FLORIS GISOLF
FGISOLF@HOTMAIL.COM
UNIVERSITY OF AMSTERDAM / DUTCH SAFETY BOARD

COMPUTER VISION IN ACCIDENT INVESTIGATION

Nowadays, many accidents are caught on video and photo. The aftermath of an accident is extensively photographed and minutely recorded by the public, press, police and safety investigators. Additionally, photos and videos are available of the accident site before the accident happened. All these data help the investigators with piecing together what caused the accident, and in case of a safety investigation, improving safety. Figure 1 gives some examples of recorded accidents.

a. Plane crash caught by dash cam

b. Ship collision caught by security cam

c. Crane collapses on houses caught by spectator

d. Crash site photo taken by journalist

FIGURE 1 ACCIDENTS CAUGHT ON CAMERA.

However, with the increasing number of photos and videos, it becomes a costly task to comb through all of the data. Investigation teams are often small, consisting of two to six members, and cannot focus solely on photos and videos. Furthermore, after seeing many photos, it becomes easy to unwittingly miss something.

The main objective of the investigator is to gain insight, to extract knowledge from the database. Using the available multimedia data, computer vision techniques can assist investigators analyzing the large number of photos and videos.

However, it should be noted that accident photos and videos are very different from those that can be found in standard databases such as ImageNet (Deng *et al.* 2009). Learned concepts such as cars, airplanes and buildings may be somewhat useful for an accident investigation, but an airplane after a crash obviously displays little similarity to one still flying.

Thus, for computer vision techniques to be useful, interaction with the accident investigator is needed. Such an interactive multimedia analytics approach has been described by Zahálka and Worring (2014). By combining several computer vision techniques and interacting with them, the investigator will be able to gain insight in the complex data.

For this essay, the MH17 aircraft crash investigation[1] is used as an example. The investigation team had collected 60,000 photos and 3,000 videos. In this case, no photos or videos of the crash itself were available. Access to the crash site was difficult due to ongoing fights in the area, which meant that the available photos and videos were of extra importance to the investigation team. Of few photos it was known up front where they were taken exactly, let alone what airplane part was visible. Two investigators analyzed all data manually, gradually gaining more insight as the investigation progressed. They were tasked with identifying the airplane parts depicted on the photos and videos, and where the airplane part was located, in order to help determining what had happened to flight MH17.

To find out how computer vision can help, a closer look will be taken at the process of an investigation, and in particular the photo analysis.

Pirolli & Card (2005) describe the process of an investigation with their sensemaking model (Figure 2). It consists of 16 actions and (preliminary) results, progressing towards insight. Computer vision can assist with these actions.

The investigation starts with accumulating data, in this case photos and videos. There are many different sources, such as the media, internet, passersby, security cameras, satellite imagery, police and investigators.

The gathering of new data never stops until the investigation is finished. The sensemaking model also shows this: there is a continuous looping of actions and preliminary results. Each action can take place many times, and results in updated (preliminary) results.

The external data sources contain many photos and videos; not all of which will be useful for the investigation. Photos and videos received directly by the investigators, such as from the media, civilians and police, are likely to contain mostly relevant data. However, there could still be failed photos, photos unrelated to the accident, and even manipulated photos.

[1] Dutch Safety Board (https://www.onderzoeksraad.nl/en/onderzoek/2049/investigation-crash-mh17-17-july-2014)

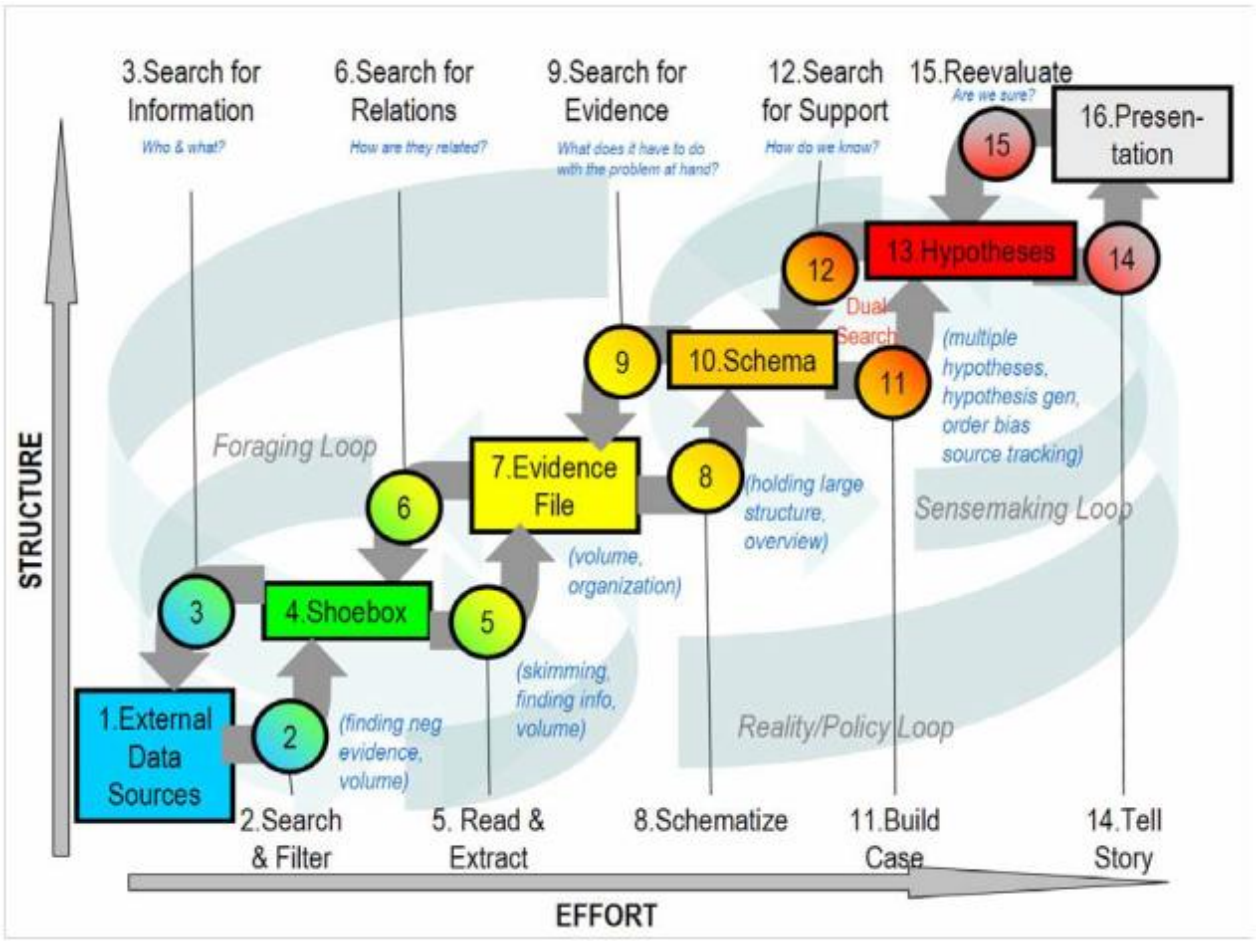

FIGURE 2 SENSEMAKING MODEL BY PIROLLI & CARD, WHERE CIRCLES DEPICT ACTIONS AND SQUARES DEPICT (PRELIMINARY) RESULTS

Action 2, searching and filtering the external data sources, means to (temporarily) discard the data not needed and results in a 'shoebox' (4) filled with images that contain information related to the accident and relevant to the investigators. Searching and filtering is one of the actions where computer vision algorithms can assist the investigator. While it can be done manually, considering the often limited availability of time and resources, a faster method is greatly appreciated.

There are several possible approaches to make step 2 easier and faster. Performing semantic multimedia analysis can filter photos based on the content as perceived by humans. It tries to answer questions such as whether there is a car or building on a photo. In this case, the investigator may want to know whether a photo contains an airplane, or part of an airplane.

A deep learning framework can be used for image classification, but as noted before, the existing classifications may only be of limited use for accident photos. Pre-existing classifications can be used to sort some of the wanted from the unwanted photos. For example, if an accident took place in the Netherlands, images depicting rainforests can be safely discarded.

New classifications could be trained. But training sets for accident photos are hard to acquire; the result of a crash looks different every time. However, just having images grouped, rather than classified, may

already provide the investigator with information to make the process of filtering faster; and it gives the investigator already a sense of the data.

The grouping of images can be achieved by looking at the extracted features of a convolutional neural network. Girshick et al. (2014) showed that visualizing the network's final convolutional layer[2], demonstrates what the network has learned. The visualization can then be used to determine interesting groups.

The result of action 2 is a 'shoebox' (4) filled with relevant photos and videos. The next step is taking a closer look at the data to make low-level inferences and start answering questions of the investigation and building hypotheses. Photos and videos significant for these inferences, questions and hypotheses will be placed in an evidence file (7). During this process, computer vision can assist.

Firstly, computer vision can be used to display the photos and videos in an intuitive way. Zahálka and Worring (2015) explore several possibilities of visualizing multimedia, such as a basic grid, similarity space and thread-based. Each one can be used and may lead to different insights.

For more in-depth exploring, the investigator may want to find more photos of interesting objects, or more photos taken at the same location. This can be achieved in several ways.

During the MH17 investigation, investigators used Microsoft PhotoDNA[3] embedded in Netclean Analyze[4]. Usually, PhotoDNA computes hash values of images to find copies of an image. However, the hash values also allowed for finding images that were similar, rather than a direct copy. This proved useful sometimes, but also resulted in a high number of false positives and negatives.

Recent research shows it is possible to train object recognition algorithms for photos and videos using just a single example (Fei-Fei et al., 2006; Tao et al., 2015; Meng et al., 2015). This may perform better than PhotoDNA, especially combined with a ranking-based system to show the results. The single example training set can be expanded, with hits found in the data set. This interaction with the investigator can improve the training set and thus the results.

Siamese neural networks are another way to find similar objects (Koch, 2015). Using these Siamese networks, similarity scores of images in the database compared to an image chosen by the investigator can be shown in a ranked fashion.

The described approaches for single example image recognition are fairly broad approaches. Some more specialized approaches may help in specific cases. Airplane parts for example, have manufacturer numbers and other text indicating their origin. Automatically finding photos containing text may help with identifying objects. Ye (2015) gives an overview of some of the available methods for text detection and recognition. Logo detection is also a possible approach (Tao et al., 2014) many airplane parts are branded with the manufacturer's logo.

In order to locate where an image was taken, buildings are a valuable source. Algorithms to detect buildings can help with quickly finding all photos with a specific building. Arandjelović and Zisserman

[2] layer $pool_5$, the max-pooled output of the Caffe implementation of the CNN described by Krizhevsky et al. (2012)
[3] https://www.microsoft.com/en-us/photodna
[4] Now Griffeye Analyze http://www.griffeye.com/

(2012) describe a combination of such algorithms that works well for retrieval of photos with buildings. Once the investigator knows where this building is located, all photos with this building are also located.

Once an evidence file is formed, the next step is to answer the questions, accept or reject hypotheses and gain insight of the database and ultimately understand what happened.

In order to achieve insight of a database as complex as that of crash investigations, many state-of-the-art methods have to be combined and applied intelligently and interactively. Current techniques can greatly assist investigations, and as techniques improve, more and more labor can be taken out of the hands of the ever time pressed investigator, achieving faster and better results.

REFERENCES

R. Arandjelovic and A. Zisserman. Multiple queries for large scale specific object retrieval. British Machine Vision Conference, 2012.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009.

L. Fei-Fei, R. Fergus, P. Perona. One-shot learning of object categories. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006.

R. Girshick, J. Donahue, T. Darrell, J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition, 2014.

G. Koch. Siamese neural networks for one-shot image recognition. MSc. Thesis – University of Toronto, 2015.

A. Krizhevsky, I. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012.

J. Meng, J. Yuan, Y.-P. Tan, G. Wang. Fast object instance search in videos from one example. Proceedings – International Conference on Image Processing, 2015.

P. Pirolli, S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proceedings of Internation Conference on Intelligence Analysis, 2005.

R. Tao, E. Gavves, C. Snoek, A. Smeulders. Locality in generic instance search from one example. IEEE Conference on Computer Vision and Pattern Recognition, 2014.

R. Tao, A. Smeulders, S. Chang. Attributes and categories for generic instance search from one example. IEEE Conference on Computer Vision and Pattern Recognition, 2015.

Q. Ye, D. Doermann. Text detection and recognition in imagery: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.

J. Zahálka, M. Worring. Towards interactive, intelligent, and integrated multimedia analytics. IEEE Proceedings on Visual Analytics, Science and Technology, 2014.