ABSTRACT
Fighting movie piracy often requires automatic content identification. The most common technique to achieve this uses watermarking, but not all copyrighted content is watermarked. Video fingerprinting is an efficient alternative solution to identify content, to manage multimedia files in UGC sites or P2P networks and to register pirated copies with master content. When registering by matching copy fingerprints with master ones, a model of distortion can be estimated. In case of in-theater piracy, the model of geometric distortion allows the estimation of the capture location. A step even further is to determine, from passive image analysis only, whether different pirated versions were captured with the same camcorder. In this paper we present three such fingerprinting-based forensic applications: UGC filtering, estimation of capture location and source identification.

INTRODUCTION
According to [15], digital forensics is the use of analytical and investigative techniques to identify, collect, examine and preserve evidence/information which is magnetically stored or encoded, usually to provide digital evidence of a specific or general activity. The first question to answer when analyzing a suspect media file is to determine whether it is a copy of a copyrighted title, and if so which title it is. In such a copy identification context, watermarking is commonly used but not all copyrighted content is watermarked. Due to voluntary or natural content manipulations, such as analog to digital conversion (camcorder capture), compression, frame removing or adding (advertising), a multimedia identification/authentication process has to withstand natural distortions. Fingerprinting solves this problem, by extracting discriminating features, called fingerprints or multimedia DNA, representative and unique for each multimedia content. Section “Image and video fingerprinting” reviews existing works in image and video fingerprinting. And section “Filtering of UGC sites” illustrates content identification with fingerprinting for User Generated Content (UGC) filtering application.

Forensic tracking aims at identifying the traitor (the dishonest user) when an illegal copy is found, to put pressure on him, and possibly to sue him in court. Watermarking techniques handle that purpose: identification marks are embedded. They either identify the customer ID (in pay-per-view or video-on-demand application, or on DVD screeners) or the theater. Pirate profiling is a broad application task which comprises the extraction of any other useful information on the pirate habits and trends. Recording location estimation in a theater is tackled in section “In-theater piracy forensics”.

Source identification (or sensor forensics) aims at identifying the acquisition device that captured an image (digital camera, cell-phone or scanner) [12]. This means associating the image with a class of devices with common characteristics (device brand and model) and eventually matching the image to an individual device. Main clues for image source identification are sensor noise and artifacts in CCD array, lens distortions, demosaicing artifacts or sensor dust. This work, detailed in section “Sensor forensics”, has recently been extended to video, to determine whether two cams came from the same camcorder or not [1].

IMAGE AND VIDEO FINGERPRINTING
Image fingerprinting extracts discriminating features, called multimedia DNA or fingerprints, specific to each image. Several techniques exist to find these unique features. Some of them are semantic (high level analysis) and other ones are more signal based (low level analysis). In this paper we focus on signal-processing based “low level” processes. Low level processes are divided in two categories: global description and local description. The global approach considers an image as a global content such color/luminance histogram, texture based. The local approach considers an image as a multitude of spatially localized characteristics [8]. The global approach is faster but suffers from collision and lack of robustness against strong attacks such as cropping, occlusion, and addition. The local approach is more resistant against strong attacks and provides interesting properties to identify the geometrical distortion. The extracted features are repeatable, inter-dependent and more discriminant than global features. Local image fingerprints are mainly used in object recognition [8] and biometry. Security and output bit length are thus the main weaknesses of fingerprinting. Perceptual hash algorithms are designed to overcome these security issues. A perceptual hash is a hash function designed for multimedia contents. The properties of a perceptual hash function are: easiness of computation, weak collision resistance, and a fixed output bit length called
perceptual digest. The perceptual digest must be resistant and robust, in other words it shall remain the same before and after attacks that do not alter the perceived content.

Image fingerprinting is an established key technology for biometry and image indexing, but some recent works only have extended image description to video description. A video fingerprint can be a global description of the video (number of scene cuts, size of the video, etc.), a set of image fingerprints of still video frames, a set of video key frames fingerprints [10] or the description of the motion vectors. For instance in [5], the authors combine a few local fingerprints, measurement of a global similarity and fingerprint database management. The measurement of the global similarity is done after a correct fingerprint matching. It is the first scientific work which combines efficient fingerprint and efficient search in a fingerprint database. In [3], the authors identify a candidate movie by its MPEG motion vectors. This method is well integrated in the broadcast workflow but is encoder dependent and thus motion vector fields dependent. In [11], authors propose efficient and fast video fingerprint. The video is divided into frames. Each frame is also divided into blocks. For each block of each frame, the mean luminance (gray level histogram) is computed and compared to adjacent blocks and to adjacent frames (spatial-temporal 2×2 Haar filter). In [10], the authors combine a visual hash function and a local fingerprinting to describe a video content. The perceptual digest of each frame captures the video content variation and detects key frames. A local image fingerprint technique characterizes the detected key frames. The set of local fingerprints for the whole video summarizes the video or fragments of the video. The algorithm proposed in [10] is more detailed in section “Filtering of UGC sites” as a solution to identify contents over community sites such as UGC sites (YouTube, DailyMotion).

FORENSIC APPLICATIONS

Filtering of UGC sites

The User Generated Content (UGC) context provides an example of automatic content identification: “What is the title of this video? Am I in copyright violation if I distribute this content?” UGC sites such as YouTube or DailyMotion are becoming the platforms of choice to exchange and view videos. Unfortunately, indec factorial users also post copyrighted contents and UGC sites become a new channel of distribution of illegal contents. The limitation of the size of uploaded content is not a deterrent mechanism. According to Digital Ethnography [15], 200,000 videos were uploaded per day in March 2008. The average video length was 2 minutes 46.17 seconds. It means that 384 days of contents, more than one year of contents, were uploaded every day in March 2008.

Ethnography estimates that the percentage of video that are probably in violation of copyright is 12%. If we consider that some uploaded videos are removed immediately by YouTube, how many copyrighted contents are really uploaded every day? Of course, content owners request UGC sites to banish their contents. Thus UGC sites need methods to detect copyrighted content.

There are four main approaches. The first method is manual reviewing. The second method automatically analyses the filenames. The third method uses watermark to detect possibly copyrighted content. The fourth method, also the most efficient one, is based on fingerprints. But the system can only identify opus that it knows; therefore, a new opus has first to be registered into the system. In a content identification context, two symmetrical processes are therefore mandatory: registration (or indexing) and detection. To register a new opus, we submit the original file to the fingerprint extractor, which generates video fingerprints. The video fingerprints are stored in a database together with the name of the opus and optional metadata. To identify the origin of a sample, we submit the test file to the fingerprint extractor to generate its visual hash(es). The decision engine retrieves them and searches the database for the potential candidate; the response is either the name of the opus, or a failure statement (Figure 1).

The decision engine does not need the complete content: just a short fragment is sufficient to identify the content. The authors of [10] have developed a two-step adaptive video identification process for the UGC context. The technique combines a visual hash function and a local fingerprinting. The first step is a global analysis of the test file. Frames that present a high degree of change are detected as shot boundaries. The most significant frame per each shot is selected. And for each selected frame, a small, fixed-length, pixel dependent descriptor is extracted. This descriptor characterizes the global picture and forms the first level of content description. This global descriptor is resistant
to many transformations such as rotation, compression, temporal cropping, etc. The second step is a local analysis of the test file to cope with more serious transformations. For each frame selected by the first step, the system extracts a set of salient points also called points of interest. The sequence of these local descriptors forms the second level of content description. Obviously the local descriptors are longer than the global ones.

Content owners provide both types of above described global and local descriptors (fingerprints), of the content they want to be filtered by the UGC site. This constitutes a reference database. Once the reference database is populated, detection of copyrighted content can start on UGC sites. Every uploaded content on UGC sites is fingerprinted. The decision engine starts with the global descriptors and, in case of non-convergence, automatically switches to the local descriptors (Figure 2).

![Figure 2: Detection process](image)

This two-step strategy provides a good compromise between speed and robustness. In case of matching, there is a high probability of upload of copyrighted content. Some companies announce 99% of hit detection. But such accuracy figure does not mean anything if the whole context is not specified: duration of the original sequence, duration of the copy, manipulations (attacks) applied on the original... False positive and false negative rates increase with the strength of the attacks and the database size.

Once a copyright content identified, the UGC site has several possible actions. The most obvious one (and most current) is filtering copyrighted content. Content is not allowed to be posted. Some UGC sites are starting some contractual and commercial agreements with some content providers. The UGC site may allow the posting and will share the corresponding advertising revenues with content owner. It may even replace an identified copyrighted upload by the official version that has a guaranteed quality.

**In-theater piracy forensics**

Video fingerprints based on local points of interest can be used for more in-depth forensic analysis than simply identifications. For example, camcorder recording in cinemas often leads to geometrically distorted images, usually through the trapezoidal effect (also known as ‘keystoning’). Wang and Farid recently proposed a method to automatically detect camcorder captures of projected movies [14]. We proposed [1] to analyze the keystoning distortions, using the derived camcorder viewing angle to estimate the approximate position in the cinema from which the movie was captured. This can be done in three steps, using the fingerprints from the pirate copy and the corresponding master. The first step identifies pirate copies in peer-to-peer networks or on UGC sites by matching digital fingerprints to the master database. Second, the cinema in which the movie was recorded illegally is tracked down by decoding a forensic cinema-identification tag. The final step involves registration and seat localization. Temporal synchronization provides pairs of reference and copy frames. Spatial registration applied to pairs of points of interest from pirate and master copies allows construction of an eight-parameter homographic model (Equation 1) of the keystoning effect from which the camcorder viewing direction can be determined.

$$\begin{cases} x' = h_{00}x + h_{01}y + h_{02} \\ y' = h_{10}x + h_{11}y + h_{12} \\ h_{20}x + h_{21}y + 1 \end{cases}$$  

The system of equations and its resolution is detailed in [1]. Intersection of the viewing axis and a 3D model of the cinema’s seating plane eventually lead to the pirate’s seat.

![Figure 3: Capture location estimation](image)

The experimental results [1] show that the approximate capture location within the theater can be extracted from the pirate copy, with an acceptable accuracy. A first result would be to automatically determine whether the copy was captured from the projection booth or
from the seating area. Additionally, with regard to the measured standard deviation, it seems attainable to divide this seating area into about a dozen of zones and to automatically assign the capture position to one of them.

**Sensor forensics**

Source identification forensics looks for the device (digital camera or camcorder) that captured an image or video. Forensic methods for digital cameras may use metadata in image header, watermarking or fingerprinting. Metadata stored in image header may be the date, serial number of the camera, camera references, etc. But the metadata are not persistent in case of post-processing and file format conversions. Watermarking is another possibility but it requires embedding a watermarking algorithm inside the camera device. Only few manufacturers only are ready to include such a feature in their products. Tracing a digital camera device only based on the output patterns generated by this device was the purpose of the pioneering work by Lukas [9]. It is based on the observation that each sensor generates a specific noise, also called sensor fingerprint. Methods using alternative features to noise pattern were proposed for camera model identification. For example Kharrazi feeds 34 image features into a multi-class classifier [6] and Swaminathan estimates the interpolation filter coefficients of the color filter array (CFA) pattern and identifies the valid CFA for the analyzed image [13].

This sensor fingerprint technology suffers however from the same security weaknesses against “copy attack” [7] than some watermark algorithms. If anybody can create a sensor fingerprint from 50 images (according to the authors) taken by a digital camera A, an adversary can copy and paste the sensor fingerprint from the camera A into a new image. A malicious adversary can also remove sensor fingerprint from an image taken by a camera A and paste a new sensor fingerprint from a camera B into this image. Manufacturers make also efforts to reduce sensor noise. When sensor noise is removed, does sensor fingerprinting still make sense?

**CONCLUSION**

Fingerprinting is a powerful tool for media forensics. Not only does it provide fast and accurate copy identification, robust to signal distortions, without the embedding constraints of watermarking, but it also enables many other forensic tasks. In this paper we first reviewed the most popular image and video fingerprinting algorithms, focusing on our own method which was initially described in [10]. Filtering of copyrighted contents on UGC sites is the application in the spotlight, but they are many other possible uses of fingerprinting. We showed that an image description based on points of interest and associated local features enables the modeling of the geometric distortion. In case of in-theater camcorder capture, the capture location can be retrieved from this distortion model. Other specific fingerprints keep the tracks of the image sensor and can be used to identify the capture device. By combining all those various passive, non-intrusive, image and video analysis techniques, into a comprehensive forensic analysis scheme, one can learn a lot from an anonymous video file.

We must keep in mind, however, that as digital media forensics is a new emerging field, most, if not all, proposed schemes have not so far really considered serious attackers [4]. Possible goals of such attackers could be fooling copy identification, hiding content tampering or erasing (or even forging) content origin. Taking this in consideration opens the door to a new research area and a never ending arm races with the pirates.

**REFERENCES**