 2005.

[74] X. Wu, C.-W. Ngo, A. G. Hauptmann and H.-K. Tan, "Real-time near-duplicate elimination for web video search with content and context," *IEEE Transactions of Multimedia vol. 11 no.2,* p. 196–207, 2009.

[75] S. Baudry, B. Chupeau and F. Lefèbvre, "A framework for video forensics based on local and temporal fingerpirnts," *Proceedings of IEEE International Conference on Image Processing (ICIP2009),* pp. 2889-2892, 2009.

[76] C. De Roover, C. De Vlesschouwer, F. Lefèbvre and B. Macq, "Robust video hashing based on radial projections of key frames," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 4020-4037, 2005.

[77] A. Massoudi, F. Lefèbvre, C. H. Demarty, L. Oisel and B. Chupeau, "A video fingerprint based on visual digest and local fingerprinting," *Proceedings of IEEE International Conference on Image Processing,* Vols. Atlanta, GA, USA, pp. 2297-2300, 2006.

[78] S. Baudry, B. Chupeau and F. Lefèbvre, "Adaptive video fingerprints for accurate temporal registration," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2010),* pp. 1786-1789, 2010.

[79] S. Baudry, "Frame-accurate Temporal Registration for Non-blind Video Watermarking," in *MM&Sec '12 Proceedings of the on Multimedia and security*, New York, USA, 2012.

[80] H. Bay, A. Ess, T. Tuytelaars and L. Van Goo, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding (CVIU),* vol. 110, no. 3, pp. 346-359, 2008.

[81] S. Wu and Z. Zhao, "A multimodal content-based copy detection approach," *International Conference on Computational Intelligence and Security (CIS),* pp. 280-

283 DOI:10.1109/CIS.2012.69, 2012.

[82] N. Chen, X. D. Xiao and W. Wan, "Audio hash function based on non-negative matrix factorisation of mel-frequency cepstral coefficients," *IET Information Security,* vol. 5, no. 1, pp. 19-25 DOI: 10.1049/iet-ifs.2010.0097, 2011.

[83] N. Chen, H. D. Xiao, J. Zhu, J. J. Lin, Y. Wang and W. H. Yuan, "Robust audio hashing scheme based on cochleagram and cross recurrence analysis," *Electronics Letters*, vol. 49, no. 1, pp. 7-8 DOI:10.1049/el.2012.3812, 2013.

[84] C. Y. Chiu and H. M. Wang, "Time-Series Linear Search for Video Copies Based on Compact Signature Manipulation and Containment Relation Modeling," *IEEE Transactions on circuits and systems for video technology,* vol. 20, no. 11, pp. 1603-1613, 2010.

[85] V. Varna and L. Wu, "Modeling an analysis of correlates binary fingerprints for content identification," *IEEE Transactions on Information Forensics and Security,* vol. 6, no. 3 part 2, pp. 1146-1159, 2011.

[86] D. Jan, C. Yoo and T. Kalker, "Distance Metric Learning for Content Identification," *IEEE Transactions on Information Forensics and Security,* vol. 5, no. 4, pp. 932-944, 2010.

[87] Netflix, "Netflix Support," 2013. [Online]. Available: https://support.netflix.com/en/node/87 and https://support.netflix.com/en/node/306. [Accessed 3 October 2013].

[88] YouTube, LLC, 2013. [Online]. Available: http://www.youtube.com/.

[89] J. Löfvenberg, "Random Codes for Digital Fingerprinting," Linköping Studies in Science and Technology Thesis No. 749, Department of Electrical Engineering Linköping University, SE-581 83 Linköping, Sweden, 1999.

[90] G. Tardos, "Optimal Probabilistic Fingerprint Codes," *Proceeding of the thirty-fifth annual ACM symposium on Theory of computing,* vol. DOI:10.1145/780542.780561, pp.

116-125, 2003.

[91] O. Blayer and T. Tassa, "Improved Versions of Tardos' Fingerprinting Scheme," *Designs, Codes and Cryptography*, vol. 48, no. 1, pp. 79-103, 2008.

[92] B. Skoric, T. U. Vladimirova, M. Celik and J. C. Talstra, "Tardos Fingerprinting is Better Than We Thought," *IEEE Transactions on Information Theory,* vol. 54, no. 8, pp. 3663-3676, 2008.

[93] E. Amiri and G. Tardos, "High rate fingerprinting codes and the fingerprinting capacity," *Proceeding SODA '09 Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms,* pp. 336-345, 2009.

[94] T. Furon, L. Pérez-freire, A. Guyader and F. Cérou, "Estimating the minimal length of Tardos code," in *Information Hiding*, Heidelberg, Springer-Verlag Berlin, 2009, pp. 176 - 190 .

[95] A. Simone and B. S. Skoric, "Accusation probabilities in Tardos codes: the Gaussian approximation is better than we thought.," *IACR Cryptology ePrint Archive 01/2010; 2010:472..*

[96] F. Xie, T. Furon and C. Fontaine, "On-off keying modulation and tardos fingerprinting," *Proceeding MM&Sec '08 Proceedings of the 10th ACM workshop on Multimedia and security,* vol. DOI:10.1145/1411328.1411347, pp. 101-106, 2008.

[97] B. Škorić, S. Katzenbeisser, H. G. Schaathun and M. U. Celik, "Tardos Fingerprinting Codes in the Combined Digit Model," *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY,* vol. 6, no. 3, pp. 906-919, Sep 2011.

[98] W. Trappe, M. Wu and K. J. R. Liu, "Anti-Collusion Codes: Multi-user and Multimedia Perspectives," in *Proc. IEEE International Conference on Image Process.,vol. 3, pp. 981–984,* Rochester, NY, 2002.

[99] M. Cheng and Y. Miao, "On Anti-Collusion Codes and Detection Algorithms for Multimedia Fingerprinting," *IEEE Transactions on Information Theory,* vol. 57, no. 7,

pp. 4843-4851, 2011.

[100] M. Cheng, L. Li and Y. Miao, "Separable Codes," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1791-1803, 2012.

[101] P. Meerwald and T. Furon, "Toward Practical Joint Decoding of Binary Tardos Fingerprinting Codes," *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, vol. 7, no. 4, pp. 1168-1180, Aug 2012.

[102] M. Kuribayashi, "Interference Removal Operation for Spread Spectrum Fingerprinting Scheme," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 403-417, 2012.

[103] A. Charpentier, C. F. Fountaine, T. Furon and I. Cox, "An asymmetric fingerprinting scheme based on tardos codes," *Information Hiding Lecture Notes in Computer Science*, Vols. 6958,, pp. 43-58, 2011.

[104] K. H. RHEE, "Evaluation of Multimedia Fingerprinting Image," in *Multimedia - A Multidisciplinary Approach to Complex Issues*, DOI: 10.5772/36370, 2012, p. Chapter 7.

[105] K. A. Saadi and A. Bouridane, "H.264/AVC Digital Fingerprinting Based on Content Adaptive Embedding," in *7th International Conference on Information Assurance and Security (IAS)*, Malacca, Malaysia, 2011.

[106] M. Chaumont, "Ensuring Security of H.264 Videos by Using Watermarking," in *Mobile Multimedia/Image Processing, Security, and Applications, Part of SPIE Defense, Security, and Sensing, DSS'2011, SPIE'2011*, Orlando, Florida, USA, Apr 2011.

[107] M. Koubaa, M. Elarbi, C. B. Amar and H. Nicolas, "Collusion, MPEG4 compression and frame dropping resistant video watermarking," *Multimedia Tools and Applications*, vol. 56, no. 2, pp. 281-301, 2012.

[108] S. P. Maity, S. Maity, J. Sil and C. Delpha, "Collusion resilient spread spectrum watermarking in M-band wavelets using GA-fuzzy hybridization," *Journal of Systems*

and Software, vol. 86, no. 1, pp. 47-59, 2013 .

[109] H. V. Zhao, W. S. Lin and K. J. R. Liu, "Cooperation and Coalition in Multimedia Fingerprinting Colluder Social Networks," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 14, no. 3, pp. 717-733, 2012.

[110] L. Caroll, Alice´s Adventures in Wonderland, simple text UTF-8 ed., http://www.gutenberg.org/ebooks/11.

7.1 Testing files references

Document	**Name**	**URL**
Alice	Alice's Adventures in Wonderland by Lewis Carroll	http://www.gutenberg.org/ebooks/11
Speech1	60 year old male american says stick it in the drive	http://www.freesfx.co.uk/download/?type=mp3&id=9792
RealNoise1	Eating an apple loudly	http://www.freesfx.co.uk/download/?type=mp3&id=10053

7.2 Video dataset references

#	**Name**	**URL**
1	Documental1	http://www.open-video.org/details.php?videoid=346
2	Documental2	http://www.open-video.org/details.php?videoid=348
3	Documental3	http://www.open-video.org/details.php?videoid=351
4	Documental4	http://www.open-video.org/details.php?videoid=354
5	Documental5	http://www.open-video.org/details.php?videoid=496
6	Documental7	http://www.open-video.org/details.php?videoid=400
7	Animated1	http://ia700401.us.archive.org/24/items/Popeye_forPresident/Popeye_forPresident_512kb.mp4
8	Animated2	http://ia600701.us.archive.org/21/items/TomAndJerryInANightBeforeChristmas/TomAndJerry-003-NightBeforeChristmas1941.mp4
9	Animated3	http://archive.org/details/woody_woodpecker_pantry_panic
10	Animated4	http://ia700406.us.archive.org/33/items/mother_goose_little_miss_muffet/mother_goose_little_miss_muffet_512kb.mp4
11	Sports2	http://ia700200.us.archive.org/3/items/TeamRyoukoPromoVideo/TeamRyoukoPromoVideo_512kb.mp4
12	Sports3	http://ia700209.us.archive.org/6/items/ScComboHunts/combo_512kb.mp4
13	TVComm1	http://www.youtube.com/watch?v=QbxEx2o8XIA
14	TVComm2	http://www.youtube.com/watch?v=lZUkEhWw0RI
15	TVComm3	http://www.youtube.com/watch?v=jSa_ZxTj6aw
16	TVComm4	http://www.youtube.com/watch?v=4KEBw6opgVk
17	TVComm5	http://www.youtube.com/watch?v=LGh0Uuo895c&feature=endscreen
18	TVComm6	http://www.youtube.com/watch?v=Fo31riY3mzM
19	TVComm7	http://www.youtube.com/watch?v=36kHzCCJkeM
20	Open1	http://www.bigbuckbunny.org/index.php/download/
21	Open2	http://orange.blender.org/
22	Open3	http://www.sintel.org/
23	Open4	http://mango.blender.org/

Appendix A. References for traitor tracing codes

Reference	Method	Code length	Results & error rates

J. Löfvenberg, "Random Codes for Digital Fingerprinting," Linköping Studies in Science and Technology Thesis No. 749, Department of Electrical Engineering Linköping University, SE-581 83 Linköping, Sweden, 1999.	Binary Random Fingerprints	$\boldsymbol{m}$
		$\geq \log V_M(c)$
		$\geq \log 2^{-\frac{1}{2}\log\frac{5M}{2}+MH(\frac{c}{M})}$
			m= 140 (length of code)
			M=50 000
			c=11
			M :number of user,
			c: number of colluders

G. Tardos, "Optimal Probabilistic Fingerprint Codes," *Proceeding of the thirty-fifth annual ACM symposium on Theory of computing,* vol. DOI:10.1145/780542.780561, pp. 116-125, 2003.	Fully randomized binary code	*Digit model, arbitrary alphabets:* $m = \Omega(c_0^2\ ln\frac{1}{\varepsilon_1})$ *randomized binary fingerpritning code* $m = 100c_0^2[\ ln\frac{1}{\varepsilon_1}]$, ***O(c2 log(n/ϱ)).*** $m = \mathrm{O}(c_0^2\ log(n/\varepsilon)$	(e1-e2)-secure with $\varepsilon_1 \ll \varepsilon_2$ That is: FN<<FP

O. Blayer and T. Tassa, "Improved Versions of Tardos' Fingerprinting Scheme," *Designs, Codes and Cryptography*, vol. 48, no. 1, pp. 79-103, 2008.	Optimization of variables on Tardos codes	**m=6.426 c^2 log(1/ϵ))**	c=20
			n=100
			e1=e2= 0.01

B. Skoric, T. U. Vladimirova, M. Celik and J. C. Talstra, "Tardos Fingerprinting is Better Than We Thought," *IEEE Transactions on*	Reevaluated the performance of the Tardos fingerprinting	$m = 4\pi^2 c_0^2[\ ln\varepsilon_1^{-1}]$ c0 large and e1 independently from e2	c>9

Information Theory, vol. 54, no. 8, pp. 3663-3676, 2008.	scheme by parameterizing its numerical constants and fixed functions.		

E. Amiri and G. Tardos, "High rate fingerprinting codes and the fingerprinting capacity," *Proceeding SODA '09 Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 336-345, 2009.	Combination of Tardos and weak fingerprinting codes	Code length to infinity
		The t-fingerprinting capacity (maximum achievable rate of t-secure fingerprint schemes)is
		$\Theta(\frac{1}{t^2})$
			t=2 R=.25 W=0.31 t: pirates
			t=3 R=0.0975 W=0.137
			t=7 R= 0.0168 W=0.025
			R&W rates (lower and upper bounds)

T. Furon, L. Pérez-freire, A. Guyader and F. Cérou, "Estimating the minimal length of Tardos code," in *Information Hiding*, Heidelberg, Springer-Verlag Berlin, 2009, pp. 176 – 190	Based on a rare event analysis	Code length is defined smaller than theoretic	$\varepsilon_2 = \varepsilon_1^{c_0/4}$
A. Simone and B. S. Skoric, "Accusation probabilities in Tardos codes: the Gaussian approximation is better than we thought.," *IACR Cryptology ePrint Archive 01/2010; 2010:472.*	Gaussian approximation to Tardos codes		Decouple e1 from e2 ≈0.5
F. Xie, T. Furon and C. Fontaine, "On-off keying	Symmetric q-ary Tardos +		2<=c<=20 Robust against

modulation and tardos fingerprinting," *Proceeding MM&Sec '08 Proceedings of the 10th ACM workshop on Multimedia and security*, vol. DOI:10.1145/1411328.1411347, pp. 101-106, 2008.	zero-bit side informed watermarking technique used with a on-off keying modulation,		block attacks, fusion and processing

B. Škorić, S. Katzenbeisser, H. G. Schaathun and M. U. Celik, "Tardos Fingerprinting Codes in the Combined Digit Model," *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, vol. 6, no. 3, pp. 906-919, Sep 2011.	Improvement of Tardos		C= 20
			m=100
			r= 0.01
			0.05
			0.1
			r: false positive

W. Trappe, M. Wu and K. J. R. Liu, "Anti-Collusion Codes: Multi-user and Multimedia Perspectives," in *Proc. IEEE International Conference on Image Process.,vol. 3, pp. 981–984*, Rochester, NY, 2002.	AND-Anti Collusion Codes	Short codes
		Spread spectrum insertion
			M=16
			C=3

M. Cheng and Y. Miao, "On Anti-Collusion Codes and Detection Algorithms for Multimedia Fingerprinting," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4843-4851, 2011.	Logical Anti Collusion Codes	Colluders tracing time = O(nM), M usuarios y n codeword length	
M. Cheng, L. Li and Y. Miao, "Separable Codes," *IEEE Transactions on Information*	Separable Codes	*~2-SCs with lengths 2 and 3*	Problems ~t-SCs with t=2 and length n>=4, and

Theory, vol. 58, no. 3, pp. 1791-1803, 2012.			t>2 and n>=t are wide open
P. Meerwald and T. Furon, "Toward Practical Joint Decoding of Binary Tardos Fingerprinting Codes," *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, vol. 7, no. 4, pp. 1168-1180, Aug 2012.	Continuous Tardos distribution	$M=10^4$ to 10^7 M=2048 $PFP=10^{-3}$	C={2,3,4,6,8} Robust against Random, majority, minority, coin flip, AWGN addition

Appendix B. References for traitor tracing codes and watermarking techniques

Reference	Method	Code parameters	Robustness
M. Kuribayashi, "Interference Removal Operation for Spread Spectrum Fingerprinting Scheme," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 403-417, 2012.	SS-spread spectrum watermarking scheme in the asymmetric fingerprinting protocol on Images	*Users* *$M=2^{20}$* *Colluders* *c=2 to 80* P $FP=10^{-4}$	Average JPEG compression QF: 75%, 50%
A. Charpentier, C. F. Fountaine, T. Furon and I. Cox, "An asymmetric fingerprinting scheme based on tardos codes," *Information Hiding Lecture Notes in Computer Science*, Vols. 6958,, pp. 43-58, 2011.	Fingerprint: Tardos + CE (commutative encryption) Watermark: SS + Composite signal representation	Same as Tardos	Same as spread spectrum
X. Li, B. Guo, F. Meng and L. Li, "A novel fingerprinting algorithm with blind detection in DCT domain for images," *AEU - International Journal of Electronics and Communications,* vol. 65, no. 11 DOI:10.1016/j.aeue.2011.03.005, pp. 942-948, 2011.	ACC modulated fingerprints embedded in DCT domain	*M=20* *c= 3*	JPEG FQ >55 Collusion attacks: Averaging Maximum Minimum Median, Minmax Randneg pd≈0.9

K. H. RHEE, "Evaluation of Multimedia Fingerprinting Image," in *Multimedia - A Multidisciplinary Approach to Complex Issues*, DOI: 10.5772/36370, 2012. Chapter 7.	Based on BIBD code.
	AND, OR, XOR → ACC
	Inserted in Y component and gray-
			Colluders traceable= n-1;
			n= number of users

	level of a color image		

K. A. Saadi y A. Bouridane, «H.264/AVC Digital Fingerprinting Based on Content Adaptive Embedding,» de *7th International Conference on Information Assurance and Security (IAS),* Malacca, Malaysia, 2011.	Tardos + SS adaptive insertion	n=100 usuarios, e1= 10^{-3}, c= 20,
		m= 92104 bits, embeds 10 bits per frame
			ave, min, max, minmax, modneg, med

M. Chaumont, "Ensuring Security of H.264 Videos by Using Watermarking," in *Mobile Multimedia/Image Processing, Security, and Applications, Part of SPIE Defense, Security, and Sensing, DSS'2011, SPIE'2011*, Orlando, Florida, USA, Apr 2011.	Tardos + SS over frame slices Additionally includes a study of watermarking and traitor tracing necessities.	Embeds 1 bit per frame slice	
M. Koubaa, M. Elarbi, C. B. Amar and H. Nicolas, "Collusion, MPEG4 compression and frame dropping resistant video watermarking," *Multimedia Tools and Applications,* vol. 56, no. 2, pp. 281-301, 2012.	SS + image mosaicing		Collusion MPEG compression Frame dropping
S. P. Maity, S. Maity, J. Sil and C. Delpha, "Collusion resilient spread spectrum watermarking in M-band wavelets using GA-fuzzy hybridization," *Journal of Systems and Software,* vol. 86, no. 1, pp. 47-59, 2013	SS optimized in M-WT. Uses a GA for hiding parameters and fuzzy for detection	C=70	fading and noise gain
H. V. Zhao, W. S. Lin and K.	Detection of colluding nodes in a social network.		

J. R. Liu, "Cooperation and Coalition in Multimedia Fingerprinting Colluder Social Networks," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 14, no. 3, pp. 717-733, 2012.	Study of cost effective colluder cooperation