MIXED REALITY IN TRAFFIC SCENES

A VIEW SYNTHESIS APPROACH

A.B. Martínez, J.P. Arboleda, E.X. Martín, C. Torrens
Dept. of Automatic Control and Computer Engineering. UPC. SPAIN

Abstract: Our research group, interested in outdoor scenes, has developed a methodology to register a CAD model of a city, with images taken with video cameras installed in a car, while driving city streets. So, we can merge video captured data and the render of elements of the model into one image, forcing real and virtual camera points of view, orientations and parameters to coincide. Consequently, we generate real-video sequences and can insert, in real time, virtual objects congruently placed in relation to real objects of the scene. With the aim of increasing realistic likeness of inserted objects, we propose an alternate method that uses view synthesis techniques instead of render techniques. From a finite set of pictures of a real object, taken from different points of view, we are able to generate any point of view object projection, to be inserted in the physical coherent location.

Key words: Outdoor registration, traffic scenes, view synthesis, mixed reality.

1. Introduction

This paper presents a work located in the augmented reality (AR) side of the taxonomy of mixed reality continuum introduced by [Milgram 99]. Our work uses AR to merge virtual images into a video image sequence with the aim of recreating a realistic traffic scenario and using it as a driving simulator.

The AR application we have developed consists of dynamically inserting stereoscopic virtual objects (i.e. traffic signals, cars, pedestrians, etc.) in previously recorded stereoscopic videos of the real world. This technology allows the creation of image sequences with dynamic computer-controlled virtual traffic situations. In this application, the objective of these

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35660-0_65
R. Nakatsu et al. (eds.), *Entertainment Computing*

stereoscopic images is to recreate in a realistic way situations that may occur on a road and that are difficult, expensive or dangerous to reproduce. The resulting merged world is seen by a driver with a head mounted device (HMD). The driver can interact partially with the system through a pair of brake and accelerator pedals and a steering wheel.

This methodology needs special techniques to correlate singular features within images with their known positions in the virtual world; registration covers this field of research. Besides, a set of position sensors: encoders, differential GPS and radio signals will give additional data to increase precision in this registration.

When the viewer position is known, and video image is selected, we are able to augment recorded reality with the insertion of virtual objects. While common techniques render object models, we propose to place real views of real objects in the correct location, synthesised at video-rate, allowing an interactive behaviour and more realistic appearance of the application.

View synthesis strategies allow us to generate any view of an object or scene, if enough pictures have been taken, and some restrictions are considered. After reviewing different view synthesis approaches, we have developed a method to obtain a minimum picture database of an object, and interactively and consistently put a physical-coherent synthetic view of the object in place of the image.

2. Video-Based Augmented Reality Stereoscopic System

2.1 Functional description. Settings.

We can identify several processes in the whole system, according to a functional division viewpoint:

- Record a stereoscopic video sequence of the real world.
- Acquire position sensor data of the car.
- Digitize video and sensor data.
- Build a virtual world.
- Make geometric registration with both real and virtual worlds.
- Show the augmented scenes to the driver.
- Record and evaluate the behaviour of the driver.

Figure 1 shows a block diagram of the steps of this process; the first step is shown in figure 1.a. Two stereo-videos are recorded in two VTR from a vehicle running on real streets of a big city. Position and velocity of the vehicle are obtained using Differential GPS (DGPS) technology, and a factory-made odometer sensor. Besides, GPS time information is used to synchronize the stereoscopic videos.

Data acquired in first step have to be pre-processed. This is the aim of the second step (figure 1.b), where the data are digitized and arranged in a proper manner. It consists of the creation of a database of pictures and spatial positions, conversion from GPS latitude-longitude data to UTM units and synchronization of the videos starting from the GPS time data.

Third step is geometric registration (figure 1.c). We use a topographic map of the streets as a partial model of the world. With this, the digital database created in the previous step, and a kinematics model of the car, the road and the vehicle's position will be known at any time.

A set of realistic stereoscopic objects is created separately, some of them are constructed entirely with CAD tools, and others are obtained from real pictures. As a result of the third step, we obtain a stereoscopic augmented world, with real views of the streets as seen by the driver and some adding (the virtual objects) placed in congruent locations with reference to the car's position (figure 1.d), that can be used to generate several applications like driver evaluation systems.

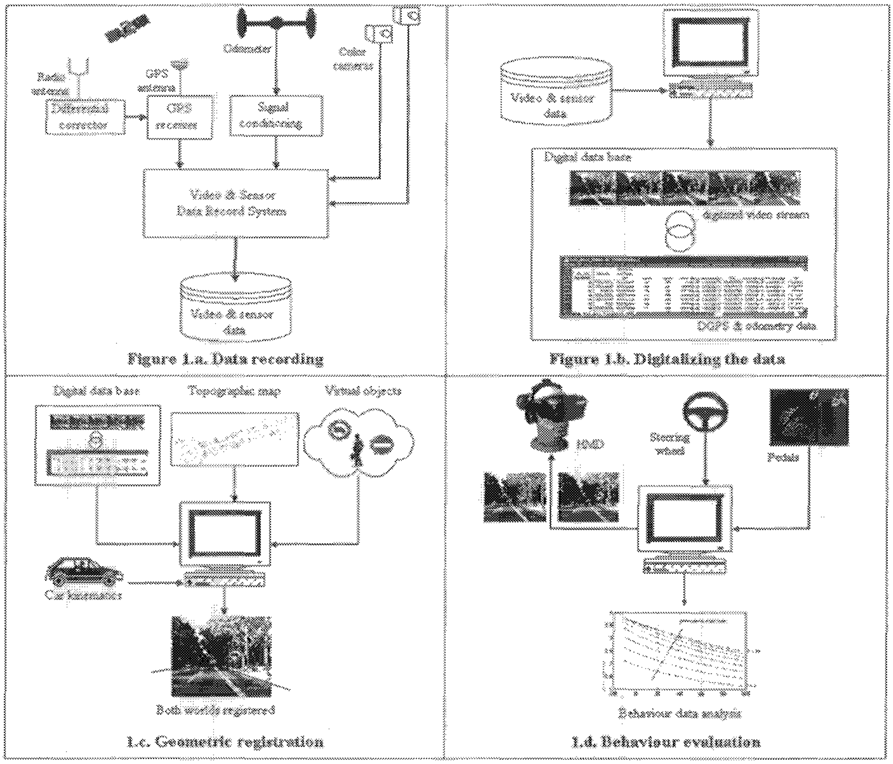

Figure 1. Functional description

2.2 Geometric registration.

The geometric registration process is based on hybrid sensor techniques, i.e. we take profit of the advantages of different sensors and compensate the drawbacks of each of them. Instead of looking for a single sensor that allows the registration, many researchers groups are have been using hybrid techniques ([Hirose 99]; an overview can be found in [Arboleda 00]).

The results obtained with the differential GPS receiver show that the geometric registration problem can't be solved based on this single sensor; there is an absolute position error, and a *multipath* error that appears when the signal follows more than one way to reach the receiver's antenna. So we need to add information from other sensors, like odometers and computer vision, to know the exact position of our car.

We use a vision system as a smart sensor to reduce the position errors of the GPS data, and cancel the odometric error of the dead reckoning process. The goal of our image-processing algorithm is to detect some special points in the image that could serve as a reference for the registration process. The algorithm analyses the colour image in a centred Region Of Interest (ROI) and finds all the white lines painted on the road (figure 2, left). After the lane detection, we follow the bus lane by means of a tracking algorithm and annotate the occurrence of all the pedestrian crossings, which serve us to cancel the odometric errors after the dead reckoning process. We use the lateral distance of the car to the bus lane as a method to estimate the steering angle caused either by action of the driver over the steering wheel, or by changes of the slope of the bus lane.

More details about the geometric registration method can be found in [Arboleda 00].

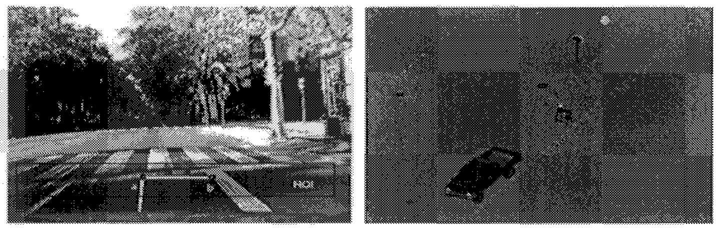

Figure 2. Image processing example and CAD model example.

2.3 Augmenting reality.

Figure 3 shows a set of images extracted from an augmented reality video and a view from the CAD model augmented with the virtual objects. A different view of the objects in the CAD model is shown in figure 2 (right).

We show a sequence of video images of an avenue, augmented with the presence of a car, a direction traffic signal and a water pump.

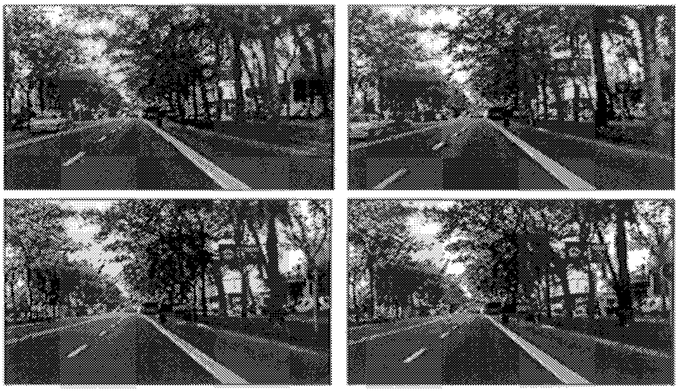

Figure 3. Sequence of augmented images.

3. View Synthesis

View synthesis is a set of techniques used for generating an image of a scene, from a point of view where no real camera has been placed. This implies that we must use information given by some real cameras and perhaps, a model of the location, to produce visual data. With regards to entertainment purposes, view synthesis gives us the ability to project an image of a real object, from any point of view, in a virtual or real world, as the object was there. This view of a real object will be more than a model-rendered image.

3.1 Overview.

Authors work in view synthesis under three different perspectives:

- Building previously a 3D model of the scene. Then, the photometric information given by the cameras will be used to render the new images, texturing the surfaces projected by the 3D model. The group of [Kanade 00] in the Carnegie-Mellon University is using this method, under the generic name of virtualised.
- Inferring the 3D structure from the images. Using stereovision techniques is possible to reconstruct dynamically the 3D structure of the scene and re-project the new view. A research group in France [Laveau 97] has

created special representation structures to optimally generate the new views.

- Dealing directly with the images, without using explicitly the three-dimensional information. Employing epipolar geometry properties and the photometric data, it is possible to interpolate directly the pixels of the new image with the pixels of the captured ones [Naemura 99].

3.2 Towards interactivity.

To produce attractive mixed reality applications, interactivity will be an essential achievement. These days, we can see in TV spots and in some movies, images of real objects from unbelievable points of view; but in both cases desired new points of view are known a priori, and is possible to spend some time in rendering a perfect image. When giving interactivity, i.e. freedom of moving dynamically our point of view, we will need to generate these views at video-rate (e.g. 20 or 16 ms per frame).

First and second methods presented in 3.1 are, nowadays, unreachable to implement at video rate, but the third family of procedures, dealing directly with images of real objects, have some features allowing hardware, software and algorithmic acceleration. The methodology basically consists in this steps: photometric data acquisition, calculation of correspondence maps between acquired images, and when the new desired point of view is known, rectification of camera points of view [Scharstein 99], to obtain a geometry were easily interpolate pixels for the new view.

Analysing method steps, we detected that is possible to make some processes in batch time, and only rectification and synthesis of the new view must be done in real time, complying video-rate constrictions. As we can save and pre-calculate correspondence maps in batch time, it appears a trade-off between the amount of data to be stored, and process time to synthesise new views. We have tried to obtain the best rate between them to generate video-rate images of objects, to be inserted in our video sequences.

4. Synthesising Images for Mixed Reality

4.1 Acquiring photometric database.

The first step to synthesising views of objects is to obtain enough information about them, so we have constructed a dome structure to take many pictures of an object. Figure 4 shows a slice of the object view database. The object, in this case a car, is seen from N points of view, in a circular distribution every 360/N degrees. Required view density is calculated from average object complexity.

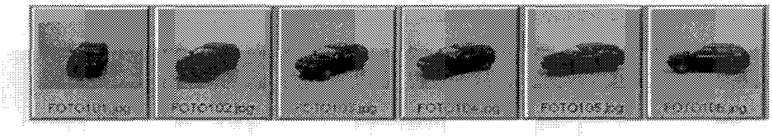

Figure 4. Object view database example.

4.2 Generating pixel correspondence maps.

We must calculate the pixel correspondence maps for every pair of neighbour even in batch time. Thus we obtain N correspondence maps. A correspondence map is a list of pixel pairs, the first figure of the pair corresponding to one image, and the second to the other one. Pixel matching is done with an minimization algorithm of absolute intensity differences [Kanade 94].

4.3 Rectifying and synthesising views.

When the desired point of view is known, we search the database for the two images closest to the required point of view and project them onto a plane parallel to that which contains the three points of views. In this way we obtain a rectified pair of images [Scharstein 99] (see figure 5), allowing pixel linear interpolation to generate the new one [Martín 01]. Figure 6 will show a synthetic view of the car, inserted in the video sequence.

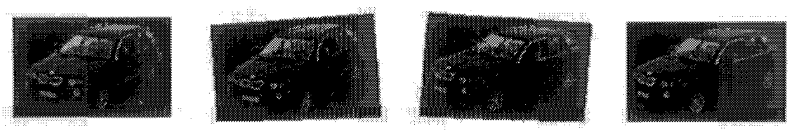

Figure 5. A pair of database images (corners) and the rectified ones (centre).

5. Conclusions: Binding View Synthesis and Augmented reality

The result of our work is that we can augment reality, or sequences of real video images, with the insertion of synthetic views of real objects, as represented in figure 6. Having registered virtual and real world, we are able to coherently blend images of rendered models and real object views.

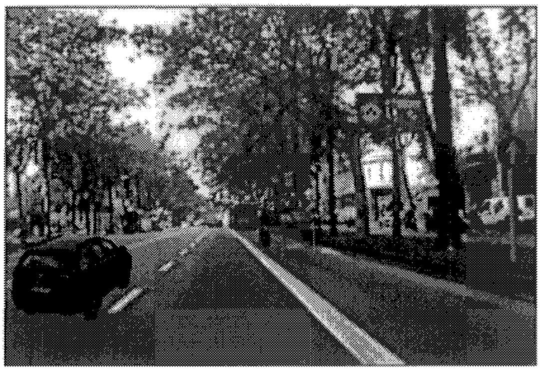

Figure 6. Augmenting reality with views of real objects.

The paper has presented a methodology to register video images with a model of the city, using a variety of sensors mounted on a car. Besides, we show a method to obtain a view database of an object, and synthesise intermediate views. Finally, we developed an application to insert the real object views in registered places at video rate.

6. References

[Milgram 99] P. Milgram, Herman Colquhoun Jr. "A Taxonomy of Real and Virtual World Display Integration", Mixed Reality - Merging Real and Virtual Worlds, Ohmsha (Tokyo)-Springer Verlag (Berlin), pp. 5-30, 1999. ISBN 3-540-65623-5 (Springer Verlag)

[Hirose 99] M. Hirose, T. Tanikawa, T. Endo, "Building a Virtual World from the Real World", Mixed Reality - Merging Real and Virtual Worlds, Ohmsha (Tokyo)-Springer Verlag (Berlin), pp. 183-197, 1999

[Arboleda 00] J.P. Arboleda, A.B. Martínez, E.X. Martín. Video-Based Augmented Reality Stereoscopic System: an application to traffic scenes. Proceedings SPIE, Ed. Merrit, Benton, Woods. PP 23-32.

[Kanade 00] Takeo Kanade, Research Group in Computer Science, Carnegie Mellon Univ., Virtualized Reality Home Page: http://www.cs.cmu.edu/~virtualized~reality/

[Laveau 97] Stéphane Laveau, Olivier Faugeras, "Oriented projective geometry for Computer Vision", Tech. Report INRIA Sophia-Antipolis, 1997

[Naemura 99] Takeshi Naemura, Hiroshi Harashima, The University of Tokyo, Japan "The Ray-Based Approach to Augmented Spatial Communication and Mixed Reality", from the book Mixed Reality, merging Real and Virtual Worlds, edited by Y. Ohta, H. Tamura, Ed. Ohmsha, Ltd. Tokyo 1999.

[Scharstein 99] Daniel Scharstein, "View Synthesis Using Stereo Vision". Ed. Springer-Verlag, Berlin 1999. ISSN 0302-9743

[Kanade 94] T. Kanade Development of a video-rate stereo machine. Image Understanding Workshop, Monterey, CA. November 1994. Morgan Kaufmann Publishers.

[Martin 01] Antonio B. Martínez, Enric X. Martín "Generation of synthetic views for teleoperation in industrial processes" Dept. ESAII. UPC. Proceedings ETFA'01. Sophia-Antipolis October 2001