Steerable Video: Generating Video-Based Environments for Driving Simulation

Michael Brogan [1], Charles Markham [2], Sean Commins [2], Catherine Deegan [1]

(1): IT Blanchardstown
Dublin 15, Ireland
+353 (0) 1 885 1093
E-mail : {michael.brogan, catherine.deegan}@itb.ie

(2): NUI Maynooth
Co. Kildare, Ireland
+353 (0) 1 708 3383
E-mail : {charles.markham, sean.commins}@nuim.ie

Abstract – This paper describes an approach that allows for a steerable environment to be generated directly from a video for the purpose of integration with a video-based driving simulator. As the range of steering motion in a driving simulator is relatively limited, a pseudo-three-dimensional approach can be taken. This method requires only a single image sequence or video, acquired by any type of imaging system along a road. No three-dimensional, stereo or visual odometry data is acquired or calculated.

Key words: Video-based, driving simulation, pseudo-three-dimensional, photo-realistic

1. Introduction

Using video as a basis for a driving simulator's visual cue stream is relatively rare, with the vast majority of simulators using a graphical-based environment. There are two primary reasons for this; the ease at which full three-dimensional graphical models can be generated, and the difficulty in introducing a form of interactive control over a pre-recorded video sequence. There are several existing video-based driving simulators, examples being [Bre1, Bro1, DeC1, Her1, Ono1, Sat1].

The simulator described [Ono1] uses a combination of photo-textures and graphical models to create a sense of high-fidelity, but relies on graphical models for population and execution of scenarios. The image data acquired are used to texture the general environment in the far-distance, but all near-distance features, such as roads, traffic and road boundaries are graphics-based.

The simulator described in [Bre1] utilises data acquired from a high-cost mobile mapping system to generate a wire-frame environment. This environment is generated using data returned from a LiDAR laser-scanning device. Once generated, this skeletal structure can be textured using multi-view imagery acquired by the mapping system's cameras. This method produces a highly detailed photo-based environment that allows users to drive around a three-dimensional environment. As the generated environment is three-dimensional the benefits associated with traditional graphical-based environments remain present; scenarios can be introduced, and full control over the driving simulator world is present.

These systems have introduced photo-based environments to driving simulation, allowing the user to steer on or around textures, as opposed to through video sequences. The steering mechanisms are similar to previous graphical-based simulators, with the graphics having been replaced with photographic textures.

The driving simulator described in [DeC1] addresses this by using dual video sequences that allows the simulator visual cue stream to switch video feeds dependent on the position of the driver in the simulator. This approach, in and of itself has a major limitation, in that to acquire the data, the road segment must be driven twice, once in the wrong direction. This is possible if the road is not open to general traffic, but in terms of general data acquisition, is not a feasible approach unless road closures are coordinated with a governing body. Again, the ability to steer through a single video remains absent; the change in video feed does not allow for a gradient change in perspective.

Research has used non-steerable videos to demonstrate high correlations among video, graphical model and ground-truth speeds [Bro1], as well as augmenting different video sequences [Her1]. However, without the ability to steer, the full dynamic behaviour of driving cannot be realised. Such a capacity would allow for driver speed, position and the effect of one on the other to be measured using a video-based visual cue stream.

This paper describes a method by which a pseudo-three-dimensional photo-realistic video-based model can be generated to allow for a steerable environment to be generated without the need for stereo images, synchronization, calibration, correspondence, or three-dimensional reconstruction. However, the absences of these introduce some constraints that will be discussed, including the introduction of distortions around road boundaries, based upon lack of sufficient data.

It is divided into seven sections; section one gives an introduction to the topic and describes video-based driving simulators. Section two describes the route selected for testing purposes and details the video camera used for data acquisition. The testing system is also described. Section three explains how the steerable environment is generated from the single video sequence. Section four details how this steerable environment is interfaced with the driving simulator. Section five gives an overview of the geometry of the generated environment. Testing of the technique and the associated results are presented in section six, and conclusions based on these results are drawn in section seven.

2. Methodology

For the purposes of developing and testing the technique described in this paper, an off-the-shelf Mio MiVue 388 witness camera was mounted on the internal side of the windshield of a standard vehicle, and a two minute video of a road was acquired. This video was in High Definition format (1920x1080 resolution with a frame rate of 29.97 fps) [Mio1].

The selected route was the link road between the M3 motorway and R147 rural road in Kells, Co. Meath, Ireland. It is shown in Fig. 1, alongside an example of the road scene as acquired by the camera.

For the purposes of testing the technique described in this paper, a Microsoft Windows 8-based notebook with 15.6 inch, 1366x768 resolution widescreen display was used, with a Thrustmaster gaming steering wheel [Thr1].

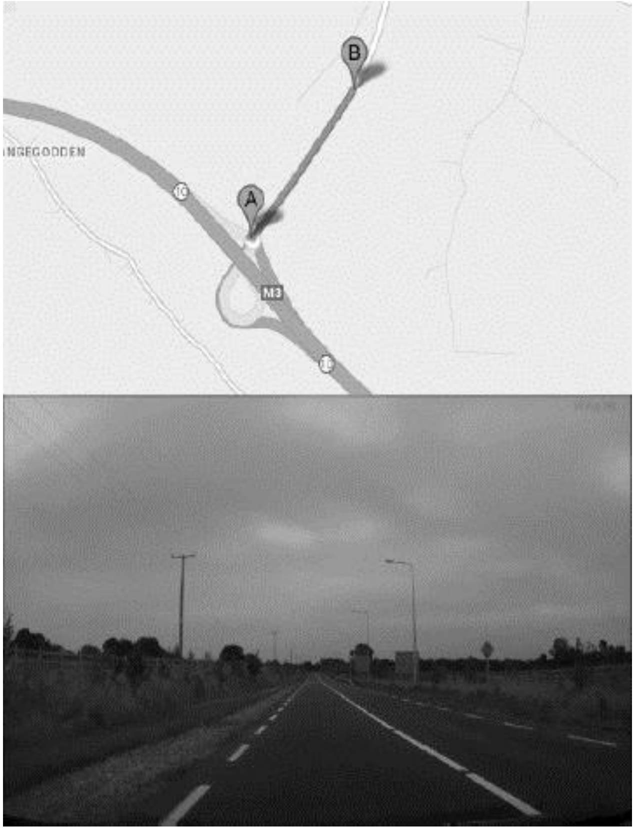

Fig. 1. Selected data acquisition route (R147).

3. Environment Generation

Generation of a full three-dimensional environment requires the acquisition of multi-view geometry of a scene, commonly using two camera views (stereo) or multiple camera views. The disparity in the resultant images can be used to generate a dense stereo description of the scene, called a depth map. This map is a greyscale description of the depth of a scene, where the lower the pixel intensity, the further the feature lays from the reference camera, and the higher the pixel intensity, the closer the feature lies to the reference camera [Har1, Tru1]. In an environment where the geometry of the scene remains relatively constant, such as the road scene video acquired by the system on the test route, a dense stereo depth map can, instead of being generated using stereo-view or multi-view image sequences, be estimated using the horizon line.

This allows for pseudo-depth perception to be inferred onto a single image. The process by which a depth map can be estimated from a single image is described next. By its nature, the line at infinity lies at an infinite distance from the acquiring camera. As the line at infinity is represented in images by the horizon, the horizon must first be identified in the single image. Once this is achieved, a greyscale gradient is generated, beginning at zero pixel intensity on the horizon, and increasing gradually to maximum pixel intensity to the bottom of the image. An example of a depth map generated using this approach is shown in Fig. 2.

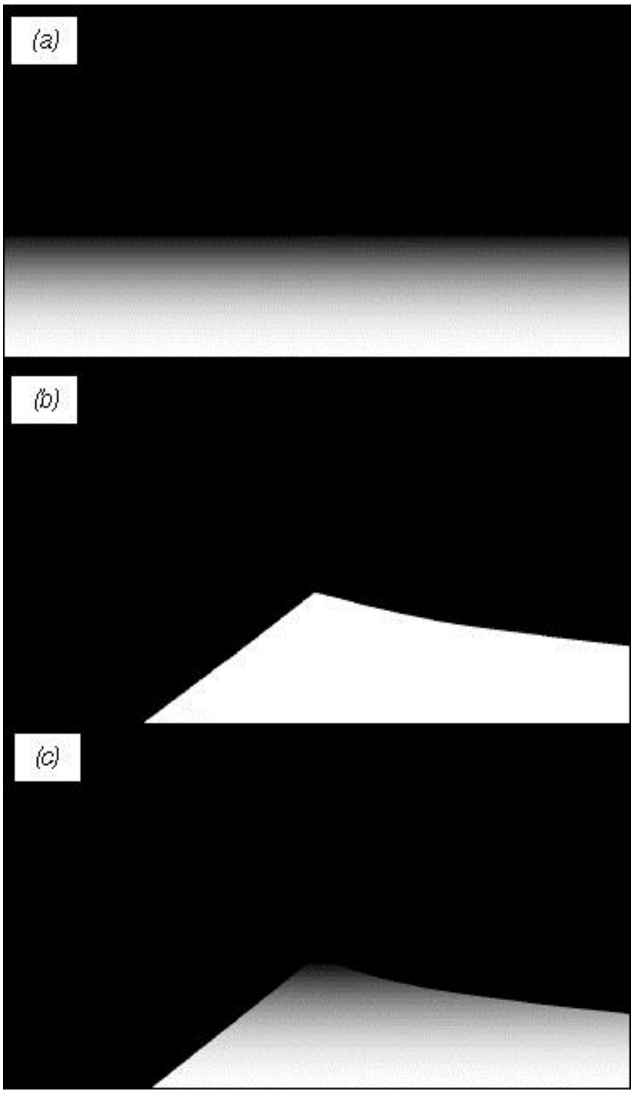

Fig. 2. (a) Horizon-defined depth map; at this stage all features below the horizon line can be assigned a depth. (b) Road area mask; identifies the road area in the image. (c) Masking (a) with (b) allows only the road to be assigned a depth.

Once this is achieved, the primary feature which requires a sense of depth is that of the road. Again, by its nature, the road will begin nearest the camera, and, as the edges of the road are parallel, will meet at the horizon [Har1]. This produces a mask whereby the depth map is masked so that a depth map describing the road only is estimated.

This allows a greyscale gradient to be applied to the corresponding road area, beginning at zero pixel intensity on the horizon, and increasing gradually to maximum pixel intensity when the area closest to the acquiring camera is reached. Once the depth map has been estimated, and the geometry of future images in the video sequence is approximately the same, the same depth map can be used for each image. The next step that must be undertaken is to infer depth perception on a video frame using this depth map. This is done by layering the original RGB video frame over the depth map, essentially creating an RGB-Depth (RGB-D) image from the original RGB image and the estimated depth map.

Each value in the depth map is assigned a corresponding distance within a virtual graphical environment, and the virtual camera is situated at a constant distance from the displayed pseudo-three-dimensional image. Lack of sufficient data is evidenced in the pseudo-three-dimensional images as warping and distortion of areas close to the road edge, where the depth map was primarily estimated. This occurs where the depth map boundary does not map the road segment to the road plane, causing distortions between the road and non-road planes.

4. Interfacing Environment

To allow for steering around the three-dimensional image, the camera view is linked to the driving simulator steering wheel position, thereby allowing for the view of the road scene to change based on the user’s input. The implementation of steering within a video sequence is completed by sequencing the video frames, such that upon any pressure applied to the driving simulator’s acceleration pedal, the next frame of the video is loaded, textured and displayed at the correct viewpoint, based upon the current orientation of the steering wheel. A sequence of such images is shown in Fig. 3.

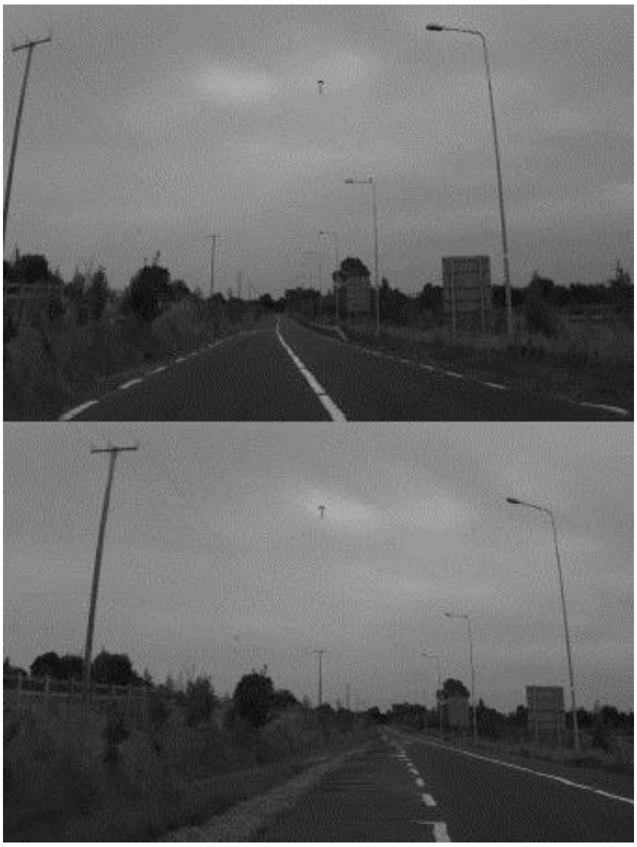

Fig. 3. Estimated depth map applied to image sequence.

5. Environmental Geometry

The generated environment consists of a local world three-dimensional co-ordinate system, which is viewed using an arcball camera, centred on the central horizon point. Two two-dimensional planes are used to generate the pseudo-three-dimensional environment; the non-road plane, $\pi_{N,}$ and the road plane, π_R. π_N will contain all non-road features, and π_R will contain the road only.

This results in π_N being coincident with the XY_{World} plane, and π_R being non-parallel with the XZ_{World} plane, and creates the pseudo-three dimensional environment that enables steerable video to be generated. The offset between the world co-ordinate system origin and the intersection of π_R with π_N is dependent on the position of the horizon line in the acquired video.

π_R is divided into 256 subsections, along the Z_{World} axis, each relating to a depth map greyscale intensity. This allows for the depth map to define which pixel intensities are to be mapped to the corresponding π_R Z_{World} axis line. The parameters of the arcball camera are defined in terms of its rotation and translation with respect to a fixed point in the viewing scene. That is, the projection of the world co-ordinate scene is defined in terms of three parameters: the fixed point, the camera's rotation around this point, and the camera's translation from this point. The fixed point can be described in terms of a 3-vector, with the position of the camera being described in terms of a 3-vector, with an *X, Y* co-ordinate describing the position of the camera relative to the surface of the viewing arc, and a *Z* co-ordinate describing the radius of the viewing arc from the point of focus. The orbit of the camera's X-axis is shown in Fig. 4.

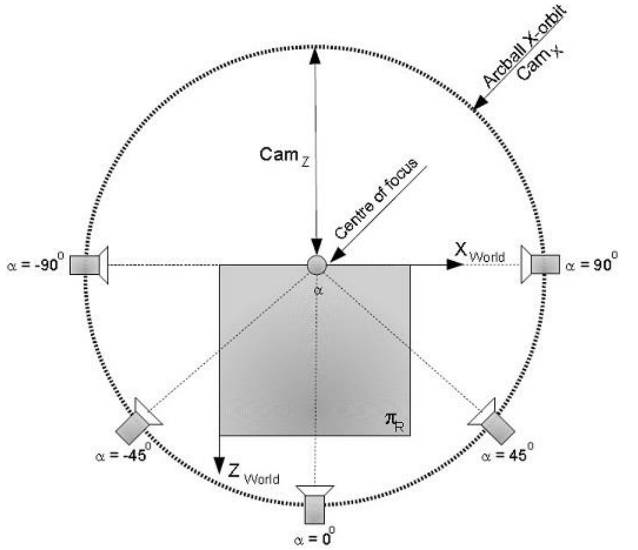

Fig. 4. Orbit of the camera's X co-ordinate, showing the movement of the camera relative to the XY plane, described by the angle α.

When α = 0º the virtual camera plane will be parallel to π_N. In this case the original image will be displayed, with no redundancy visible. Redundancy can be considered symmetrical across the -90º ≤ α ≤ 0º and 0º ≤ α ≤ 90º ranges. Redundancy in the viewed image plane is dependent on the value of α, and can be quantified in terms of the percentage of the road plane that increases with the increase in α.

The orientation of the steering wheel is relayed to the PC as a 16-bit value, giving a range of 0 to 65,535, with a 0 value representing the wheel at the leftmost position and the 65,535 value representing the wheel at the rightmost position. The value is normalised in the range of ±180º, representing the full range across 360º. In the case where the arcball camera position is neutral (i.e. it lies in the same position relative to the imaged road as the acquiring camera did), the projected scene will be the same as that of the original video. As the arcball camera's rotation changes, the viewpoint of the road scene will change also, resulting in a steerable scene.

6. Testing and Results

A timer was displayed on screen, and reset every ten seconds. Participants were instructed to change lane when the counter reset to zero.

For the purposes of testing, acceleration was disabled. The video advanced automatically, allowing each participant's steering response to be measured independently of speed. The lane that the driver was in was recorded, with the normalised average position of the ten participants shown in Fig. 5.

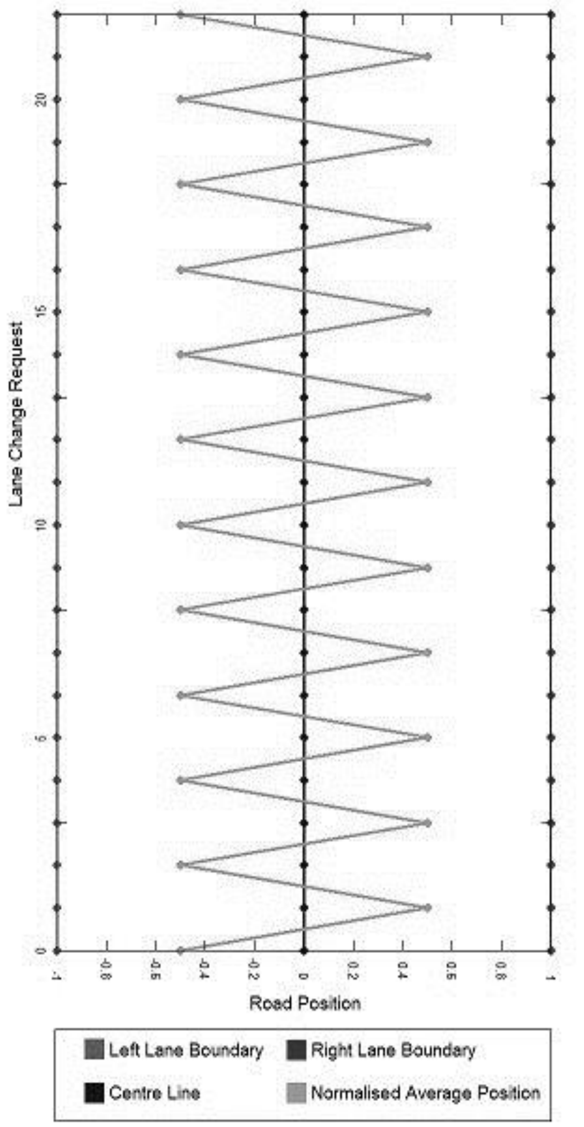

Fig. 5. Normalised average lane position (22 lane changes per participant with 10 participants).

The number of lane changes requested from the 10 participants was 22 each. Of these 220 total requests, 186 were completed successfully (87%).

The ten participant data set had an average success rate of 87%, ranging from 64% to 100% per participant, with a mean of 18.6 lane changes and standard deviation of 2.67.

7. Conclusions and Future Work

This paper has shown that a single video sequence, acquired by a standard witness camera, can be adapted for use in a video-based driving simulator by estimating a single depth map that represents a road with an assumed constant geometry. Average driver response to the change lane instruction was 87%. The failure rate of 13% may be attributed to participants missing the resetting of the onscreen counter.

Future work will consist of extending the existing approach such that a depth map can be estimated for each frame in the video sequence using the parallelism of the road edges and the horizon line. This will enable the approach to be implemented using any road geometry, whether it is constant or changing constantly. A comparison of this technique against the genuine depth maps generated using the stereo image data of the original mapping system will then be undertaken to compare and contrast the two methods. Behavioural testing of drivers will be the focus of further testing using the driving simulator once the full video steering component has been integrated with it.

Acknowledgments

The authors wish to thank the support of the Irish Research Council (IRC) for their continuing support, and also to the participants for their time.

References

[Bre1] Bredif, M., "Image-based rendering of LOD1 3D city models fortraffic-augmented immersive street-view navigation," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Antalya, Turkey, 2013, pp. Volume II-3/W3.

[Bro1] Brogan, M., Kaneswaran, D., Commins, S., Markham, C., and Deegan, C. "Automatic generation and population of graphics-based driving simulator using mobile mapping data for the purpose of behavioral testing of drivers," *Proceedings of the Transportation Research Board Annual Meeting*, Washington DC, January 2014.

[DeC1] De Ceunynck, T., et al., "Proactive evaluation of traffic signs using a traffic sign simulator," in *Annual Meeting of the Transportation Research Board*, Washington, DC, 2014.

[Har1] Hartley R., Zisserman, A., *Multiple View Geometry in Computer Vision*, 2nd ed.: Cambridge University Press.

[Her1] Heras, A.M., Breckon, T.P. and Tirovic, M., "Video Re-sampling and Content Re-targeting for Realistic Driving Incident Simulation," in *Proc. 8th European Conference on Visual Media Production*, 2011, pp. sp-2.

[Mio1] Mio MiVue 388 Witness Camera, [Online : 22 May 2014]

http://eu.mio.com/en_gb/mivue-388.htm

[Ono1] Ono, S., et al., "A photo-realistic driving simulation system for mixed-reality traffic experiment space," in *IEEE Symposium on Intelligent Vehicles Symposium*, Las Vegas, Nevada, 2005, pp. 747-752.

[Sat1] Sato, R., Ono, S., Kawasaki, H., and Ikeuchi, K., "Real-time image-based rendering system for virtual city based on image compression technique and eigen texture method," in *19th International Conference on Pattern Recognition (ICPR)*, Tampa, Florida, 2008, pp. 1-4.

[Thr1] Thrustmaster Steering Wheel [Online : 22 May 2014]

http://www.thrustmaster.com

[Tru1] Trucco, E., A. Verri, A., *Introductory Techniques for 3-D Computer Vision*, 1st ed.: Prentice Hall, 1998.