Investigating visualisation techniques for rapid triage of digital forensic evidence

Gavin Hales

Ethan Bayne

Investigating Visualisation Techniques for Rapid Triage of Digital Forensic Evidence

Gavin Hales, Ethan Bayne

Division of Cyber Security, Abertay University, Dundee, UK
{gavin.hales, e.bayne}@abertay.ac.uk

Abstract. This study investigates the feasibility of a tool that allows digital forensics (DF) investigators to efficiently triage device datasets during the collection phase of an investigation. This tool utilises data visualisation techniques to display images found in near real-time to the end user. Findings indicate that participants were able to accurately identify contraband material whilst using this tool, however, classification accuracy dropped slightly with larger datasets. Combined with participant feedback, the results show that the proposed triage method is indeed feasible, and this tool provides a solid foundation for the continuation of further work.

1   Introduction

Digital forensics (DF) is considered an essential practice of law enforcement, often employed when crimes may have been committed with the aid of an electronic device. During the evidence collection phase of a DF investigation, data is recovered from digital devices that are suspected to be involved in a crime. Devices seized may include computers, external storage devices, storage media, and mobile devices—such as a smartphone or tablet. The evidence collection phase includes the recovery of deleted data that may be of forensic significance to the investigation, in a process known as file carving. A frequent objective of file carving when used in law enforcement is to recover photographic evidence that may be important to the case.

The file carving stage of an investigation is generally a hands-off process where software is used to recover data. During recovery, the investigator cannot proceed with reviewing results until the file carving operation is complete. Currently available software that is used to complete file carving operations tend to be slow and often provide the investigator with very little feedback about the content found on a device during processing. Previous work has been conducted that utilised Graphics Processing Units (GPUs) to accelerate computationally intensive pattern matching operations in the file carving process, significantly reducing the time required to perform the reconstruction

of evidence. However, visual feedback of evidence discovered during processing is an area that remains relatively unexplored and may benefit from further research.

This work examines the possibility of applying information visualisation techniques to the collection stage of the investigation, in unison with GPU accelerated file carving, to allow the DF investigator to rapidly triage datasets. The motivation behind this work is the increasing average workload for DF investigators, specifically in the area of law enforcement. These ever-growing workloads lead to longer investigation times and delays in the processing of cases. Through allowing the investigator to triage and prioritise relevant datasets at an earlier stage of the investigative process, it is hoped that this will reduce the time required to investigate.

In this study, participants were shown 6 videos of a GPU accelerated file carving tool that recovers images from a disk and displays these in near real-time (assuming a small delay caused by on-screen rendering). To simulate a case where a suspect is in possession of illicit images, participants were informed that any photos of birds were to be considered "illegal" content. Half of the videos that were shown contained images of birds in varying proportions, and the other half of the videos only showed non-bird images. This methodology was utilised to assess the ability of the participants (n=30) to accurately identify datasets with potentially "illegal" content.

The following sections present; the methodology that the study followed in order to answer whether visualisation techniques could be applied to the collection phase of a DF investigation, the results that were gathered from participants who undertook the experiment, followed by an analysis and conclusions based on these results.

2 Background

Over recent years, the number of computing devices owned by people has increased due to the growing adoption of portable devices such as smartphones, smart home products and digital assistants [1]. The ubiquity of such devices implies that they often play a central role in criminal activity such as extortion, cybercrime, identity theft, etc. Even where these devices are not the primary instruments of a crime, they frequently prove to be an invaluable source of information in other cases, providing information such as location history, call records, photo metadata, etc. Therefore, it is not uncommon for law enforcement to seize all digital devices from suspects. These devices, however, have not only become more prevalent, but have also continually grown in terms of storage capacity as shown by combining data in [2, 3 & 4] (depicted in Fig 1.). Devices with a storage capacity of 200GB+ would have previously been restricted to larger devices, such as desktop computers and laptops, but recently developed smartphones regularly come with storage capacity options of 256GB or more. However, the technology used to examine devices in DF investigations has not kept pace with these increases in device storage, which presents significant challenges for investigators. A report by Her

Majesty's Inspectorate of Constabulary (HMIC) in 2015 reported that police investigations involving digital devices were commonly delayed up to 12 months or more after a review of 124 digital forensic cases from six UK police forces [5].
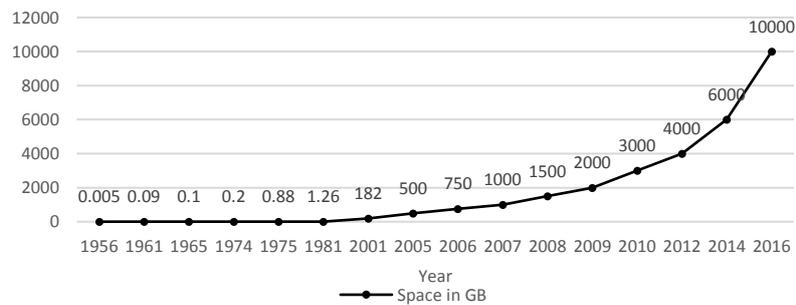

Fig. 1. Consumer hard drive capacity timeline

The digital investigative methodology, as defined by the Digital Forensics Research Workshop [6], proposes 6 stages of DF investigation. The stages, as shown in Fig. 2, include the Examination stage, where the investigator will explore all of the data recovered from devices and identify evidence that is relevant to the case. It is this stage of the investigative process that can take significant human input and time, which may be exacerbated by the increasing volumes of data involved in these cases. The tools frequently used by the investigator to explore the data are traditionally text-based tools that do not lend themselves well to the efficient exploration of large datasets due to the variety of media found in these datasets. Hales [7] suggests the use of exploratory information visualisation techniques to construct a visual timeline of events on the digital device that could assist the investigator in creating a narrative of behaviour. The developed tool – Insight – takes the information from a popular DF tool – Autopsy – and presents it to the user as a visual timeline. Different types of information such as web browsing history and EXIF metadata are all presented on the same timeline; allowing the investigator to see a detailed picture of what the device owner was using the device for at any point in time. The research shows that participants reacted more positively to software that provides a timeline visualisation of the device data when comparing to traditional textual interfaces commonly used. Statistical analysis shows that the use of visualisations improve the ability of the investigator to make accurate conclusions regarding a narrative of suspect behaviour. User experience when utilising visualisation software is also judged to be equal to traditional DF software and significantly better when performing tasks such as corroborating evidence or determining user behaviour at a specific point in time. Visualisation techniques have also been successfully applied to other areas of computer security; such as network security [8, 9] and malware analysis [10, 11].
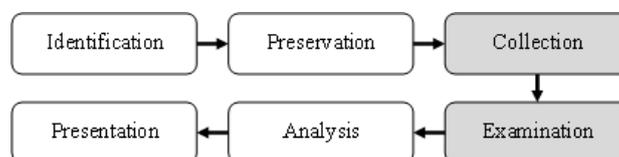
Fig. 2. DF investigative model [6]

The previous research conducted into the application of information visualisation techniques suggests further exploration of the application of these techniques to earlier stages of the DF investigative process; specifically, the Collection phase. During this phase, file carving is performed on data that is retrieved from a device, attempting to recover data using physical information stored on the drive. Tools available to perform this step are generally automatic and require minimal human intervention, with output during the process generally limited to little more than a progress bar. Historically, this has also been a time-consuming process as all data from a device has to be processed on the CPU for patterns that indicate the beginning and end of files. Recent research conducted by Bayne et al. [12] demonstrates how pattern matching – a computationally intensive processes vital to file carving – can be moved to GPUs using an asynchronous processing approach to greatly reduce processing time when compared to traditional CPU processing approaches employed by other DF tools. The tool developed by this research – OpenForensics[1] – showed time reductions of up to 97% when compared to performing the same file carving operation on other established DF tools that employ CPU processing.

To explore the application of information visualisation techniques whilst performing file carving on the data under investigation, the OpenForensics tool was modified to create a visualisation prototype for this work. This prototype displays the images discovered by the tool in near to real-time to the end user in a 4x4 grid on screen (Fig. 3). Due to the fast processing speed of the tool, the images displayed within this grid change quickly. The authors hypothesise that the display of near to real-time results within the application could be used to allow investigators to perform rapid triage of a dataset thereby reducing the examination time required to identify datasets that may contain information of forensic interest.

Researchers have suggested machine learning solutions to perform automatic triage of data collected in DF cases by classifying any images found [13, 14]. Whilst it is acknowledged that indeed this may be a worthwhile route to pursue, for the purposes of law enforcement use, it would not be appropriate to allow a computer to be the sole decision maker as to whether datasets contain potentially valuable evidence, or whether any images should be examined in detail by a human or not. There are a number of legal and ethical ramifications potentially raised by such an approach. It should also be noted that whilst advances in machine learning have been significant, there are still areas where there are challenges. One challenge that remains problematic is the ability

---

[1]　https://github.com/ethanbayne/OpenForensics

of these systems to accurately differentiate between children, adolescents and adults. When used in law enforcement settings, in many cases these tools will be used to look for possession of illegal images, such as those depicting minors [15]. The automated recognition of images is an area of ongoing research and, as such, machine learning techniques are not explored as a part of this work at this time.
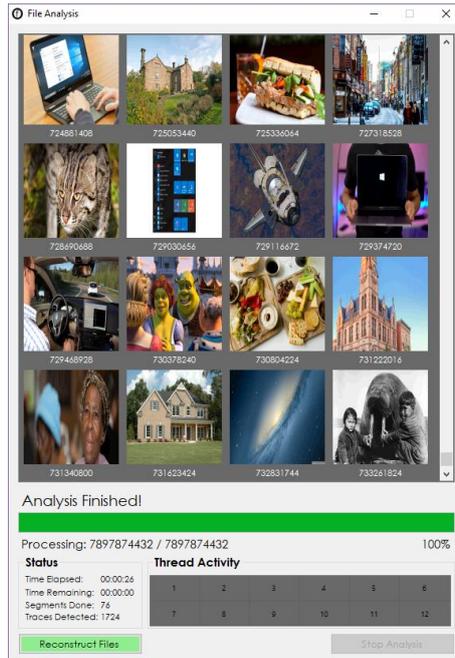

Fig. 3. OpenForensics Visualisation Interface

3 Methodology

The purpose of this body of work is to ascertain whether end users can accurately identify "illegal" images in a dataset as they are being recovered and shown to them by a GPU accelerated file carving tool. The tool displays the images to the participant as it processes them in a 4x4 grid filling each row from left to right and moving rows upwards as the last row is filled. The approach followed by this research used the Open-Forensics tool, which was modified to display acquired images to the end user in near real-time. This allowed the accuracy of the participant's answers to be examined, in terms of the number of correct and incorrect responses.

3.1 Dataset Design

Six datasets were developed for use in this experiment, with the aim of closely reflecting the type of content that may be found on a small USB storage device in an investigation. To create these datasets, a tool was used to scrape random images from Google Images, using several different keywords, including "birds". A USB storage device was securely wiped, and then a sample of these images was placed on the drive. The images were subsequently deleted, and a clone (disk image) of the USB storage device was acquired, as is standard practice in a DF investigation. Following this sequence provided a dataset in a format that the file carving tool could analyse and recover the deleted images from. The nature of the tool means that subsequent runs on different hardware would produce a visualisation where the images change at different speeds. Allowing the participants to use the tool directly would thus produce skewed results depending on the hardware the experiment was conducted on. In order to counteract this, a video of the tool running on each dataset was created to ensure that each participant observed the visualisation in the exact same way, with images appearing onscreen at the same speed for each participant. During the video, each image remains on the screen for an average of 800ms. As the time taken for a human to perceive the content of an image is around 1/3 of a second (~333ms) [16], the images are ensured not to be appearing at a speed that is outside of the bounds of what can be perceived.

These datasets were designed to have an increasing number of non-bird images, therefore decreasing the ratio of "illegal" bird images in the dataset. In 3 of the 6 datasets, no bird images were included, and in the other 3 datasets, there were varying amounts of bird images included, as shown in Table 1. The ratio of bird images included in each of these 3 disk images was decreased to determine participant accuracy with increasing levels of visual "noise". When creating the datasets, the images were manually checked by both authors to ensure that there were no images that were ambiguous (e.g. a cartoon image of the Linux penguin mascot "Tux") or contained birds that were not the main subject of the image (e.g. a landscape photograph with a few small birds flying across the sky).

Table 1. Data File Summary

Dataset	Total Number of Images	"Illegal" Images Incl?
1	100	Yes – 10 (10%)
2	100	No
3	1000	No
4	1000	Yes – 10 (1%)
5	5000	Yes – 5 (0.1%)
6	5000	No

3.2 "Illegal" Image Detection

Participants (n=30) were required to have a good knowledge of computing, with regular use of computers as part of their career. Due to the use of pre-recorded videos of the tool, a significant technical ability such as a background in computer science was not required. Each participant reported normal or corrected to normal vision and were asked if they were colour-blind. The participants were given information at the start of the experiment that informed them about the fact that images of birds were to be considered to be "illegal" during the experiment, and they should look out for any images of these in each video. They were told that they were not required to count the number of images they saw, or to try and remember what each image looked like; only that they would be asked to identify if there had been any images of birds, which would indicate it would be worthwhile for law enforcement to investigate further. The videos were played in full-screen mode to eliminate other distractions on screen and were conducted on displays with similar sizes and resolutions at as close to eye level as was feasible.

Each participant was asked to watch each video only once, and to refrain from rewinding or stopping the video once it had started. A video player was used to blank the screen after each video was played so that no images remained on the screen. To control for fatigue, a Latin Square design for the order of the videos was adopted (Table 2). This was essential, as both videos 5 and 6 were over 4 minutes in length, which could lead to fatigue when concentrating on the images appearing. Participants were allocated to a group when signing the consent form, and all groups had a total of 5 participants each.

Table 2. Participant Group Video Order

Group	Video Order					
A	1	2	3	4	5	6
B	6	1	2	3	4	5
C	5	6	1	2	3	4
D	4	5	6	1	2	3
E	3	4	5	6	1	2
F	2	3	4	5	6	1

After watching each video, the participants were asked to complete 2 questions relating to the video they had just watched. The first of these questions was simply "Did you see any pictures of birds?" with possible responses of simply "Yes" or "No". The results of the question were used to determine metrics such as the hit rate and false alarm rate. They were also asked to give an indication of their confidence in their answer. This was presented in the format "How confident are you that you saw/did not see pictures of birds in Video X?" and a response in the format of a Likert scale where 1 represented "Not Very Confident" and 5 represented "Very Confident". The results from this question were used to determine whether there was any correlation between

the number of images being shown to the participants and the confidence in their answer, or between false alarms/misses and confidence. Finally, to gauge the participant perception of the tool, and provide additional context to the triage results, the participants were asked to optionally respond to the question: "Now that you have watched all of the videos, how did you find the experience of searching for pictures of birds as they appeared?" Thematic analysis was conducted on the responses to this question.

3.3   Interface Preference

Upon the conclusion of the experiment, the participants were shown a few simple interfaces and asked to indicate their preferred interface. These questions were proposed in order to inform future visualisation design based on the preference of the users. The first interface comparison displayed 2 interfaces to the participant (Fig. 4) and asked; "If you had a large gallery of photos on your computer and wanted to find a specific photo, how would you prefer to do this?" They were presented with 2 responses: "(Interface 1) Show one full-screen photo at a time, and keep pressing the next arrow until you found the photo" or "(Interface 2) Show all of your photos in a file browser with thumbnails and find it from there." This question was asked to determine whether the participants were likely to prefer a "flashcard" approach or view multiple photos at the same time.
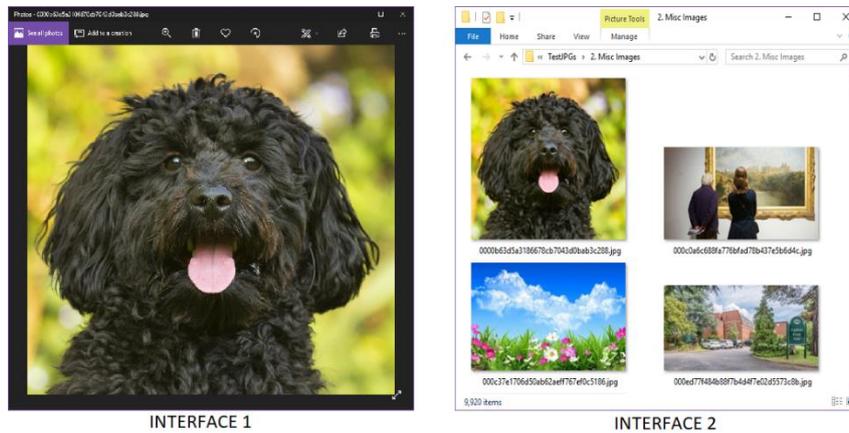

Fig. 4. Flashcard vs Grid layout preference

The second question presented the participant with 3 interfaces (Fig. 5). Each interface depicted an image gallery, with varying sizes of images in each interface. This question was presented to determine whether the user is likely to prefer larger images with fewer shown at a time, or smaller images with more shown on screen at one time. The question was presented as "Which of the following interfaces do you prefer?" with a simple option to select Interface 1 or Interface 2.
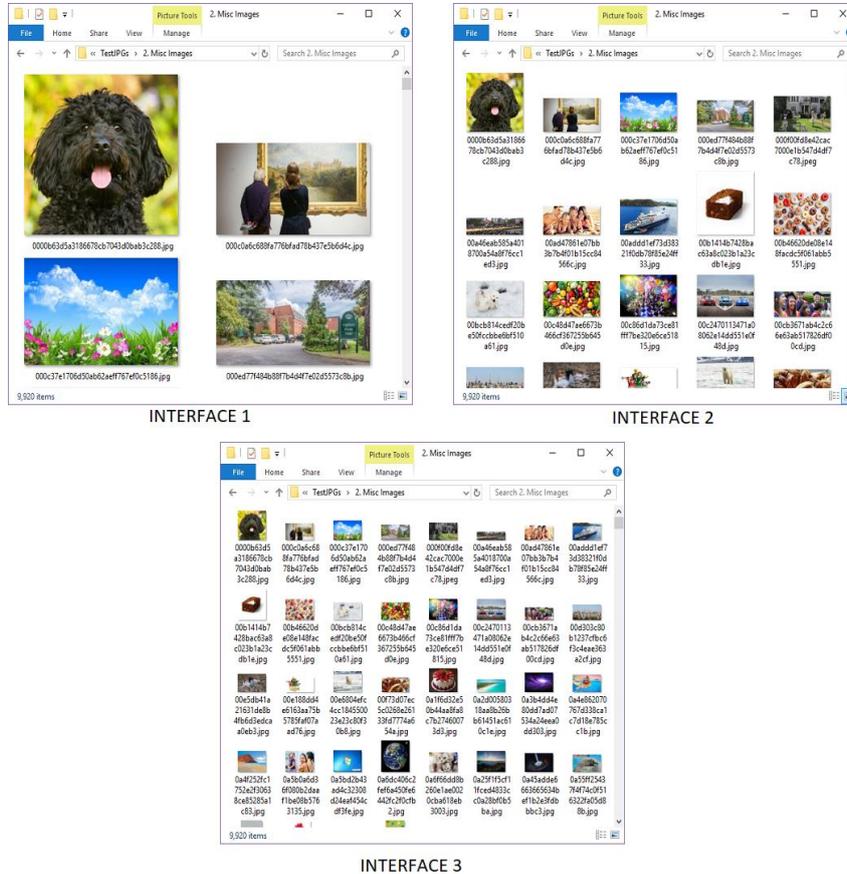
Fig. 5. Image browser size preference

4   Results

In this section, results from the experiment where participants were asked to detect images of birds will be presented and discussed, alongside the confidence ratings of participant responses after watching each video. Responses provided on interface design preferences will also be discussed, together with a thematic analysis of the perceptual feedback provided by participants at the conclusion of the experiment.

4.1   File Carving Triage Results

Accuracy

    When exploring the results from the file carving tool experiment, the videos were grouped into pairs of datasets with the same number of images; one video with bird

images present and one video without bird images present. The pairings, therefore, as per Table 1, are Videos 1 & 2, 3 & 4 and 5 & 6.

The first of these pairs, which contained 100 images, 10% of which were birds in Video 1 showed a hit rate of 93%. The correct rejection rate for this pair is also 93%. These figures are high, as was expected by the researchers, as the relatively high number of bird images in the dataset led to many participants verbally commenting that they were "obvious". The breakdown of participant responses to this video pair can be seen in Table 3. This shows that participants can accurately differentiate between the datasets with and without bird images in a small dataset.

Table 3. Video 1 & 2 detection results

	Responded "Yes"	Responded "No"
Video 1 (Birds Present)	28 (93%)	2 (7%)
Video 2 (Birds Absent)	2 (7%)	28 (93%)

The second of the video pairs, containing 1000 images and 1% bird images in Video 3 shows a hit rate of 83%. The correct rejection rate for this pair was 90%. The breakdown of the participant responses for this pair can be seen in Table 4. Although these figures still show a relatively high level of accuracy amongst the participants in identifying bird images, this is a noticeable decrease from the previous video.

Table 4. Video 3 & 4 detection results

	Responded "Yes"	Responded "No"
Video 4 (Birds Present)	25 (83%)	5 (17%)
Video 3 (Birds Absent)	3 (10%)	27 (90%)

The final pair of videos that contain a total of 5000 images, 0.1% of which are of birds in Video 5, show again a hit rate of 83% and but a lower correct rejection rate of 77%. As can be seen in Table 5, this means that the false alarm rate for this pair of videos is 23%. This value is relatively high and indicates a slightly decreasing level of accuracy from the participants in the longer videos with fewer bird images, this is also shown in Fig. 6.

Table 5. Video 5 & 6 detection results

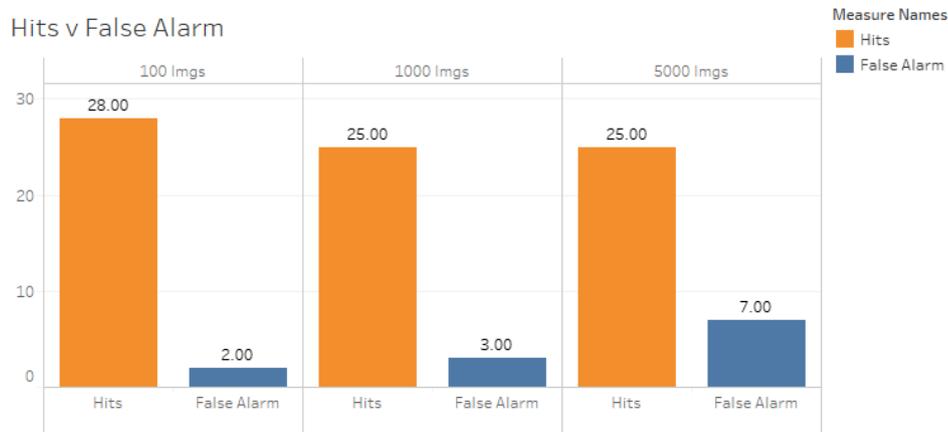|	Responded "Yes"	Responded "No"
Video 5 (Birds Present)	25 (83%)	5 (17%)
Video 6 (Birds Absent)	7 (23%)	23 (77%)

Fig. 6. Hits and false alarm rates for each dataset size

Participants' answers between varying sizes of datasets were analysed. An exact McNemar's test determined that there was no statistically significant difference in the proportion of answers given between videos with birds present within them—videos 1 and 4 ($p = .37$), videos 4 and 5 ($p = 1.00$), and videos 1 and 5 ($p = .45$). An exact McNemar's test similarly determined that there was no statistically significant difference in the proportion of answers given between videos without birds present within them—videos 2 and 3 ($p = 1.00$), videos 3 and 6 ($p = .29$), and videos 2 and 6 ($p = .12$).

It is interesting to note that, although not a large enough sample to analyse in depth, one participant that declared themselves as colour-blind (deuteranopia) gave as accurate answers in the experiment as other participants. It was found that colour-blindness did not affect their ability to recognise the bird images. This participant succeeded at correctly identifying all videos with a 100% hit and correct rejection rate. It is not known whether the colour-blindness of the participant influenced their results, or whether this is coincidence. This may be an area for future research.

4.2 Participant Confidence

After watching each video and indicating whether they saw images of birds in the video, the participants were asked to rate on a Likert scale of 1 to 5, the confidence they had in the answer that they had given, where 1 indicated that they were "Not Very Confident" in their answer and 5 indicated that they were "Very Confident" in their answer. Fig. 7 shows a breakdown of the responses given by participants for this question for each video. Overall, for many of the videos, participants generally responded that they were confident with their responses. All participants were confident with their answer to Video 1, reinforcing the verbal comments that this video was "obvious", although it was found that 2 participants answered this incorrectly with "no".
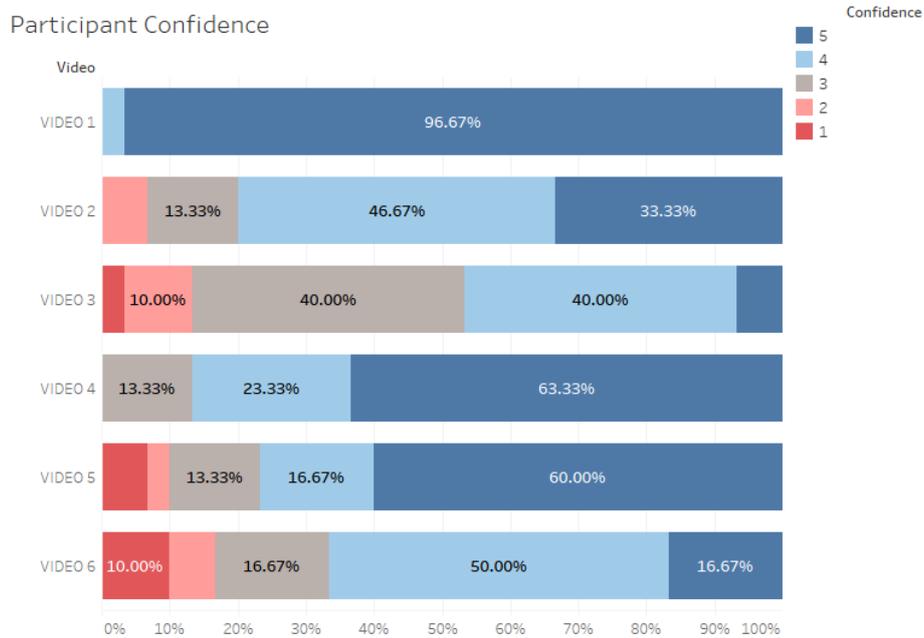
Fig. 7. Participant confidence levels (From left: 1 – Not Very Confident to 5 – Very Confident)

Using Friedman's ANOVA test, the confidence ratings of the participants significantly differed over the six videos ($x^2 (5) = 8.00, p < .001$).

Wilcoxon Signed Rank tests were used to measure the confidence of a participant's answers between videos with bird images to those without. It appeared that a participant's confidence significantly differs between videos 1 and 2 ($T = -3.91, N$ for Test $= 30, p < .001, r = -.5$) , between videos 3 and 4 ($T = -3.96, N$ for Test $= 30, p < .001, r = -.51$) and between videos 5 and 6, ($T = -2.57, N$ for Test $= 30, p = .01, r = -.33$). It can be concluded that a participant's confidence significantly differs between videos with birds and videos without birds in them.

Wilcoxon Signed Rank tests were also used to test a participant's confidence in their answers between shorter and longer videos. It appeared that a participant's confidence significantly differs between videos 1 and 4 ($T = -2.89, N$ for Test $= 30, p = .004, r = -.37$) and videos 2 and 3 ($T = -3.40, N$ for Test $= 30, p = .001, r = -.44$). However, each participant's confidence did not differ significantly between videos 4 and 5 ($T = -1.20, N$ for Test $= 30, p = .23$) and videos 3 and 6 ($T = 1.04, N$ for Test $= 30, p = .30$). It can be concluded that a participant's confidence in

their answers weakened between watching the videos with 100 images and 1000 images, but each participant remained as confident with their answers between the videos with 1000 images and 5000 images.

4.3   Interface Preference

At the end of the experiment, the participants were shown different interfaces and asked to indicate which one they would prefer to use if they were looking for a specific photo on their computer (these interfaces can be seen in Figs. 4 & 5). The first of these questions presented the user with a single large full-screen image versus a grid of images. The response to this question was that all participants unanimously preferred Interface 2 that depicted a grid of images. As this was the format used for the visualisation of images in the file carving tool, it reinforces that this design decision was the correct to make instead of rapidly showing the users images one at a time.

The second question was related to the sizing of the images displayed in a grid format, the purpose of which was to determine whether users prefer an interface with relatively large images, average sized images (similar to the interface in the tool), or an interface with relatively small images. The results of this (Fig. 8) show that the majority of participants (86.67%) preferred the interface that showed averagely sized thumbnails, which was similar to the format and size of the images presented by the tool. Again, this reinforces the design decisions made when creating this tool. It should, however, be noted that these preferences may be due to a response bias as the participants were asked these questions after using the tool. In future, it may be of interest to test if these responses vary before exposure to the software.
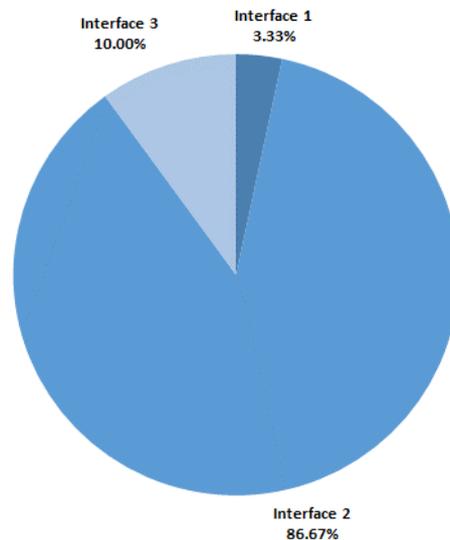

Fig. 8. Interface thumbnail size preference

4.4 Qualitative Feedback

After watching all of the videos in the experiment, the participants were asked to give feedback on how they found the experience of searching for pictures of birds as they appeared in the video. This was a free text field where they could leave any comments they felt were relevant. Of the 30 participants, 29 provided responses to this question. Thematic analysis was conducted on the responses, where five common themes were identified. These themes can be seen in Table 6.

Table 6. Themes in Participant Feedback

Theme	Participants commenting on theme	Example of comments
Difficult	53%	"Like looking for a needle in a haystack" "Quite difficult"
Length of task	30%	"Longer videos were difficult" "Difficult to keep track of all the images and the longer each video is the harder it was."
Fast image speed	30%	"...rather difficult to keep up with the image flow" "Very hard because the pictures came and went so fast"
Fatigue	23%	"Cognitively demanding" "Quite straining on the eyes during the longer videos."
Variable rate of image appearance	13%	"The rate differences made it more challenging in some of the videos than others - I suspect the images appeared as they were carved rather than at a consistent rate." "Because the bulk image elections *[sic]* were not a consistent pace I felt it was much harder to see the birds, it would be smooth then jarring"

As can be seen from the analysis of themes, many of the participants commented on how difficult they found the task to be. This comment was often made in combination with one of the other themes; specifically, participants also frequently identified that the difficulty they had with the task was caused by the fast speed that the images appeared onscreen. The difficulty was also caused by the length of the 5000 image videos, which were over 4 minutes long. The length of the videos also appears as a theme in several responses where the participants mentioned a level of fatigue whilst undertaking the task.

Some participants also commented that the inconsistent rate that the images were displayed on the screen posed a challenge and, in some cases, made it much harder to see the images of birds. This inconsistent rate at which the images appear is due to the way that the tool and file carving work. As images are being recovered from deleted data on a disk, there are times where the tool may find multiple images clustered in co-located blocks of data; other times the tool may not find images for a period of time. This leads to points where the images stop populating, pause for a short time, and then continue populating rapidly. A buffer could be used to attempt to smoothen the display of these images to the end user, however, this was not implemented in the tool as it would add an artificial overhead to the process and make the process longer, which contradicts the motivation for this research.

5   Analysis

As can be seen from the ability of the participants to detect images of birds in a dataset, and correctly triage this dataset, generally the results are indicative of a potential benefit to the digital forensic investigative process, given further refinements based on the outcomes of this work. Participants were found in most cases able to correctly identify the presence of 'illegal' images, however, as also shown by these results, the accuracy of the responses given starts to decline slightly in datasets with more images, and fewer bird images. In the videos with 5000 images, the number of correct hits is still relatively high at 83%, however, as shown in Fig. 6, the number of false alarms is also relatively high at 23%. The threshold at which the detection accuracy is too low is highly dependent on the scenario that the tool is being utilised in. If the tool is being used to triage the data of someone accused of possession of inappropriate images of minors for example, then law enforcement authorities are likely to reasonably demand a hit rate of very close to 100%. However, if the tool is being used to investigate a device in an organisation for an employee accused of photographing confidential property, for example, a slightly lower accuracy rating may be acceptable.

In this study, it was hypothesised that the length of time the investigator was required to watch a video for would have a direct impact on the confidence of their answer; with longer videos leading to lower confidence ratings. However, the findings do not support this hypothesis, as the confidence rating between videos with 1000 and 5000 images were not significantly different. Instead, it was found that there were consistent and

significant differences between the confidence ratings of videos including images of birds, and those without, across all dataset sizes. Combined with the increasing false positive rate, it could be conjectured that as the participants become more uncertain whether they may have missed a bird image in the large datasets that did not contain bird images, they are more likely to respond indicating that they did see a bird image so that a dataset potentially containing evidence is not incorrectly dismissed, signifying a possible response bias. This is not necessarily an issue in law enforcement scenarios, as a suggestion of a false positive during the triage stage will lead to further investigation of an irrelevant dataset, thus simply decreasing the efficiency of the investigation. Incorrectly dismissing datasets that contain evidence (a "miss") is more of an issue as in a law enforcement scenario, this could lead to failure to prosecute or acquit a suspect.

Exploration of the feedback given by participants yield several themes that highlight some of the issues encountered by the participants. Notably, a significant number of participants (53%) indicated that they found the tasks "difficult" in some way. Exploring the themes further revealed further core reasons for this difficulty. The speed that the images appeared on screen was mentioned by 30% of the participants, who generally said that the flow of images was difficult to keep up with. Artificially reducing the rate that the images appeared could be a solution to this, however, as mentioned previously, this would delay the overall investigation, which is contrary to the aim of this research. The relatively long time where the participants are required to look at the screen, and the high level of concentration that they must apply to the task led to 23% of the participants mentioning that some form of fatigue was experienced during the experiment. Some participants indicated that this was in the form of high cognitive demand, whilst others indicated it was in the form of eye strain. This would be exacerbated in lengthy DF investigations. Thus, an alternative visualisation may need to be explored that could allow the user to look away from the screen without the risk of missing large volumes of data.

6   Conclusions and Future Work

This work has explored a method of visualising the results of a GPU accelerated file carving tool in real time, to provide a rapid triaging tool for use in DF investigations. It was hypothesised that this tool would provide the user with a way to make accurate and timely decisions about the kind of images contained on a device, and thus decide as to whether investigating it at length would be necessary.

In testing this tool with 30 participants, it was shown that the accuracy of the participants in detecting contraband images is generally good, however, it begins to decline slightly (albeit not statistically significantly) with larger datasets containing fewer contraband images. This has indicated that the underlying idea of rapid triage using such a tool is feasible, however, further work will be required in refining the format of the visualisation tool. This conclusion is reinforced by participant feedback indicating that the task was difficult due to the high speed at which the images appeared and then

disappeared from the screen, along with the fatigue experienced from having to concentrate on the screen for an extended period of time. Ideally, future work will allow a visualisation method to be developed where the information is displayed to the investigator for a longer period of time, allowing them to look away from the screen, while maintaining or improving upon the accuracy levels seen in this study.

It would be of interest to apply machine learning techniques to aid identification of contraband material—not as a replacement for human analysis, but to augment it. This approach could allow for certain pieces of data to be drawn more clearly to the attention of the human investigator if deemed by an image recognition algorithm to potentially be of interest. Such an approach would hypothetically reduce the strain on the investigator and allow them to make more accurate and confident triage decisions, whilst still ensuring that no data is hidden; an important feature in DF tools.

7   Acknowledgements

8   References

1. OFCOM: Decade of Digital Dependency, https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/decade-of-digital-dependency (2018).
2. Vaughan-Nichols, S.J.: Hard drive technology reaches a turning point. Computer (Long. Beach. Calif). 36, 21–23 (2003).
3. Farrance, R.: Timeline: 50 Years of Hard Drives, https://www.pcworld.com/article/127105/article.html, (2006).
4. Mah Ung, G.: Seagate's 10TB Barracuda Pro is the world's largest consumer hard drive, https://www.pcworld.com/article/3096292/storage/seagates-10tb-barracuda-pro-is-the-worlds-largest-consumer-hard-drive.html, (2016).
5. Her Majesty's Inspectorate of Constabulary: Online and on the edge: Real risks in a virtual world. (2015).
6. Palmer, G.: A Road Map for Digital Forensic Research. In: First Digital Forensic Research Workshop. pp. 27–30. , Utica, New York (2001).
7. Hales, G.: Visualisation of device datasets to assist digital forensic investigation. In: 2017 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (Cyber SA). IEEE (2017).
8. Angelini, M., Blasilli, G., Catarci, T., Lenti, S., Santucci, G.: Vulnus: Visual Vulnerability Analysis for Network Security. IEEE Trans. Vis. Comput. Graph. 25, 183–192 (2019).
9. Liu, X., Sun, Y., Fang, L., Liu, J., Yu, L.: A Survey of Network Traffic Visualization in Detecting Network Security Threats. Trust. Comput. Serv. ISCTCS 2014. Commun. Comput. Inf. Sci. 520, 91–98 (2015).
10. Nataraj, L., Manjunath, B.S.: SPAM: Signal Processing to Analyze Malware [Applications Corner]. IEEE Signal Process. Mag. 33, 105–117 (2016).

11. Quist, D.A., Liebrock, L.M.: Visualizing compiled executables for malware analysis. In: 2009 6th International Workshop on Visualization for Cyber Security. pp. 27–32. IEEE (2009).
12. Bayne, E., Ferguson, R.I., Sampson, A.T.: OpenForensics: A digital forensics GPU pattern matching approach for the 21st century. Digital Investigation. 24, S29–S37 (2018).
13. Perez, M., Avila, S., Moreira, D., Moraes, D., Testoni, V., Valle, E., Goldenstein, S., Rocha, A.: Video pornography detection through deep learning techniques and motion information. Neurocomputing. 230, 279–293 (2017).
14. Platzer, C., Stuetz, M., Lindorfer, M.: Skin sheriff. In: Proceedings of the 2nd international workshop on Security and forensics in communication systems - SFCS '14. pp. 45–56. ACM Press, New York, New York, USA (2014).
15. Mayer, F., Steinebach, M.: Forensic Image Inspection Assisted by Deep Learning. In: Proceedings of the 12th International Conference on Availability, Reliability and Security - ARES '17. pp. 1–9. ACM Press, New York, New York, USA (2017).
16. Potter, M.C.: Meaning in visual search. Science. 187, 965–966 (1975).