Forensic Science International 202 (2010) 26–35

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Tracking people and cars using 3D modeling and CCTV

Gerda Edelman, Jurrien Bijhold *

Netherlands Forensic Institute, Postbus 24044, 2490 AA Den Haag, Netherlands

ARTICLE INFO

Article history:
Received 27 July 2009
Received in revised form 9 April 2010
Accepted 13 April 2010
Available online 2 May 2010

Keywords:
Reconstruction
Surveillance camera
Video analysis
3D modeling
Tracking
Tracing

ABSTRACT

The aim of this study was to find a method for the reconstruction of movements of people and cars using CCTV footage and a 3D model of the environment. A procedure is proposed, in which video streams are synchronized and displayed in a 3D model, by using virtual cameras. People and cars are represented by cylinders and boxes, which are moved in the 3D model, according to their movements as shown in the video streams. The procedure was developed and tested in an experimental setup with test persons who logged their GPS coordinates as a recording of the ground truth. Results showed that it is possible to implement this procedure and to reconstruct movements of people and cars from video recordings. The procedure was also applied to a forensic case. In this work we experienced that more situational awareness was created by the 3D model, which made it easier to track people on multiple video streams. Based on all experiences from the experimental set up and the case, recommendations are formulated for use in practice.

1. Introduction

In forensic casework, CCTV footage can provide useful information about the crime, perpetrators or witnesses. With the growing number of security cameras, the amount of information increases rapidly. However, this makes the analysis of all data more time consuming and labor intensive.

A typical complicating factor in the forensic analysis of CCTV footage is that many different CCTV systems often use proprietary formats for capturing and recording the data, and come with different video player software for viewing the footage. Also, because of a limited storage capacity, footage is usually recorded time-lapsed, with varying frame rates, and in a low resolution. These factors make the interpretation of CCTV footage difficult and almost impossible when footage from different systems has to be viewed simultaneously. Next to this, a certain amount of situational awareness is needed to be able to track people through different video streams. This means that all people involved in a criminal investigation would have to make themselves familiar with the scenes that are visible in the footage at a considerable cost of time.

The question arises whether CCTV footage could be used more effectively with the help of three-dimensional models (further referred to as 3D models), projected onto the footage. 3D models are expected to give more insight in the situation, which helps analysts understanding the contents of the evidence [1]. Also, a 3D reconstruction could be useful for the visualization of events, as was shown for other forensic cases [2–4]. Instead of writing viewing reports, a 3D animation could be produced to note observations. Assumptions could be added to the animation to test different hypotheses. This approach could especially work very well if all video footage could be converted into video files with a standard frame rate and a fixed starting time.

Based on these ideas, a procedure was set up for tracking people and cars in multiple video streams using 3D models of the environment. The procedure was further developed and tested in an experimental setup, in which GPS coordinates of test persons were logged. Results of the experiment are discussed and compared to the GPS data. After this experiment, the procedure was tested in a forensic case. Based on all experiences, recommendations are formulated for use in practice.

2. General method

The general procedure to track people and cars in multiple video streams using 3D models of the environment consists of four steps:

(1) synchronization of all video streams in the CCTV footage;
(2) 3D modeling of the area;
(3) creation of virtual cameras in a 3D model;
(4) 3D reconstruction of people and cars.

These steps are described in more detail in the following sections.

* Corresponding author. Tel.: +31 70 888 6411, fax: +31 70 888 6559.
E-mail address: j.bijhold@nfi.minjus.nl (J. Bijhold).

0379-0738/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2010.04.021

2.1. Synchronization of the video streams

To be able to easily switch between different camera views of the same time interval, all video streams are synchronized. This means that CCTV footage has to be processed such, that all video streams have the same start time and the same frame rate.

An important step to reach that result needs to be taken when the footage is collected. At that moment, the timing of the video system should be checked with the real time and date, so that it can be adjusted if necessary. Without checking the system clocks, different video streams can only be synchronized by using an event that is visible on all these streams. However, when there is no overlap between the camera views, this is usually impossible and an assumption has to be made about the time difference between the video streams.

Next, time-lapsed video streams are processed such, that the time interval between two successive frames corresponds with the time between the recordings. To this, recording times of all frames are needed, which can often be found either in the data of the video stream, or as a timestamp on top of each image. In the first case, software tools have to be found, or created, to extract the information needed. In the latter case, Optical Character Recognition (OCR) [5] can be used to read the timestamp in the video files.

At the Netherlands Forensic Institute (further referred to as NFI), the software MediaSync is developed, which uses OCR or the digital time sources to order frames chronologically and with the right amount of time between two frames. Small gaps in a video stream are filled with a copy of the last image, until all video streams have the same, constant frame rate. Big gaps are filled with blank images. The software stores the time stamps resulting from OCR in a list with the frame numbers. As an option, this list can be corrected manually before it is used for the production of a synchronized video stream.

2.2. 3D modeling of the area

A 3D model of the area surrounding the security cameras is created. Apart from using tape measurements, there are different techniques to obtain a 3D model [1,6]. Since many decades (stereo) photogrammetry has been used to determine the 3D coordinates of objects. More recently, different 3D scanning technologies have been developed, e.g. phase-shift laser scanners and time-of-flight laser scanners. Difference can also be made between aerial and terrestrial scanning methods. The method chosen depends on the size of the area and on the level of detail needed.

2.3. Creation of virtual cameras in a 3D model

The 3D modeling software 3dsMax (Autodesk) is used for the analysis of the CCTV footage. The 3D model is imported in 3dsMax and a video stream is played in a viewport background. To be able to reconstruct the positions of people in the camera images, virtual cameras are created which have the same field of view as the real security cameras. Fig. 1 shows a user screen of the modeling software. The left viewports show a perspective view and a top view on the 3D model. In this example, the buildings are gray objects with a low level of detail. A satellite image was used to texture the ground plane. The virtual cameras are shown white. The upper right viewport shows one of the video streams. The lower right viewport shows the view of the corresponding virtual camera. Note that the satellite image not only shows the sidewalks and the street furniture but also the cars that were parked during its recording.

For creating a virtual camera, seven parameters have to be estimated: 3 for the position, 3 for the orientation (pan, tilt, roll), and the focal length. 3dsMax provides the user with an automated procedure that estimates these parameters by minimizing the mean square distance between a number of user selected points in a video image with corresponding points in the 3D model which have to be identified by the user. This procedure is referred to as camera matching. This method is explained in more detail in publications on photogrammetric analysis of CCTV images [7,8].

In case not enough match points can be found for an automatic camera match, the virtual camera parameters have to be estimated interactively by trial and error. In that case, the locations,

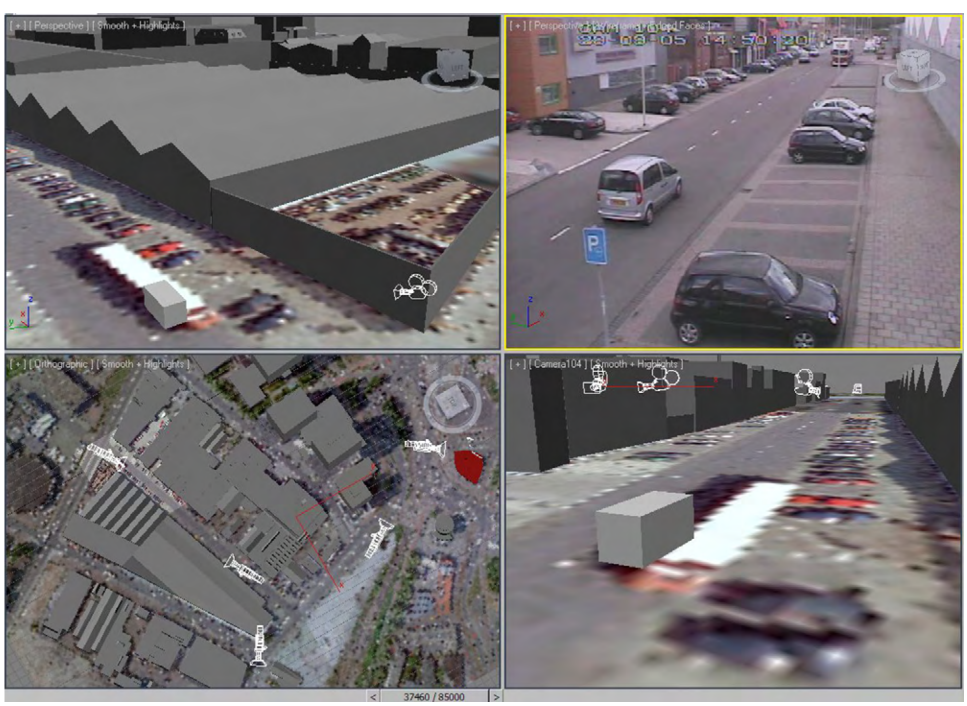

Fig. 1. Viewports of the software 3dsMax, in which the 3D model and the virtual cameras are shown from different perspectives. The top right viewport shows a camera image, which is reconstructed by a virtual camera in the bottom right viewport. The car in the image is reproduced as a box in the 3D model. By using the slider one can step through different frames of the video stream and the animation of the box.