Analysing large and complex image collections during a safety investigation

Floris Gisolf *is active for the Dutch Safety Board as an investigator and data analyst for the maritime department and was part of the MH17 crash investigation team. He studied biopsychology (BSc) and forensic science (MSc), and is pursuing his PhD in interactive multimedia analysis at the University of Amsterdam. Floris also worked at the Netherlands Forensic Institute as researcher at the Digital Technology and Biometrics department, where he worked on several publications on digital camera identification.*

Zeno Geradts *is a senior forensic scientist at the Netherlands Forensic Institute of the Ministry of Security and Justice at the Forensic Digital Biometrics Traces department. He is expert witness in the area of forensic (video) image processing and biometrics. Within the team Forensic Big Data Analysis he works in research on deep learning and images and video. He is President of the American Academy of Forensic Science 2019-2020 and chairman of ENFSI Forensic IT Working group. From September 1st 2014, he is full professor on Forensic Data Science by special appointment at the University of Amsterdam.*

Marcel Worring *is a professor in computer science at the University of Amsterdam, working on multimedia analytics, bringing together multimedia analysis, interactive visualizations, and AI techniques. Among other domains, he has been applying such techniques in collaboration with law enforcement agencies for more than 15 years. He is director of the Informatics Institute and co-founder of the Innovation Center for Artificial Intelligence.*

Introduction

On 17 July 2014 flight MH17 crashed due to the detonation of a warhead launched from the eastern part of Ukraine using a Buk missile system. Because the remains of the airplane were located in an area of on-going armed conflict, it was not possible to secure the physical investigation material and an extensive investigation could not be conducted at the crash site right away. For this reason, an important part of the investigation by the Dutch Safety Board consisted of the manual analysis of the photos and videos acquired by Ukrainian and Malaysian investigators, the Australian Federal Police, the OCSE, journalists and local people. In total, approximately 20,000 photos and 3,000 videos were collected (1).

In this paper the photo and video analysis performed by the Dutch Safety Board will be described, the lessons learned, as well as the ways we have sought to improve the efficiency of the analysis of large and complex image collections to support future investigations. Finally, the application developed based on these lessons, and a use case to show how the application can be used, are presented in this paper.

The image analysis in the MH17 crash investigation

The overarching question of the investigation was: what happened to flight MH17? The main goal of the photo and video analysis was to find which wreckage pieces were found where. Once access was gained to the crash site, having this information made it more efficient to decide which pieces to get, and where to get them. Furthermore, it assisted in answering the main question by providing information about the breakup sequence and the state of the wreckage pieces right after the crash.

The analysis started fairly simplistic, by filtering out unwanted files (e.g., non-image files, thumbnails, low resolution images) and by sorting images into folders with several categories, such as engines, wings, cockpit, etc. Due to the complexity and large number of images, this quickly became hard to manage. Many images contained multiple objects; many images could not be classified right away, as it was unclear what was actually shown on the image; it was hard to keep track of which images had been seen by the investigators; our PCs and Windows Explorer could not keep up with displaying large folders with many images; and going through all the images one by one was very time consuming.

A software tool called Netclean Analyze (now known as Griffeye Analyze) (2) was suggested by Team High Tech Crime of the Dutch Police. This tool allows for quickly browsing through a large number of images, by generating thumbnails, and it allows for tagging images with multiple tags, which can subsequently be used to filter. It solved several of the aforementioned problems. However, it was still time consuming, and the interface to tag images was somewhat cumbersome, requiring multiple clicks.

The image and video analysis resulted in several 'products' for the investigation team and for the report. First, an overview was created of the whole crash site, subdivided in smaller areas, both based on their location and on the airplane parts that were found there (Figure 1). Together with the sideview (Figure 2), it gave a quick and clear idea of the general break-up sequence.

Figure 1 Overview of wreckage area showing the six smaller sites. (source of satellite images: Google Earth/Digital Global

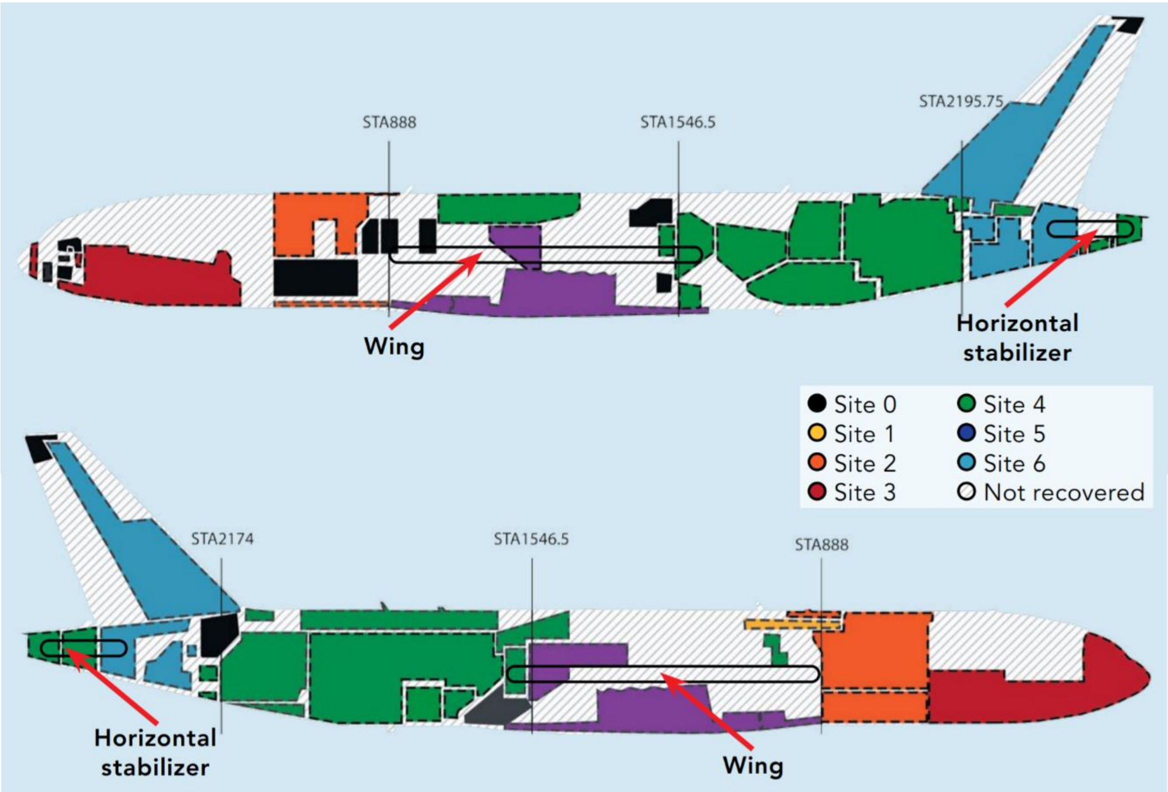

Figure 2 Side view left (top) and right (bottom). Identification of wreckage retrieved from the wreckage sites. (Source: Dutch Safety Board)

Then, for each area a more detailed map was created with the exact location of all the identified wreckage pieces (Figure 4). In some cases it was easy to find the exact location, due to GPS data included with the photo. In most cases however, no GPS data was available, and satellite imagery and the linking of multiple photos was needed to pinpoint the location. For example, the piece in Figure 3 was found by finding the two houses in the background, one with a green roof, a slight extension to the side, and an extension to the back on the left; and the house with the grey roof on the right.

Figure 3 Example of finding the location through satellite images. (Source of top image: Rob Stothard; source of bottom image: Google Earth)

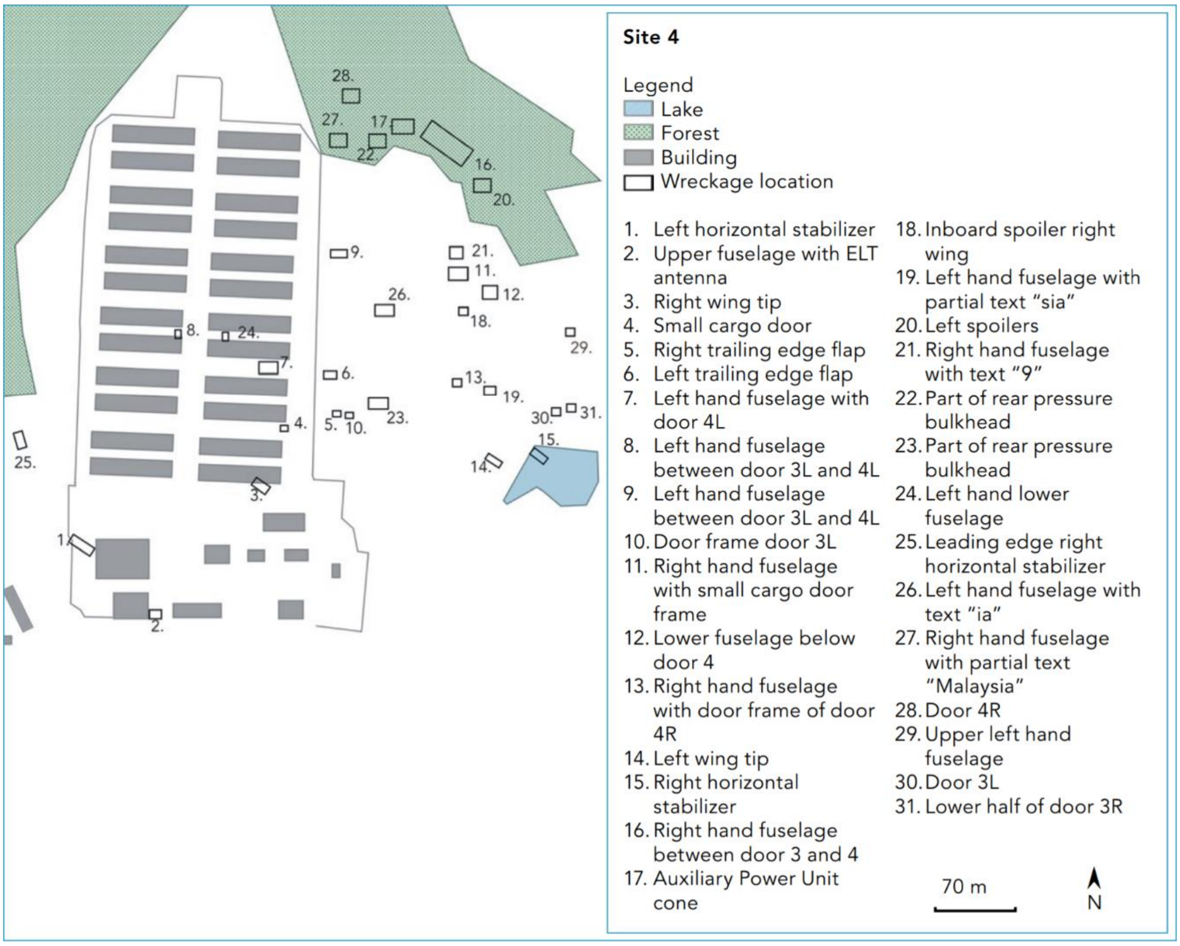

Figure 4 Overview of wreckage site 4 and the location of the wreckage pieces. (Source: Dutch Safety Board)

Improving the image analysis process

Where the above explained how the actual investigation took place, we now switch to the post analysis in which we consider methods that could have made the process more efficient and which could form the basis for potential future investigations.

The analysis process of a large image collection generally consists of two phases (3):

- **Exploration**, applicable when the investigator is faced with a collection she does not know much about beforehand, and wants to discover what is inside and/or how the data are structured. An exploratory session typically takes time and involves a dynamic model of the data, continuously refined as the analyst iteratively gains knowledge.
- **Search**, applicable when the investigator has a clear idea what she is looking for and queries the system for items relevant to certain attributes. A search session is then a sequence of query-response pairs, and the analyst expects fast response. The data model is static, since the investigator knows exactly what she is looking for, and this can be communicated to the system through a query.

Tasks related to these phases can be placed on an explore-search axis (Figure 5), with tasks on the left generally preceding tasks more towards the right, but with a lot of switching back and forth between the different tasks.

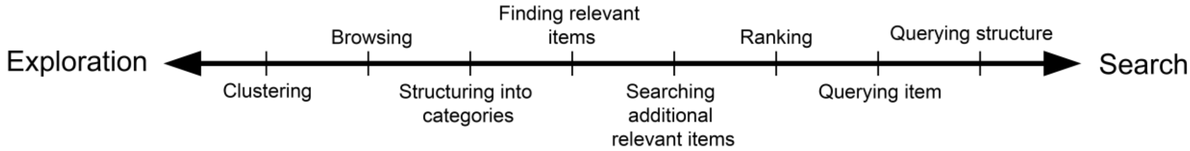

Figure 5 The exploration-search axis with example multimedia analytics (sub)tasks, adapted from (3)

Clustering: Group images based on similarity, to make it easy for the investigator to find structure in the collection and relations between images.
Browsing: Allow the investigator to quickly and intuitively view the image collection.
Structuring into categories: Bring relevant structure to the image collection, to more easily make sense of the data.
Finding relevant items: Find those images that give information that support hypotheses or answer questions of the investigator.
Searching additional relevant items: Images of the same object or location, but from a different angle can give new information.
Ranking: Sort images based on content or meta-data.
Querying item: Search for a specific image.
Querying structure: Use the created structure to test hypotheses and answer questions.

To improve efficiency, several of these tasks can be (partially) automated using computer vision. Computer vision is the interdisciplinary scientific field of how computers can make sense of digital images or video. It tries to automate tasks that humans can do with their visual systems. Computers nowadays can learn to recognize objects and locations, such as cats, cars, Paris, the beach, etc. However, in order to do so, the computer needs a lot of examples to train on (approximately 1000 images per object) (4). While this is no problem for everyday objects and locations, airplane crashes and other accident sites are often unique in location and the type and state of objects. State of the art computer vision techniques are thus not yet able to do some of the most difficult parts of the analysis: determining what the object is and where it is located. In combination with the expertise of an investigator, however, it can make several tasks much easier.

In order to assist the investigator in the analysis, and with the limitations of the current state of the art in computer vision in mind, the following tasks were sought to be automated and developed into an app:

- Cluster images with similar content into groups;

- Query an image, sorting all images based on their similarity with the queried image;
- Query part of an image, sorting all images based on their similarity with the queried part of the image.

Furthermore, the investigator should be able to:

- Browse fluidly through the images;
- Place images in user defined category ‘buckets’ to structure the image collection;
- Retrieve and filter images based on these buckets;
- Gain information about progress made in the structuring of the image collection.

We developed ImEx (Incident Image Explorer) with these tasks and features in mind, in order to assist investigators in investigations with large image collections.

In order to cluster images based on similarity, ImEx makes use of a convolutional neural network (5). A brief description follows, as a full explanation of neural networks goes beyond the scope of this paper. In short, convolutional neural networks are the current state of the art in computer vision. By making use of large collections of labeled training data, a neural network is trained to discriminate between categories (such as cats, dogs, houses, cars etc.) by extracting features from images, such as shapes and textures. Features extracted from an image are represented by a value, where a higher value means the feature is present more frequently and more clearly in the image. In the training phase, the neural network learns which features are best to discriminate between categories. By finding these features, it can decide to which category an image belongs.

ImEx works slightly different. As noted, training a neural network requires a lot of training examples, which are usually not available for crash sites or other accident sites. Therefore, rather than classifying images (deciding to which category an image belongs), ImEx only calculates whether images look similar or not. ImEx still makes use of a neural network trained to classify everyday objects and scenes (such as different types of animals, sceneries, intact airplanes, other modes of transportation, instruments etc.).

The neural network used in ImEx (Resnet152 (6)) extracts 2048 features per image. The similarity between two images can then be calculated by correlating the 2048 features of one image with the 2048 features of another image. If this correlation is higher than a user defined threshold, the two images are placed in the same cluster. If other images also correlate higher than this threshold, these images are also placed in the same cluster. This process is repeated until all images are placed in a cluster.

A high threshold will result in many small clusters, whereas a low threshold will result in fewer, but larger clusters. This threshold can be changed by the investigator to suit the task and preferences. A cluster overview can be generated, which shows the most relevant image of each cluster. This enables the investigator to quickly find relevant clusters.

It is then up to the investigator to classify the clusters. In ImEx, the images in a cluster are displayed in a scrollable canvas at the bottom of the screen. The display size of the images can be adjusted. The investigator can create buckets for holding whole clusters or a selection of images, in order to structure the image collection. Relevant images or parts of images can be queried to find additional images. By generating buckets, and by adding images to these buckets, the image collection is given structure by the user.

Figure 6 shows the main user interface (note that screenshots may differ from the actual application, since development is still ongoing), with descriptive buttons, as well as an explanation canvas on the right, which guides the investigator through the app, and can display a description of the function of each button by right-clicking the button. The figure also shows an example of a cluster generated by ImEx containing 79 images of the vertical tail.

A second window shows the progress of structuring the image collection and a Sankey diagram to show relations between the buckets. Based on images contained in multiple buckets, the Sankey diagram shows a breakdown of each bucket. E.g., upon close inspection, the Sankey diagram in Figure 7 shows that parts of the wings were located both in site 4 and site 6, because images containing wings were placed in the “wings” bucket, but also in the “site 4” and “site 6” buckets.

Successful use case

While the application was developed in light of the MH17 investigation, it is not the only investigation with a large number of photos, as nowadays almost everyone has a camera on their phone with them at all times, and thus crashes and other incidents and their aftermath are sometimes captured by many people. To see how the application generalizes to other cases the following use case is discussed.

This use case focused mainly on the search part of the exploration-search axis. After a large bonfire during new year’s eve on the beach got out of hand, the Dutch Safety Board started an investigation (7). Approximately 4,000 images were collected from citizens, police and journalists. ImEx was then used to find all images showing the tower of pallets before it was set on fire. By making use of the cluster overview, relevant clusters were quickly identified, allowing for efficient browsing of the image collection. ImEx greatly reduced the time needed to find the relevant images, as only a small part of the image collection needed to be inspected in close detail.

Figure 6 ImEx user interface with example cluster of vertical tail

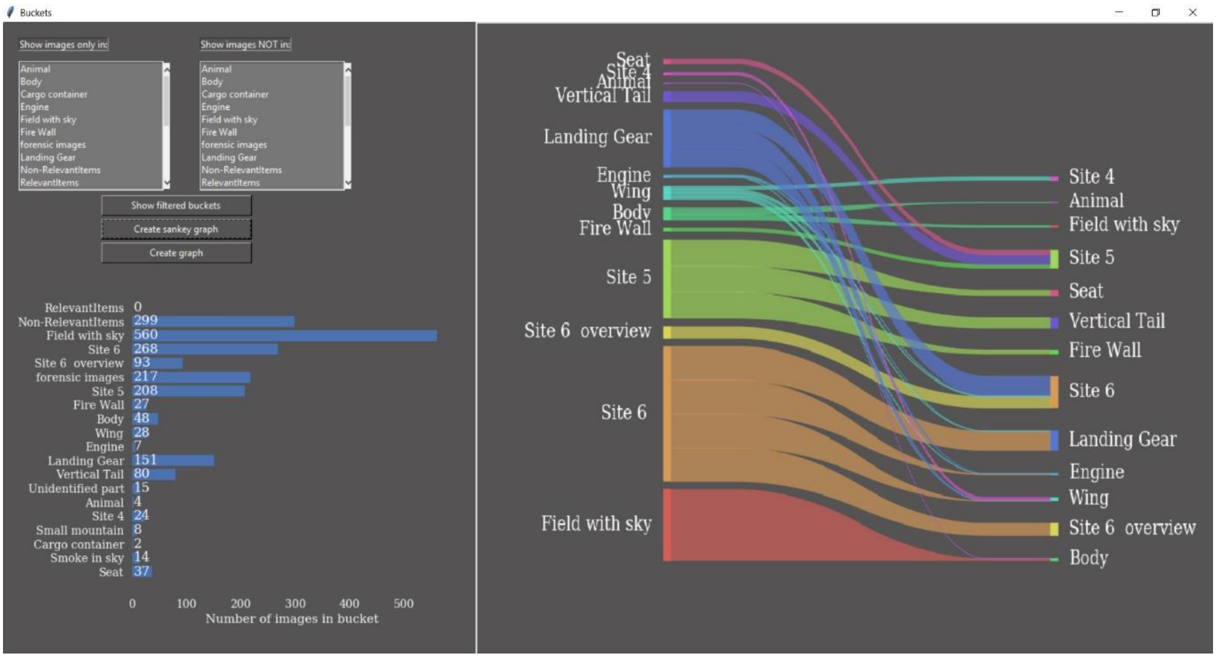

Figure 7 ImEx user interface, showing overall progress (left part) and a Sankey diagram to show the progress of the buckets capturing the categorization performed by the investigator (right part).

Conclusion

In this paper, an application is presented that has been developed after the MH17 investigation showed its necessity. ImEx generates clusters with images containing similar content, based on features extracted with a neural network. This allows the investigator to efficiently explore and search through a large image collection, bring structure to the data by placing images into user defined buckets, and show the relationships between the buckets through the Sankey diagram. This makes analyzing a large and complex image collection achievable through an efficient and clear process.

The app can be downloaded for free from https://tinyurl.com/imexapplication.

References

(1) Dutch Safety Board, Investigation crash MH17, 17 July 2014, 2015. Retrieved June 12, 2019 from https://www.onderzoeksraad.nl/en/page/3546/mh17-crash-17-juli-2014

(2) Griffeye Analyze. Retrieved June 12, 2019 from https://www.griffeye.com/the-platform/analyze-di/

(3) J. Zahálka and M. Worring, Towards interactive, intelligent, and integrated multimedia analytics. *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2014

(4) Olga Russakovsky et al., ImageNet Large Scale Visual Recognition Challenge, *Int J Comput Vis,* 2015

(5) Y. LeCun, Y. Bengio and G. Hinton, Deep Learning, 2015

(6)] K. He et al., Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016
(7) Dutch Safety Board, Bonfires The Hague (investigation ongoing), 2019. Retrieved June 12, 2019 from https://www.onderzoeksraad.nl/en/page/13141/bonfires-the-hague