Visual Analytics for Semi-Automatic 4D Crime Scene Reconstruction

Niklas Weiler* University of Konstanz
Matthias Kraus* University of Konstanz
Timon Kilian* University of Konstanz
Wolfgang Jentner* University of Konstanz
Daniel A. Keim* University of Konstanz

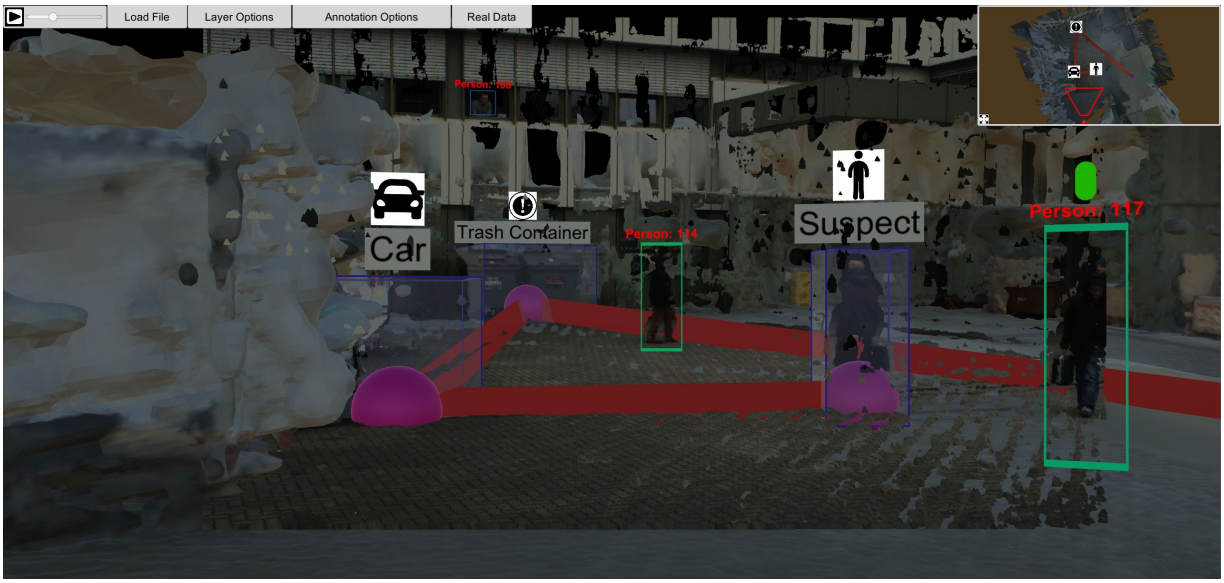

Figure 1: The interactive visualization displays point clouds which are reconstructed from multiple videos. The user is able to navigate the reconstructed scene spatially and temporarily. Automatic detections based on the imaging material are visible in the reconstructed scene. The user can further add manual annotations and animate them using an included path editor.

Abstract

During criminal investigations, every second saved can be valuable to catch a suspect or to prevent further damage. However, sometimes the amount of evidence that needs to be investigated is so large, that it can not be processed fast enough. Especially after incidents in public, the law enforcement agencies receive a lot of video and image material from persons and surveillance cameras. Currently, all these videos are viewed manually and annotated by criminal investigators. The goal of our tool is to make this process faster by allowing the investigators to watch a combination of several videos at the same time and giving them a common spatial and temporal reference.

Index Terms: Human-centered computing—Visual analytics

1 Introduction

After criminal incidents, analysts are often confronted with large amounts of video and image data collected by different people and surveillance cameras. Watching and annotating all these videos one after another often takes thousands of working hours. The video material typically originates from various sources including mobile phones, which makes it difficult to find temporal and spatial references. We report on a work-in-progress tool that provides an interactive visualization of point-clouds that are generated from multiple, various video sources.

In the VICTORIA-project [1], we are provided with point-cloud data and additional metadata. Project partners extract the point-clouds from videos. SLAM techniques such as ORB-SLAM [2] are used to extract 3D point-clouds from videos (stream of point-clouds for each video). Subsequently, the extracted point-clouds of the temporally aligned videos are merged to one single timeline. I.e., for an arbitrary scene, recorded by several sources, one common time series of point-clouds is generated. For efficiency purposes, static contents are extracted and stored for the entire timeline, whereas dynamic contents are stored frame-wise. Like this, investigators are able to browse the scene spatially and temporally, taking all sources into account. Neural networks, e.g., YOLO [3] allow to identify various objects in a scene and track them over time. The visual analytics tool developed throughout the VICTORIA project can include this information in the 4D (3D + time) scene which allows the investigator to easily track objects across several videos while allowing her an intuitive spatial and temporal navigation within the generated scene. Furthermore, our tool allows the criminal investigator to access the original material while analyzing the scene. Using this, analysts can verify their findings in the reconstruction of the source material and forward all relevant material to the responsible department.

*e-mail: firstname.lastname@uni.kn

2 Scene Reconstruction

The main view of the tool consists of a 4D reconstruction of the crime scene (Figure 1). All dynamic elements are encoded in the point-cloud. For each frame, one separate point-cloud is loaded (possibly from multiple video sources). Showing these point-clouds in rapid succession creates the impression of viewing a 3D video which the investigator can navigate through. Spatial navigation is realized by common interaction techniques using the mouse and keyboard. A mini-map (top right) of the scene is used to aid the investigator to navigate the scene. It implements standard interaction techniques such as panning and zooming, but it can also be used to track objects which are represented by user-defined icons. Temporal

Figure 2: The left image shows a static mesh of the reconstructed scene. The right side shows the original video frame. The analyst is able to seamlessly switch between the reconstructed scene and the imaging material.

navigation is implemented with a time slider at the top left corner.

Depending on the resolution of the involved videos, the point-clouds can grow quite large. However, to offer the investigator a smooth, video-like impression, it is necessary to show around 30 point-cloud-frames per second. To reduce this large amount of data, static parts of the point-clouds are removed and stored separately, i.e., areas in the point-clouds that do not change from frame to frame. Furthermore, the static elements are merged to one mesh which also leads to a more consistent view as the point-clouds often contain visual artifacts in the form of gaps.

3 Object Tracking

Objects in the scene are identified by deep neural networks. These networks are trained to identify several object classes which may be interesting during criminal investigations. For example, one network can identify cars in the scene and give information on what brand it is. These networks operate on the original video footage that was used to generate the point-clouds. Therefore, all detections are assigned to positions in the 2D image space of the videos. These 2D coordinates are then propagated into 3D coordinates to position them in the 4D reconstruction. Objects that are found in successive frames can also be tracked as a single object by the networks. This allows us to create persistent annotations for a specific object.

Automatic detections sometimes fail, or in other cases, a crime scene is only recorded after the incident. For such cases, our tool provides a manual annotation function that can be used to place annotations anywhere in the scene. These manual annotations can then be assigned to a user-defined path. Using this feature, the object can be tracked over time which allows answering further analysis tasks like analyzing which persons have met each other during the investigated time-frame.

Our tool also offers visual analytics techniques that assist in keeping track of how many instances of each class are identified at every second of the video. Depending on the incident that is being inspected, the analyst can decide which object classes are displayed. All other classes are not shown in the visualization to reduce visual clutter. Each interesting class is then represented by a heatmap showing the temporal distribution at the bottom of the screen. One possible use of this feature is the detection of interesting time points. For example, if many people rapidly move in the scene at the same time, they may be fleeing from something.

Another visualization offered by the tool allows tracking of individual object instances. A detection, for example, a suspicious person, can be marked as interesting. The tool then shows at which points in time this object can be found in the scene. This allows to quickly search an object across all videos used to create the scene. Especially for long-term investigations, where hour-long surveillance footage might be used, this can speed up the analysis.

4 Incorporating the Real Data

Insights gained while using the tool usually need to be verified using the source material. To assist with this task, the tool provides an intuitive way to access the source material, i.e., the actual video footage and not the extracted artificial digital model (3D mesh / point-cloud). The user can select any point in the digital scene and retrieve the original video footage from which the respective point was extracted and modeled. For example, in Figure 2, the user selected the wood stack in the virtual environment (left) and the original video footage is displayed (right). The original footage delivers more detail on certain parts as the digital model is distorted and abstracted. This is particularly helpful if there are multiple video sources or location-changing video sources for the considered point-cloud. Another way to retrieve the source material is the best-shot functionality. In a preprocessing step, a neural network detects the best-shot for each object, i.e., the image or video where the objects can be seen best. During the inspection of the 4D scene, the best-shot from the original footage can be viewed for any object that is tracked in the scene. This can be used to quickly find suitable images for further criminal investigations.

5 Future Work

Since criminal investigations are often rather complex, it is unlikely that the object detection approach works well in each scenario. Therefore, it will be helpful to give the analyst the capability to steer the machine learning models and provide feedback to the object detection algorithms. This is especially useful for retrieval- and query-by-example-tasks. Finally, the tool could be incorporated into reporting processes of the law enforcement agencies as it may be used as a valuable resource for reasoning-, sensemaking-, and justification-purposes; outcomes of the visual analysis should be reproducible and explainable in order to use the results in legal proceedings.

Acknowledgments

This work has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 740754.

References

[1] VICTORIA: Video analysis for Investigation of Criminal and TerrORIst Activities. `www.victoria-project.eu/`. Accessed: 2018-09-18.

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.