Generation of Virtual Display Surfaces for In-vehicle Contextual Augmented Reality

Srinath Sridhar*
Max-Planck-Institut Informatik

Victor Ng-Thow-Hing†
Honda Research Institute USA, Inc.

Abstract

In-vehicle contextual augmented reality (I-CAR) has the potential to provide novel visual feedback to drivers for an enhanced driving experience. To enable I-CAR, we present a parametrized road trench model (RTM) for dynamically extracting display surfaces from a driver's point of view that is adaptable to constantly changing road curvature and intersections. We use computer vision algorithms to analyze and extract road features that are used to estimate the parameters of the RTM. GPS coordinates are used to quickly compute lighting parameters for shading and shadows. Novel driver-based applications that use the RTM are presented.

1 Introduction

The concept of augmented reality (AR) windshields in vehicles holds great potential for increased safety and an enhanced driver experience. A majority of in-vehicle display systems are currently located in the central/steering wheel dashboard areas, requiring drivers to take their eyes off the road. We present a system for overlaying relevant virtual elements in the same 3D spatial context of the scenery in the driver's field of view. Our system uses a sparse parametrized model that can estimate a continuously changing simplified representation of road structure, called the *road trench model* (RTM). The RTM creates *virtual display surfaces* (VDSs) that provide reference coordinate frames for augmentation. Since there are few parameters, fitting and tracking of the model over continuously changing straight and curved roads can be maintained at interactive rates. To complement the RTM, we provide methods for fast shading and shadows using GPS information and multi-cue road intersection detection. Novel applications enabled by our model are presented, including virtual roadside address labels and dynamic advertising at intersections. Our system is designed such that the RTM can be used with a conformal HUD or conventional video overlays.

2 The Road Trench Model (RTM)

Virtual display surfaces provide a frame of reference that can be used for overlaying 3D virtual scenes on real scenes. We propose this as a general concept applicable to any scene. The RTM, which is a simplified representation of the road scene, is a specific instance of a virtual display surface. The RTM simultaneously records the geometry, lighting and other semantic attribute information about the road scene which can be used for augmentation.

The rationale behind choosing the RTM as the virtual display surface for the road scene is intuitive. A typical road scene as seen by a camera mounted on a vehicle can be represented simply by three dominant planes (Figure 1). First, the flat ground plane denotes the surface on which the vehicle is moving. The RTM also consists of two side walls that lie along the left and right edges of the road. All other objects in the scene can thought of as

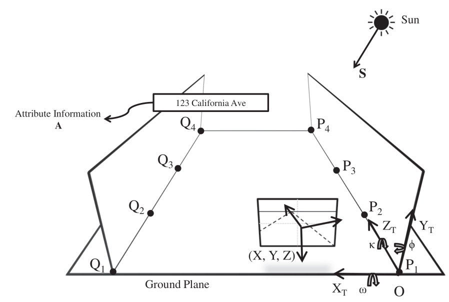

Figure 1: The Road Trench Model.

being attached and located in coordinate systems associated with one of these major planes. By locating these VDSs, we can establish a simplified context to place augmented data in our scene. In addition to the surfaces, the RTM also captures the position and orientation (pose) of the camera on the moving vehicle. The surface geometry and the pose of the camera constitute the geometric parameters of the RTM, $\mathbf{G}$. The set $\mathbf{G}$ can be written as, $\mathbf{G} = \{\mathbf{P}, \mathbf{Q}, X, Y, Z, \omega, \phi, \kappa\}$, where $\mathbf{P}$ and $\mathbf{Q}$ denote points P_k (right road edge) and Q_k (left road edge) respectively for $k = \{1, 2, 3, 4\}$; X, Y and Z denote the x-, y- and z-coordinates of the camera with respect to the coordinate system O; ω, ϕ and κ denote the rotation of the camera about the x-, y- and z-axes respectively.

The RTM also captures some lighting information about the scene from which shading and shadowing of the virtual scene can be performed. We assume that the only source of light on the road is the Sun. The RTM encodes the position of the Sun as a direction vector, $\mathbf{S} = [s_1, s_2, s_3]^T$ with respect to the coordinate system O. If the vehicle has access to the internet then location-based information (attributes) can also be incorporated in the RTM. The RTM has a variable set of attributes denoted by $\mathbf{A} = \{a_1, a_2, \ldots, a_i\}$, where i denotes the maximum number of attributes available. Examples of attributes include street address labels or business names of buildings.

Thus, any particular configuration of the RTM at a given time instance, t, can be denoted by the set of parameters given by $\mathbf{T}_t = \{\mathbf{G}_t, \mathbf{S}_t, \mathbf{A}_t\}$. The goal is to estimate values for $\mathbf{G}_t, \mathbf{S}_t$ and provide values for $\mathbf{A}_t$ for the RTM at every time instant $t > 0$ while our system is running. The entire system pipeline is shown in Figure 2.

3 Generation of a Dynamic RTM

3.1 Onboard Measurements

Raw sensor data is collected onboard the vehicle from a single Point Grey Flea 2 RGB camera and a 10Hz QStarz Q1000X GPS receiver. The camera is mounted on a rigid frame on the roof of the vehicle. We cache secondary (attribute) information from web services such as Google Maps for different types of location-based information display.

3.2 Estimating RTM Parameters

We use feature detectors for road edges to estimate various RTM parameters.

*e-mail: ssridhar@mpi-inf.mpg.de
†e-mail: vng@honda-ri.com

IEEE International Symposium on Mixed and Augmented Reality 2012
Science and Technology Proceedings
5 - 8 November 2012, Atlanta, Georgia
978-1-4673-4662-7/12/$31.00 ©2012 IEEE

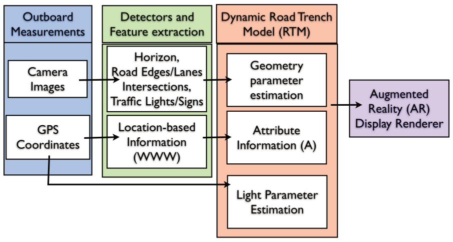

Figure 2: System Pipeline.

3.2.1 Geometric Parameters ($\mathbf{G}_t$)

- **Horizon Detector** The horizon encodes useful information about the scene which can be used to estimate the geometric parameters ω and κ of the RTM. We assume that the camera's height from the ground is a constant, h. Through a calibration procedure, we enforce the roll of the camera with respect to the horizon to be zero which makes the angle κ zero as well. We use this information along with the vanishing point and horizon estimated using the method proposed by [4] to compute the angle $\omega = \tan^{-1}(\frac{v_h}{f_y})$. Here, the v-coordinate of the horizon on the (u,v) plane of the image is v_h and the focal length of the camera along the v-direction is f_y.

- **Road Edge and Lane Detector** The road edge and lane detector is used to estimate the geometric parameters $\mathbf{P}, \mathbf{Q}, X$ and ϕ of the RTM. Our method for road edge detection is based on [1]. We first obtain an Inverse Perspective Mapping (IPM) [2] of the road scene. Then we filter the transformed image using a separable 2D Gaussian filter tuned to detect vertical edges (lane markers). We then use a RANSAC-based approach to fit a Bézier spline to the detected edges. The spline control points for road edges when transformed to ground coordinates correspond to the geometric parameters $\mathbf{P}$ and $\mathbf{Q}$ of the RTM (Figure 1). The angle ϕ can also be computed by considering the angle made by the line joining the points P_1 and P_2 on the IPM image. We perform *Kalman filtering* on X and Z (from GPS) to avoid jitter. Y is fixed since the height h of the camera from the ground is considered constant. We also connect multiple RTMs together for smooth visual motion.

3.2.2 Lighting Parameters ($\mathbf{S}$)

In order to properly illuminate and shade virtual objects so that they appear part of the real scenery, we estimate the positions of lights (*i.e.* the Sun) in the real world. GPS is used to obtain the position and time of the car's current location so that the relative position of the Sun can be estimated accurately. In our system we adopt the empirical method proposed by [3] which gives the altitude and azimuth of the Sun at any given time for the specified location to one-tenths of a degree. We then transform the spherical coordinates into a direction vector, $\mathbf{S} = [s_1, s_2, s_3]^T$ which is used for real-time shading and shadow casting.

3.2.3 Attribute Information ($\mathbf{A}$)

In order to drive the content of the virtual scene we need some attribute information ($\mathbf{A}$) such as addresses. We obtain these from online map query services and make them part of the RTM. Apart from addresses, we also incorporate intersections, stop signs and traffic lights as attribute information. In order to detect intersections we have an intersection detector which uses a multi-cue approach for detection. The cues include gray road areas, stop signs and traffic lights, all detected using HSV color segmentation.

4 Results and Conclusion

We created two scenarios where the dynamic RTM can be used to augment the driver's view. The first scenario is the display of addresses as the vehicle moves on a road (Figure 3 (bottom left)), allowing drivers to keep their eyes on the road. The addresses are obtained by reverse geocoding GPS coordinates using mapping services. The label is displayed in 3D correctly in perspective. As the vehicle moves forward, the label moves smoothly with the RTM defined for the current visible road scene. In Figure 3 (top images) the shadows are correctly positioned relative to the Sun as the car changes orientation and viewpoint while turning onto a new road.

Figure 3: Top Left and Right: Correct shadows for road; Bottom Left: Virtual address labels; Bottom Right: Virtual advertising and signage at intersections.

The second scenario demonstrates the display of important traffic information to aid drivers at intersections. When a car approaches an intersection in Figure 3 (bottom right), the detected red light triggers a virtual STOP sign. While the car is stopped at the red light, selected video advertisements are displayed like virtual billboards. Once the light turns green, the advertisement automatically disappears and a green GO sign is displayed to remind the driver it is now safe to proceed.

We have demonstrated that with a relatively simple model for roadside scenery, the RTM can adapt smoothly without jitter to changes in scenery while driving. The computational performance of our detectors allows us to achieve interactive frame rates of 30fps on a computer with a 2.66 GHz Intel Xeon X650 processor, 8 GB of RAM and an Nvidia Quadro FX5800 graphics card. We show how the various parameters of the RTM can be extracted including geometric and lighting parameters. We also show how attribute information is used to display information relevant to the driver in straight roads and intersections. Finally, several interesting in-vehicle AR applications are prototyped and demonstrated.

Acknowledgements

This work was done when the first author was an intern at the Honda Research Institute USA, Inc., Mountain View, CA, USA.

References

[1] M. Aly. Real time detection of lane markers in urban streets. In *IEEE Intelligent Vehicles Symposium 2008*, pages 7–12.

[2] M. Bertozzi and A. Broggi. Real-time lane and obstacle detection on the gold system. In *Proceedings of the 1996 IEEE Intelligent Vehicles Symposium*, pages 213–218.

[3] D. Laney. Approximate Position of the Sun from any Location at any Time. `http://goo.gl/gyZaq`, 2009.

[4] J. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1250–1257.