Semi-interactive construction of 3D event logs for scene investigation
Dang, T.K.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Introduction

1.1 Motivation

The increasing availability of cameras and the reduced cost of storage have encouraged people to use images and videos in many aspects of their life. Instead of writing a diary, nowadays many people capture their daily activities with a camera. When such capturing is continuous this is known as “life logging”. This idea goes back to Vannevar Bush’s Memex device [21] and is still a topic of active research [7, 150, 37]. Similarly, professional activities can be recorded with videos to create professional logs. For example, in home safety assessment, an investigator can walk around, examine a house and record speech notes at the same time. Another interesting professional application is crime scene investigation. Instead of looking for evidence, purposely taking photos and writing notes, investigators can just wear a head-mounted device and focus on finding the evidence, while everything is automatically recorded in a log. These professional applications all share a similar setup, namely a first person view video log recorded in a typically static scene. In this thesis we focus on this group of professional logging applications which we call scene investigation.

Our proposed scene investigation framework includes three phases:

- **Capturing.** An investigator records the scene and all objects of interest it contains using various media including photos, videos, and speech. The capturing is a complex process in which the investigator performs several actions to record different aspects of the scene. In particular, the investigator records the overall scene to get an overview, walks around to search for objects of interest, and then examines those specific objects in detail. Together these actions form the events in the capturing process.

- **Processing.** In this phase, all data are analyzed to yield information about the scene, the objects, and the events in the capturing phase.

- **Reviewing.** In the reviewing phase, an investigator uses the collected data to perform various tasks: assessing the evidence, getting an overview of the case, measuring specific scene characteristics, or evaluating hypotheses.
One important characteristic of the capturing phase is that it is difficult if not impossible to repeat this phase as the assumption of remaining static has a limited time-span. In crime scene investigation, for example, it is very hard to keep the crime scene untouched for a long time. Thus, it is best that the investigator captures the scene as completely as possible to avoid the need for further visits.

Depending on the specific reviewing task, the requirements on completeness and accuracy in the capturing phase vary. For example, measuring specific characteristics requires high accuracy but not completeness, while evaluating a hypothesis requires both completeness and sufficient accuracy.

Events of the investigation process constitute valuable information. Important events lead us quickly to important facts, while relations between events suggest connected facts. But all this information is only implicitly present in the data. The processing phase should uncover this information in a way that it is completely transparent to the user. This is the main subject of our research, how to turn data from the capturing phase into information that supports the various tasks in the reviewing phase.

In the traditional investigation process, experts take photos and notes of the scene and objects they find important. This standard way of capturing does not provide sufficient basis for the later processes of processing and reviewing. In particular, a collection of photos cannot give a good overview of the scene. Thus, it is hard to imagine the relations between objects, come up with hypotheses and assert them based on the photos alone. To get a better overview, in some cases, investigators use the pictures to create a panorama. But since the viewpoint is fixed for each panorama, it does not give a good spatial impression. The measuring task is also not easily performed using those photos when the investigator has not planned for this in advance. More complicated tasks like making a hypothesis on how the suspect moved are very difficult to perform using a collection of photos without a good sense of space. Finally, a collection of photos and notes hardly captures investigation events, which are important for understanding the investigation process. For supporting advanced tasks, new ways of capturing the scene are needed.

A potential solution to enhance scene investigation is to log the whole investigation process using a handheld or head-mounted camera, and to create 3D models of the scene. Research on techniques to capture crime scenes as 3D models has pointed out that 3D models can be used in many investigation tasks [72]. 3D models make discussion easier, hypothesis assessment more accurate, and court presentation much clearer. Using a video log to capture investigation events is straightforward in the capturing phase. Instead of taking photos and notes, investigators film the scene with a handheld camera. All moves and observations of investigators are thus recorded in video logs. However, in order to take the benefit, it is crucial to have a method to extract events from the logs for reviewing. When combined, 3D models and video logs have great potential to improve information accessibility. For example, a 3D model can help visualize the spatial relations between events, or details of certain parts of the model can be checked by reviewing events captured in that part of the scene. Together, a 3D model of the scene and log events form a 3D event log of the case. Such a log, apart from direct application like event-based navigation in 3D, will enable other applications such as knowledge mining of expert moves, finding correlation among cases, or training.

There are several paths towards creating 3D events logs: interactive, automatic, or semi-interactive. 3D models can be built interactively using 3D authoring software such as
Blender *. This, however, requires a lot of time measuring the scene, and interactive model building. As an alternative, 3D models can also be acquired automatically using laser scanners. However, laser scanners are expensive and not flexible, hence there are limits on their large scale deployment. Another automatic approach is modeling scene from image data. Under specific conditions, e.g. for certain outdoor scenes, it is also possible to automatically reconstruct a 3D model from video [121, 40]. This suggests that we could reconstruct a 3D model from an investigation log, and somehow combine it with the event analysis. However, as there are inherent limits on the accuracy one can obtain and under certain circumstances it might not even succeed [121, 23, 146]. As we will show in this thesis this is especially true when the aim is to do reconstruction of indoor scenes. A wiser approach is to handle the two problems, building 3D models and analyzing events, separately and leverage user interaction to overcome the limitation of the automatic algorithm, i.e. taking a semi-interactive approach.

Semi-interactive approaches for this problem are not commonly found, and there are several questions that need to be answered before this is feasible. Those questions are discussed in the following section.

## 1.2 Problem Statement

As discussed, fully automatic methods for 3D reconstruction would be ideal, but not always realistic because of their limitations in terms of accuracy and robustness. Yet we should not ignore them. Automatic methods hold technical elements for creating an effective and affordable method of 3D reconstruction. Studying them in the scene investigation context will help us understand more what computers can do in this application, and what the difficulties are that need user interaction to solve. In addition, automatic reconstruction from videos is directly relevant for estimation of camera motion and scene structure, which are both important features for analyzing the investigation log. Hence, our first question is:

**Q1:** To what extent can automatic methods be used for accurately reconstructing a 3D model from video logs?

As we have argued, using interaction is the best solution in practical applications. Users rather spend a dozen minutes to have an accurate result than waiting hours for a mediocre automatic result. The problem is that we should not overuse interactions and create just another manual tool. The proper way is to keep the interactions as limited as possible. To be effective, they should also be simple and intuitive. Once we answer the first question about what computers can do in 3D reconstruction, we will have the basis to answer the next question:

**Q2:** How to semi-interactively construct 3D models such that interaction is simple, intuitive and as limited as possible?

Once having 3D models of investigated scenes, our next interest is to analyze video logs of investigations, turning them into sequences of events. More specifically we are interested
in finding and analyzing events capturing the expert knowledge of investigators. Detecting events from videos is known as a challenging problem in literature since the information to be extracted highly depends on the context. Also most of existing works are about detecting events in the content, i.e. what is captured in the video, while in scene investigation we are also interested in how the content is captured. So our next question is:

Q3: How to detect and understand investigation events in investigations logs?

Finally, we consider how to use the results of analysis for navigation of the scene. To that end we should note that events are not only related in time, but also in space. Given that we have the 3D model of the scene (by answering Q2), the investigation process will be more comprehensible if its events are visualized in that model. Hence the last question is:

Q4: How to connect investigation events to the 3D model of the scene?

1.3 Organization

The following chapters of the thesis aim to answer the questions raised.

To answer Q1, we investigate in Chapter 2 available tools and methods in literature for fully automatic 3D reconstruction from image sequences or videos. We identify opportunities and challenges one faces when applying those automatic methods in the practical application of scene investigation. in chapter 3, using a theoretical model, we analyze one aspect that affects the accuracy of automatic 3D reconstruction, namely the feature location error. In Chapter 4, we fully investigate Q1 in practice. Every required step for automatic 3D reconstruction from indoor investigation videos are sketched and experimentally verified. This gives a clear picture of the performance a fully automatic method can deliver in scene investigation applications, and pave the way to answer other questions.

Drawing from the knowledge and experience in doing 3D reconstruction, Chapter 5 presents a semi-interactive system for reconstructing indoor scenes. This system is the answer to Q2, delivering an effective and accurate method that hopefully will bring more use of 3D models into scene investigation.

Chapter 6 aims to answer Q3 and Q4. We present a framework to analyze a video log of an investigation, and visualize it so that its events can be navigated in 3D. Together, Chapter 5 and 6 provides a complete methodology for building a 3D event log.

The process of creating and accessing 3D event logs is summarized in Figure 1.1. Our contribution is mainly in studying the processing phase for a scene investigation solution using handheld cameras to enhance users access to the captured scene.
Our proposed solution

1. Capturing
   - Automatic 3D reconstruction
   - Images
   - Video log

2. Processing
   - Semi-interactive 3D reconstruction
   - 3D model
   - Automatic video log segmentation
   - Event log
   - Semi-interactive matching

3. Reviewing
   - 3D event log

**Figure 1.1**: The proposed solution for improving scene investigation.