

Videntifier™ Forensic: Robust and Efficient Detection of Illegal Multimedia

Friðrik Ásmundsson[†] Herwig Lejsek^{†‡} Kristleifur Daðason^{†‡} Björn Þ. Jónsson[‡] Laurent Amsaleg[§]
[†]Eff2 Technologies ehf. [‡]School of Computer Science [§]IRISA–CNRS
Kringlan 1 Reykjavík University Campus de Beaulieu
IS-103 Reykjavík Kringlan 1, IS-103 Reykjavík 35042 Rennes
Iceland Iceland France
{fridrik, herwig, kristleifur}@eff2.net bjorn@ru.is laurent.amsaleg@irisa.fr

ABSTRACT

A large portion of the video material available on the Internet is distributed illegally. In this demonstration we present Videntifier™ Forensic, a new law enforcement solution for automatically identifying videos and images. Videntifier™ Forensic is very robust and efficient, even at a very large scale. We encourage ACM Multimedia participants to bring original videos and modified (yet visually acceptable) copies to challenge the capabilities of the system.

Categories and Subject Descriptors: K.4.1 [Computers and Society]: Public Policy Issues—*Abuse and crime involving computers*; H.2.4 [Database Management]: Systems—*Multimedia Databases*

General Terms: Algorithms, Performance.

Keywords: Video Detection; Robustness; Scalability.

1. INTRODUCTION

The Internet has revolutionized the consumption and distribution of video material. While most of the material is harmless and acceptable, a significant portion is illegal. When police suspect someone of distributing illegal material, they seize the suspect’s computer and all storage devices for investigation. Each video file must then be manually opened and viewed. Not only is this work expensive and time-consuming, but also very demanding and stressful.

The major challenge in automating the identification process is that videos are often modified and distributed in many different versions. For example, they may be resized to fit portable devices and remade for marketing purposes. For offensive material, such as child pornography, further modifications are often used to prevent discovery by the police, such as embedding the material within other videos.

In this demonstration we present Videntifier™ Forensic, a new system developed by Eff2 Technologies for automatically identifying videos and images. Videntifier™ Forensic has two key features that make it interesting to the multimedia community. First, it is very *robust*, accurately detecting difficult video transformations. Second, it is very *efficient*, even at a very large scale. We encourage ACM Multimedia participants to bring original videos and modified (yet visually acceptable) copies and challenge the system.

2. SYSTEM ARCHITECTURE

To identify videos, Videntifier™ Forensic uses a combination of fine-grained local image descriptors [2] and a large-scale multidimensional NV-tree index [5]. As soon as the system has “seen” a video once and its descriptors have been stored, it is capable of identifying future copies of that video. Due to the fine-grained description of the content, Videntifier™ Forensic is tolerant to many visual changes, including compression, camrips, subtitles, and mirroring. The system has three main components: client interface; secure fingerprint extraction unit; and central database server.

2.1 Client Interface

The interface to the Videntifier™ Forensic service is very straightforward. The detective connects the suspected storage device and instructs the software to investigate its content. Videntifier™ Forensic now starts the search process by splitting each video file into frames which are sent through a fingerprint extraction unit to the database server for identification. The results are then categorized as legal, illegal or unidentified. The detective can then export these results and/or print a report. The detective can also insert the unidentified clips into the collection for future reference.

2.2 Secure Fingerprint Extraction Unit

Videntifier™ Forensic provides a secure fingerprint extraction unit to its clients. The extraction of the visual fingerprints (local image descriptors) takes place inside this unit, which is a small server that acts as a mediator between police workstations and the central multimedia database server. The unit is equipped with a state-of-the-art NVIDIA graphics processor, which consumes video frames and extracts up to 250 GPU-Eff² descriptors from each frame. The GPU-Eff² descriptors are a new highly parallelized descendant of the SIFT family implemented with the CUDA programming environment. They have shown to be very tolerant to a large variety of image modifications [2].

In order to query for videos, Videntifier™ Forensic takes a set of samples from the video (720 frames per hour by default) and generates the fingerprints. As a single NVIDIA GTX 280 graphic card can extract descriptors from 40–60 video frames per second, this descriptor generation process is more than 100 times faster than real-time. The descriptors within each scene are grouped together and sent as a single query over an encrypted connection to the central NV-tree database server, which aggregates results from all scenes.

During insertion of videos into the central database server, a slightly different process is used. In this case, the client submits all the frames of the video to the FEU, which generates the fingerprints for each frame as before. Since subsequent frames are typically very similar, however, the fingerprints are likely to also be very similar. The FEU therefore applies a filtering step to the descriptor stream, to remove such redundant descriptors and store only very representative descriptors in the descriptor collection. Through this filtering step, the majority (90–98%) of all descriptors are removed, allowing the service to describe a typical hour of video content using only 100,000–300,000 descriptors.

In comparison, other systems focusing on video copy detection (such as [3]) typically try to identify so-called key frames (or stable scenes) and only extract descriptors from those frames. VidentifierTM Forensic uses instead a much more thorough process applied to the whole video, which selects the *descriptors* most representative of the content.

2.3 Multimedia Database Server

The central database server receives query requests from all fingerprint extraction units and searches both the legal and illegal collections. The results for all descriptors from each scene are aggregated and sorted by the number of votes each video receives. The result lists from consecutive scenes are then processed and the correlation between the queries and the result scenes is calculated. When the correlation is high, the query clips are marked as successfully identified. The name of the identified video along with its metadata is fetched from a relational database and sent back through the fingerprint extraction unit to the client interface.

The underlying NV-tree index is a high-dimensional index structure supporting efficient approximate nearest neighbor queries, which has been shown to outperform its competitors for large-scale multimedia retrieval [5]. While older versions of the NV-tree stored descriptor identifiers redundantly to improve recognition and were disk-bound as a result, we have since eliminated this redundancy while improving the quality of the results. This change allows the index to reside completely in main memory, even for very large descriptor collections, resulting in very significant efficiency gains. The NV-tree also supports efficient insert operations and basic write-ahead logging prevents critical loss of data in the event of crashes and allows recovery to a consistent state.

3. PERFORMANCE

In 2007, we demonstrated an early prototype at ACM Multimedia [1]. That prototype did not have GPU-based descriptor extraction and could handle fewer modifications. Furthermore, it had much more redundancy, as there was no filtering of descriptors and the NV-tree also stored descriptors redundantly. The system demonstrated here is therefore much more efficient, as described in the following.

First, the GPU-based descriptor extraction is more than 100 times faster than real-time. As the most computationally intensive part of the descriptor filtering is also performed on the GPU, real-time filtering for a single inserted video can be performed using less than 10% of the computation time.

As many fingerprint extraction units may connect to the database server, the NV-tree index must have both low response times and high query throughput. Compared to the previous (redundant) version of NV-Tree [5], which was able to perform around 100 descriptor queries per second, the

new non-redundant NV-tree is no longer I/O dependent. It can therefore perform up to 5,000 queries per second per CPU core, and scales nearly linearly with additional cores.

Insertion into the NV-tree, on the other hand, is I/O bound. When leaf nodes of the growing NV-tree are split, the insertion process must actually read the relevant descriptor data from disk in order to re-index those descriptors. As a result, insertion performance degrades linearly, albeit slowly, with index size. Nevertheless, our system is able to insert 2 million descriptors per hour into a collection of 2.5 billion descriptors (representing about 20,000 hours of video) using a single hard drive per index. This yields an overall insertion speed of 5–12 times faster than real-time.

VidentifierTM Forensic has already been adopted by the Icelandic police. In a performance evaluation performed on-site by police detectives, 112 video clips from previous investigations were inserted, and then modified with 33 different modifications, such as cropping, rotation, flipping, compression, rescaling, contrast/brightness changes, aspect ratio changes, subtitles and a picture-in-picture modification. In this evaluation, the system showed an identification rate of over 98% [4].

4. DEMONSTRATION SCENARIO

We will demonstrate the performance and capabilities of the entire system. The fingerprint extraction unit will be connected to a database server indexing more than 7,000 hours of video material. Using a prepared set of modified video clips, we will demonstrate the robustness and efficiency of the system, and also demonstrate how new videos can be inserted.

Furthermore, we will allow ACM Multimedia participants to interactively scan their USB sticks (or even hard drives) using VidentifierTM Forensic, in order to identify the video files stored on these devices. In particular, participants are encouraged to bring different modified versions, which are still visually acceptable, of the same video clip. By inserting the original clip and scanning for the modified versions, they can interactively test the system’s effectiveness.

5. REFERENCES

- [1] K. Dadason, H. Lejsek, F. H. Ásmundsson, B. T. Jónsson, and L. Amsaleg. Eff² Videntifer: Identifying pirated videos in real-time. In *Proc. ACM Multimedia (demo paper)*, Augsburg, Germany, 2007.
- [2] K. Dadason, H. Lejsek, B. T. Jónsson, and L. Amsaleg. Full GPU acceleration of local descriptors using CUDA. Technical report, Reykjavík University, 2009.
- [3] A. Joly, O. Buisson, and C. Frélicot. Content-based copy detection using distortion-based probabilistic similarity search. *IEEE Transactions on Multimedia*, 9(2):293–306, 2007.
- [4] H. Lejsek, F. H. Ásmundsson, K. Dadason, Á. T. Jóhannsson, B. T. Jónsson, and L. Amsaleg. Videntifer Forensic: A new law enforcement service for automatic identification of illegal video material. In *Proc. Multimedia in Forensics (MiFor)*, Beijing, China, 2009.
- [5] H. Lejsek, F. H. Ásmundsson, B. T. Jónsson, and L. Amsaleg. NV-tree: An efficient disk-based index for approximate search in very large high-dimensional collections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):869–883, 2009.