

Videntifier™ Forensic: Large-Scale Video Identification in Practice

Herwig Lejsek,^{†‡} Hrönn Þormóðsdóttir,[†] Friðrik Ásmundsson,[†] Kristleifur Daðason,^{†‡}

Ársæll Þ. Jóhannsson,[†] Björn Þ. Jónsson,[‡] Laurent Amsaleg[§]

[†] Videntifier Technologies ehf.
Ofanleiti 2
IS-103 Reykjavík
Iceland
herwig@videntifier.com

[‡] School of Computer Science
Reykjavík University
Menntavegi 1, IS-101 Reykjavík
Iceland
bjorn@ru.is

[§] IRISA–CNRS
Campus de Beaulieu
35042 Rennes
France
laurent.amsaleg@irisa.fr

ABSTRACT

Identifying videos on seized hard drives and other storage devices is a very tedious and time consuming task for forensic investigators. In particular, the vast amount of available material on the Internet and the large storage capacities of today's hard drives have become a strong headache for them. Videntifier™ Forensic is a recent service for forensic video identification, which is based on state-of-the-art high-dimensional descriptors and high-dimensional indexing. In this paper we describe how Videntifier™ Forensic tackles very large collections of video material and how robust it is towards standard modifications. We then present measurements that involve four different datasets and three collection sizes of up to 25,000 hours of video content. Our results show that Videntifier™ Forensic scales very well, both in terms of the efficiency and effectiveness of the service.

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: Videntifier™ Forensic; Video Identification; Service; Robustness; Scalability.

1. INTRODUCTION

Law enforcement agencies have been confronted with investigating vast amount of video material during the last few years. The amount of content distributed over the Internet is still increasing, as are the storage capacities of hard drives, resulting in private multimedia collections that typically contain many thousands of hours of video.

These collections must be investigated by the forensic investigation force of the police in case a person is suspected of consuming or distributing video material that is considered

to be opposed to the local law, especially in cases involving child abuse material, videos used for radical political propaganda or terroristic education videos.

Despite massive efforts from police forces in various countries, the amount of material that must be identified in each case is growing beyond their capacity. The investigator must manually open each and every file and inspect its contents. Of course, the same videos are found over and over again, but still they must be inspected in each separate case, as file names and contents are regularly reorganised and changed.

In order to reduce the workload for manual identification, several police forces have adapted the use of MD5 or SHA checksums to their investigation process. A checksum is a hash value that is computed from a video file and compared to a collection of known checksums – for both legal and illegal material. Unfortunately, however, checksum matching is prone to failure. For example, by changing even a single bit of a file, the checksum becomes different.

Indeed, practical evaluations made by the Danish police have shown that maintaining such a database is a very tedious task, and that the improvements by employing such a database are rather limited. The two major reasons are a) that too few resources are allocated for building up and maintaining a large database of hash values, and b) that hash values are simply not robust enough as they can only identify exactly the same file and there are often many different versions of the video content in distribution. This applies especially to illegal videos, such as videos containing child abuse material, where people are reorganizing their video files and stitching together their preferred scenes.

1.1 Non-Governmental Activities

Developing strategies to combat child abuse material on the Internet is a complex global problem. Not only police departments can benefit from the development of tools helping to automatically identify and/or classify multimedia content, but NGOs are also involved in the fight against the distribution of offensive multimedia content on the Internet. One of the most prominent organisations in Europe is INHOPE (Internet Hotlines Providers in Europe) founded in 1999 under the EC Safer Internet Action Plan, which represents Internet hotlines all over the world. Internet hotlines offer Internet users a way of reporting (anonymously if required) something they suspect to be illegal on the Internet and the hotlines investigate these reports to determine if

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MiFOR'10, October 29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-4503-0157-2/10/10 ...\$10.00.

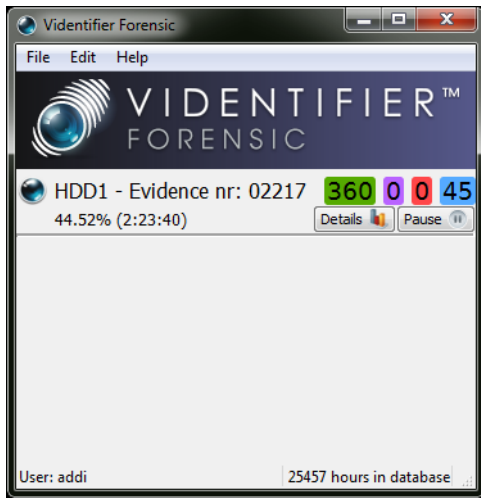


Figure 1: VidentifierTM Forensic Client Window.

the content is indeed illegal, and if so, trace the origin of the content and report to the law enforcement agency. Other NGOs such as “Save the Children” and the “National Center for Missing and Exploited Children” take a more general approach to the problem but also participate in INHOPE’s fight against online distribution of child abuse material.

Besides police and private NGOs, Internet Service Providers are third parties that carry responsibility in the fight against illegal usage of the Internet infrastructure. EuroISPA is the world’s largest association of Internet Service Providers representing over 1700 ISPs across the EU and EFTA countries. In November 2003, EuroISPA and INHOPE signed a Memorandum of Understanding between the two organisations which committed them to concentrate on issues of mutual interest, such as illegal and harmful use of Internet services.

1.2 VidentifierTM Forensic

VidentifierTM Forensic is a service which proposes to radically improve the forensic video identification process, by providing law enforcement agencies with a large-scale, yet robust, video identification system. In order to issue an identification task, the storage device to investigate needs to be connected to the computer; then its icon can be dragged into the VidentifierTM Forensic client window (see Figure 1). Then VidentifierTM Forensic scans the device and automatically identifies all videos by watching their visual content.

To view the results of the identification scan, the Details button on the Identification panel must be selected; the report window then displays a list of all videos found on the device along with the title and category of each identified video. A separate summary page shows statistics such as the time of the identification process and the total length of all videos found. Furthermore, it offers extended printing features and the possibility to import the identification results and their categorization into EnCase[®] Forensics, one of the most popular commercial Forensic tools available.

1.3 Underlying Technology

VidentifierTM Forensic is based on state-of-the-art video identification technology, and it performs its identification based on the actual contents of the videos. The service extracts 100–150 thousand local feature points from each hour of video. The visual signal around these points is then en-

coded into a sequence of numbers, the so-called visual descriptor or fingerprint of the point. VidentifierTM Forensic uses the GPU-Eff² descriptors [2] for its service, which are a member of the popular SIFT image descriptor family [4].

This method of video identification is very secure against data theft and data misuse, as the fingerprints themselves have no recollection of the actual content they represent and can never be connected to a specific person or case. This is of high concern to law enforcement agencies as they deal with highly sensitive data that they may neither share over any kind of public network, nor even expose to risks that such sensitive material could be stolen by finding leaks in the organisation’s firewall.

1.4 Contribution

In this paper, we focus on the scalability features of the service. In order to be usable in practise, a video identification system, such as the VidentifierTM Forensic, must be able to cover many thousands of hours of video content and be able to successfully identify all those thousands of hours of content. The two major issues for the scalability of such a system are *efficiency* and *robustness*. The efficiency is guaranteed by the constant search time of the NV-Tree index [3], as well as the efficient computation of the GPU-Eff² descriptors [2].

For the robustness, on the other hand, the key question is whether the identification capability, which has been proven at a small scale [4, 6], still holds at such a large scale. The major contribution of this paper is to demonstrate, through detailed experimentation, that the VidentifierTM Forensic service is indeed robust at a very large scale.

The details of the VidentifierTM Forensic service were presented in [6], and will not be repeated here due to space constraints. The remainder of the paper is then organized as follows. We first describe the details of the collections and hardware used in our experiments in Section 2. We then describe three different experiments, in Sections 3 through 5. Finally, we give an overview on related work and conclude.

2. SCALING VIDENTIFIERTM FORENSIC

The VidentifierTM Forensic service was presented in [6], which described the architecture of the system and presented the individual components. The acceptance test presented in that work showed that 98.6% of the modified video clips could be successfully identified. That test was performed on a very small database of only 100 hours of video content, which is neither representative nor useful in practise. In this paper, however, we present results for a collection of more than 25,000 hours of video. In this section we give an overview of our experimental setup, including the video collection, query collections, and hardware settings.

2.1 Scaling the Collection

Suspects of downloading illegal video material are typically large-scale collectors of all kinds of content. Such collections contain content of different types, ranging from commercial Hollywood movies, through popular TV shows, to different kinds of adult video content, sometimes even including child abuse material. VidentifierTM Forensic has therefore been focusing not only on indexing descriptors of so-called “illegal” video content, but rather a broad range of commonly available video content. This is useful because it varies significantly between different countries which video

content is considered “illegal”, and even varies according to the individual regulations within these countries.

The main focus of this paper is to present results for large scale video identification using the VidentifierTM Forensic service. In order to do so, we have calculated visual fingerprints from about 25,457 hours of video material, consisting of 12,197 hours of published movies and 13,260 hours of TV series and TV shows. All this material yielded a total collection of 3,176,275,670 (or more than three billion!) 72-dimensional GPU-Eff² descriptors. During the creation of this collection, we have run experiments evaluating the identification capability for various collection sizes, and some of those results are reported here.

2.2 Overview of Experiments

In this paper, we present three different experiments, designed to measure a) the impact of scale on the identification capability for regular police work, b) the robustness of the system against various video modifications, and c) the robustness of the system against false positives.

For the first experiment, the Icelandic police chose four different evaluation hard drives containing many different video files (mostly non-offensive content, however) which VidentifierTM Forensic was supposed to automatically identify. Since the identification capability is expected to be growing with the growing collection of video fingerprints, this experiment was performed at three different points in time, with collections of 5,005 hours, 16,247 hours, and finally with the entire collection of 25,457 hours of video material. For each collection we measured the percentage of the video material on the hard drives that could be correctly identified, as well as whether any false positives appeared in the identification process.

In the second experiment, we repeated the acceptance test of [6]. We used the final reference collection, and inserted 112 different 10-minute long reference clips prepared by the police into the central database. We then selected a single 1 minute clip out of each of the 112 samples, and applied 33 different modifications to each clip. The modified versions were then used to query the collection, and the exact recall was measured. This experiment is completely compatible with the results reported in [6]. The goal was to learn how much a 250 times larger reference collection would affect the robustness of the VidentifierTM Forensic system.

In the third and final experiment, we extracted 2,576 one-minute clips from the MUSCLE Video Copy Detection benchmark, which contains 101 videos from different sources, with a total length of 80 hours [5]. This material was likely to not be contained in our central database as it was comprised of freeware videos, mostly footage from amateur or semi-professional video producers. It was therefore very unlikely that clips from this collection would have a representation in our large fingerprint database, allowing us to gauge the likelihood of false positives in our service.

2.3 Hardware Configuration

We used three different computers for our experiments. A Windows desktop machine was used for the VidentifierTM Forensic client. The hard drives to investigate were connected to this machine and the client was used to measure quality and performance of the identification procedure.

The second computer was a desktop machine, equipped with a 3GHz Intel Core 2 E8400 processor, 4GB RAM and

Disk	Video Files	Total Length	Total Size
1	179	210 hours	100 GB
2	935	594 hours	293 GB
3	384	228 hours	115 GB
4	213	151 hours	84 GB

Table 1: Summary of evaluation disks selected by the Icelandic police.

an NVIDIA GTX 275 GPU. It acted as a Fingerprint Extraction unit, responsible for extracting the fingerprints from the video frames and sending them to the third computer.

The third computer was a server computer equipped with 2 Xeon 8400 Quad-Core processors and 72 GB DDR3-RAM, which was used to store and query the fingerprint collection. When running the largest of our three collections, approximately 75% of the available memory was used, keeping the whole search index (NV-Tree) in main memory for very fast nearest neighbour retrieval of the descriptors.

3. EXPERIMENT I: IMPACT OF SCALE

3.1 Experimental Setup

As mentioned above, the Icelandic police selected four different hard disks from previous cases for our analysis. Table 1 gives an overview of the contents of the selected disks.

There are three observations worth noting for these disks. First, disks 1 and 2 were seized from the same person, and can thus be expected to yield similar results. Disks 3 and 4, on the other hand, are completely independent.

Second, while most of the video material found on these disks consists of well-known movies and TV shows, a small percentage was Icelandic material, including popular shows and movies that are produced in Iceland each year. This does not bias results much as our collection has a very limited scope of Icelandic content, and in fact the Icelandic material was identified less frequently than the international material. National material is, however, an important concern for many European countries, which produce many popular TV shows and movies.

Third, only disk 4 contained any significant fraction of content that can be categorized as adult content (about 15% of the files). Currently our reference collection does not include fingerprints of adult content (though we will be adding it in the near future), so we expect the results for this disk to be significantly lower than the other disks. Note, however, that in the robustness study of Section 4, several of the inserted 10 minute clips contain adult material.

These four hard drives were investigated three times over the course of the last year, as the size of the central fingerprint database grew over time. The first run was performed in November 2009 on a collection of 5,005 hours of video content (mostly movies); the second run in March 2010 on a collection of 16,247 hours of video content (both movies and TV shows); and the third run in May 2010 on the entire collection of 25,457 hours of video content.

For all the measurements, a standard sampling rate of 10% of the video content was used. Each sample was up to 1 minute in length, containing 60 query frames (at 1 fps) and 1000 nearest neighbours per query descriptor from the central NV-Tree database. The starting point of each sample was selected randomly from the file, but care was taken to avoid overlap between different samples. With this configu-

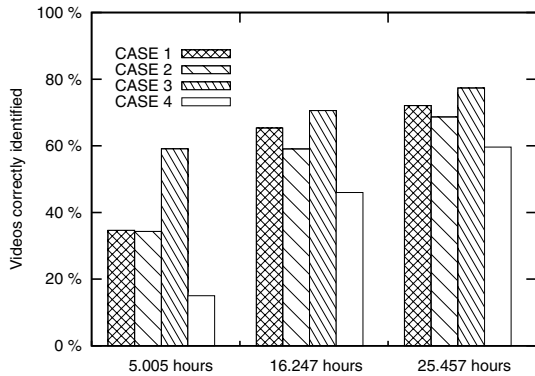


Figure 2: Identification rate for four selected disks over three different collection sizes.

ration, the evaluation speed on the largest dataset was between 47 and 67 faster than realtime. The differences in the search time can be explained by several factors: a few large video files can be processed faster than many smaller ones; the content itself affects the number of descriptors extracted, which in turn affects speed; and the reading speed of the hard drive affects the speed, as well as the codecs/resolution used for the various video files.

3.2 Results

The results of these evaluation runs, which were performed by the Icelandic police, can be found in Figure 2. Overall, we observe that the identification rate grows steadily as content is added to the central database. The growth rate is higher when growing from 5 to 16 thousand hours, than for the second growth period from 16 to 25 thousand hours. As we expect a logarithmic relation between the collection size of the central database and the identification rate, this is to be expected (the collection size tripled in the first step, while growing only about 60% during the second step).

Figure 2 also shows clearly that disks 1 and 2 show a similar identification rate, as they were seized from the same person, while disks 3 and 4 were completely independent. The slightly lower identification rate of disk 4 is (partially) explained by the share of adult content.

In summary, these results show that with the currently available collection of 3,1 billion fingerprints originating from over 25 thousand hours of video content, the Videntifier™ Forensic service can already save about 70% of the time that was previously spent in the manual identification process. The trend also shows that as the collection grows, that fraction is likely to get even larger.

4. EXPERIMENT II: ROBUSTNESS

4.1 Experimental Setup

To verify the robustness of Videntifier™ Forensic, we ran the acceptance test that was performed by the Icelandic police forces in June 2009 [6]. To recall the configuration of the acceptance test, the Icelandic police selected 112 videos from previous cases, containing conventional video material (such as Hollywood movies and TV series), as well as a set of adult videos. From each of these 112 videos, a 10-minute clip was randomly selected and inserted into the full database.

A 1-minute query clip was then randomly selected from each of the 10 minute clips. Each of these 112 original query

clips was then modified with 33 predefined modifications, resulting in a total of $112 \times 33 = 3,696$ query clips. Most of the modifications are very common video transformations, a detailed description can be found in [6].

For these short query clips, we chose again a coverage of 100% (as was done in [6]), and ran each query clip until completion. In order to better understand the robustness of the system, however, we varied two parameters that impact both query time and detection capability.

First, we varied the sampling rate from 2 frames per second (2 fps, resulting in ~ 120 query frames for a 1-minute clip; this was the setting used in [6]), through 1 fps (~ 60 query frames), down to only 0.5 fps (~ 30 query frames). Reducing the number of frames per second by a factor of 2 each time obviously yields an immediate reduction in identification time. The number of query descriptors is cut roughly in half, but the reduction in evaluation time is not as dramatic, due to overheads in network traffic and general frame processing. Still, a 30–40% reduction in query processing time is seen when going from 2 fps to 1 fps, and again from 1 fps down to 0.5 fps.

Second, we varied the number of nearest neighbours (NN) retrieved from the NV-Tree database. In [6], each query descriptor retrieved 500 NN, but we expected that in a collection that was about 250 times larger, more neighbors would increase the accuracy of the results (resulting in fewer misses). We therefore ran the experiment with 500 NN, 1,000 NN, and 2,500 NN. Note that pure retrieval time from the database was increased by about 20% from 500 to 1,000 NN, and by about 33% going from 1,000 to 2,500 NN. The final timing measurements presented in Table 2 include furthermore overhead spent on network traffic and calculations within the decision module.

4.2 Results

Table 2 shows the robustness of the Videntifier™ Forensic system, for the various parameter settings. For comparison, the column on the far left shows the results reported in [6] (100 hours of video, 500 NN, 2 fps). Then, from left to right, the table shows the results for the experiments over the collection of 25,457 hours. The three last rows summarize the results in terms of identification time, detection capability and missed clips (false negatives). In our analysis, we focus particularly on the missed clips, as this is the key measure to evaluate robustness.

Overall, we observe that both parameters impact the robustness of the system. In particular, the difficult transformations, such as significant changes in aspect ratio (descriptors of the SIFT family are known to tolerate only moderate affine modifications), severe cropping, or a major decrease in brightness, benefit from the additional neighbours that are retrieved when going from 500 NN to 2,500 NN. The best results, obtained with settings of 2,500 NN and 2 fps, show that only 4.3% of the query clips are missed. While this is about three times more than reported in [6], it is still an excellent result for a collection that is 250 times larger.

When analysing the detailed results, we observed that the two parameters benefit robustness in a quite different way. This is due to the decision process and the way it concludes whether a query clip matches a reference clip in the database. It is based on two major principles. First, a sequence of query frames must yield a match within the database. In order for a frame to yield a match, a minimum

Modification	MiFor 2009	500 NN			1000 NN			2500 NN		
		0.5 fps	1 fps	2 fps	0.5 fps	1 fps	2 fps	0.5 fps	1 fps	2 fps
ADD BORDERS	112	77	96	98	94	100	105	99	109	109
BLUR	112	106	111	111	108	110	112	110	112	112
BRIGHTNESS 50	96	31	45	52	39	54	64	50	70	75
BRIGHTNESS 150	112	105	111	111	108	110	111	111	112	112
CONTRAST 50	112	101	109	109	103	111	112	109	112	112
CONTRAST 150	112	103	110	111	107	109	111	109	112	112
CROP 10	112	71	100	102	96	103	104	101	109	111
CROP 25	111	33	57	72	52	80	89	79	97	101
FLIP HORIZONTAL	112	106	111	112	108	112	112	111	112	112
FLIP VERTICAL	112	106	111	112	108	111	112	112	112	112
GAMMA 70	112	81	101	105	96	105	109	104	110	110
GAMMA 180	112	80	100	102	94	106	108	107	110	112
GRAYSCALE	112	107	111	112	110	111	111	111	112	112
HORIZONTAL REDUCE	105	0	2	2	1	11	27	29	63	78
INFO	106	85	97	100	91	100	104	94	106	109
NOISE	112	101	111	111	105	110	112	109	112	112
PAL DEINTERLACE	112	106	111	112	108	112	112	110	112	112
PIP CENTER	103	26	54	72	41	82	90	73	98	102
PIP TOP RIGHT	109	57	83	89	78	95	101	92	107	111
RAIN	112	98	108	109	101	108	111	110	112	112
RESCALE 25	109	89	101	104	97	104	108	103	109	109
RESIZE 50	112	104	112	111	109	111	112	111	112	112
REVERSE	112	107	111	112	108	111	112	112	112	112
ROTATE 90 LEFT	112	93	104	107	100	110	112	106	112	112
ROTATE 90 RIGHT	112	87	102	106	98	110	111	107	111	112
ROTATE 10	112	94	103	104	99	108	112	103	110	112
FRAMERATE 12	112	107	111	112	109	111	112	111	112	112
FRAMERATE 2	111	107	110	109	109	111	111	109	111	112
SHARPEN	112	107	112	112	108	112	111	110	112	112
SHIFT	112	91	103	105	98	108	109	106	108	111
SPOTLIGHT	112	65	87	94	76	93	98	90	97	103
SUBTITLES	112	96	107	109	103	108	111	106	111	112
VERTICAL REDUCE	106	0	0	0	1	8	19	23	53	69
Identification Time	3.9 h	3.5 h	5.1 h	8.1 h	3.7 h	6.0 h	9.3 h	4.8 h	8.4 h	14.0 h
Total Recognition	98.6%	73.7%	83.9%	86.3%	80.2%	87.8%	90.8%	87.3%	93.9%	95.7%
False Negatives	1.4%	26.3%	16.1%	13.7%	19.8%	12.2%	9.2%	12.7%	6.1%	4.3%

Table 2: A quality comparison of the descriptor schemes for different image modifications.

percentage of the query descriptors must match with a particular scene (collection of frames) in the central database. Second, the individual frame matches must be correlated over time, which means that a minimum number of consecutive frames must yield a match with a set of consecutive scenes in the database.

Turning back to the impact of the two measured parameters, increasing the numbers of nearest neighbours helps the system to find more matches to individual frames, as the frames have a higher likelihood of surpassing the identification threshold for a frame match. When whole scenes of difficult frames arise, the increased number of consecutive frame matches helps the correlation-based decision process to find a stable correlation between the query and reference frames, resulting in better identification.

Increased sampling density, on the other hand, mostly benefits the detection of very short matches. In very fast-moving (action) scenes or in scenes with very dim lighting (where there are few query descriptors!), it is difficult to get successful identification. But even fast-moving action scenes are sometimes interrupted by short scenes that are slower and more stable, and then increased sampling density helps to capture those scenes. Also, most dark scenes occasionally contain a bit more detail, where the increased sampling density kicks in and can, for a short period of time, push the matching signal above the matching threshold.

5. EXPERIMENT III: FALSE POSITIVES

5.1 Experimental Setup

The previous experiments showed that the VidentifierTM Forensic system is highly robust to various video modifications. What remains, however, is to investigate the potential for false matches, when the query clip is not in the indexed collection. When running the experiments presented in Section 4, we already monitored the results to find potential false positive matches within the 3,696 query clips. The 10 minute samples inserted into the database were assigned to a special category, so that a false match would automatically stand out by falling into one of the other categories. The result was that in fact no false positives were found.

As an additional proof of the extremely low false positive rate we performed a third experiment. We extracted 2,576 one-minute clips from the MUSCLE Video Copy Detection benchmark, which contains 101 videos with a total length of 80 hours [5]. We then queried the large collection of 25,457 hours, using 2,500 NN and 2 fps.

5.2 Results

We expected that none of these clips would be found in our reference collection. The results did indeed contain no false positive matches to any of the queries. Surprisingly, however, there were some real matches, which we verified

by comparing the query clips and the matched clips visually, and by inspecting the video titles.

First, we discovered that one of the 101 freeware movies from the MUSCLE benchmark had a representative in our database, as 52 of the clips matched the movie “La Figlia di Frankenstein”. Furthermore, there were three other query clips that yielded a match; this time, however, just in a small part of the actual clip. We inspected also these three clips and discovered that each clip contained partial footage of the 9/11 disaster. The respective scenes matched with the movie “The Flight That Fought Back”, a docudrama about the United Airlines Flight 93 containing exactly the same footage and can thus not be considered a false positive.

Beyond these 55 clips that yielded an exact match, there were two other clips which indicated a potential match. The first of these two clips yielded a score of 13.8% on a scale from 0% (no match) to 100% (exact match). Investigating the content of the query, we saw that it contained archive footage from the attack on Pearl Harbour in 1941. Watching at the matching scene from the movie in our reference database, we found that it was the title “Killing Hitler”, containing exactly the same footage, but in better quality and with a different aspect ratio. The second clip with a potential match was in fact a “false positive”. The query video consisted solely of text (credit lists) and individual letters matched a scene from the TV series “Penn & Teller: Bullshit!” containing the same letters in the same font, sometimes even in the same order. The score of this potential match is very low, or 1.4% (there is a high barrier before yielding an actual false match). In particular the query only matched in 4 individual frames and the correlation between the matching frames was only moderate.

All remaining 2,519 clips yielded no signal for any of the videos in the indexed collection, and thus no false positives.

6. RELATED WORK

There are two further projects we are aware of that are concerned with developing tools for the same or similar purposes as Videntifier™ Forensic. The FIVES (Forensic Image and Video Examination Support) project, a EU project coordinated by the University of Karlstad, focuses on the development of a software tool-set for law enforcement organizations to handle large amounts of image and video material. These tools incorporate a whole set of different technologies. Besides image/video fingerprinting, they include sound processing, face detection and the extraction of text from images and video frames. Another focus is the presentation of the visual material in order to speed up manual investigation, such as creating short summaries of video files and assigning a porn detection score to the videos. The tool-set is completed with a set of basic file operation, such as hashing, filename and file fragment matching.

The I-DASH (Investigator’s Dashboard) project coordinated by the University of Amsterdam has a goal and technology most similar to Videntifier™ Forensic. The I-DASH solution is built on VizXview, a tool developed by the company ZiuZ to create short summaries of larger video files, and aims to integrate video fingerprinting in order to allow automatic detection of known video snippets containing child abuse content. In contrast to Videntifier™ Forensic, I-DASH’s solution focuses especially on building up a database of video fingerprints of child abuse material and distributing this database to the individual police investiga-

tion departments while Videntifier™ Forensic aims to build up a central database containing fingerprints of both popular video material and illegal content.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented extensive experimental results for the Videntifier™ Forensic system. Our results show that within a large-scale reference collection of more than 25 thousand hours of video, Videntifier™ Forensic can save law enforcement agencies significant time and costs by automatically identifying most of the video material that is currently being distributed over the Internet.

We have presented measurements performed by the Icelandic police on four different hard drives, which show that more than 70% of the material from their benchmark cases could be identified with this collection. We are planning to grow our reference collection even further, in order to increase the identification ratio of the service.

We also showed that, even for such a large-scale collection, the Videntifier™ Forensic system is very robust towards attacks on the video contents, through common video transformations, while not detecting any false positives in any of our experiments. We believe, however, that the robustness can be improved even further, in particular with respect to aspect ratio changes and affine distortions, which are very common video modifications. One possible strategy is to occasionally take short query samples, modify them by distorting the aspect ratio, and check whether the modified video finds a match. As the query evaluation of the central database is very fast, this might be a valid strategy to overcome this bottleneck. Another strategy, of course, is to improve the description to better handle these attacks. Studying these transformations is part of our future work.

Finally, our experiments showed that due to the increased size of the collection, some tuning of the frame sampling and the number of nearest neighbors was necessary, in order to achieve results of such high quality. These results also indicate, however, that the system can offer a variety of tradeoffs between efficiency and effectiveness, which may benefit different customers of the Videntifier™ Forensic service.

8. REFERENCES

- [1] F. H. Ásmundsson, H. Lejsek, K. Dadason, B. T. Jónsson, and L. Amsaleg. Videntifer Forensic: Robust and efficient detection of illegal multimedia. *Proc. ACM Multimedia (demo paper)*, Beijing, China, 2009.
- [2] K. Dadason, H. Lejsek, B. T. Jónsson, L. Amsaleg. Full GPU acceleration of Eff2 descriptors using CUDA. *Proc. ACM Multimedia*, Firenze, Italy, 2010.
- [3] 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 TPAMI*, 31(5):869–883, 2009.
- [4] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [5] MUSCLE Video Copy Detection Evaluation Benchmark. www-rocq.inria.fr/imedia/civr-bench.
- [6] H. Lejsek, Á. Jóhannsson, F. Ásmundsson, B. Jónsson, K. Dadason, and L. Amsaleg. Videntifer™ Forensic: A new law enforcement service for automatic identification of illegal video material, *Proc. MiFor*, Beijing, China, 2009.